BACHELOR OF COMPUTER APPLICATIONS (BCA)

CSB-1101 (Computer Fundamental and Office Automation)

Block 1: Introduction to Computer and Storage Devices

Unit 4: Primary Storage Technologies and Capacities

4.0 Introduction

The purpose of this unit is to teach you about the basic concept of computer memory and its significance. Understanding the processing of data in a computer system is not sufficient. The system must have the data available at all times before processing it. We will be discussing how data is stored in a computer in this unit. You will become familiar with the different storage devices that exist, such as SRAM, DRAM, ROM, and cache memory. Additionally, the course will cover particular varieties of magnetic storage devices, such as magnetic tapes and magnetic disks, as well as disk devices like floppy disks and hard disks. The organization of data on a magnetic disk will be taught to you as well.

This unit will teach you about optical and solid-state storage devices. James Russell is the inventor of the Compact Disk. Molded polymer is the material used to make a CD, a device that is both small and portable and easy to use. The use of it involves recording, storing, and playing back audio, video, text, graphics, and other types of content. in a digital form.

4.1 Objectives

After completing this unit, you will have gained the following abilities:

- Distinguish between memory types and the requirements of the memory system.
- Identify the distinctions between different types of memories, such as semiconductor, magnetic, and optical memory.
- Describe the importance of the memory hierarchy
- Recognize the different types of storage devices.
- Describe the importance of magnetic storage devices
- Explain how data is arranged on a magnetic disk
- Talk about the different types of disk devices
- Familiarize yourself with the use of tape drives and measure their performance.
- Provide an explanation of the different types of optical and solid storage devices

4.2 Introduction to Computer Memory System

Memory is an electronic circuitry that facilitates the storage and retrieval of data when necessary. The memory unit is commonly known as the primary memory. There is a direct interaction between it and the CPU. Backup storage is provided by storage devices, which are also known as auxiliary memory.

It is categorized as a storage device. It falls under the category of storage devices. It stores sets of instructions, i.e. Programs, data, and the results of computations' intermediate stages. It is classified into two categories called main (or primary) memory and auxiliary (or secondary) memory. The main or primary memory is referred to as the memory unit, which has already been mentioned.

The following key characteristics are used to categorize computer memory:

(1) **Location**

- CPU
- Internal memory (main memory)
- External memory (secondary memory)

(2) Access Method

- Sequential access
- Direct access
- Random access

(3) Performance

- Access time
- Cycle time
- Transfer rate

(4) Physical Type

- Semiconductor
- Magnetic surface

(5) Physical Characteristics

- Volatile/non-volatile
- Erasable/non-erasable

(6) Capacity

- Internal
- External
- Addressable unit

(7) Unit of transfer

- Word
- Block

4.3 Types of Storage Devices

The storage unit of the computer system's storage unit is characterized and evaluated using the following common properties:

1. Storage Capacity: Represents the size of the memory. The storage unit's data storage capacity. Compared to secondary storage units, primary storage units have less storage capacity. Internal and main memory's capacity can be measured in terms of words or bytes. Internal or secondary storage capacity is measured in bytes, but external or secondary storage capacity is measured in bits.

2. Storage Cost: Cost is another key factor that is of prime concern in a memory system. Bits are the norm when expressing it. It is clear that it is desirable to have lower costs. It's worth remembering that the cost of memories decreases with increased access time.

3. Access Time: The duration needed to locate and retrieve data from a storage unit is influenced by its physical attributes and the access mode utilized. The efficiency and speed of accessing data are determined by factors such as the storage device's technology, its physical components, and the chosen access method.

In direct comparison, primary storage units typically offer faster access times compared to secondary storage units. Primary storage, often referred to as "main memory" or "RAM" (Random Access Memory), provides swift access to data for the processor. This immediacy in accessing information is due to the direct electronic access method employed by primary storage, allowing the CPU to rapidly retrieve and manipulate data stored within it.

On the other hand, secondary storage devices like hard disk drives (HDDs) or solid-state drives (SSDs) generally have relatively slower access times when compared to primary storage. Secondary storage is characterized by its capacity to retain large volumes of data even when the device is powered off. However, accessing data from secondary storage often involves mechanical processes (in HDDs) or electronic procedures (in SSDs) that are comparatively slower than the direct electronic access of primary storage.

4. Access Mode: xxx Certainly! Memory encompasses various access modes, each offering distinct ways to retrieve information from memory devices:

(a) Random Access Memory (RAM): RAM facilitates accessing any memory location in any order within the same duration of time. It's the primary storage in computers, usually composed of ferrite and semiconductor memories. This type of memory enables swift access to data without needing to follow a specific sequence.

(b) Sequential Access: Memory systems operating in a sequential access mode allow access to data only in a predetermined order. Access time is contingent on the location, requiring sequential traversal through other locations before reaching the desired one. For instance, magnetic tapes follow this sequential access pattern, similar to songs stored on a cassette, accessed one after another.
(c) Direct Access: This mode lies between random and sequential access. It involves accessing information serially on separate tracks where each track has its own read/write head. Magnetic disks, employing this semi-random-access method, enable accessing data on each track individually.

