Unit - 2

I/O Organization

Mode of Transfer

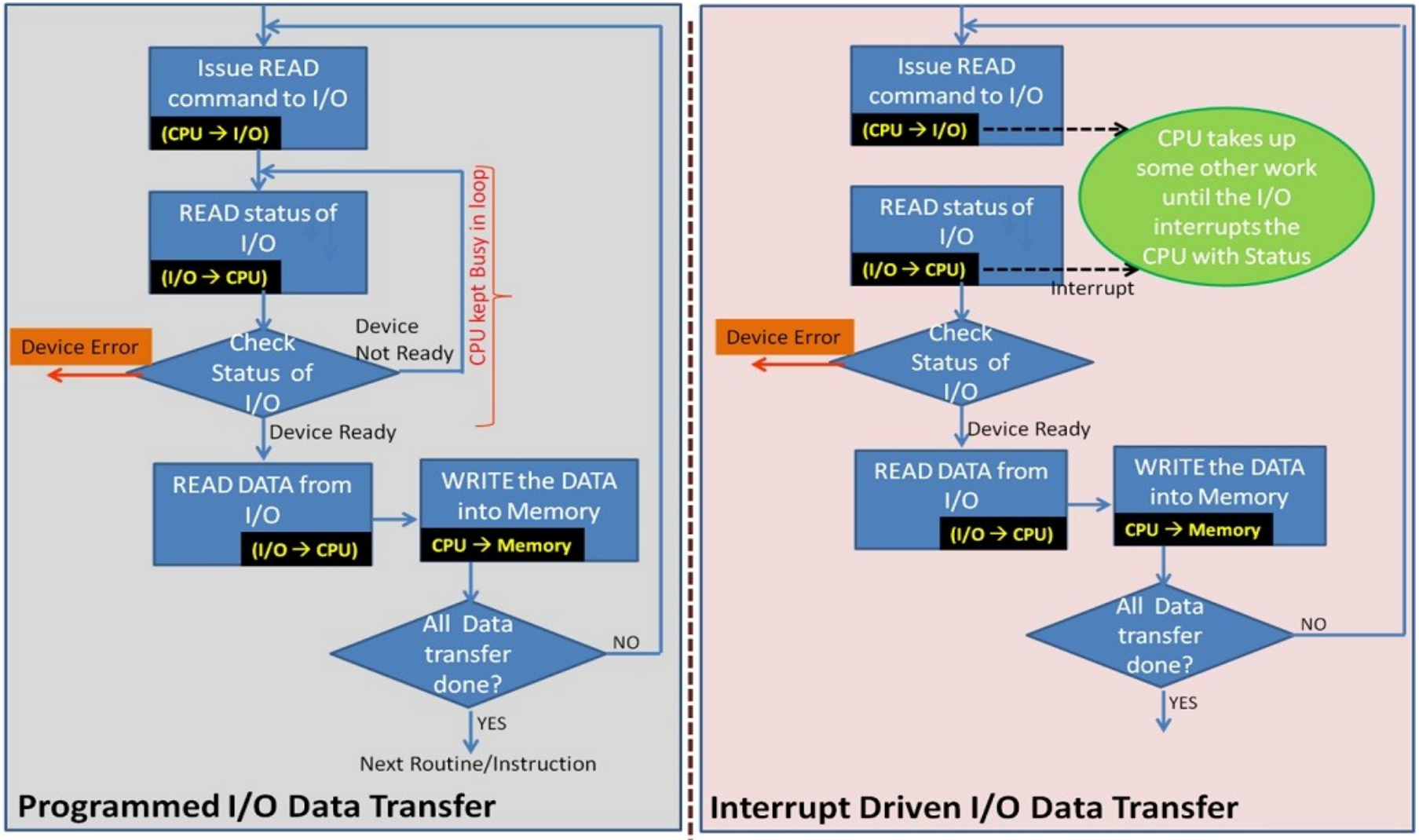
The binary information that is received from an external device is usually stored in the memory unit. The information that is transferred from the CPU to the external device is originated from the memory unit. CPU merely processes the information but the source and target is always the memory unit. Data transfer between CPU and the I/O devices may be done in different modes.

