COA

CH1

Introduction

BY

DR. RAGHAD SAMIR AL NAJIM

2017@ SPU

Textbook

Computer Organization and Architecture: Designing for Performance, 8th Edition By William Stallings, Prentice Hall

M. Murdocca and V. Heuring, Computer Architecture and Organization, an integrated approach,Wiley, 2007.

Linda Null and Julia Lobur, The Essentials Of Computer Organization and Architecture, 2nd edition, **Jones and Bartlett Publishers** ,2006.

Contents

Computers

what are they?

How do they look like?

How it works?

?

?

?

From Outside

Processor
Memory
Hard disk
DVD ROM
Graphics card
……

From Inside

What We Are Going To Study?

- How these internal components look like?
  - Top-down approach with schematics
- How do they fit together?
- How to program them?
- How to improve their performance?

Computer Organization & Architecture - COA

Computer technology has made incredible improvement in the past half century . Today, a personal computer has more computational power, more main memory,more disk storage, smaller in size and it is available in effordable cost.

In this course we will mainly deal with the computer design.

The task that the computer designer handles is a complex one：

Determine what are important for a new machine, then design a machine to maximize performance while staying within cost constraints

This task has many aspects, including instruction set design, functional organization, logic design, and implementation.

While looking for the task for computer design, both the terms computer organization and computer architecture come into picture.

Computer architecture

Refers to those parameters of a computer system that are visible to a programmer or those parameters that have a direct impact on the logical execution of a program. Examples of architectural attributes include

- the instruction set ,
- the number of bits used to represent different data types,
- I/O mechanisms,
- and techniques for addressing memory.

Computer organization

Refers to the operational units and their interconnections that realize the architectural specifications. Examples of organizational attributes include those

- hardware details transparent to the programmer, such as control signals ,
- interfaces between the computer and peripherals ,
- the memory technology used

Computer architecture & organization

Computer Architecture = ISA + MO

- Instruction Set Architecture
- What the executable can “see” as underlying hardware
- Logical View

Machine Organization

- How the hardware implements ISA ?
- Physical View

- Architecture is those attributes visible to the Programmer
  - Instruction set, number of bits used for data representation, I/O mechanisms, addressing techniques.
    - e.g. Is there a multiply instruction?

- Organization is how features are implemented
  - Control signals, interfaces, memory technology.
    - e.g. Is there a hardware multiply unit or is it done by repeated addition?

- All Intel x86 family share the same basic architecture.
- The IBM System/370 family share the same basic architecture since 1970
- This gives code compatibility

- Organization differs between different versions

Computer architecture

There are two types of architecture: **Von Neumann Architecture & Harvard Architecture**

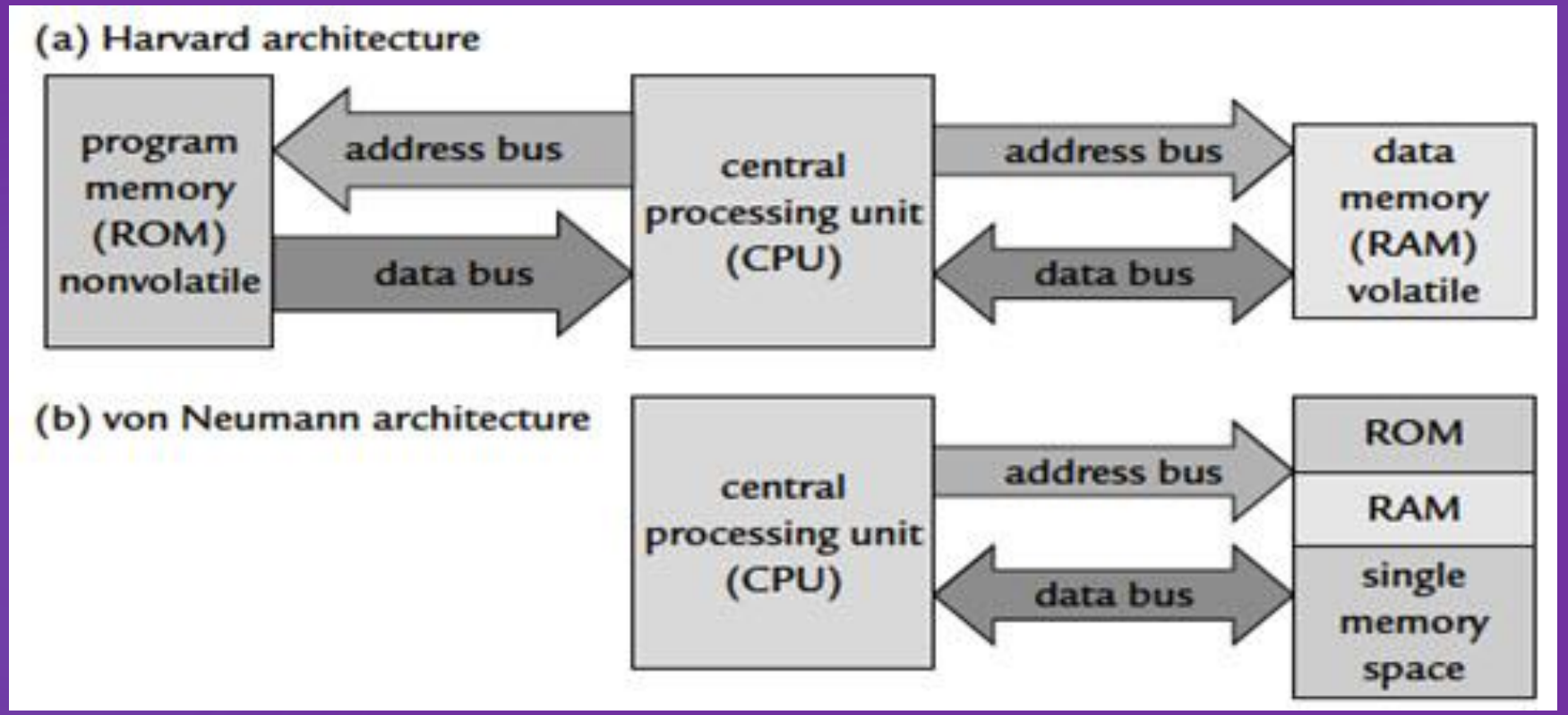
Von Neumann Architecture

1- It is named after the mathematician and early computer scientist John Von Neumann.
2- The computer has single storage system(memory) for storing data as well as program to be executed.
3- A single set of address/data buses between CPU and memory.