The Permanence of Storage also Delineates Memory Types:

- Non-volatile storage retains data even when the power is turned off or interrupted, making it a preferred choice for storing large volumes of data.
- Volatile storage, in contrast, loses data when power is disrupted, typical of primary storage units in computer systems.

Static and Dynamic RAM Represents Distinctions within the Main Memory:

Static RAM (SRAM): It is a type of RAM that uses a bistable latching circuit to store each bit. SRAM is fast and does not need to be refreshed regularly but is more expensive and consumes more power.

Dynamic RAM (DRAM): This type of RAM stores each bit of data in a separate capacitor within an integrated circuit. DRAM needs to be refreshed periodically to retain data and is more affordable but slower compared to SRAM.

The primary storage in a computer system serves as the central storage unit, accommodating programs and data during computer operations. Its characteristics and access modes determine the efficiency, speed, and volatility of memory, influencing the overall performance of computing systems.

The primary memory technology is built on semiconductor integrated circuits. Integrated circuit RAM chips can be used in two different modes. These modes are: static and dynamic.

Binary information is stored by the Static RAM (SRAM) through the use of clocked sequential circuits. Only when power is applied to the unit will the stored information remain valid. Dynamic RAM (DRAM) stores binary information by applying electric charges to capacitors inside the chip. Refreshing the dynamic memory periodically is necessary to recharge the stored charge on the capacitors, which tends to discharge with time. By using dynamic RAM, you can increase storage capacity and decrease power consumption. Dynamic RAM is the preferred choice for large memories, while static RAM is primarily utilized for specialized applications.

Both read and write types of memory are discussed above. What about a memory where only one of the operations is possible, e.g. If we restrict memory access solely to reading and refuse to alter the information stored there? The memory might have some major importance; like an important bit of the computer’s operating system which normally does not change can be stored in this kind of memory. Such a memory is called ROM (Read Only Memory).

Read Only Memory (ROM): Most of the memory in a general-purpose computer is made of RAM-integrated circuit chips, but a portion of the memory may be constructed using ROM chips. RAM was originally referred to as random-access memory, but now we use read/write memory to differentiate it from read-only memory (since ROM is also random access). RAM is used for storing the bulk of the programs and data that are subject to change, while ROM is used to store programs that are permanently resident in the computer and do not change once the production of the computer is completed.

Within the main memory, the Read-Only Memory (ROM) plays a critical role in storing essential programs and instructions that remain intact even when the power is turned off. One primary function of ROM is housing an initial program known as the "bootstrap loader." This program is fundamental as it kickstarts the operation of computer software when the power is switched on.

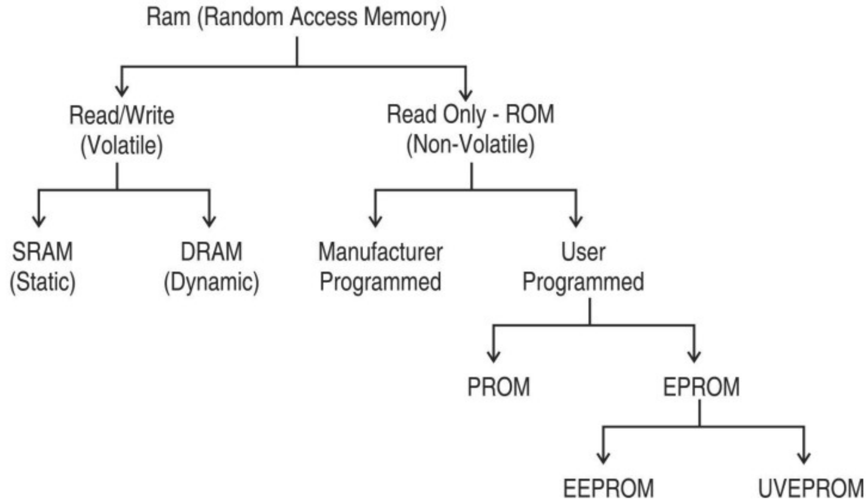
Unlike Random Access Memory (RAM), which is volatile and loses its contents when power is cut off, ROM retains its stored information regardless of power cycles. This non-volatile nature of ROM ensures that its contents remain unchanged and preserved even after multiple instances of power being turned on or off.

The bootstrap loader, residing in ROM, holds instructions necessary to initiate the computer's operating system. When the system powers up, the processor refers to the bootstrap loader in ROM to initialize essential functions and load the initial processes that enable the computer to start running its software. Therefore, while the contents of RAM are temporary and volatile, requiring constant power to maintain data integrity, the contents of ROM persistently hold vital instructions, making it a crucial component in the startup process of a computer system.

Programming read-only memories is achievable for both the manufacturer and the user. Manufacturer-programmed ROM is the term used for data that is burned into the computer device circuitry by the manufacturer. A private computer manufacturer has the option to stock up the boot program permanently in the ROM fragment of computers they manufacture. The manufacturer supplies these chips and they cannot be modified by users. This is an inflexible process and requires mass production. A new type of ROM was created, which is called Programmable Read-only Memory (PROM). The nature of this is non-volatile. The use of special equipment is necessary to write it once. The writing process in PROM is capable of performing electrically by either the supplier or the customer.

