COMPUTER ORGANIZATION AND ARCHITECTURE (COA)

EET 2211

4TH SEMESTER – CSE & CSIT

CHAPTER 4, LECTURE 15

CHAPTER 4 – CACHE MEMORY

TOPICS TO BE COVERED

- Computer Memory System Overview
- Cache Memory Principles
- Elements of Cache Design
- Cache Organization

LEARNING OBJECTIVES

After studying this chapter, you should be able to:

- Present an overview of the main characteristics of computer memory systems and the use of a memory hierarchy.
- Describe the basic concepts and intent of cache memory.
- Discuss the key elements of cache design.
- Distinguish among direct mapping, associative mapping and set-associative mapping.
- Explain the reasons for using multiple levels of cache.
- Understand the performance implications of multiple levels of memory.

INTRODUCTION

- Computer memory exhibits the widest range of type, technology, organization, performance and cost of any feature of computer system.

- Till now the technologies are not optimal in satisfying the memory requirements of a computer system, and so the typical computer system is equipped with a hierarchy of memory sub-systems.

- This chapter primarily focuses on internal memory elements.

4.1. COMPUTER MEMORY SYSTEM OVERVIEW

Characteristics of Memory Systems

The complex subject of computer memory is made more manageable if we classify memory sub-systems according to their key characteristics. The most important of these are listed below.

- Location
- Capacity
- Unit of transfer
- Access method
- Performance
- Physical type
- Physical characteristics
- Organisation

Location

- The term location refers to whether memory is internal or external to the computer.
- Internal memory is often equated with main memory but there are other forms of internal memory also.
- The processor and also the control unit requires its own local memory in the form of registers.
- Cache is a type of internal memory.
- external memory consists of peripheral storage devices, such as disk and tape that are accessible to the processor via I/O controllers.

Capacity

- For internal memory it is expressed in terms of Bytes (1 byte = 8 bits) or Words.
- Common word lengths are 8, 16, and 32 bits.
- External memory capacity is typically expressed in terms of Bytes.

Unit of Transfer

- Internal Memory
  - Usually governed by data bus width
- Word
  - The natural unit of organisation
- Addressable unit
  - Smallest location which can be uniquely addressed
  - Word internally
  - Addressable units is $2^A=N$.

  A is length in bits in an address, N is addressable units.
- Unit of transfer
  - Usually blocks which is much larger units than a word

Access Methods

- Sequential
  - Start at the beginning and read through in order
  - Access time depends on location of data and previous location
  - e.g. tape
- Direct
  - Individual blocks have unique address
  - Access is by jumping to vicinity plus sequential search
  - Access time depends on location and previous location
  - e.g. disk

Cont.

- Random
  - Individual addresses identify locations exactly
  - Access time is independent of location or previous access
  - e.g. RAM
- Associative
  - Data is located by a comparison with contents of a portion of the store
  - Access time is independent of location or previous access
  - e.g. cache

Performance

- Access time
  - Time between presenting the address and getting the valid data
- Memory Cycle time
  - Time may be required for the memory to "recover" before next access
  - Cycle time is access + recovery
- Transfer Rate
  - Rate at which data can be moved

Physical Types

- Semiconductor
  - RAM
- Magnetic
  - Disk & Tape
- Optical
  - CD & DVD
- Others
  - Bubble
  - Hologram

Physical Characteristics

- Volatility

  Information lost when power is switch off.

- Non-volatility

  Information not lost when power is switch off.

  e.g. magnetics-surface memory.

- Non-Erasable

  e.g. ROM.

