UNIT

04

Introduction to Memory System

Names of Sub-Units

Basic Concepts of Memory System, Semiconductor RAM Memories, Read Only Memories (ROM), Speed, Size, Cost.

Overview

The unit begins by discussing about the concept of memory system. Next, the unit discusses the semiconductor RAM memories. Then, the unit discusses the Read Only Memories (ROM). Towards the end, the unit discusses the speed, size and cost of memories.

Learning Objectives

In this unit, you will learn to:

- Explain the concept of computer memories
- Discuss the concept of semiconductor
- Explain about Random Access Memory (RAM)
- Describe the concept of Read Only Memory (ROM)
- Discuss the speed, size, and cost of memories

Learning Outcomes

At the end of this unit, you would:

- Evaluate the importance of computer memories
- Analyse the concept of semiconductor
- Assess the importance of Random Access Memory (RAM)
- Evaluate the use of Read Only Memory (ROM)
- Analyse the speed, size, and cost of memories

Pre-Unit Preparatory Material

- https://www.uobabylon.edu.iq/eprints/publication_12_21274_1610.pdf

4.1 INTRODUCTION

One of the most crucial components of a computer system is its memory. It saves the data and instructions needed for data processing and output outcomes. Storage may be necessary for a short amount of time, immediately, or for a long time. The electrical storing area for instructions and data that the processor can read rapidly is referred to as computer memory. Computer memory is divided into two categories – primary and secondary. Primary memory is directly accessed by a processor to execute instructions. An example of primary memory is Random Access Memory (RAM). Secondary memory, such as hard disks drives and Solid State Drive (SSD), are used to store and retrieve data from a computer. Both memories are an integral part of a computer. The failure of any of the two memories stops a computer from functioning. The storage capacity of computer memory is measured in bits, bytes, Kilobyte (KB), Megabyte (MB), Gigabyte (GB), Terabyte (TB), and Petabyte (PB).

4.2 BASIC CONCEPTS OF MEMORY SYSTEM

The addressing method determines the maximum size of memory that may be used in every machine. A 16-bit computer, for example, that generates 16-bit addresses, may address up to 216=64K memory locations. Machines that produce 32-bit addresses, on the other hand, may use a memory with up to 232=4G memory locations. Byte-addressable computers make up the majority of today's computers. Figure 1 shows the block diagram of memory system:

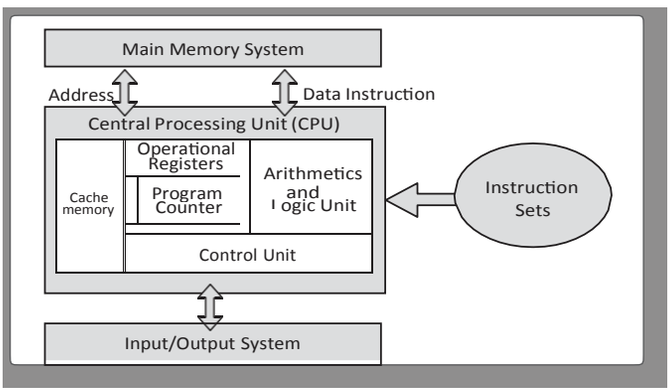

Figure 1: Block Diagram of Memory System

The time it takes for an operation to start and finish is a valuable indicator of memory unit speed. Memory access time is what it's called. Another significant metric is the memory cycle time, which is the shortest time between two memory operations. If any place can be reached for a Read or Write operation in some defined period of time, regardless of the location's address, the memory unit is known as Random Access Memory (RAM). The system's bottleneck is its memory cycle time.

Using a cache memory is one approach to minimise memory access time. Cache memory is a tiny, quick memory that sits between the CPU and the bigger, slower main memory. Virtual memory is used to make the actual memory appear larger. Data is addressed in a virtual address space that is as big as the processor's addressing capacity. However, only the active fraction of this space is mapped onto physical memory locations at any one time. The remaining virtual addresses are mapped to the bulk storage devices (such as magnetic drives) that are utilised.