Both ROM and PROM have a limit on how many write operations can be performed, and there is no way to modify anything that has already been written. But what about the cases where you mostly read but also write a few times? To handle these situations, a memory chip called EPROM (Erasable Programmable Read Only Memory) was created. EPROMs are frequently employed by R&D personnel to test their efficiency by modifying micro-programs on the computer system.

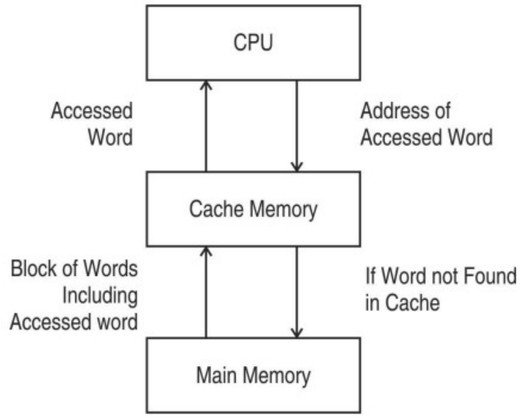
Further, EPROM chips are of two types: EEPROMs (Electrically EPROM) in which high voltage electric pulses are used to erase stored information and UVEPROM (Ultra Violet EPROM) in which stored information is erase by revealing the chip for a while to the ultraviolet light.



Cache Memory: Cache memories are small, fast memories placed between CPU and main memory. Their access times are closer to the CPU's speed, making them faster than main memory. Caches have a fast speed, but they are also very expensive. So, they are used only in small quantities. As an example, caches of size 64K, and 128K are normally used in PC-386 and PC-486, which can have 1 to 8 MB of RAM or even more. The size of the memory is not compromised by cache memories, which provide fast memory retrieval.

Given the small size of the memory, would it be advantageous to increase its overall speed? This can be answered with the help of a phenomenon known as locality of reference. Let us examine what this means.

Locality of Reference: If we analyse a large number of typical programs, we would find that memory references at any given interval of time tend to be confined to a few localized areas in the memory. This phenomenon is known as the property of reference location. Most programs have iterative loops (such as 'for' or 'while' loops), which is why this is true. During the execution of such programs, the same set of instructions (within the loop) are executed many times. The CPU keeps referring to the set of instructions in memory that make up the loop. The set of instructions for a specific subroutine is fetched from memory every time it is called. As a result, loops and subroutines tend to localize references to memory for retrieving instructions.



We may deduce from the proximity of reference that the cache holds a copy of some of the main memory. The cache is the initial stage in detecting memory read or write processes, and if there is enough data in the cache, the CPU uses it directly. If not, a block of words is read from primary memory and stored in the cache, and the CPU uses the cached word.

Secondary Storage Devices: RAM, as previously stated, is a volatile memory that has restricted storage capacity. RAM is somewhat more expensive than secondary memory. The rationale necessitated that a reasonably affordable medium with some form of store permanence be employed. External or supplementary memory, often known as secondary storage, is utilized for extra memory in most computers.

The magnetic media became a suitable solution for huge storage requirements due to its extended lifespan and low cost. Magnetic tapes and disks are popular storage medium. As optical technology progresses, optical disks have grown into one of the most important secondary storage systems.

4.4 Memory Types and Storage Devices

Memory is a vital component of a computer in which every piece of information and data are stored as binary digits (combinations of 0s and 1s) and retrieved as needed.

Memory is separated into two basic functions: storing applications, data, and information within the system and keeping track of calculation results.

A computer system stores instructions and data via a variety of devices. A computer program must be kept in memory in order to operate. Any input data required for that program's processing ought to be preserved in memory. Memory is where any of the program's initial outcomes and outputs is saved until the computer is switched off.

In the hierarchy of computer storage devices, several criteria are used to rank and assess their performance:

Access Time: This refers to the duration required to locate and retrieve stored data in response to a program instruction. It signifies the period between requesting data and its availability. Faster access times are preferred as they contribute to quicker data retrieval.

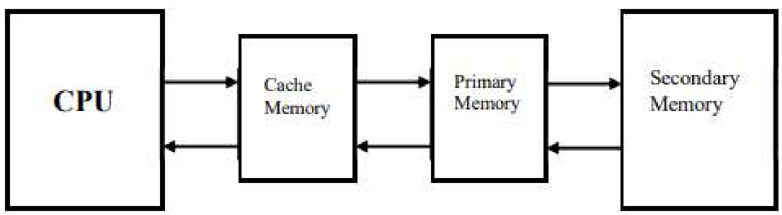
Storage Capacity: It indicates the volume of data a storage unit can hold. A larger storage capacity is favourable as it allows for the retention of more data. Systems with higher storage capacities can accommodate a greater volume of information.

Cost per Bit of Storage: This denotes the expense associated with a storage unit concerning its storage capacity. Minimizing this cost is a primary objective, aiming for a lower cost per unit of storage to ensure economical and efficient storage solutions.

