CIS11032 Logic Designing & Computer Organization

Lesson 07 Computer Memory Systems

AR Fathima Shafana
Department of ICT
Faculty of Technology
South Eastern University of Sri Lanka

Lesson Learning Outcomes

At the completion of this lesson students should be able to,

- Identify the Key Characteristics of Computer System
- Understand the need of Memory Hierarchy

Understand how Replacement Algorithms work

COURSE OUTLINE

- Computer Memory System
- Key Characteristics
- Memory Hierarchy
- **Replacement Algorithms**

Overview

- Computer memory exhibits wide range of type, technology, organization, performance, and cost of any feature of a computer system.
- The typical computer system is equipped with a hierarchy of memory subsystems, some internal to the system and some being external

Key Characteristics of Computer Memory System

Location
- Internal (e.g., processor registers, cache, main memory)
- External (e.g., optical disks, magnetic disks, tapes)

Capacity
- Number of words
- Number of bytes

Unit of Transfer
- Word
- Block

Access Method
- Sequential
- Direct
- Random
- Associative

Performance
- Access time
- Cycle time
- Transfer rate

Physical Type
- Semiconductor
- Magnetic
- Optical
- Magneto-optical

Physical Characteristics
- Volatile/nonvolatile
- Erasable/nonerasable

Organization
- Memory modules

01. Location

- Refers to whether memory is **internal or external** to the computer
- **Internal memory** is often equated with main memory.
- However, it includes, registers and cache as well.
- **External memory** consists of peripheral storage devices, such as disk and tape, that are accessible to the processor via I/O controllers.

02. Capacity

- It is an obvious characteristic of memory
- Typically expressed in terms of bytes (1 byte = 8 bits) or words for Internal Memory
- Common word lengths are 8, 16, and 32 bits.
- External memory is generally expressed in Bytes

03. Unit of Transfer

- For internal memory, the unit of transfer is equal to the number of electrical lines into and out of the memory module.
- This may be equal to the word length, but is often larger, such as 64, 128, or 256 bits.
- **Word, Addressable Units, Unit of Transfer**.
    - Are they all same?

03. Unit of Transfer contd.

WORD

- The “natural” unit of organization of memory.
- The size of a word is typically equal to the number of bits used to represent an integer and to the instruction length.
- However, this may vary.

ADDRESSABLE UNITS

- Many systems allow addressing at the byte level, but generally it is the word.

$$\boxed{2^A = N.}$$

A=The length in bits
N=Number of Addressable Units

03. Unit of Transfer contd.

UNIT OF TRANSFER

- The number of bits read out of or written into memory at a time.
- The unit of transfer need not equal a word or an addressable unit.
- For external memory, data are often transferred in much larger units than a word, and these are referred to as **blocks.**

04. Access Method

I. **SEQUENTIAL ACCESS**

- Memory is organized into units of data, called records.
- Access must be made in a specific linear sequence.
- Stored addressing information is used to separate records and assist in the retrieval process.
- A shared read–write mechanism is used, and this must be moved from its current location to the desired location, passing and rejecting each intermediate record.
- Thus, the time to access an arbitrary record is highly variable.
- **EXAMPLE:** *Tape Units*