4.2.1 CPU-Main Memory Connection – A Block Schematic

The Main Memory (MM) unit can be thought of as a "block box" in terms of the system. The MAR (Memory Address Register) and MDR (Memory Data Register) are two CPU registers that transmit data between the CPU and the MM (Memory Data Register). If MAR is K bits long and MDR is 'n' bits long, the MM unit can hold up to 2k addressable locations, each of which is 'n' bits wide and has a word length of 'n' bits. n bits of data may be exchanged between the MM and the CPU during a "memory cycle."

The CPU loads the address into MAR, sets READ to 1, and sets additional control signals as needed for a read operation. MDR receives the data from the MM and MFC is set to 1. For a write operation, the CPU loads MAR and MDR appropriately, sets write to 1, and sets the other control signals appropriately. The data is loaded into suitable places by the MM control circuitry, which also sets MFC to 1. The following block diagram depicts this arrangement.

Connection of the Memory to the Processor

Some basic concepts related to memory system are as follows:

- Memory access time: It is a valuable indicator of the memory unit's speed. It is the amount of time that passes between the start of an operation and its completion (for example, the time that passes between READ and MFC).
- Memory cycle time: It's a crucial metric for the memory system. It is the shortest time interval between two consecutive memory operations (for example, the time between two consecutive READ operations). In most cases, the cycle time is slightly longer than the access time.
- Transfer rate: The pace at which data may be moved in and out of a memory unit.

4.2.2 Units of Memory

A computer stores data internally in the form of binary numbers, 0 (OFF/low voltage) and 1 (ON/high voltage). All the digits (0–9), alphabet (a to z or A to Z), special characters ($, %, @, * etc.) are stored in the computer in the binary form. ASCII (American Standard Code for Information Interchange) assigned a unique number or code to all the alphabet and special characters. Later on, this number/code could be converted into the binary form to store the corresponding letter/character in the computer.

In a computer, characters are represented by a group of bits that depends upon the encoding scheme being used. The encoding scheme is the manner of specifying the binary code for each character. There are two types of encoding scheme: ASCII and Unicode. ASCII represents a character as a group of 8 bits or 1 byte. Unicode represents a character as a group of 16 bits or 2 bytes. For example, if you want to store the word "bottle" in the computer memory, then as per the ASCII scheme, it is 6 bytes (6 characters of 1

byte each); and as per the Unicode scheme, the same word is stored as 12 bytes (6 characters of 2 bytes each). Apart from the byte, there are various other units of memory. The following is the description of the units of memory in a computer:

- Byte (B) = 8 bits
- Kilobyte (KB) = 1,024 bytes
- Megabyte (MB) = 1,048,576 bytes = 1024 KB
- Gigabyte (GB) = 1,024 megabytes
- Terabyte (TB) = 1,024 gigabytes
- Petabyte (PB) = 1,024 terabytes
- Exabyte (EB) = 1,024 petabytes
- Zettabyte (ZB) = 1,024 exabytes
- Yottabyte (YB) = 1,024 zettabytes
- Brontobyte = 1,024 yottabytes
- Geopbyte = 1,024 brontobytes

4.3 SEMICONDUCTOR RAM MEMORIES

Any electronics assembly that employs computer processing technologies uses semiconductor memory. Semiconductor memory is a critical electronic component for any PCB assembly based on a computer.

Memory cards have also become widespread for temporarily storing data, ranging from portable flash memory cards used for file transfers to semiconductor memory cards used in cameras, mobile phones, and other devices. As the need for greater and larger quantities of storage has grown, the use of semiconductor memory has expanded, as has the size of these memory cards.

Many different types and methods are employed to fulfil the rising need for semiconductor memory. New memory technologies are being launched as demand rises, and old types and technologies are being improved.