Additionally, there's a specialized high-speed memory called Cache Memory, strategically used to enhance processing speed. Cache memory stores frequently used instructions or data, providing rapid access to the CPU. Positioned between the CPU and main memory, cache memory acts as a high-speed buffer, decreasing the access time for critical information.

The access time for cache memory is notably swift, typically ranging between 0.5 to 2.5 nanoseconds, significantly faster than that of main memory, which usually operates at about 50-70 nanoseconds. However, due to its higher speed and specialized function, cache memory has a limited capacity, usually around 2 to 3 percent of the main memory's capacity, making it more expensive.

Cache memory's role is crucial in reducing the latency between the CPU and main memory, ensuring that frequently accessed data or instructions are readily available for swift processing, despite its limited capacity and higher cost compared to main memory.



4.4.1 Semiconductor (Main) Memory

Presence of Semiconductor and Magnetic Memory: With the exception of very compact computing devices, most computers utilize both semiconductor and magnetic memory types in their architecture.

Main Memory in Modern Computers: The primary memory in modern computers is semiconductor-based and is commonly referred to as Random Access Memory (RAM). Its distinguishing feature lies in the ability to access any part of the memory for both reading and writing purposes.

Storage of Programs and Data: The CPU stores and accesses the programs and data currently in use within the main memory, ensuring quick accessibility and retrieval for immediate processing.

Role of Read Only Memory (ROM): The main memory includes Read Only Memory (ROM), which, unlike RAM, does not allow writing data to it. ROM retains data in a read-only format, making it only possible to read from it, not write to it.

Composition of Computer Main Memory: The primary memory of a computer comprises both RAM and ROM, each serving distinct purposes in data storage and retrieval.

CPU's Dependence on Main Memory: To execute its operations, the CPU relies on the programs and data stored in the main memory, accessing and manipulating this information during processing.

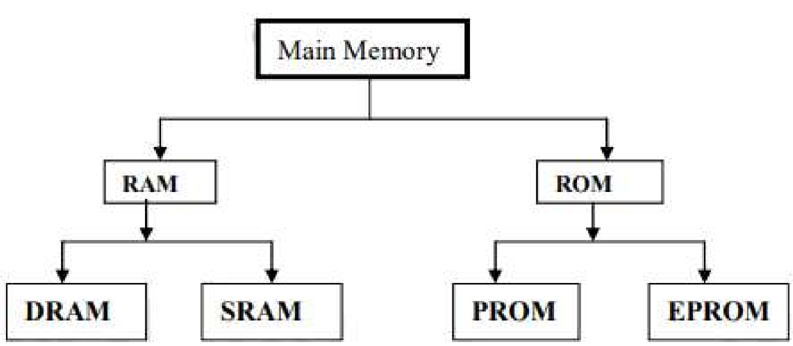
Construction of Primary Storage: The primary storage, composed of semiconductor materials, utilizes integrated circuits. Data within the main memory is stored in binary form, representing both numeric and non-numeric information.

Binary Representation and Data Capacity: Binary form represents data in computers, enabling the representation of various characters and digits. For instance, two binary digits (bits) can represent four different characters, while eight bits form a byte, capable of representing 256 characters or symbols. A byte serves as a fundamental unit for digital information storage and processing in computer systems.
Bytes serve as a fundamental unit of measurement for quantifying the capacity of a computer's memory. Within computer systems, codes like ASCII (American Standard Code for Information Interchange) utilize different configurations of bits to form bytes that represent a wide range of characters, symbols, and numerical values.

ASCII, one of the most commonly used character encoding standards, employs a seven or eight-bit binary code to represent characters. By combining bits in various patterns, ASCII creates bytes that correspond to different characters. These bytes can represent numeric digits (0 to 9), uppercase and lowercase letters of the alphabet, punctuation marks, special characters, and control characters used in computing.
For example, with seven bits, ASCII can represent 128 characters, including the English alphabet (both upper and lower case), numbers, punctuation marks, and control characters. When using eight bits (one byte), ASCII extends its character set to include additional symbols, foreign characters, and extended punctuation, resulting in a total of 256 possible characters.

Bytes, being a standardized unit of measurement in computing, facilitate the representation and manipulation of data in various forms, allowing for the encoding and interpretation of alphanumeric characters and symbols in computer systems through codes like ASCII.

Storage capacities in computing are commonly measured using various units such as Kilobytes (KB), Megabytes (MB), Gigabytes (GB), or Terabytes (TB), reflecting different levels of data volume.
In terms of main memory, two primary types are widely recognized: Random Access Memory (RAM) and Read-Only Memory (ROM).



RAM (Random Access Memory): RAM refers to the volatile memory in a computer system that facilitates both reading and writing operations (R/W memory). Users can input data into RAM and retrieve information from it. The term "random access" signifies the capability to access any memory location within RAM for both reading and writing purposes. This characteristic ensures that each memory location in RAM has an equal access time, allowing for swift and random access.

RAM is often described as "temporary" memory since data stored in RAM is volatile and gets lost when the system is powered off or shut down. Its temporary nature means that it serves as a workspace for the CPU to actively use during computing operations, but it doesn't retain data once the power is turned off. Due to its quick access times and ability to read and write data, RAM plays a crucial role in facilitating the efficient functioning of applications and the operating system by temporarily storing information that the CPU requires for immediate processing.