Data transfer to and from the peripherals may be done in any of the three possible ways

1. Programmed I/O.
2. Interrupt- initiated I/O.
3. Direct memory access(DMA).

Programmed I/O

- It is due to the result of the I/O instructions that are written in the computer program. Each data item transfer is initiated by an instruction in the program. Usually the transfer is from a CPU register and memory. In this case it requires constant monitoring by the CPU of the peripheral devices.
- In this case, the I/O device does not have direct access to the memory unit. A transfer from I/O device to memory requires the execution of several instructions by the CPU, including an input instruction to transfer the data from device to the CPU and store instruction to transfer the data from CPU to memory. In programmed I/O, the CPU stays in the program loop until the I/O unit indicates that it is ready for data transfer. This is a time consuming process since it needlessly keeps the CPU busy. This situation can be avoided by using an interrupt facility.

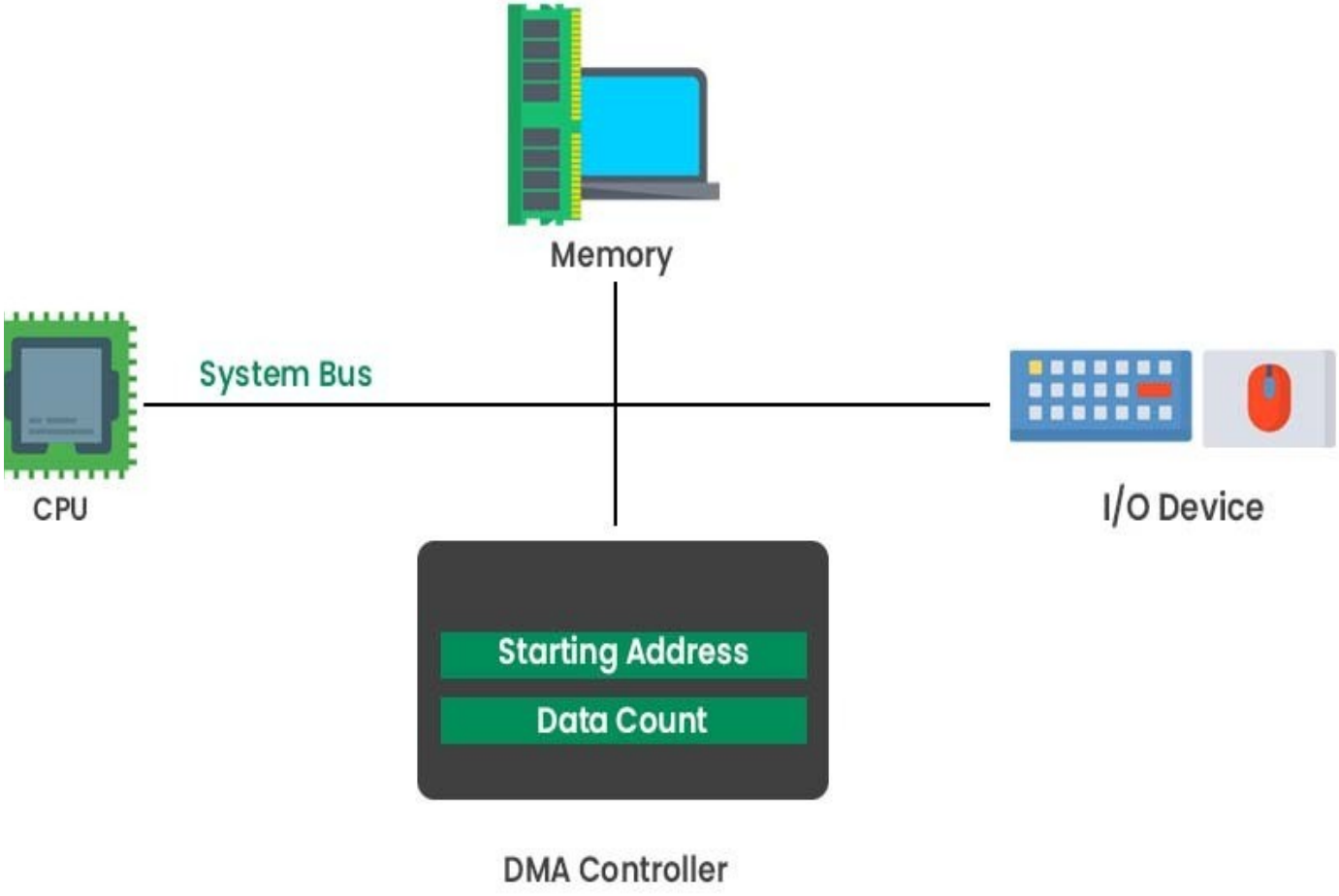

Issue READ command to I/O
(CPU → I/O)
READ status of I/O
(I/O → CPU)
Device Error
Check Status of I/O
Device Not Ready
CPU kept Busy in loop
Device Ready
READ DATA from I/O
(I/O → CPU)
WRITE the DATA into Memory
CPU → Memory
All Data transfer done?
NO
YES
Next Routine/Instruction
Programmed I/O Data Transfer
Issue READ command to I/O
(CPU → I/O)
CPU takes up some other work until the I/O interrupts the CPU with Status
READ status of I/O
(I/O → CPU)
Interrupt
Device Error
Check Status of I/O
Device Ready
READ DATA from I/O
(I/O → CPU)
WRITE the DATA into Memory
CPU → Memory
All Data transfer done?
NO
YES
Interrupt Driven I/O Data Transfer

Interrupt Intiate I/O

- Since in the above case we saw the CPU is kept busy unnecessarily. This situation can very well be avoided by using an interrupt driven method for data transfer.
- By using interrupt facility and special commands to inform the interface to issue an interrupt request signal whenever data is available from any device.