There is a multitude of memory technologies available, each with its own set of advantages and disadvantages. There are many different forms of memory, including ROM, RAM, EPROM, EEPROM, Flash memory, DRAM, SRAM, SDRAM, F-RAM, and MRAM, and new varieties are being created to increase performance. DDR3, DDR4, DDR5, and a variety of other terms are used to describe different types of SDRAM semiconductor memory.

Furthermore, semiconductor devices are accessible in a variety of formats, including integrated circuits for printed circuit boards, USB memory cards, Compact Flash cards, SD memory cards, and even solid state hard drives. Semiconductor memory is even used as on-board memory in several microprocessor processors.

4.3.1 Types of Semiconductor Memory

There are two primary sorts or categories of semiconductor technology that can be employed. These memory kinds or categories distinguish the memory in terms of how it works:

- Random Access Memory (RAM): RAM is a type of memory that is used to store data. It is a semiconductor-based memory where the CPU or the other hardware devices can read and write

data. In this memory, the data is temporarily stored since it is a volatile memory. Once the system turns off, it loses the data. As a result, RAM is used as a temporary data storage area. This form of memory is used to store and read data multiple times.

- Read Only Memory (ROM): A ROM is a type of semiconductor memory in which data is written once and then is not altered. As a result, it is employed in situations where data must be kept permanently, even when the power is turned off (many memory technologies lose data when the power is turned off).

4.3.2 Types of RAM

The following are the different types of RAM:

- Static Random Access Memory (SRAM): It is a type of memory that stores data in a fixed location. The data in this type of semiconductor memory does not need to be updated dynamically, unlike DRAM.
- Dynamic Random Access Memory (DRAM): It is a kind of random access memory. Each bit of data in DRAM is stored on a capacitor, and the charge level on each capacitor determines whether the bit is a logical 1 or 0. However, because these capacitors can not keep their charge permanently, the data must be updated on a regular basis. It gets its moniker as a dynamic RAM as a consequence of this dynamic updating. DRAM is a kind of semiconductor memory that is often found in equipment such as personal computers and workstations, where it serves as the computer's primary RAM. Semiconductors are often supplied as integrated circuits in the form of surface mount devices or, less commonly, as leaded components for use in PCB assembly.
- Synchronous DRAM (SDRAM): It is a kind of DRAM that is synchronised. This type of semiconductor memory can operate at higher rates than standard DRAM. It is synchronised with the processor's clock and can maintain two sets of memory addresses open at the same time.
- Ferroelectric Random Access Memory (F-RAM) is a random access memory technology that is quite comparable to regular DRAM. The main distinction is that it has a ferroelectric layer rather than the more common dielectric layer, which gives it its non-volatile properties. F-RAM is a direct rival to Flash since it has a non-volatile capability.
- Double Data Rate SDRAM (DDR SDRAM): Refers to the memory that transfers data at high speed. The clock speed of DDR SDRAM ranges from 133 MHz to 2133 MHz.
- Rambus DRAM (RDRAM): It is the fastest among all the random memory types with the data transfer speed of 1 GHz. Generally, RDRAM is used for the purpose of video memory on graphics accelerator cards. Dynamic RDRAM is an improvement to the existing RDRAM. The RDRAM chip provides high bandwidth and therefore used by workstations and servers. This memory chip places under the RIMM (Rambus Inline Memory Module) module. In addition, the number of chips placed under the module completely relies on the bus width of the RAM. RDRAM (RAM Bus DRAM) of 160 or 184 Pins operates at 300-400 MHz.
- Magnetic RAM (MRAM): It is a kind of magneto-resistive RAM. It's a non-volatile RAM memory that stores data using magnetic charges rather than electric charges.
- Phase Change Random Access Memory (P-RAM) or Phase Change Memory (PCM): It is based on a phenomena in which a chalcogenide glass changes state or phase from amorphous (high resistance) to polycrystalline (low resistance) (low resistance). It is feasible to detect the status of a single cell and so use this information to store data.