Harvard Architecture

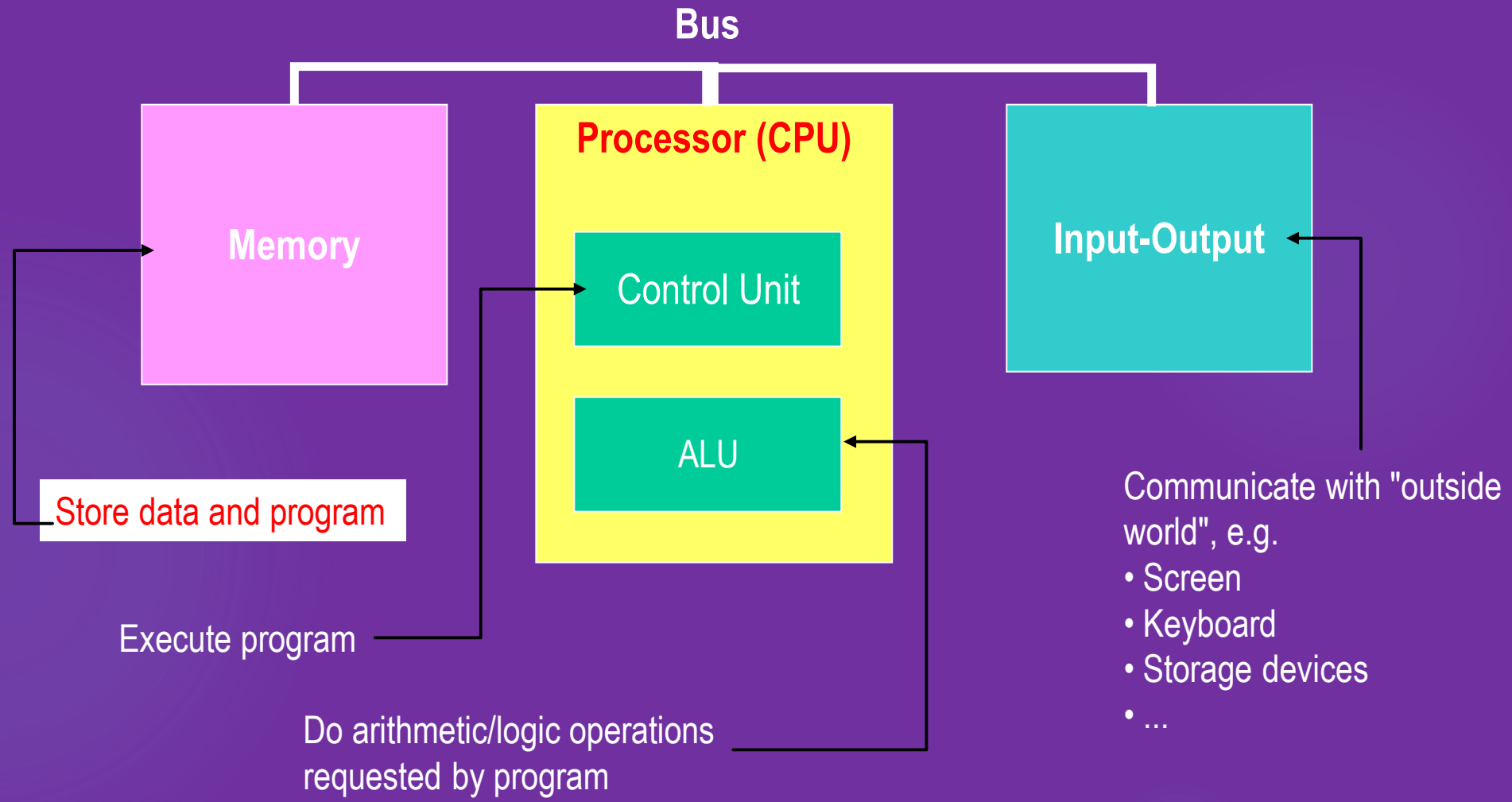
1-The name is originated from "Harvard Mark " a relay based old computer,which stored instruction on punched tape(24 bits wide) and data in electo-mechanical counters.
2- The computer has two separate memories for storing data and program.
3- Two sets of address/data buses between CPU and memory.