- In the meantime the CPU can proceed for any other program execution. The interface meanwhile keeps monitoring the device.
- Whenever it is determined that the device is ready for data transfer it initiates an interrupt request signal to the computer.
- Upon detection of an external interrupt signal the CPU stops momentarily the task that it was already performing, branches to the service program to process the I/O transfer, and then return to the task it was originally performing.

DMA Transfer

- The data transfer between a fast storage media such as magnetic disk and memory unit is limited by the speed of the CPU. Thus we can allow the peripherals directly communicate with each other using the memory buses, removing the intervention of the CPU. This type of data transfer technique is known as DMA or direct memory access. During DMA the CPU is idle and it has no control over the memory buses. The DMA controller takes over the buses to manage the transfer directly between the I/O devices and the memory unit.


Memory
System Bus
CPU
I/O Device
Starting Address
Data Count
DMA Controller

DMA - Direct Memory Access

DMA is io technique that provides Direct access to the main memory while CPU is temporary disabled to speed up the memory operations

The process is managed by a chip known as DMA controller

i/o devices are connected to system bus via a special interface circuit called DMA controller

In DMA both CPU and DMA controller have access to main memory via shared system bus having data address and control lines

During DMA transfer the CPU is idle and has no control of the system bus or allocated memory bus

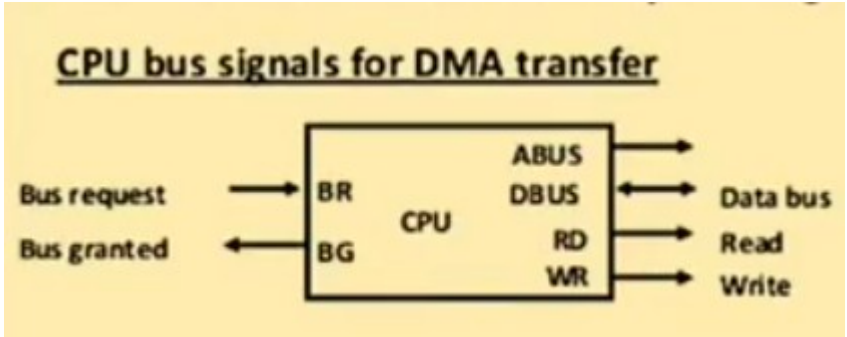
A DMA controller temporary borrows the address bus data bus and control bus from the CPU and transfers data between input output device and main memory

The DMA transfer is also used to do high speed memory to memory transfer. for example USB drive