4.4 READ ONLY MEMORY (ROM)

A read-only memory, or ROM, is a form of semiconductor memory in which data is written once and then not changed. As a result, it's used in circumstances when data has to be saved indefinitely, even after the power is switched off (many memory technologies lose data when the power is turned off).

As a result, this form of semiconductor memory technology is extensively employed to store programmes and data that must survive the power down of a computer or processor. The BIOS of a computer, for example, is stored in ROM. Data cannot be easily written to ROM, as the name indicates. Writing data into the ROM may require specific hardware at first, depending on the ROM's technology. Although data may frequently be changed, this benefit necessitates the use of specific technology to delete the data so that new data can be written in.

4.4.1 Types of Read-Only Memory (ROM)

Following are the different types of ROM:

- Programmable Read Only Memory (PROM): It is a type of memory that can be programmed. It's a type of semiconductor memory that can only have data written to it once before it becomes permanent. These memories are purchased in a blank state and are programmed with a PROM programmer.
- Erasable Programmable Read Only Memory (EPROM): It is a type of memory that can be erased and reprogrammed. These semiconductor devices have the ability to be programmed and then deleted at a later date. Normally, this is accomplished by exposing the semiconductor device to UV radiation. To make this possible, the EPROM packaging has a circular window that allows light to pass through to the device's silicon. This window is usually covered with a label when the PROM is in use, especially if the data has to be maintained for a long time.

  Flash memory may be thought of as an advancement of EEPROM technology. Data may be written to it and deleted from it in blocks, but data can only be read from individual cells.
- Electrically Erasable Programmable Read-On Memory (EEPROM): It is a kind of electrically erasable programmable read-only memory. Data may be written and deleted on these semiconductor devices using an electrical voltage. This is usually applied to a chip's erase pin. EEPROM, like other forms of PROM, keeps the memory's contents even after the power is switched off. EEPROM, like other kinds of ROM, is slower than RAM.

4.5 MEMORY INTERLEAVING

This method splits the memory system into a number of memory modules and organizes the addressing so that consecutive words in the address space are assigned to distinct modules. Memory access requests involving successive addresses will be sent to separate modules.

The average pace of obtaining words from the Main Memory can be improved since parallel access to these modules is available.

Memory cells are commonly arranged in an array, with each cell capable of storing one bit of information. Each row of cells represents a memory word, and each row's cells are linked together by a common line known as the word line.

Two bit lines link the cells in each column to the Sense/Write circuit. The Sense/Write circuits are linked to the chip's data input or output lines. The sense / write circuit receives input information and stores it in the cells of the chosen word during a write operation.

4.6 SPEED, SIZE, AND COST OF MEMORY

Figure 2 shows the speed, size and cost of memory:

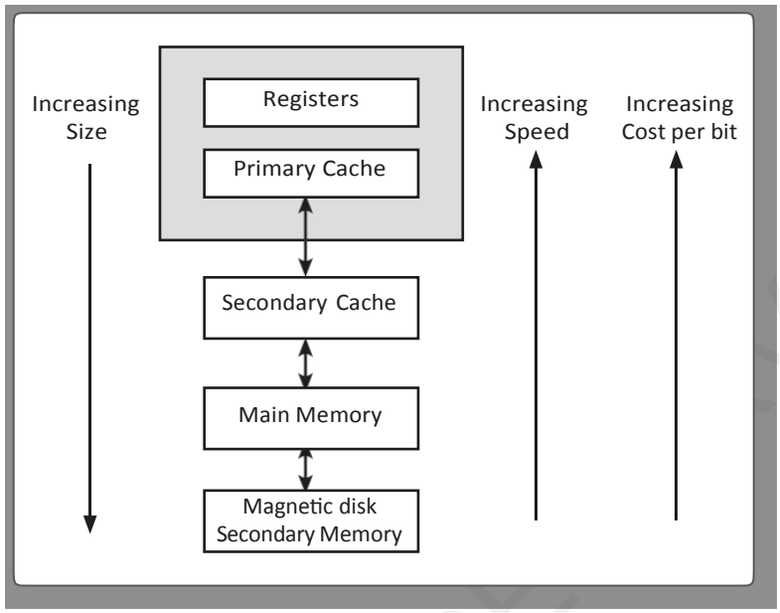

Figure 2: Speed, Size, and Cost of Memory

4.6.1 Speed

A perfect memory would be quick, large, and cheap. SRAM chips can be used to create a very fast memory. However, these chips are costly because their basic cells have six transistors, which prevents a large number of cells from being packed onto a single chip. As a result, using SRAM chips to build a large memory is impractical due to cost. Dynamic RAM chips, which have much simpler basic cells and are thus much less expensive, are an alternative. However, such memories are much slower.

4.6.2 Size

A very small amount of memory can be implemented directly on the processor chip at the next level of the hierarchy. This memory, known as a processor cache, stores copies of instructions and data that are stored in a much larger external memory. Caches are frequently divided into two tiers. On the CPU chip, there is always a primary cache. Because it competes for space on the processor chip, which must also implement many other functions, this cache is tiny. The primary cache is known as the level 1 (L1) cache. Between the main cache and the rest of the memory lies a bigger secondary cache.

The main memory is the next level in the hierarchy. Dynamic memory components, including as SIMMs, DIMMs, and RIMMs, are used to create this fairly big memory. The main memory is substantially larger than the cache memory, but it is a lot slower. The access time to the main memory in a typical computer is about ten times longer than the access time to the L 1 cache.

Disk drives can store a lot of data at a low price. When compared to the semiconductor devices used to implement the main memory, they are extremely sluggish. A hard disc drive (HDD; sometimes known as a hard drive, hard disc, magnetic disc, or disc drive) is a storage and retrieval device for digital data, particularly computer data. It is made up of one or more rigid (thus “hard”) fast spinning discs (commonly referred to as platters) that are covered with magnetic material and have magnetic heads

positioned to write and read data from the surfaces. The speed with which a software may access memory is critical during execution.

4.6.3 Cost

Although dynamic memory units in the hundreds of megabytes range can be implemented for a reasonable price, the size is still small when compared to the demands of large programmes with large amounts of data. To implement large memory spaces, a solution is provided by using secondary storage, primarily magnetic discs. Large discs can be purchased for a reasonable price and are widely used in computer systems. They are, however, significantly slower than semiconductor memory units. So Magnetic discs can provide a huge amount of cost-effective storage. Dynamic RAM technology can be used to create a large, yet affordable main memory.

4.7 CONCLUSION

- One of the most crucial components of a computer system is its memory.
- The electrical storing area for instructions and data that the processor can read rapidly is referred to as computer memory.
- Computer memory is divided into two categories – primary and secondary. Primary memory is directly accessed by a processor to execute instructions.
- The storage capacity of computer memory is measured in bits, bytes, Kilobyte (KB), Megabyte (MB), Gigabyte (GB), Terabyte (TB), and Petabyte (PB).
- The addressing method determines the maximum size of memory that may be used in every machine.
- Cache memory is a tiny, quick memory that sits between the CPU and the bigger, slower main memory.
- The Main Memory (MM) unit can be thought of as a "block box" in terms of the system.
- The MAR (Memory Address Register) and MDR (Memory Data Register) are two CPU registers that transmit data between the CPU and the MM (Memory Data Register).
- Memory access time is a valuable indicator of the memory unit's speed.
- Memory cycle time is a crucial metric for the memory system.
- The pace at which data may be moved in and out of a memory unit is called transfer rate.
- ASCII (American Standard Code for Information Interchange) assigned a unique number or code to all the alphabet and special characters.
- In a computer, characters are represented by a group of bits that depends upon the encoding scheme being used.
- There are two types of encoding scheme: ASCII and Unicode.
- ASCII represents a character as a group of 8 bits or 1 byte. Unicode represents a character as a group of 16 bits or 2 bytes.
- Any electronics assembly that employs computer processing technologies uses semiconductor memory.
- Semiconductor memory is a critical electronic component for any PCB assembly based on a computer.