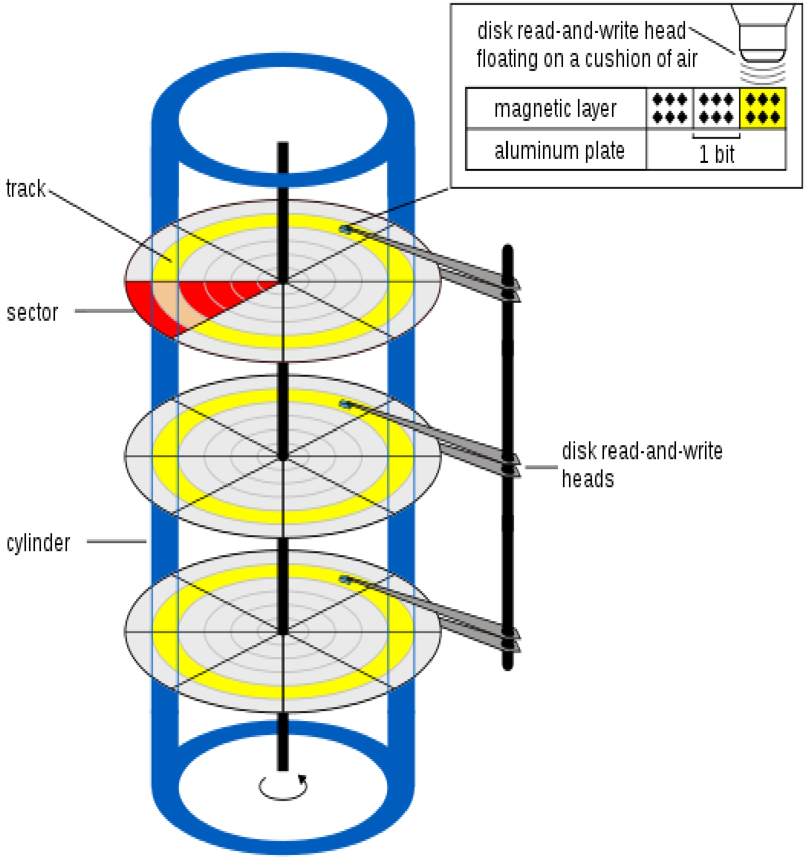
04. Access Method

II. **DIRECT ACCESS**

- Direct access involves a shared read–write mechanism.
- However, individual blocks or records have a unique address based on physical location.
- Access is accomplished by direct access to reach a general vicinity plus sequential searching, counting, or waiting to reach the final location.
- EXAMPLE: Disk units

How data is read from disks?

- Disk is a combination of random-access and sequential-access memory.
- Data is stored on a rotating disk or drum.
- The disk (drum) is organized into tracks onto data is sequentially stored.
- The read/write heads are then positioned (random-access) over the appropriate track and the data is then read/written (sequentially) on the appropriate place on the rotating disk.



04. Access Method

III. RANDOM ACCESS

- Each addressable location in memory has a unique, physically wired-in addressing mechanism.
- The time to access a given location is independent of the sequence of prior accesses and is constant.
- Thus, any location can be selected at random and directly addressed and accessed.
- EXAMPLE: Main memory and some cache systems

04. Access Method

IV. ASSOCIATIVE ACCESS

- This is a random access type of memory that enables one to make a comparison of desired bit locations within a word for a specified match.
- Thus, a word is retrieved based on a **portion of its contents** rather than its address.
- Also known as *Content Addressable memory*.
- Similar to ordinary random-access memory, each location has its own addressing mechanism, and retrieval time is constant independent of location or prior access patterns. It is a special type of random-access method.

05. Performance

ACCESS TIME OR LATENCY

For random-access memory,

The time it takes to perform a read or write operation, i.e. the time from the instant that an address is presented to the memory to the instant that data have been stored or made available for use.

For non-random-access memory,

The time it takes to position the read–write mechanism at the desired location.

05. Performance

MEMORY CYCLE TIME

- Refers to the access time plus any additional time required before a second access can initiate.
- This additional time may be required for transients to die out on signal lines or to regenerate data if they are read destructively.

05. Performance

TRANSFER RATE

- The rate at which data can be transferred into or out of a memory unit.

Bus Width: Number of data lines (e.g., 64 bits for DDR RAM)
Clock Frequency: Speed of memory clock (in Hertz or MHz)
Transfers per Clock:
SDR = 1
DDR = 2
DDR5 = 2 (but with dual channels per module, doubling effective bandwidth)
Divide by 8: To convert bits to bytes

$$\text{Transfer Rate (Bytes/sec)} = \frac{\text{Bus Width (bits)} \times \text{Clock Frequency (Hz)} \times \text{Transfers per Clock}}{8}$$

06. Physical Types

- Different type of materials are used where the most common today are

  1. Semiconductor-Based Memory
  Silicon-based integrated circuits
  Examples: SRAM, DRAM

  2. Magnetic-Based Memory
  Store data via magnetized domains
  Examples: Hard Disk Drive, Magnetic Tapes

  3. Optical-Based Memory
  Stored using light and lasers
  Examples: CD, DVD, Blu-ray

07. Physical Characteristics

VOLATILE MEMORY

Information decays naturally or is lost when electrical power is switched off.