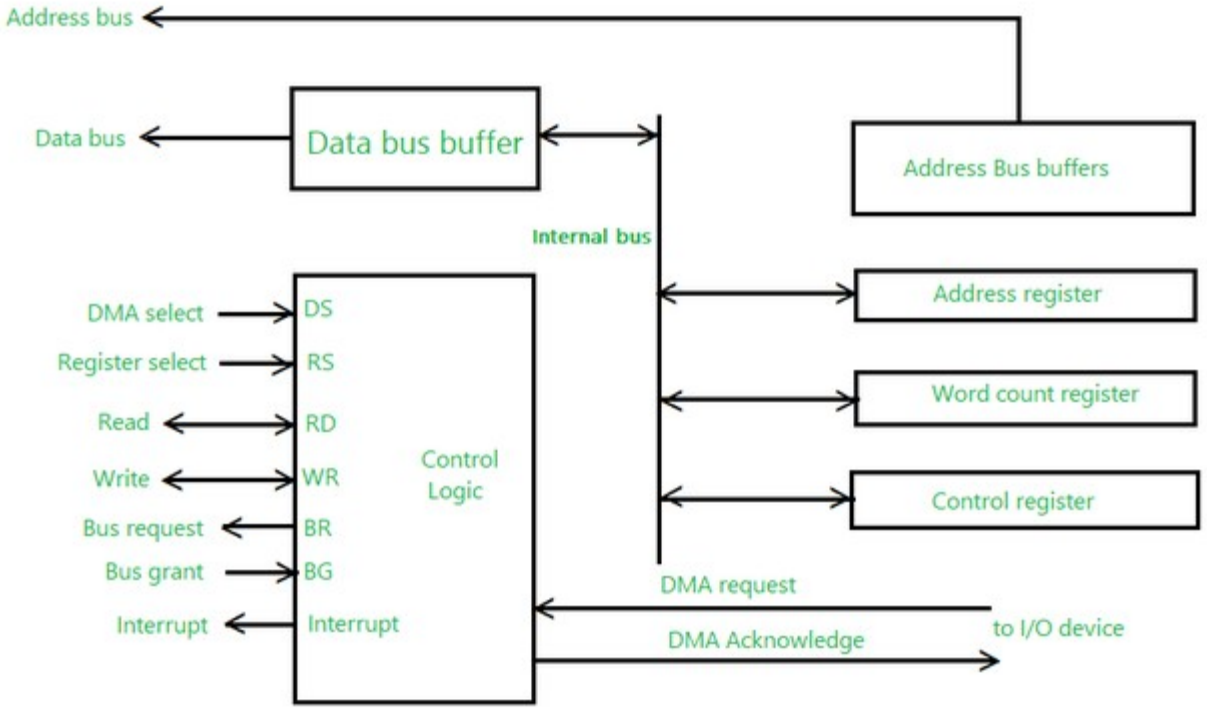
DMA is sometimes referred to as DMA channel in an alternative configuration DMA controller maybe incorporated directly into input output device

DMA Working

1. I/O wants to transfer data with main memory
2. I/O sends DMA request to DMA controller
3. DMA sends bus request to CPU for buses
4. DMA waits until CPU sends bus grant signal to DMA
5. CPU relinquishes control of buses and place address bus data bus read and write lines into high impedance state
6. CPU activate the bg signal and becomes in idle state
7. DMA takes control of the buses to conduct direct memory transfer without CPU intervention
8. When DMA transfer terminates , it disables bus request line
9. The CPU disables the bus grant and takes control of the buses and return to its normal operation



DMA Controller



DMA Controller

DMA Controller is a hardware device that allows I/O devices to directly access memory with less participation of the processor.

Figure shows the block diagram of the DMA controller. The unit communicates with the CPU through data bus and control lines. Through the use of the address bus and allowing the DMA and RS register to select inputs, the register within the DMA is chosen by the CPU. RD and WR are two-way inputs. When BG (bus grant) input is 0, the CPU can communicate with DMA registers. When BG (bus grant) input is 1, the CPU has relinquished the buses and DMA can communicate directly with the memory.

The DMA controller has three registers as follows.

- **Address register –** It contains the address to specify the desired location in memory.
- **Word count register –** It contains the number of words to be transferred.
- **Control register –** It specifies the transfer mode.