Von Neumann & Harvard Architecture

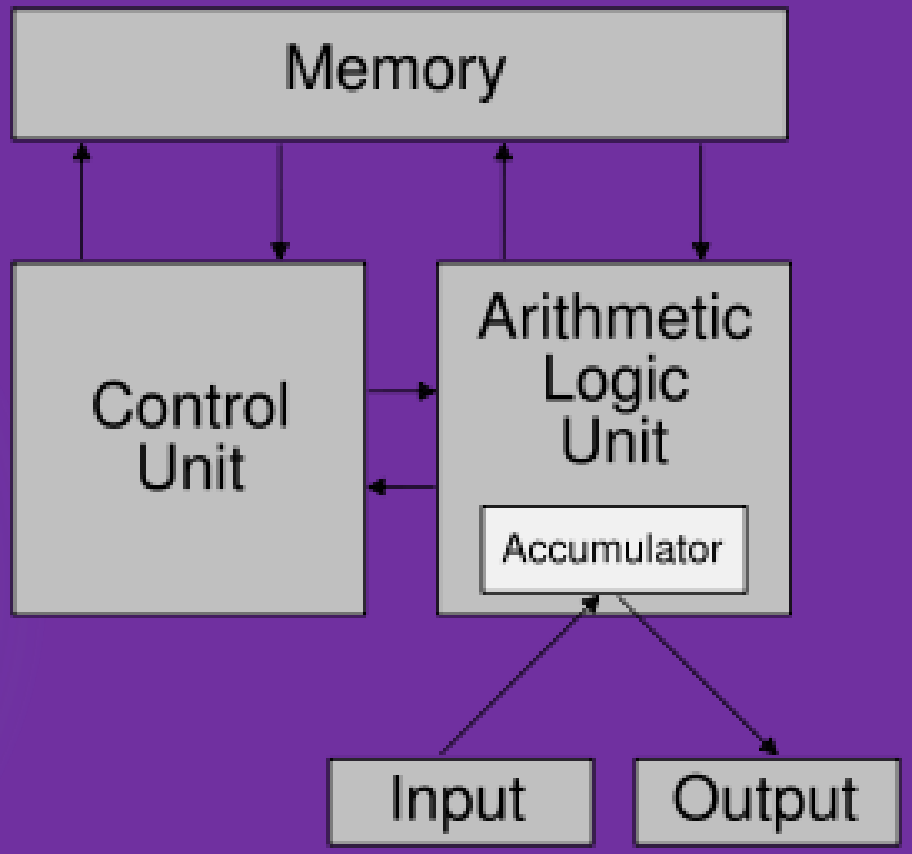


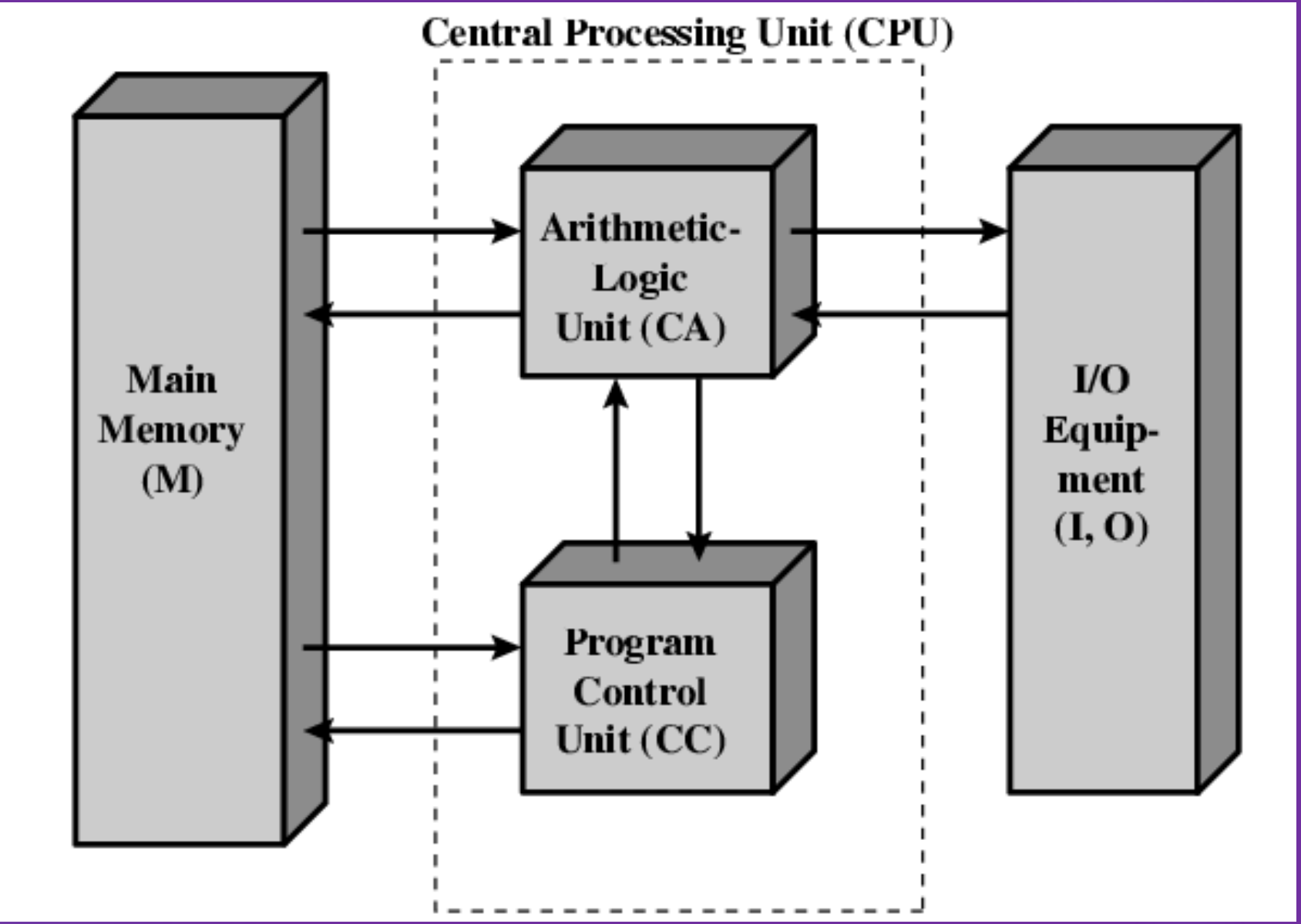
Designing Computers

All computers more or less based on the same basic design, the **Von Neumann Architecture**!



Simplified Architecture




Central Processing Unit (CPU)
Main Memory (M)
Arithmetic-Logic Unit (CA)
Program Control Unit (CC)
I/O Equip-ment (I, O)

Structure & Function

- Structure is the way in which components relate to each other
- Function is the operation of individual components as part of the structure.

