Unit 2: Computer Hardware

Introduction

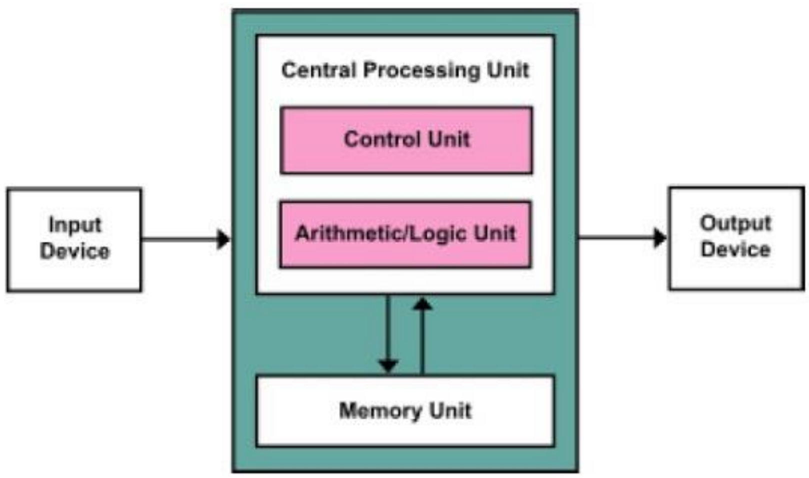
A computer consists of three main components—

(1) Input/output (I/O) Unit,

(2) Central Processing Unit (CPU), and

(3) Memory Unit.

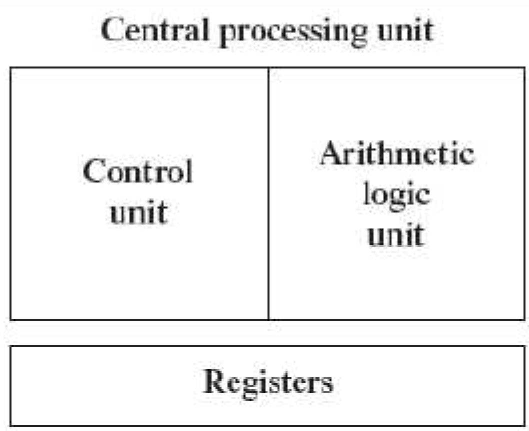
The computer user interacts with the computer via the I/O unit. The purpose of I/O unit is to provide data and instructions as input to the computer and to present relevant information as output from the computer. CPU controls the operations of the computer and processes the received input to generate the relevant output. The memory unit stores the instructions and the data during the input activity, to make instructions readily available to CPU during processing. It also stores the processed output.



Central Processing Unit (CPU):

Central Processing Unit (CPU) or the processor is also often called the brain of computer. CPU consists of Arithmetic Logic Unit (ALU) and Control Unit (CU). In addition, CPU also has a set of registers which are temporary storage areas for holding data, and instructions. ALU performs the arithmetic and logic operations on the data that is made available to it. CU is responsible for organizing the processing of data and instructions. CU controls and coordinates the activity of the other units of computer. CPU uses the registers to store the data, instructions during processing.

CPU executes the stored program instructions, i.e. instructions and data are stored in memory before execution. For processing, CPU gets data and instructions from the memory. It interprets the program instructions and performs the arithmetic and logic operations required for the processing of data. Then, it sends the processed data or result to the memory. CPU also acts as an administrator and is responsible for supervising operations of other parts of the computer.



Arithmetic Logic Unit

- ALU consists of two units—arithmetic unit and logic unit.
- The arithmetic unit performs arithmetic operations on the data that is made available to it. Some of the arithmetic operations supported by the arithmetic unit are—addition, subtraction, multiplication and division.
- The logic unit of ALU is responsible for performing logic operations. Logic unit performs comparisons of numbers, letters and special characters. Logic operations include testing for greater than, less than or equal to condition.
- ALU performs arithmetic and logic operations, and uses registers to hold the data that is being processed.

Registers

- Registers are high-speed storage areas within the CPU, but have the least storage capacity. Registers are not referenced by their address, but are directly accessed and manipulated by the CPU during instruction execution.
- Registers store data, instructions, addresses and intermediate results of processing. Registers are often referred to as the CPU's working memory.
- The data and instructions that require processing must be brought in the registers of CPU before they can be processed. For example, if two numbers are to be added, both numbers are brought in the registers, added and the result is also placed in a register.
- Registers are used for different purposes, with each register serving a specific purpose. Some of the important registers in CPU are as follows—
- ❖ Accumulator (ACC) stores the result of arithmetic and logic operations.
- ❖ Instruction Register (IR) contains the current instruction most recently fetched.
- ❖ Program Counter (PC) contains the address of next instruction to be processed.
- ❖ Memory Address Register (MAR) contains the address of next location in the memory to be accessed.
- ❖ Memory Buffer Register (MBR) temporarily stores data from memory or the data to be sent to memory.
- ❖ Data Register (DR) stores the operands and any other data.