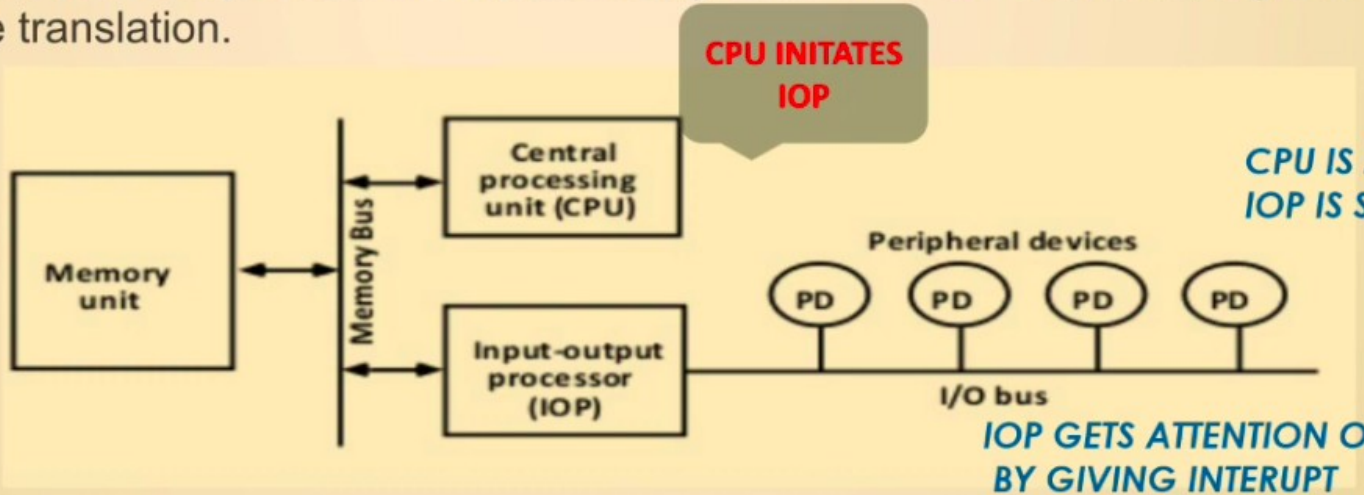
DMA Controller

The CPU initializes the DMA by sending the given information through the data bus.

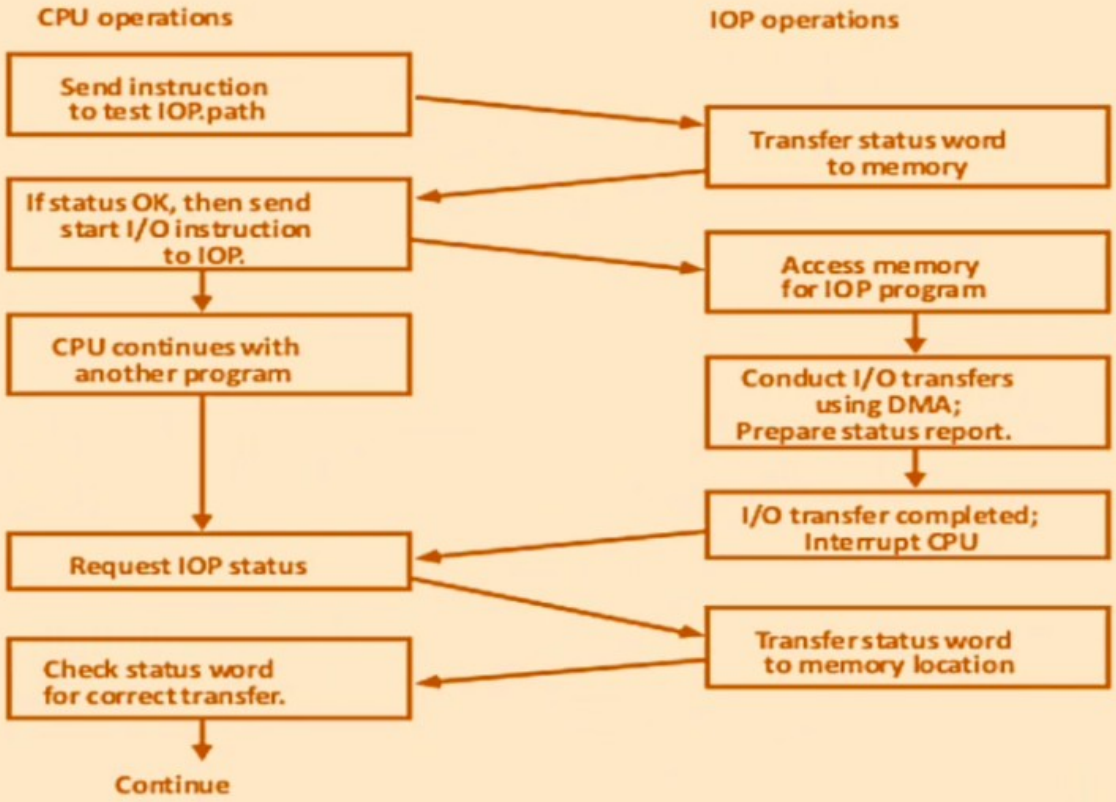
- The starting address of the memory block where the data is available (to read) or where data are to be stored (to write).
- It also sends word count which is the number of words in the memory block to be read or write.
- Control to define the mode of transfer such as read or write.
- A control to begin the DMA transfer.