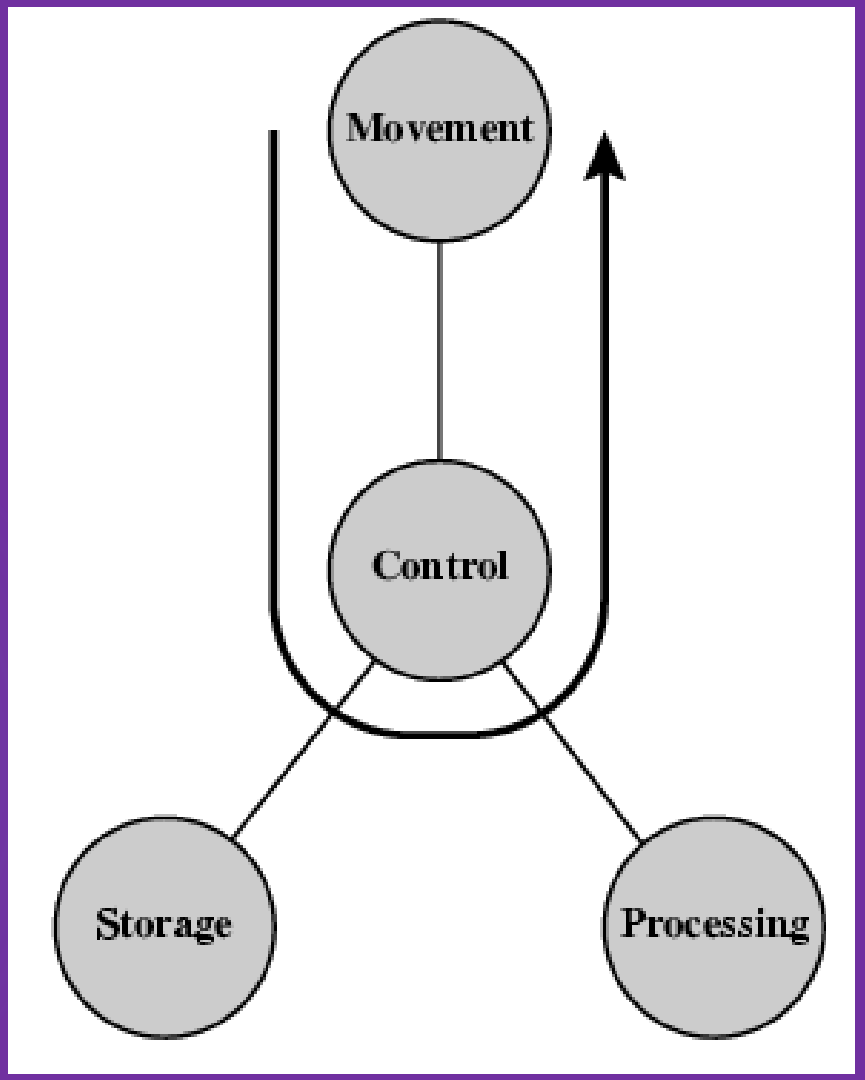
Computer Functions

Function

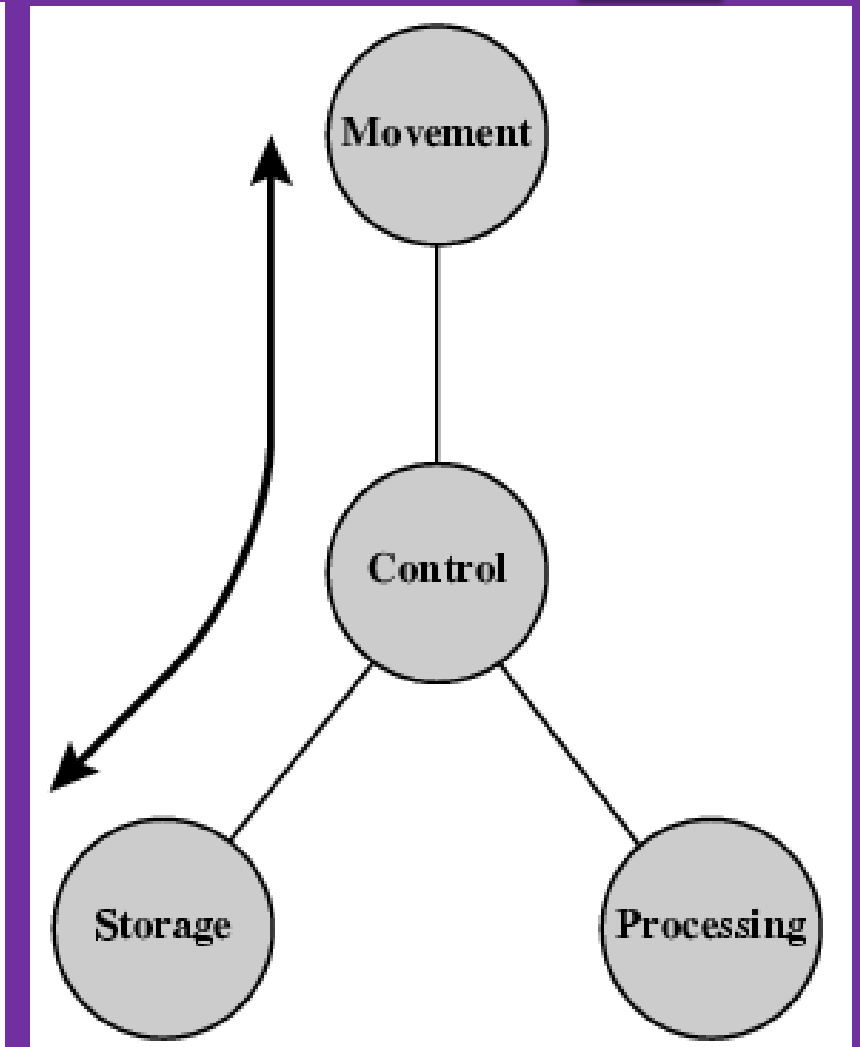
- All computer functions are:
  - **Data processing**
  - **Data storage**
  - **Data movement**
  - **Control**

The basic functions that a computer can perform are:

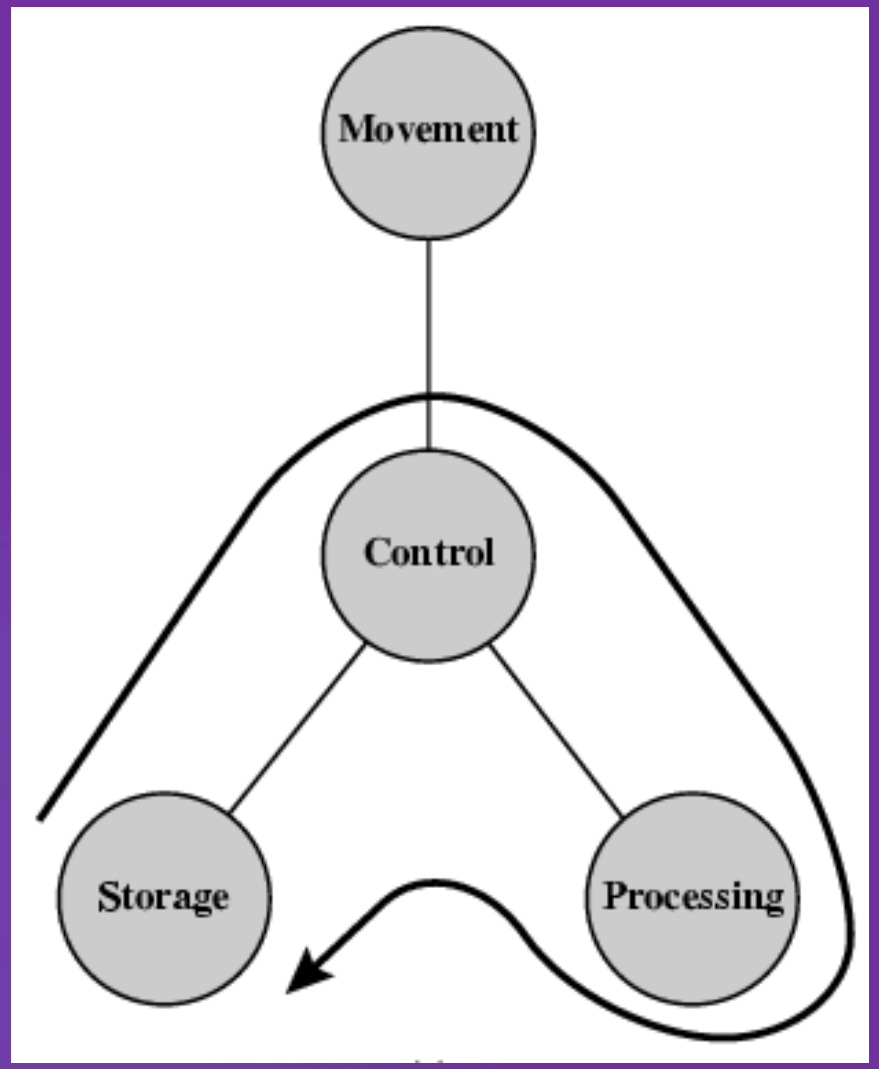
- Data processing
- Data storage
- Data movement
- Control