Organisation

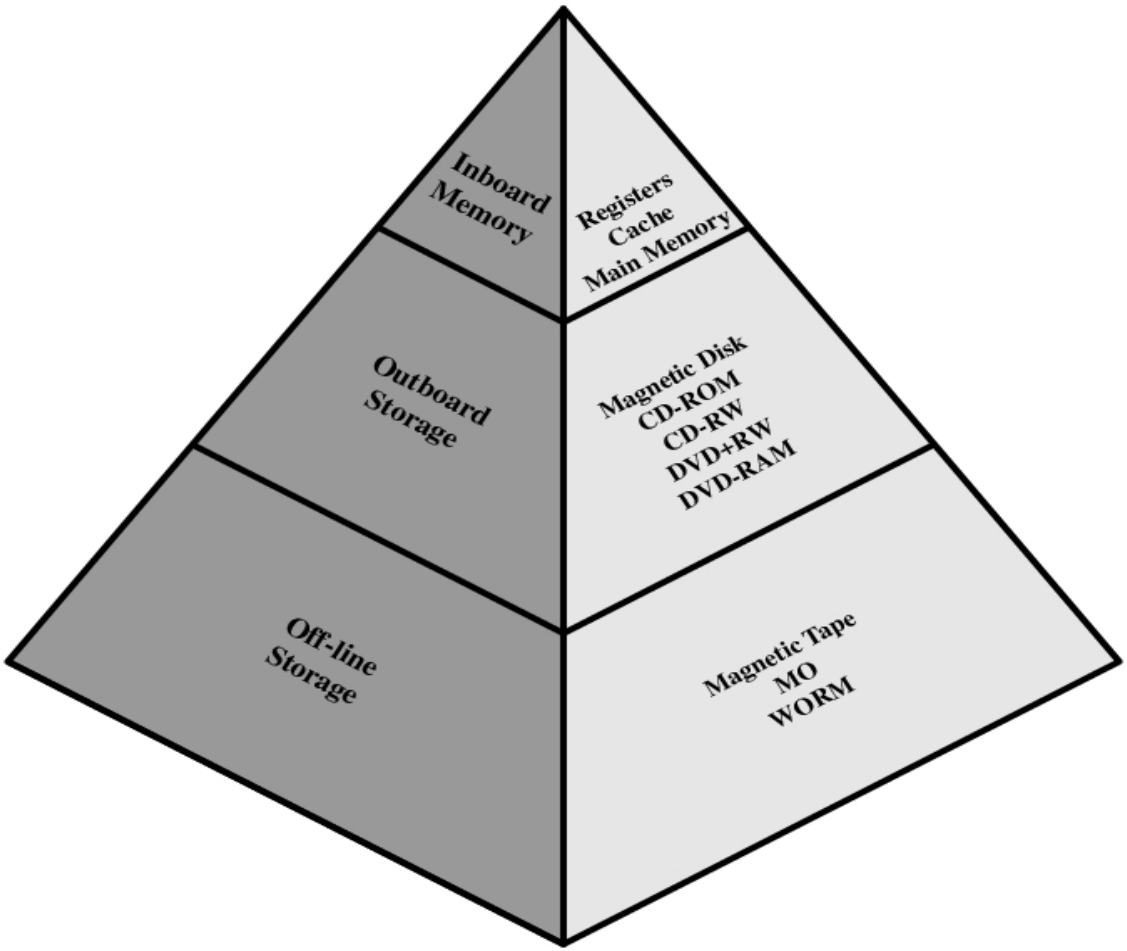
- Physical arrangement of bits into words
- Not always obvious used
- e.g. interleaved

The Memory Hierarchy

- How much?
  - Capacity
- How fast?
  - Time
- How expensive?
  - Money

NB

Memory Hierarchy - Diagram



- As one goes down the hierarchy:
  1. Decreasing cost per bit
  2. Increasing capacity
  3. Increasing access time
  4. Decreasing frequency of access of the memory by the processor.

Memory Hierarchy

- Registers
  - In CPU
- Internal or Main memory
  - May include one or more levels of cache
  - “RAM”
- External memory
  - Backing store

Hierarchy List

- Registers
- L1 Cache
- L2 Cache
- Main memory
- Disk cache
- Disk
- Optical
- Tape

So you want fast?

- It is possible to build a computer which uses only static RAM
- This would be very fast
- This would need no cache
  - How can you cache cache?
- This would cost a very large amount

THANK YOU