INPUT-OUTPUT PROCESSOR(IOP)

- IOP is similar to a CPU except that it is designed to handle the details of I/O processing.
- Unlike the DMA controller that must be setup entirely by the CPU, the IOP can fetch and execute its own instruction.
- IOP instructions are specifically designed to facilitate I/O transfers.
- In addition, IOP can perform other processing tasks such as arithmetic, logic branching and code translation.



CPU – IOP COMMUNICATION



Memory Organisation

Memory

- Essential part of digital computer.
- Used for storing programs,data and result.
- Commonly used memory
  - **Semiconductor memory:** faster , light, compact, consume less power.
  - **Magnetic memory & Optical memory:** slow and cheap as compared with semiconductor memory

Classification of Memory

- Classification is based upon the following factors:
  - Access Method
  - Capability
  - Functionality or Role
  - Technology

By Functionality

- Memory is classified into the following depending on the funtion it performs:
  - Main Memory
  - Auxilary Memory
  - Cache Memory
  - Virtual Memory

Main Memory

- It is the central storage of computer system.
- Relatively large and fast memory to store the program and data during computer operation.
- Most of the main memory in a general purpose computer is made up of **RAM** integrated circuits chips, but a portion of the memory may be constructed with **ROM** chips.
  - **RAM**– Random Access memory
    - Integrated RAM are available in two possible operating modes, ***Static*** *and* ***Dynamic.***
  - **ROM**– Read Only memory
- Main Memory has a equal access time for all its location irrespective of its address.Thus it is known as Random Access Memory.

Main Memory (continue)

Characteristics of Main Memory

- These are semiconductor memories
- It is also known as Random Access Memory.
- Usually volatile memory.
- Data is lost in case power is switched off.
- It is working memory of the computer.
- Faster than secondary memories.
- A computer cannot run without primary memory.

Main Memory (continue)

- Limitation of Main memory
  - Speed
  - Capacity

Auxilary Memory

- This type of memory is also known as external memory or non-volatile or Secondary Memory.
- It is slower and cheaper than main memory.
- These are used for storing data/Information permanently.
- CPU directly does not access these memories instead they are accessed via input-output routines.
- Contents of secondary memories are first transferred to main memory, and then CPU can access it.
- For example : disk, CD-ROM, DVD etc.

Characteristic of Secondary Memory

- These are magnetic and optical memories
- It is known as backup memory.
- It is non-volatile memory.
- Data is permanently stored even if power is switched off.
- It is used for storage of data in a computer.
- Computer may run without secondary memory.
- Slower than primary memories.

Cache Memory

- Has small capacity compared to main memory, but is relatively faster and expensive.
- Cache memory is a very high speed semiconductor memory which can speed up CPU.
- It acts as a buffer between the CPU and main memory.
- It is used to hold those parts of data and program which are most frequently used by CPU.
- The parts of data and programs are transferred from disk to cache memory by operating system, from where CPU can access them.
- The cache memory can be of two types:
  - Unified cache : It is a common cache memory that can store both instruction and data
  - Split cache : It has a separate cache to store instruction and data.

Cache Memory (Continue)

Advantages

- Cache memory is faster than main memory.
- It consumes less access time as compared to main memory.
- It stores the program that can be executed within a short period of time.
- It stores data for temporary use.

Disadvantages

- Cache memory has limited capacity.
- It is very expensive.