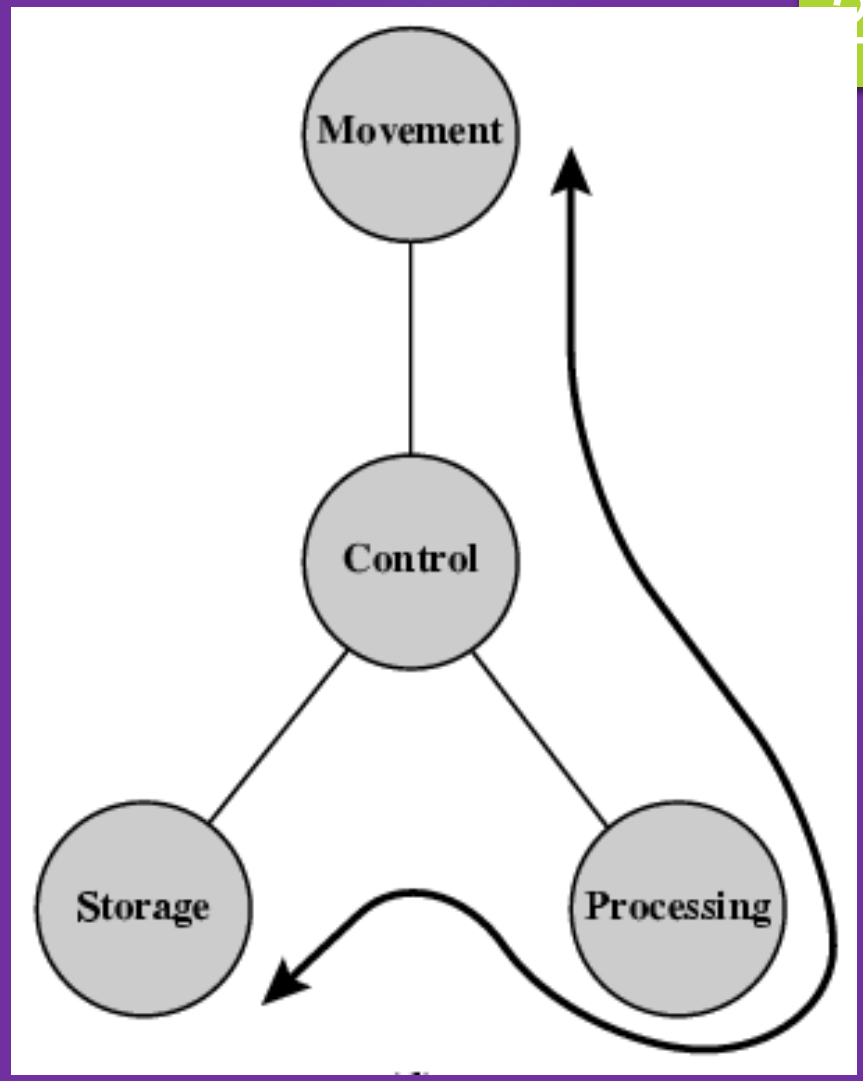
Data movement



Data Storage



Processing from/to storage



Processing from storage to I/O

What is a computer?

- Simply , a computer is a sophisticated electronic calculating machine that:
  - Accepts input information,
  - Processes the information according to a list of internally stored instructions and
  - Produces the resulting output information.
- Functions performed by a computer are:
  - Accepting information to be processed as input.
  - Storing a list of instructions to process the information.
  - Processing the information according to the list of instructions.
  - Providing the results of the processing as output.
- What are the functional units of a computer?

Information in a computer - Instructions

- Instructions specify commands to:
  - Transfer information within a computer (e.g., from memory to ALU)
  - Transfer of information between the computer and I/O devices (e.g., from keyboard to computer, or computer to printer)
  - Perform arithmetic and logic operations (e.g., Add two numbers, Perform a logical AND).
- A sequence of instructions to perform a task is called a program, which is stored in the memory.
- Processor fetches instructions that make up a program from the memory and performs the operations stated in those instructions.
- What do the instructions operate upon?