Fig: CPU Registers

- The number of registers and the size of each (number of bits) register in a CPU helps to determine the power and the speed of a CPU.
- The overall number of registers can vary from about ten to many hundreds, depending on the type and complexity of the processor.
- The size of register, also called word size, indicates the amount of data with which the computer can work at any given time. The bigger the size, the more quickly it can process data. The size of a register may be 8, 16, 32 or 64 bits.
- 32-bit processor and 64-bit processor are the terms used to refer to the size of the registers. Other factors remaining the same, a 64-bit processor can process the data twice as fast as one with 32-bit processor.

Control Unit

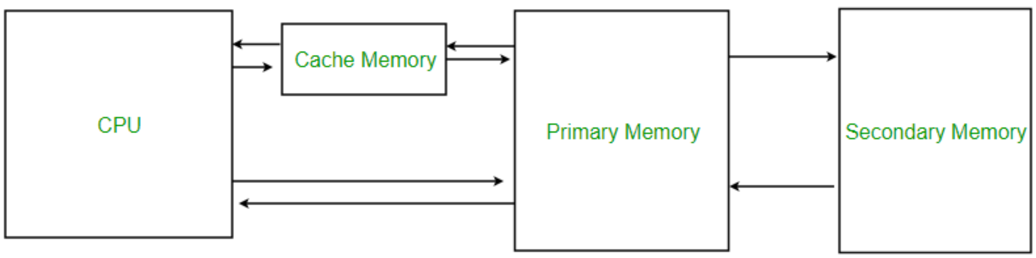
- The control unit of a computer does not do any actual processing of data. It organizes the processing of data and instructions. It acts as a supervisor and, controls and coordinates the activity of the other units of computer.
- CU coordinates the input and output devices of a computer. It directs the computer to carry out stored program instructions by communicating with the ALU and the registers.
- CU uses the instructions in the Instruction Register (IR) to decide which circuit needs to be activated. It also instructs the ALU to perform the arithmetic or logic operations.
- When a program is run, the Program Counter (PC) register keeps track of the program instruction to be executed next.
- CU tells when to fetch the data and instructions, what to do, where to store the results, the sequencing of events during processing etc.
- CU also holds the CPU's Instruction Set, which is a list of all operations that the CPU can perform.

Memory Unit

The memory unit consists of cache memory and primary memory. Primary memory or main memory of the computer is used to store the data and instructions during execution of the instructions. Random Access Memory (RAM) and Read Only Memory (ROM) are the primary memory. In addition to the main memory, there is another kind of storage device known as the secondary memory. Secondary memory is non-volatile and is used for permanent storage of data and programs. A program or data that has to be executed is brought into the RAM from the secondary memory.

Cache Memory

- The data and instructions that are required during the processing of data are brought from the secondary storage devices and stored in the RAM. For processing, it is required that the data and instructions are accessed from the RAM and stored in the registers. The time taken to move the data between RAM and CPU registers is large. This affects the speed of processing of computer, and results in decreasing the performance of CPU.
- Cache memory is a very high-speed memory placed in between RAM and CPU. Cache memory increases the speed of processing.
- Cache memory is a storage buffer that stores the data that is used more often, temporarily, and makes them available to CPU at a fast rate. During processing, CPU first checks cache for the required data. If data is not found in cache, then it looks in the RAM for data.



Primary Memory

- Primary memory is the main memory of computer. It is used to store data and instructions during the processing of data.
- Primary memory is of two kinds—Random Access Memory (RAM) and Read Only Memory (ROM).
- RAM is volatile. It stores data when the computer is on. The information stored in RAM gets erased when the computer is turned off. RAM provides temporary storage for data and instructions.
- ROM is non-volatile memory, but is a read only memory. The storage in ROM is permanent in nature, and is used for storing standard processing programs that permanently reside in the computer. ROM comes programmed by the manufacturer.
- CPU accesses the data and the instructions from RAM, as it can access it at a fast speed than the storage devices connected to the input and output unit.