Random Access Memory (RAM) stands as the primary storage location where programs and software loaded into a computer are stored. When the CPU executes a program, it retrieves its instructions from the RAM and carries out the operations accordingly. Likewise, if the CPU needs to retain the results of computations, it stores this information back into the RAM. This characteristic enables the CPU to both read data from and write data to RAM, facilitating seamless interaction and data manipulation.

RAMs are commonly categorized into two significant types: Static RAM (SRAM) and Dynamic RAM (DRAM).

Static RAM (SRAM): SRAM preserves stored information only as long as the power supply is active. This type of memory retains data as long as the computer is powered on. It is faster but more expensive compared to DRAM, consuming more power.

Dynamic RAM (DRAM): DRAM loses stored information relatively quickly, typically within a few milliseconds, even while the power supply is active. Despite its volatility, DRAM is more cost-effective and consumes less power compared to SRAM. It is commonly used as the primary memory (main memory) in computer systems due to its affordability.

SRAM's higher cost is balanced by its faster access times, making it suitable for specialized functions. It finds its application in cache memory, which acts as a high-speed buffer between the CPU and main memory. SRAM's faster speed significantly enhances the performance of the cache memory.

An essential distinction lies in the data retention mechanisms between SRAM and DRAM. DRAM requires periodic refreshing of data to maintain its stored information, while SRAM does not necessitate such refreshing processes due to its design, adding to its advantage in terms of simplicity and speed.

4.4.2 Magnetic Memory

Semiconductor RAMs, known for their high-speed operations, do come with drawbacks such as limited capacity and volatility, prompting the need for alternative storage solutions that are both cost-effective and non-volatile. Magnetic material serves as a viable choice due to its affordability, durability, and permanence.

Challenges of Semiconductor Memory:

Limited Capacity: Today's semiconductor or primary memory in computers often falls short in meeting the data storage requirements of organizations dealing with vast volumes of data. The capacity constraints of semiconductor memory become evident in scenarios demanding extensive data processing.
Volatile Nature: Semiconductor memory is volatile, meaning it loses its stored data when power is turned off. However, there's a constant need to retain data permanently, necessitating a non-volatile storage solution.

Magnetic Memory as an Alternative:

Magnetic memory presents an economical and enduring storage medium that doesn't require power to maintain stored information. Though floppy disks are obsolete, modern computers employ two types of magnetic memory:

(1) **Magnetic Disks:** This category includes hard disks and obsolete floppy disks. Hard disks are prevalent secondary storage devices, offering substantial storage capacities and relatively fast access at reasonable costs.
(2) **Magnetic Tapes:** While less commonly used in personal computing, magnetic tapes serve as an alternative for large-scale archival storage due to their cost-effectiveness and durability.

Storage Mechanism of Magnetic Disk Drives:

Construction: Metal disks coated with iron oxide serve as the foundation of magnetic disk drives. These disks, spun at speeds ranging from 3600 to 7600 revolutions per minute (rpm), consist of multiple platters stacked on a vertical shaft.

Data Reading/Writing: Electromagnetic read/write heads situated between the disks, separated by access arms, enable the reading and writing of data on concentric circular tracks. These tracks contain tiny magnetized spots that encode binary digits representing computer codes. Each track holds thousands of bytes, and each disk surface contains hundreds of data tracks, providing vast storage positions for software and data.

Magnetic disks, with their numerous tracks and storage density, offer substantial storage capacities, allowing for the retention of extensive volumes of software and data, meeting the requirements of modern computing needs.

4.4.3 Optical Memory

Optical memory and disks represent alternative mass storage systems with substantial capacities, reaching up to 20 GB. These systems rely on laser beams for both writing and reading data, offering a durable and cost-effective storage solution.

Characteristics of Optical Disks:

Single-Surface Data Storage: Optical disks store data on a single surface, utilizing laser technology for data recording and retrieval. This singular surface is where all the data is housed.

Cost and Longevity: Optical disks are relatively inexpensive and boast a lifespan of at least 15 years. This longevity makes them suitable for archiving purposes and long-term data storage.

Non-Contact Read/Write Process: Unlike traditional disk systems, optical disks operate without physical contact between the read/write head and the disk surface. This feature minimizes the risk of disk attrition or head crashes, ensuring durability and data integrity.

Types of Optical Disks:

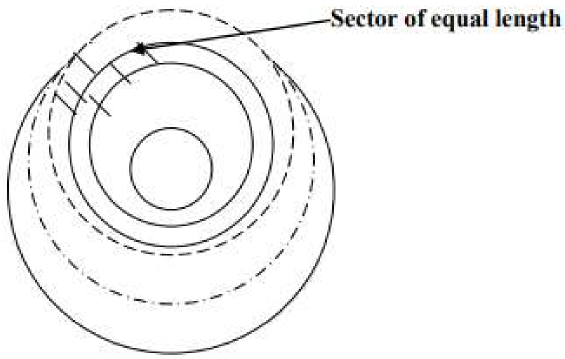
1) **CD-ROM (Compact-Disk Read Only Memory):** These disks are designed for read-only access, allowing users to retrieve data stored during the manufacturing process. They are useful for distributing software, music, or reference material that doesn't require modification.