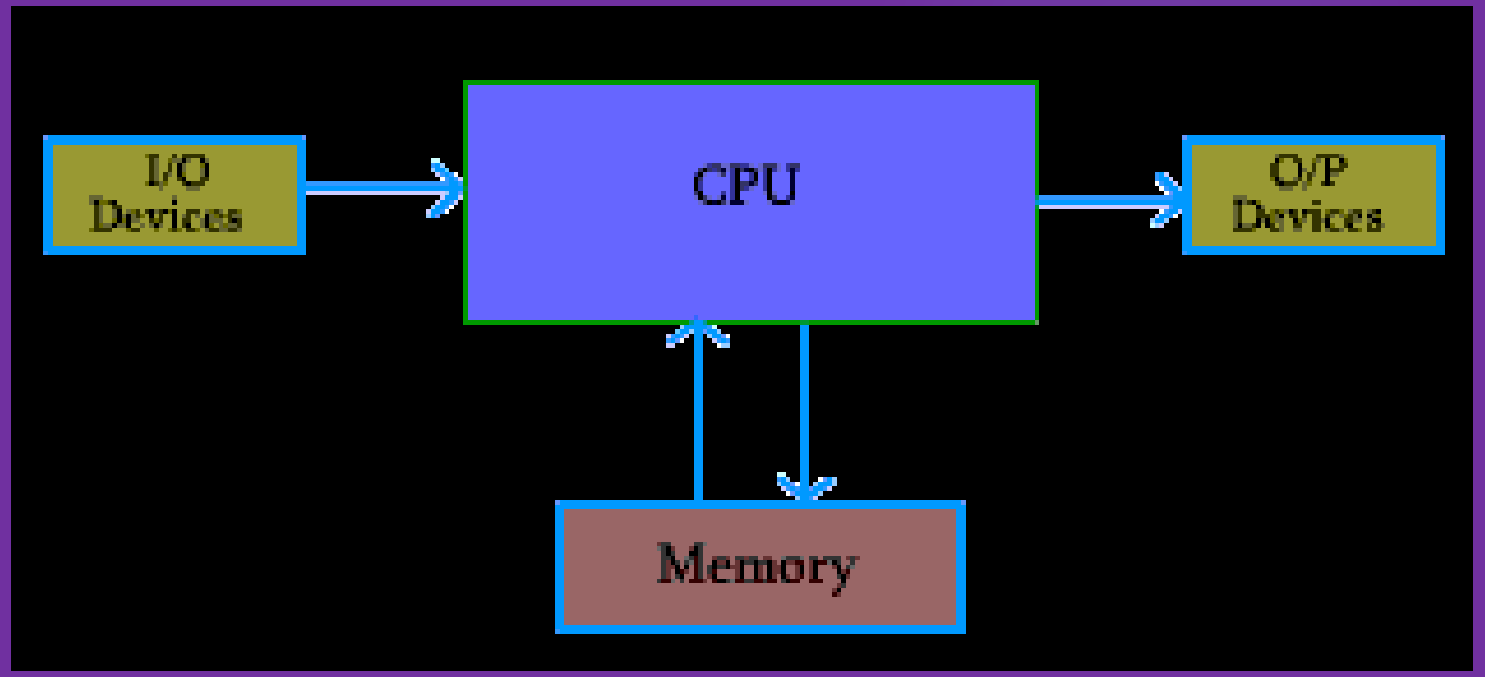
Information in a computer - Data

- Data are the “operands” upon which instructions operate.
- Data could be:
  - Numbers,
  - Encoded characters.
- Data, in a broad sense means any digital information.
- Computers use data that is encoded as a string of binary digits called bits.

Basic Computer Model

The model of a computer can be described by four basic functional units::

Central Processor Unit
Input Unit
Output Unit
Memory Unit



A. Central Processor Unit [CPU]:

Central processor unit consists of two basic blocks:

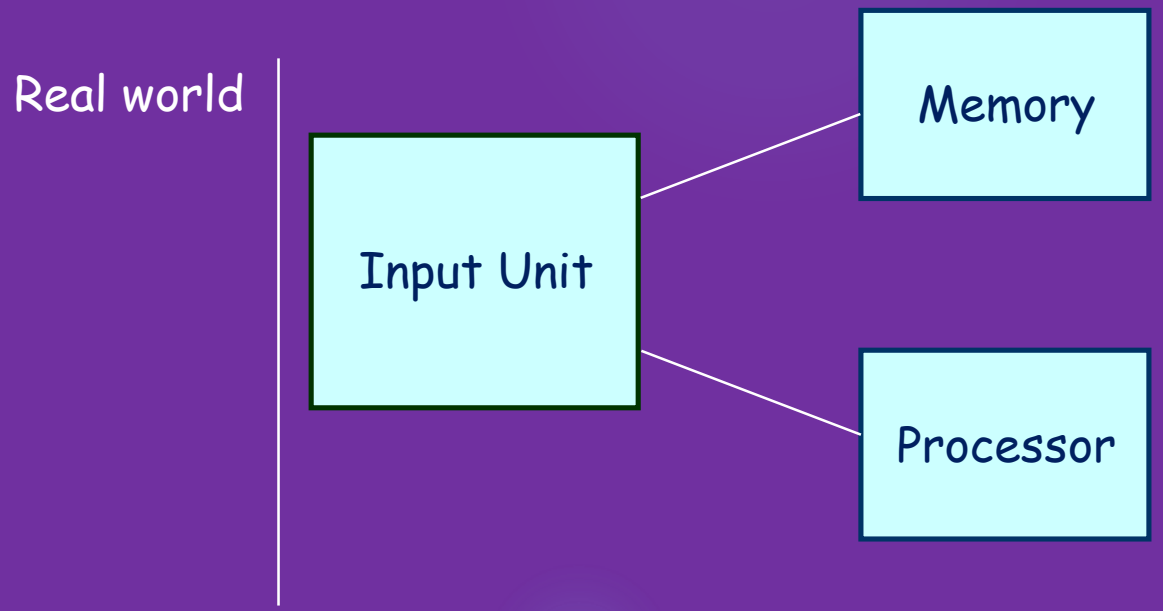
- The program control unit has a set of registers and control circuit to generate control signals.
- The execution unit or data processing unit contains a set of registers for storing data and an Arithmatic and Logic Unit (ALU) for execution of arithmatic and logical operations. In addition, CPU may have some additional registers for temporary storage of data.

B. Input unit

With the help of input unit data from outside can be supplied to the computer. Program or data is read into main storage from input device or secondary storage under the control of CPU input instruction.
Example of input devices: Keyboard, Mouse, Hard disk, Floppy disk, CD-ROM drive etc.

Binary information must be presented to a computer in a specific format.
This task is performed by the input unit:

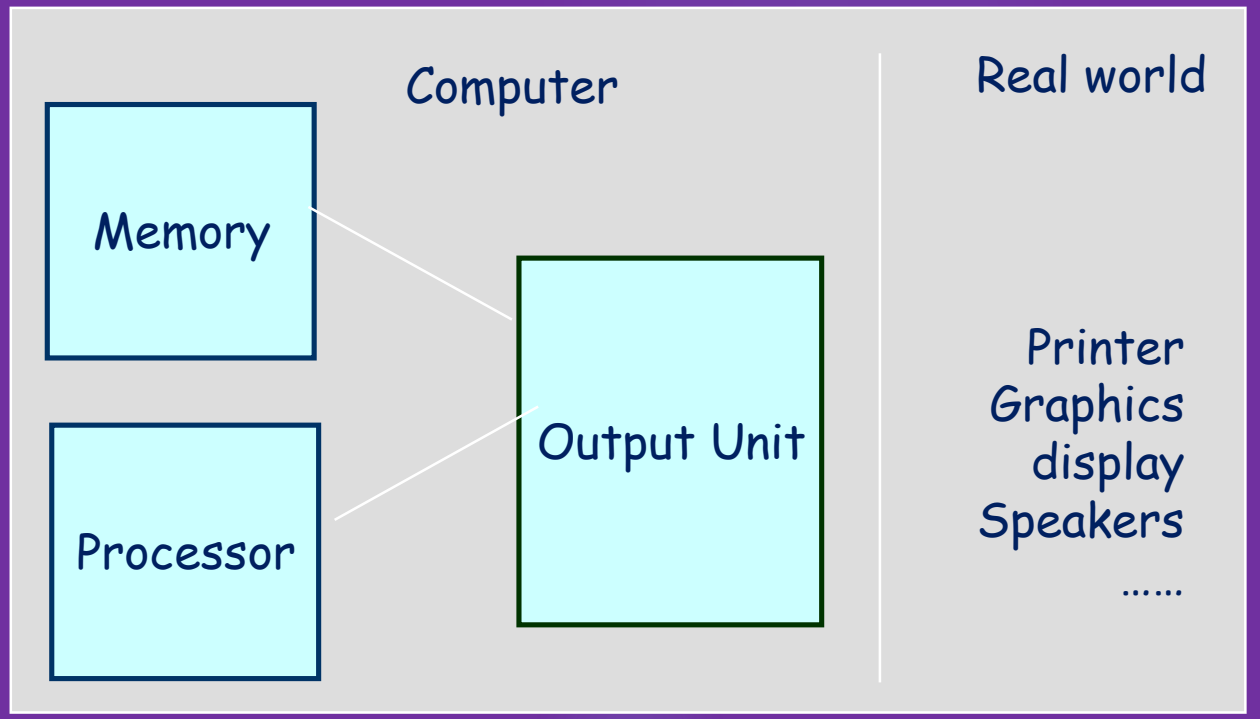
- Interfaces with input devices.
- Accepts binary information from the input devices.
- Presents this binary information in a format expected by the computer.
- Transfers this information to the memory or processor.