- There are many different forms of memory, including ROM, RAM, EPROM, EEPROM, Flash memory, DRAM, SRAM, SDRAM, F-RAM, and MRAM, and new varieties are being created to increase performance.
- RAM is a semiconductor-based memory where the CPU or the other hardware devices can read and write data. In this memory, the data is temporarily stored since it is a volatile memory. Once the system turns off, it loses the data.
- A ROM is a type of semiconductor memory in which data is written once and then is not altered.
- This method splits the memory system into a number of memory modules and organizes the addressing so that consecutive words in the address space are assigned to distinct modules.
- Memory cells are commonly arranged in an array, with each cell capable of storing one bit of information.

4.8 GLOSSARY

- ROM: It is a form of semiconductor memory in which data is written once and then not changed.
- EEPROM: It is a kind of electrically erasable programmable read-only memory.
- RAM: RAM is a semiconductor-based memory where the CPU or the other hardware devices can read and write data. In this memory, the data is temporarily stored since it is a volatile memory. Once the system turns off, it loses the data.
- Computer memory: The electrical storing area for instructions and data that the processor can read rapidly.
- Cache memory: It is a tiny, quick memory that sits between the CPU and the bigger, slower main memory.
- Semiconductor memory: It is a critical electronic component for any PCB assembly based on a computer.

4.9 SELF-ASSESSMENT QUESTIONS

A. Essay Type Questions

1. One of the most crucial components of a computer system is its memory. Discuss.
2. The storage capacity of computer memory is measured in bits, bytes, Kilobyte (KB), etc. What do you understand by units of memory?
3. Any electronics assembly that employs computer processing technologies uses semiconductor memory. Discuss.
4. RAM is categorised in different types. Discuss the RDRAM.
5. ROM is categorised in different types. Discuss one of the categories of ROM, namely, EPROM.

4.10 ANSWERS AND HINTS FOR SELF-ASSESSMENT QUESTIONS

A. Hints for Essay Type Questions

1. Memory saves the data and instructions needed for data processing and output outcomes. Storage may be necessary for a short amount of time, immediately, or for a long time. The electrical storing area for instructions and data that the processor can read rapidly is referred to as computer memory.

   Refer to Section Introduction

2. A computer stores data internally in the form of binary numbers, 0 (OFF/low voltage) and 1 (ON/high voltage). All the digits (0–9), alphabet (a to z or A to Z), special characters ($, %, @, * etc.) are stored in the computer in the binary form. ASCII (American Standard Code for Information Interchange) assigned a unique number or code to all the alphabet and special characters.

   Refer to Section Basic Concepts of Memory System

3. Semiconductor memory is a critical electronic component for any PCB assembly based on a computer. Memory cards have also become widespread for temporarily storing data, ranging from portable flash memory cards used for file transfers to semiconductor memory cards used in cameras, mobile phones, and other devices.

   Refer to Section Semiconductor RAM Memories

4. RDRAM is the fastest among all the random memory types with the data transfer speed of 1 GHz. Generally, RDRAM is used for the purpose of video memory on graphics accelerator cards. Dynamic RDRAM is an improvement to the existing RDRAM.

   Refer to Section Semiconductor RAM Memories

5. EPROM is a type of memory that can be erased and reprogrammed. These semiconductor devices have the ability to be programmed and then deleted at a later date. Normally, this is accomplished by exposing the semiconductor device to UV radiation.

4.11 POST-UNIT READING MATERIAL

- https://www.studocu.com/in/document/psg-college-of-technology/computer-architecture/m-morris-mano-solution-manual-computer-system-architecture/10775236
- https://www.jbiet.edu.in/coursefiles/cse/HO/cse2/DLD1.pdf

4.12 TOPICS FOR DISCUSSION FORUMS

- Do the online research and discuss about real application of memory system with your classmates.