CD-ROMs use a laser beam with varying intensities for reading and writing data. The writing process demands a higher intensity of 25mW from the laser system, whereas reading requires only 5mW.

These CDs store data in long spiral tracks that run serially across the disk. These tracks are divided into blocks of equal sizes, enabling the organized storage and retrieval of information.

During the reading process, the laser beam reads the pits, which are tiny indentations or reflective areas on the CD's surface. This reading occurs at a consistent linear speed. However, the CD-ROM disk rotates at a variable speed to facilitate constant linear velocity, ensuring that the track passes under the read/write head at a uniform rate.

The adjustment of the disk's rotational speed helps maintain a steady movement of the track beneath the read/write head. This synchronization between the disk's rotational speed and the laser's reading speed ensures accurate data reading and writing, allowing for efficient and reliable access to the stored information on the CD-ROM.



2) **WORM (Write Once Read Many) or CD-R (CD-Recordable):** These disks enable users to write data only once, after which it becomes read-only. They offer a way to permanently store information, suitable for archiving important data.

3) **Erasable Optical Disk:** These disks allow for data to be written, erased, and rewritten multiple times. They provide flexibility for data modification and reuse, similar to rewriteable formats.

4) **DVD-ROM, DVD-R, and DVD-RAM:** DVDs, an evolution from CDs, offer higher storage capacities than CDs. DVD-ROMs are read-only, while DVD-Rs allow for one-time data writing. DVD-RAMs permit data to be written, erased, and rewritten, resembling the erasable optical disk's functionality.

Disadvantage - Slow Access Time: One drawback of optical disk technology is its relatively slow average access time. While offering substantial storage capacities and durability, the slower access speed compared to some other storage systems can be a limitation in scenarios requiring rapid data retrieval.
Despite the slower access speed, optical disks, with their varied formats catering to different data needs, serve as reliable and cost-efficient options for long-term data storage, archival purposes, and content distribution.

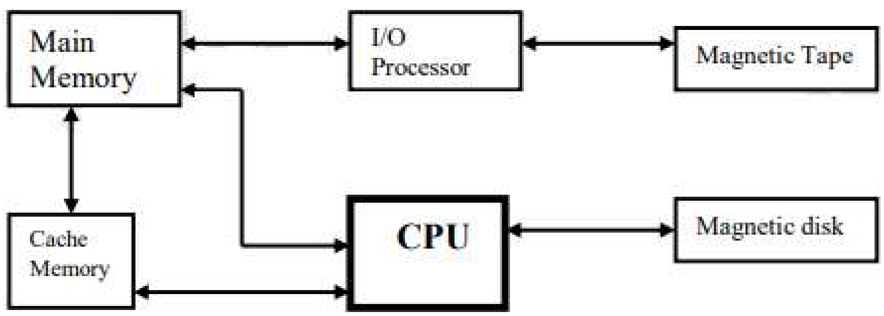
4.5 Memory Hierarchy & Their Needs

Programmers seek memory that balances both size and speed, although these requirements often conflict. Faster memories tend to be smaller and more expensive, while slower memories are larger and more cost-effective. To address this, modern computer memory systems are structured hierarchically, aiming to offer users an illusion of both speed and size.

At the apex of this hierarchy are the CPU registers, which provide extremely fast but limited storage directly accessible by the CPU. To bridge the gap between CPU registers and main memory, a swift cache memory is implemented. This cache memory serves as an intermediary, enabling faster data access compared to the main memory.

To expand the capacity of the main memory, systems use the hard disk as virtual memory, allowing for more extensive storage capabilities but at the expense of speed due to the hard disk's slower access times.
Most computer systems adopt a hierarchy of memory technologies because a singular type of memory fails to meet all requirements efficiently. This arrangement is referred to as the memory hierarchy, optimizing the trade-offs between speed and size.
In this hierarchy, faster technologies like semiconductor memory, despite being costly, require a constant power supply to maintain stored information. On the other hand, memory types with lower costs, such as optical memory, exhibit significantly higher access times. This longer access time translates to slower CPU operations, impacting overall system performance.

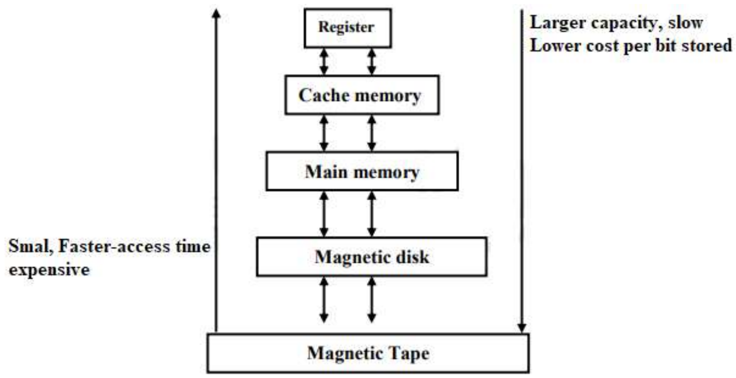
The memory hierarchy aims to strike a balance between cost and access time, striving to achieve the highest possible access speed while maintaining the overall cost of the memory system at an acceptable level. By combining various memory technologies within this hierarchy, systems aim to optimize performance and cost-effectiveness for diverse computing needs.