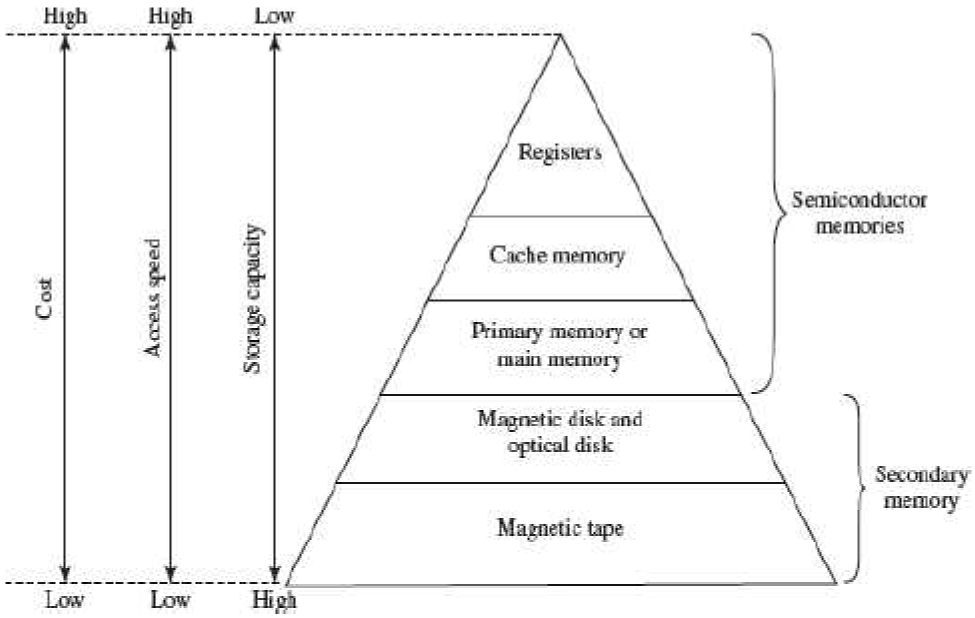
Secondary Memory

- The secondary memory stores data and instructions permanently. The information can be stored in secondary memory for a long time (years), and is generally permanent in nature unless erased by the user. It is a non-volatile memory.
- It provides back-up storage for data and instructions. Hard disk drive, floppy drive and optical disk drives are some examples of storage devices.
- The data and instructions that are currently not being used by CPU, but may be required later for processing, are stored in secondary memory.

- Secondary memory has a high storage capacity than the primary memory and is also cheaper than the primary memory. It takes longer time to access the data and instructions stored in secondary memory than in primary memory.

Memory Hierarchy

The memory hierarchy separates computer storage into a hierarchy based on response time.



Magnetic Disk:

- A magnetic disk is a storage device that uses a magnetization process to write, rewrite and access data. It is covered with a magnetic coating and stores data in the form of tracks and sectors. Hard disks and floppy disks are common examples of magnetic disks. With the development of solid-state drives (SSDs), magnetic disks are no longer considered the only option, but are still commonly used.

Optical Disk:

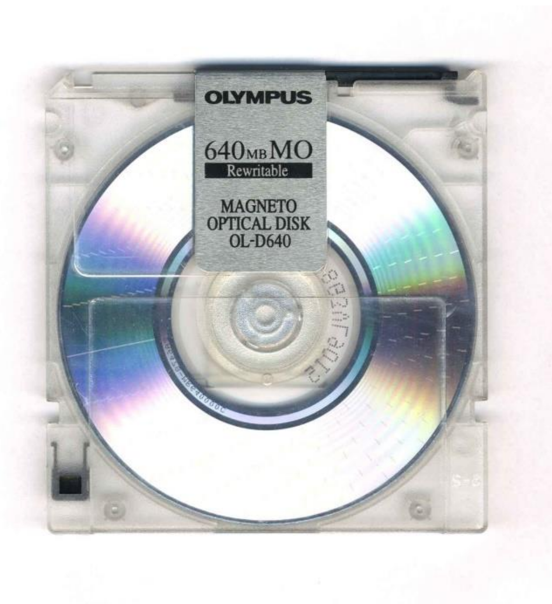
- An optical disk is any computer disk that uses optical storage techniques and technology to read and write data.
- It is a computer storage disk that stores data digitally and uses laser beams (transmitted from a laser head mounted on an optical disk drive) to read and write data.
- An optical disk is primarily used as a portable and secondary storage device.
- Compact disks (CD), digital versatile/video disks (DVD) and Blu-ray disks are currently the most commonly used forms of optical disks.

Magnetic Tape:

- Magnetic tape is a medium for magnetic recording, made of a thin, magnetizable coating on a long, narrow strip of plastic film.

Magneto-Optical Disk:

- A magneto-optical disk is a rewritable disk that makes use of both magnetic disk and optical technologies.
- Magneto-optical disks are also known as magneto-optical drives and MO drives.



Microprocessor:

A microprocessor is an integrated circuit that contains all the functions of a central processing unit of a computer. A processor's instruction set is a determining factor in its architecture. On the basis of the instruction set, microprocessors are classified as—Reduced Instruction Set Computer (RISC), and Complex Instruction Set Computer (CISC).