C. Output Unit:

With the help of output unit computer results can be provided to the user or it can be stored in stograge device permanently for future use. Output data from main storage go to output device under the control of CPU output instructions. Example of output devices: Printer, Monitor, Plotter, Hard Disk, Floppy Disk etc.

- Computers represent information in a specific binary form. Output units:
  - Interface with output devices.
  - Accept processed results provided by the computer in specific binary form.
  - Convert the information in binary form to a form understood by an output device.



D. Memory Unit:

Memory unit is used to store the data and program. CPU can work with the information stored in memory unit. This memory unit is termed as primary memory or main memory module. These are basically semi conductor memories.

There are two types of semiconductor memories–

- Volatile Memory : RAM (Random Access Memory).
- Non-Volatile Memory : ROM (Read only Memory), PROM (Programmable ROM , EPROM (Erasable PROM), EEPROM (Electrically Erasable PROM).

Secondary Memory:
There is another kind of storage device, apart from memory ,
which is known as secondary memory .

Secondary memories are non volatile memory and it is used for permanent storage of data and program.

Example of secondary memories:

Hard Disk, Floppy Disk, Magenetic Tape --- These are magnetic devices,
CD-ROM ------ is optical device

Memory unit

- Memory unit stores instructions and data.
  - Recall, data is represented as a series of bits, so , memory unit thus stores bits.
- Processor reads instructions and reads/writes data from/to the memory during the execution of a program.
  - In theory, instructions and data could be fetched one bit at a time.
  - In practice, a group of bits is fetched at a time.
  - Group of bits stored or retrieved at a time is termed as "word"
  - Number of bits in a word is termed as the "word length" of a computer.
- In order to read/write to and from memory, a processor should know where to look:
  - "Address" is associated with each word location.

- Processor reads/writes to/from memory based on the memory address:
  - Access any word location in a short and fixed amount of time

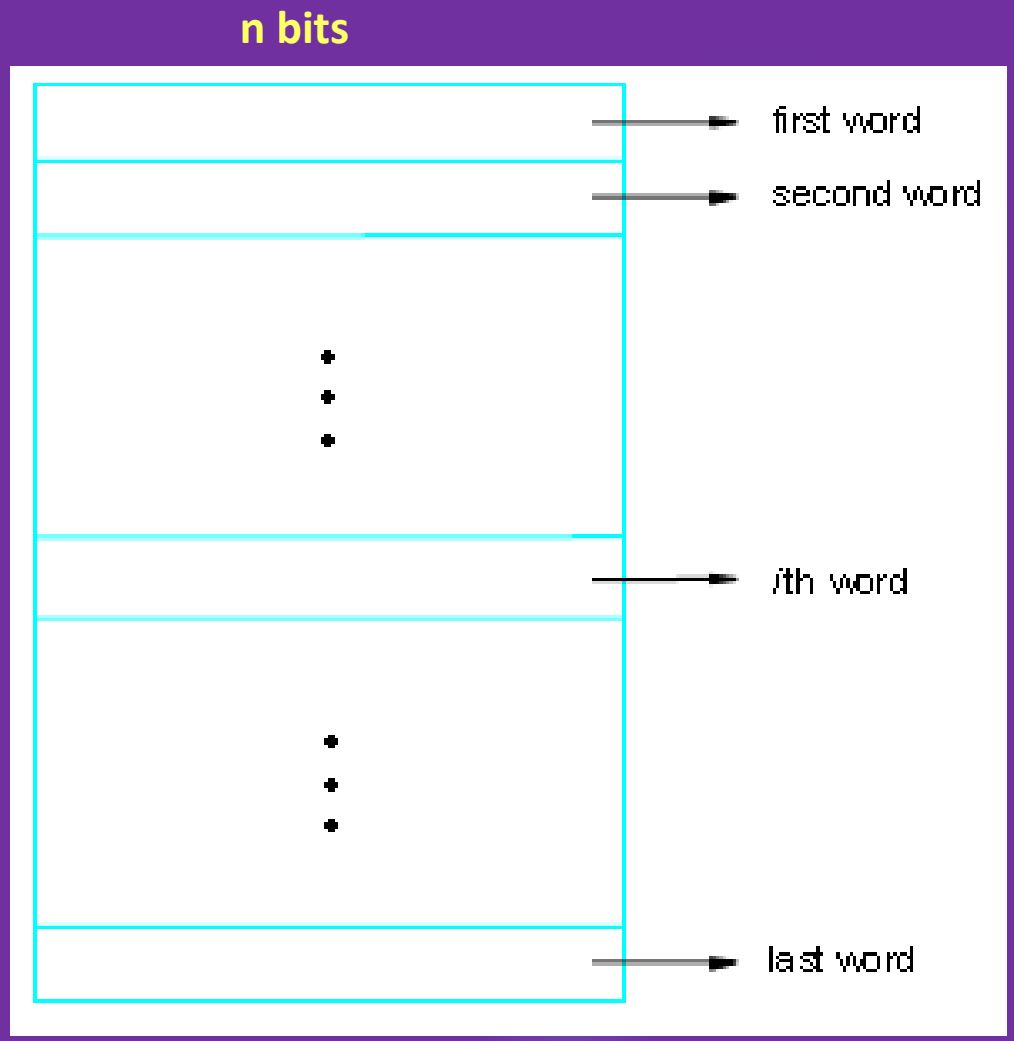
  based on the address.
  - Memory provides fixed access time independent of the location of the word. Access time is known as "Memory Access Time".
- Memory and processor have to "communicate" with each other in order to read/write information.
  - In order to reduce "communication time", a small amount of RAM (known as Cache) is tightly coupled with the processor.
- Modern computers have three to four levels of RAM units with different speeds and sizes:
  - Fastest, smallest known as Cache
  - Slowest, largest known as Main memory.