In a computer system, data and instructions are stored across various devices, each characterized by different access times, storage capacities, and costs per bit. These criteria categorize memory systems into three distinct groups:

1. **<u>Processor's Internal (CPU) Memories:</u>** This category encompasses a small set of high-speed registers that reside within the processor itself. These registers serve as temporary storage locations where actual processing takes place. They facilitate rapid access for the CPU during computation, aiding in the swift execution of instructions.
2. **<u>Primary (Main) Memory:</u>** Main memory is a fast and sizable memory, albeit slower than the processor's internal memory. It features faster access times than secondary memory but comes with a higher cost per bit of storage. The CPU has direct access to this memory, storing programs and data currently in use. Despite its speed, the high cost of main memory often restricts its size, leading to a relatively smaller capacity compared to secondary memory.
3. **<u>Secondary (Auxiliary) Memory:</u>** Secondary memory primarily serves as a repository for bulk storage or mass storage of programs, data, and various other information. While it boasts a larger capacity than the main memory, it operates at a slower speed. Its primary role lies in storing system software, compilers, assemblers, essential software packages, extensive data files, and similar items that do not require frequent access during immediate CPU operations.

Each category of memory serves distinct functions within the hierarchy of a computer system. Processor's internal memories offer rapid processing capabilities, while primary memory provides direct and quick access for the CPU, albeit at a higher cost per unit of storage. Secondary memory focuses on extensive data storage, sacrificing speed for larger capacities, making it suitable for archiving and storing less frequently accessed information.



4.6 Conclusion

This unit encompasses the organization of storage, various memory systems, and the critical role of memory hierarchy, categorizing memory based on access time, storage capacity, and cost per bit.

Two primary types of memory are classified based on these criteria:

1. **Primary Memory:** Modern computers primarily utilize semiconductor memory as their main memory. This type of memory is commonly referred to as Random Access Memory (RAM). Despite its speed, semiconductor memory is costly and limited in capacity, which falls short for the extensive data processing needs of contemporary computers.
2. **Secondary Memory:** In contrast to primary memory, secondary memory serves as an additional memory component that is inexpensive, non-volatile, and offers larger storage capacities. Magnetic memory, comprising magnetic disks (like hard disks and floppy disks) and magnetic tapes, falls under this category. Among these, magnetic disks are popular due to their high storage capacities, reasonable costs, and swift access times.

Additionally, secondary storage includes Optical memories or Optical disks, offering immense storage capacities of up to 20 GB at a comparatively lower cost. Optical disks utilize laser beams for writing and reading data, providing durable storage that can last at least 15 years. They also eliminate concerns like head crashes since the read/write head doesn't physically touch the disk surface. However, the main drawback of optical disk systems lies in their relatively slower average access times.

The introduction of CD-ROM technology has significantly impacted multimedia development in computing. Software companies distribute operating systems and applications on CD-ROMs, fostering multimedia technology's growth. These advancements extend to multimedia external devices like video recorders and Digital Audio Tape (DAT) recorders, amplifying their usage in multimedia systems. This technology integration has played a pivotal role in the evolution of multimedia in computing environments.

4.7 Unit Based Questions / Answers

Qus.1. What are the key storage locations and addresses in primary storage technologies?

Ans. Primary storage technologies encompass various storage locations and addresses, including Random Access Memory (RAM), Read-Only Memory (ROM), Programmable Read-Only Memory (PROM), Erasable Programmable Read-Only Memory (EPROM), and Cache memory. Each of these storage locations serves different purposes in the data processing and storage hierarchy, with RAM serving as the main memory for actively running programs and storing temporary data, while ROM, PROM, EPROM, and Cache memory are used for storing permanent or semi-permanent data and instructions.

Qus.2. How does the storage capacity of RAM compare to other primary storage technologies?

Ans. RAM typically offers larger storage capacities compared to other primary storage technologies such as ROM, PROM, EPROM, and Cache memory. RAM capacities can range from a few gigabytes to several terabytes in modern computing systems, providing ample space for storing program instructions and data temporarily during execution. In contrast, ROM, PROM, and EPROM usually have fixed or limited storage capacities, as they are designed to store firmware or boot instructions permanently.

Qus.3. What is the function of RAM in primary storage technologies?

Ans. RAM, or Random Access Memory, plays a crucial role in primary storage technologies as it serves as the main memory for storing program instructions and data that are actively being used or processed by the computer's CPU. RAM provides fast and temporary storage for data and program instructions, allowing the CPU to access and manipulate data quickly during program execution. Unlike other storage technologies such as ROM or PROM, RAM is volatile, meaning that its contents are lost when the computer is powered off, making it ideal for storing temporary data and working memory.

Qus.4. How does ROM differ from RAM in terms of storage characteristics and usage?