Virtual Memory

- It is a concept used to construct large programs though the physical memory has limited space.
- The OS keeps the large programs in the secondary memory and brings only a part of the program in main memory.
- The address generated by the CPU is known as virtual address.
- Each virtual address is mapped to physical address in main memory.
- The mapping or translation is handled by the OS and the CPU hardware.

Advantages

- Allows Processes whose aggregate memory requirement is greater than the amount of physical memory, as infrequently used pages can reside on the disk.
- Virtual memory allows speed gain when only a particular segment of the program is required for the execution of the program.
- This concept is very helpful in implementing multiprogramming environment.

DISADVANTAGES

- Applications run rather slower when they are using virtual memory.
- It takes more time to switch between applications.
- Reduces system stability.

By Access Method

- Method to access the different location of memory
  - Random Access
  - Sequential
  - Semi Random

Access Methods

- Each memory is a collection of various memory location. Accessing the memory means finding and reaching desired location and than reading information from memory location. The information from locations can be accessed as follows:

1. **Random access**
2. **Sequential access**
3. **Semi Random**

- **Random Access**: It is the access mode where each memory location has a unique address.

Access Methods

- **Sequential Access**: If storage locations can be accessed only in a certain predetermined sequence, the access method is known as serial or sequential access.
  - Opposite of RAM: **Serial Access Memory** (SAM). SAM works very well for memory **buffers**, where the data is normally stored in the order in which it will be used (a good example is the texture buffer memory on a video card , magnetic tapes, etc.).
- **Semi Random:** In this access, information is stored on tracks and each track has a separate

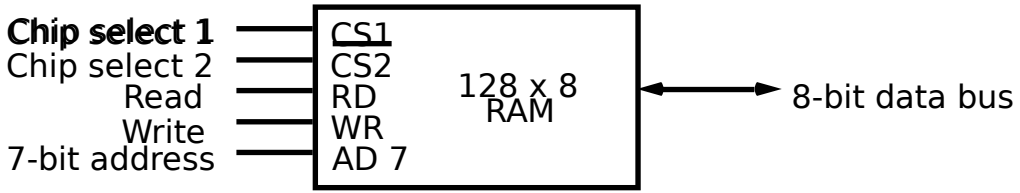
By Capabilty

- Memory is classified into two types based on read/write capabilty
  - Read/write Memory (RAM)
  - Read Only Memory (ROM)

Read/write Memory (RAM)

- RAM is used for storing bulk of programs and data that is subject to change.
- It allows both read and write operation.
- Also known as RAM

Typical RAM chip



READ ONLY MEMORY (ROM)

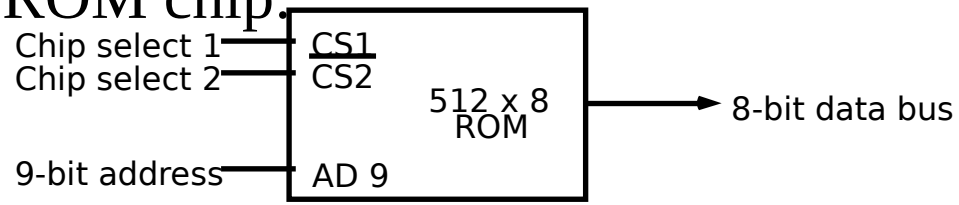
- It is non-volatile memory, which retains the data even when power is removed from this memory. Programs and data that can not be altered are stored in ROM.

- ROM is used for storing programs that are **PERMANENTLY** resident in the computer and for tables of constants that do not change in value once the production of the computer is completed.

- The ROM portion of main memory is needed for

READ ONLY MEMORY (ROM)

- Typical ROM chip:



- Since the ROM can only READ, the data bus can only be in output mode.

By Technology

- Magnetic Core
- Magnetic Bubble
- Semi Conductor

Megnetic Core

- Earlier, main frames and minicomputers used magnetic core memory but nowadays they are replaced by semoconductor memories.
- Megnetic core memories required more power and were slow.
- Non voletile in nature.

Semiconductor Memory

Devided into two parts:

1. Read/Write Memory

SRAM and DRAM

2. Read Only Memory

ROM , PROM , EPROM , EEPROM , FLASH memory

Megnetic Bubble

- Magnetic bubble memories use small magnetized areas, called bubbles to represent data bits.
- It is non volatile memory