- Primary storage of the computer consists of RAM units.
  - Fastest, smallest unit is Cache.
  - Slowest, largest unit is Main Memory.
- Primary storage is insufficient to store large amounts of data and programs.
  - Primary storage can be added, but it is expensive.
- Store large amounts of data on secondary storage devices:
  - Magnetic disks and tapes,
  - Optical disks (CD-ROMS).
  - Access to the data stored in secondary storage in slower, but take advantage of the fact that some information may be accessed infrequently.
- Cost of a memory unit depends on its access time, lesser access time implies higher cost.

Memory Location, Addresses, and Operation

Memory consists of many millions of storage cells, each of which can store 1 bit.

Data is usually accessed in n-bit groups, n is called word length.



Memory words.

To retrieve information from memory, either for one word or one byte (8-bit), addresses for each location are needed.

- **A k-bit address memory has 2^k memory locations, namely 0 –2^{k-1}, called memory space/ address space.**
- 24-bit memory: 2^{24}= 16,777,216 = 16M (1M=2^{20})
- 32-bit memory: 2^{32}= 4G (1G=2^{30})
- 1K(kilo)=2^{10}
- 1T(tera)=2^{40}

It is impractical to assign distinct addresses to individual bit locations in the memory.

- The most practical assignment is to have successive addresses refer to successive byte locations in the memory – byte-addressable memory.

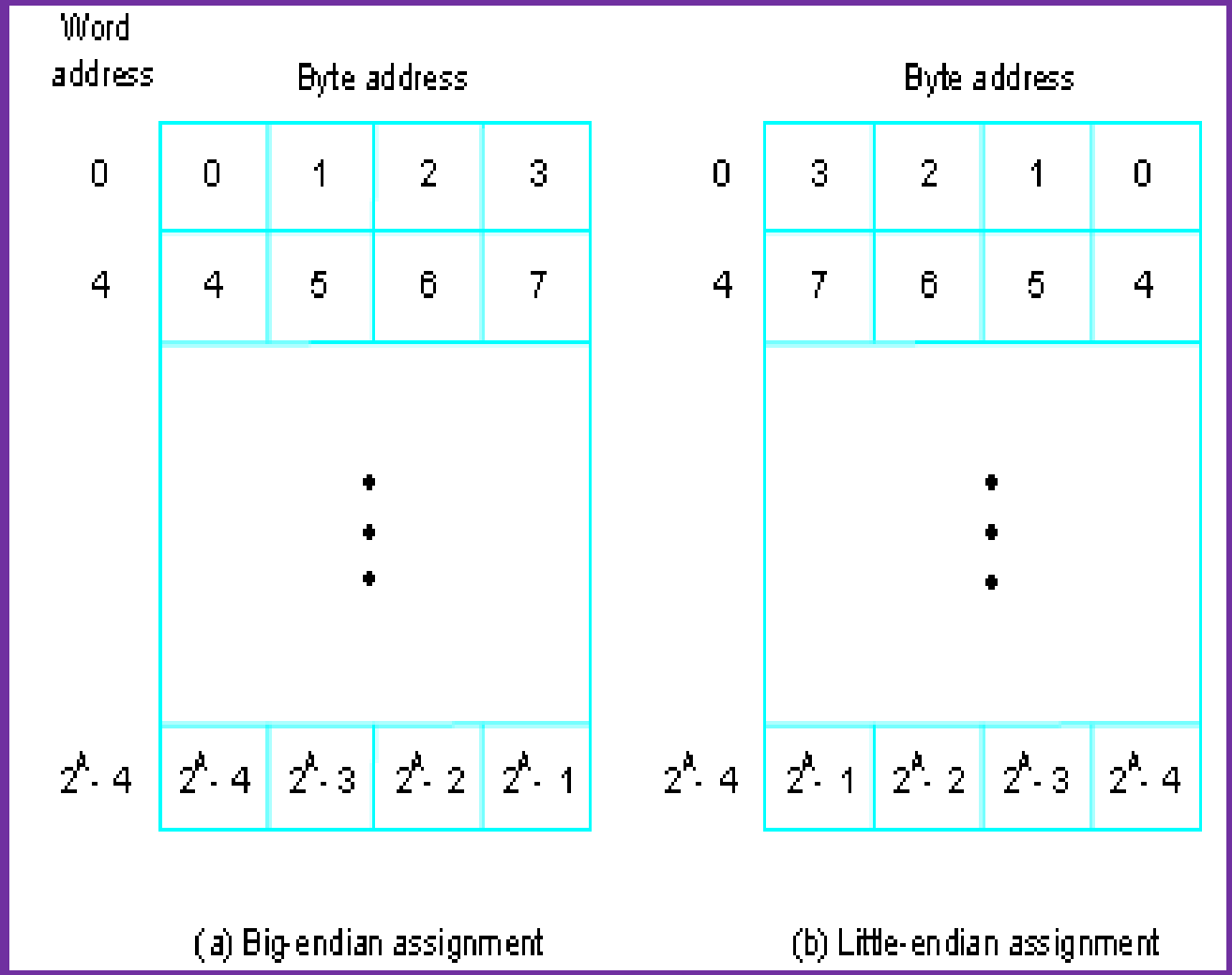
- Byte locations have addresses 0, 1, 2, …

If word length is 32 bits, they successive words are located at addresses 0, 4, 8,…(**memory blocks**)

Big-Endian and Little-Endian Assignments

Big-Endian: lower byte addresses are used for the most significant bytes of the word.

Little-Endian: opposite ordering, lower byte addresses are used for the less significant bytes of the word



Address ordering of bytes

- Word alignment – Words are said to be aligned in memory if they begin at a byte addr. that is a multiple of the num of bytes in a word.

- 16-bit word: word addresses: 0, 2, 4,….(block of 2 bytes)

- 32-bit word: word addresses: 0, 4, 8,….(block of 4 bytes)

- 64-bit word: word addresses: 0, 8,16,….(block of ……?)

The end of Chapter One