Eg: RAM

NON-VOLATILE MEMORY

Information once recorded remains without deterioration until deliberately changed.

However, no electrical power is needed to retain information

Eg: Magnetic-surface memories

07. Physical Characteristics contd.

NON-ERASABLE MEMORY

Non-erasable memory cannot be altered, except by destroying the storage unit.

Semiconductor memory of this type is known as *read-only memory* (ROM).

Data is usually fixed in the words of a ROM either by the manufacturer or by some off-line manner.

A practical non-erasable memory must also be non volatile.

08. Organization

- *Organization* refers to the physical arrangement of bits to form words

MEMORY HIERARCHY

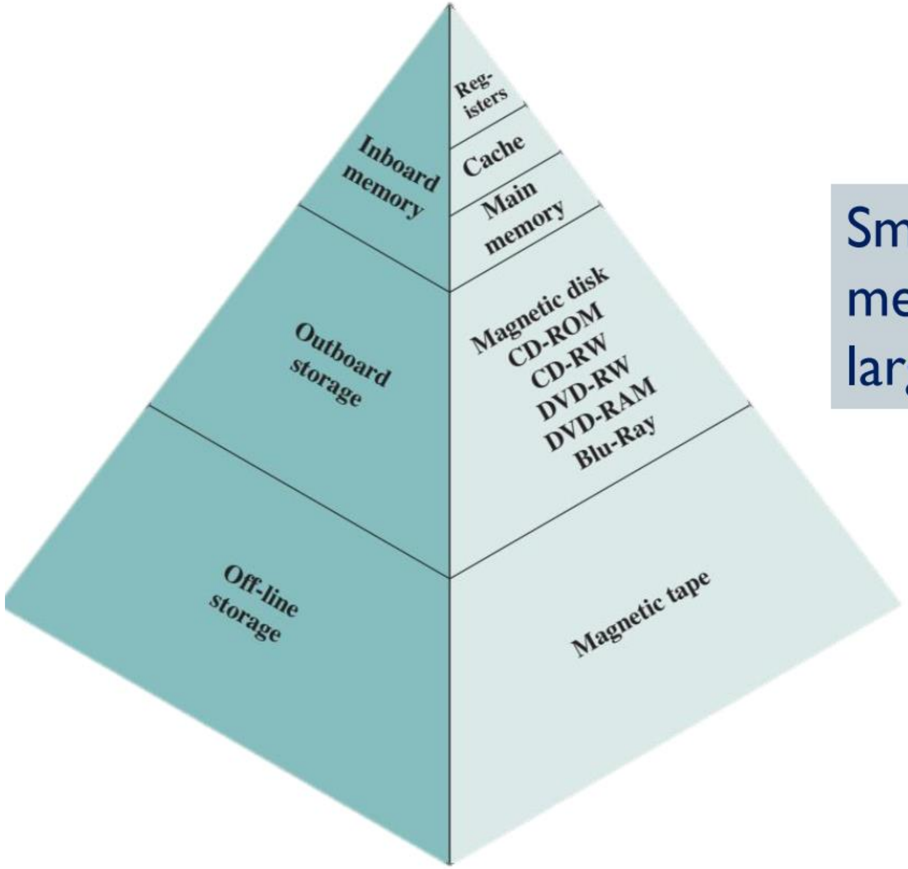
- How much?
- How fast?
- How expensive?
- There is a trade-off among the three key characteristics of memory: **capacity, access time, and cost**

Faster access time, greater cost per bit
Greater capacity, smaller cost per bit
Greater capacity, slower access time

MEMORY HIERARCHY contd.