Ans. Read-Only Memory (ROM) differs from Random Access Memory (RAM) primarily in terms of storage characteristics and usage. Unlike RAM, which is volatile and allows data to be read from and written to, ROM is non-volatile and retains its contents even when the power is turned off. ROM typically stores firmware or boot instructions that are essential for initializing the computer's hardware and operating system during the boot-up process. ROM is read-only, meaning that its contents cannot be modified or overwritten after they are programmed onto the chip, making it ideal for storing permanent or semi-permanent data and instructions.

Qus.5. What are the primary differences between PROM, EPROM, and Cache memory in terms of storage behavior and functionality?

Ans. Programmable Read-Only Memory (PROM), Erasable Programmable Read-Only Memory (EPROM), and Cache memory differ in terms of storage behavior and functionality within primary storage technologies. PROM is programmable by the user and allows data to be written to it once, after which the contents become permanent and cannot be modified. EPROM, on the other hand, can be erased and reprogrammed multiple times using ultraviolet light, making it semi-permanent and more flexible than PROM. Cache memory, meanwhile, is a high-speed buffer memory located between the CPU and main memory, designed to store frequently accessed data and instructions for faster access by the CPU, thereby improving overall system performance.

Qus.6. How does Cache memory enhance system performance in primary storage technologies?
Ans. Cache memory plays a crucial role in enhancing system performance in primary storage technologies by reducing the latency associated with accessing data from main memory or secondary storage devices. Cache memory stores copies of frequently accessed data and instructions closer to the CPU, allowing the CPU to access them more quickly than if they were retrieved from main memory or disk storage. By minimizing the time spent waiting for data to be fetched from slower storage devices, cache memory helps improve overall system responsiveness and throughput, making it an essential component of modern computer architectures.

Qus.7. What are the advantages and disadvantages of using RAM as primary storage in computing systems?
Ans. RAM offers several advantages as primary storage in computing systems, including fast access times, high bandwidth, and volatile nature, which allows for quick read and write operations. Additionally, RAM provides a large storage capacity compared to other primary storage technologies, making it suitable for storing temporary data and program instructions during execution. However, the volatile nature of RAM means that its contents are lost when the power is turned off, necessitating the use of secondary storage devices such as hard disk drives or solid-state drives for long-term data storage. Moreover, RAM can be expensive compared to other storage technologies, limiting the amount of memory that can be installed in a system.

Qus.8. How do primary storage technologies contribute to overall system performance and responsiveness?
Ans. Primary storage technologies, such as RAM, ROM, PROM, EPROM, and Cache memory, play a crucial role in determining overall system performance and responsiveness in computing systems. These technologies provide fast access times and high bandwidth, allowing the CPU to retrieve and manipulate data quickly during program execution. By storing frequently accessed data and instructions closer to the CPU, primary storage technologies help minimize the latency associated with accessing data from slower storage devices, such as hard disk drives or solid-state drives. This improves overall system responsiveness and throughput, enabling users to perform tasks more efficiently and effectively.

Qus.9. What factors influence the choice of primary storage technologies in computing systems?
Ans. Several factors influence the choice of primary storage technologies in computing systems, including performance requirements, cost considerations, and the intended use case or workload. For example, high-performance computing systems may require large amounts of fast and volatile memory, such as RAM, to support intensive data processing tasks or simulations. Embedded systems or devices with limited power or space constraints may opt for non-volatile memory technologies like ROM or EPROM for storing firmware or boot instructions permanently. Additionally, the cost per gigabyte and scalability of different storage technologies may influence the decision-making process, particularly for organizations with budget constraints or evolving storage needs.

Qus.10. How do advancements in primary storage technologies impact computing systems and applications?

Ans. Advancements in primary storage technologies have a significant impact on computing systems and applications, driving improvements in performance, reliability, and efficiency. For example, increases in RAM capacity and speed enable computing systems to handle larger datasets and more complex workloads, leading to faster data processing.

4.8 Further Readings

1. "Computer Organization and Design: The Hardware/Software Interface" by David A. Patterson and John L. Hennessy. Publisher: Morgan Kaufmann; 5th edition (May 10, 2013). ISBN-10: 9780124077263.

2. "Computer Architecture: A Quantitative Approach" by John L. Hennessy and David A. Patterson. Publisher: Morgan Kaufmann; 6th edition (May 9, 2017). ISBN-10: 0128119055.

3. "Inside the Machine: An Illustrated Introduction to Microprocessors and Computer Architecture" by Jon Stokes. Publisher: No Starch Press; 1st edition (November 3, 2007). ISBN-10: 1593271042.

4. "Memory Systems: Cache, DRAM, Disk" by Bruce Jacob, Spencer Ng, and David Wang. Publisher: Morgan Kaufmann; 1st edition (December 7, 2007). ISBN-10: 9780123797513.

5. "Computer Systems: A Programmer's Perspective" by Randal E. Bryant and David R. O'Hallaron. Publisher: Pearson; 3rd edition (March 12, 2015). ISBN-10: 9780134092669.

6. "Modern Operating Systems" by Andrew S. Tanenbaum and Herbert Bos. Publisher: Pearson; 4th edition (March 3, 2014). ISBN-10: 013359162X.

****