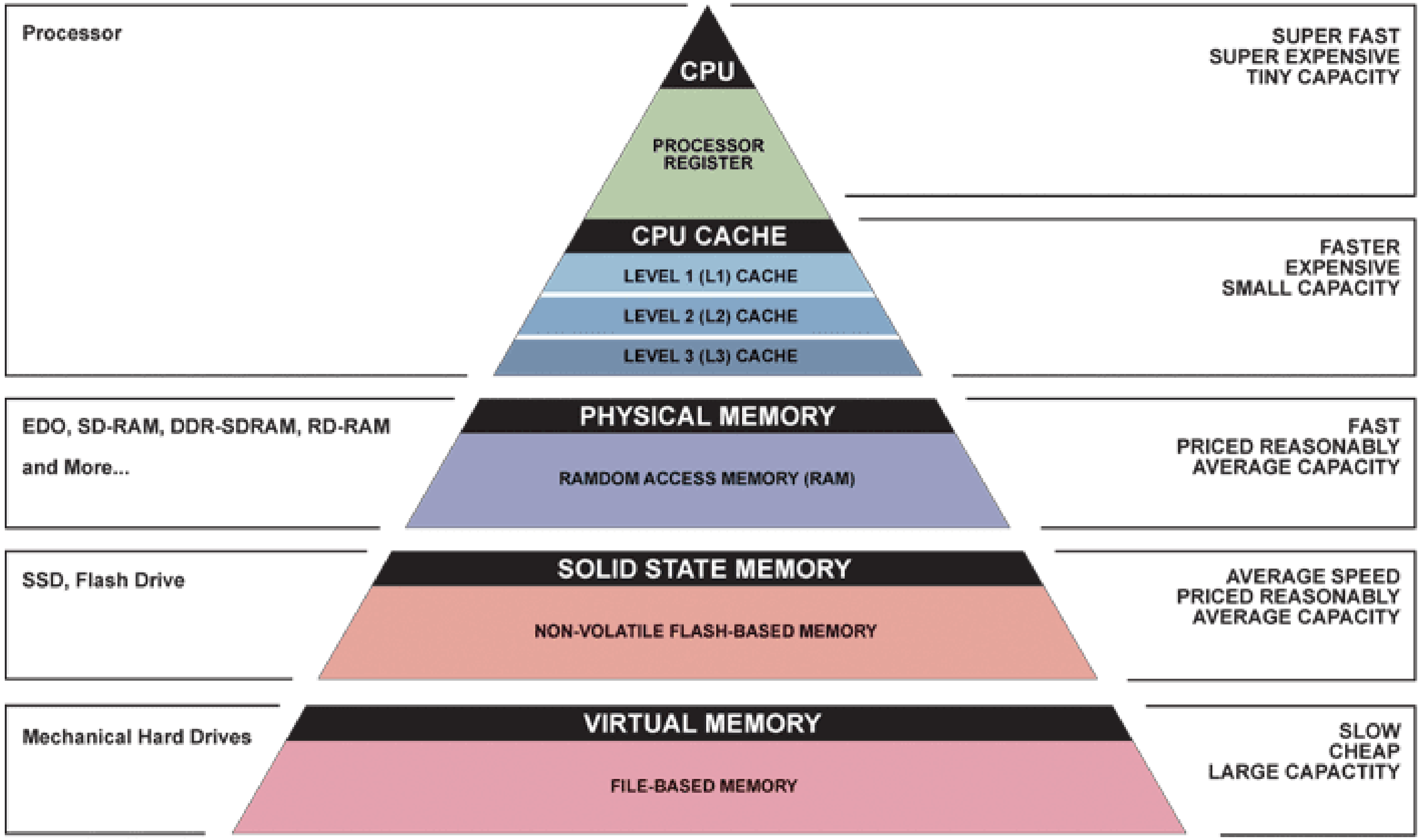
- Therefore, rather than relying on single component or single technology, **<u>Memory Hierarchy</u>** is used instead. As one goes down the hierarchy, the following would occur:
  a) Decreasing cost per bit
  b) Increasing capacity
  c) Increasing access time
  d) Decreasing frequency of access of the memory by the processor

MEMORY HIERARCHY contd.



Smaller, more expensive, faster memories are supplemented by larger, cheaper, slower memories.

MEMORY HIERARCHY contd.



▲ Simplified Computer Memory Hierarchy
Illustration: Ryan J. Leng

Replacement Algorithms

1. Least Recently Used (LRU)
2. First In First Out (FIFO)
3. Least Frequently Used (LFU)

Thank you