<u>CISC:</u>

- CISC architecture hardwires the processor with complex instructions, which are difficult to create otherwise using basic instructions. CISC combines the different instructions into one single CPU.
- CISC has a large instruction set that includes simple and fast instructions for performing basic tasks, as well as complex instructions that correspond to statements in the high-level language.
- An increased number of instructions (200 to 300) results in a much more complex processor, requiring millions of transistors.
- Instructions are of variable lengths, using 8, 16 or 32 bits for storage. This results in the processor's time being spent in calculating where each instruction begins and ends.
- With large number of application software programs being written for the processor, a new processor has to be backwards compatible to the older version of processors.
- AMD Processor is based on CISC.

<u>RISC:</u>

- RISC has simple, single-cycle instructions, which performs only basic instructions. RISC architecture does not have hardwired advanced functions. All high-level language support is done in the software.
- RISC has fewer instructions and requires fewer transistors, which results in the reduced manufacturing cost of processor.

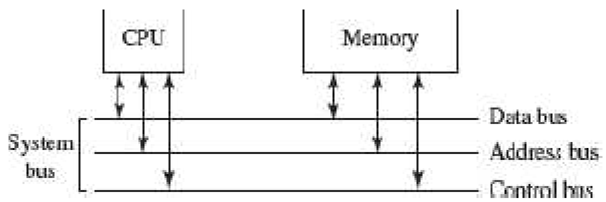
- The instruction size is fixed (32 bits). The processor need not spend time in finding out where each instruction begins and ends.
- RISC architecture has a reduced production cost compared to CISC processors.
- The instructions, simple in nature, are executed in just one clock cycle, which speeds up the program execution when compared to CISC processors.
- RISC processors can handle multiple instructions simultaneously by processing them in parallel.
- Apple Mac G3 is based on RISC.

Interconnecting the Units of a Computer:

CPU sends data, instructions and information to the components inside the computer as well as to the peripherals and devices attached to it. Bus is a set of electronic signal pathways that allows information and signals to travel between components inside or outside of a computer. The different components of computer, i.e., CPU, I/O unit, and memory unit are connected with each other by a bus. The data, instructions and the signals are carried between the different components via a bus. The features and functionality of a bus are as follows:

- A bus is a set of wires used for interconnection, where each wire can carry one bit of data.
- A bus width is defined by the number of wires in the bus.
- A computer bus can be divided into two types—Internal Bus and External Bus.
- The Internal Bus connects components inside the motherboard like, CPU and system memory. It is also called the System Bus.

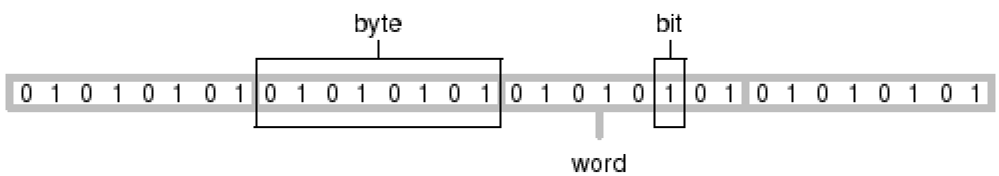
Figure below shows interaction between processor and memory.



- The External Bus connects the different external devices, peripherals, expansion slots, I/O ports and drive connections to the rest of computer. The external bus allows various devices to be attached to the computer. It allows for the expansion of computer's capabilities. It is generally slower than the system bus. It is also referred to as the Expansion Bus.
- A system bus or expansion bus comprise of three kinds of buses: data bus, address bus and control bus.
- The interaction of CPU with memory and I/O devices involves all the three buses.
  - ❖ The command to access the memory or the I/O device is carried by the control bus.
  - ❖ The address of I/O device or memory is carried by the address bus.
  - ❖ The data to be transferred is carried by the data bus.

Memory Representation:

- The computer memory stores different kinds of data like input data, output data, intermediate results, etc., and the instructions. Binary digit or bit is the basic unit of memory.
- A bit is a single binary digit, i.e., 0 or 1. A bit is the smallest unit of representation of data in a computer. However, the data is handled by the computer as a combination of bits. A group of 8 bits form a byte.
- One byte is the smallest unit of data that is handled by the computer. One byte can store 28, i.e., 256 different combinations of bits, and thus can be used to represent 256 different symbols.
- In a byte, the different combinations of bits fall in the range 00000000 to 11111111.
- A group of bytes can be further combined to form a word. A word can be a group of 2, 4 or 8 bytes.



1 bit = 0 or 1

1 Byte (B) = 8 bits

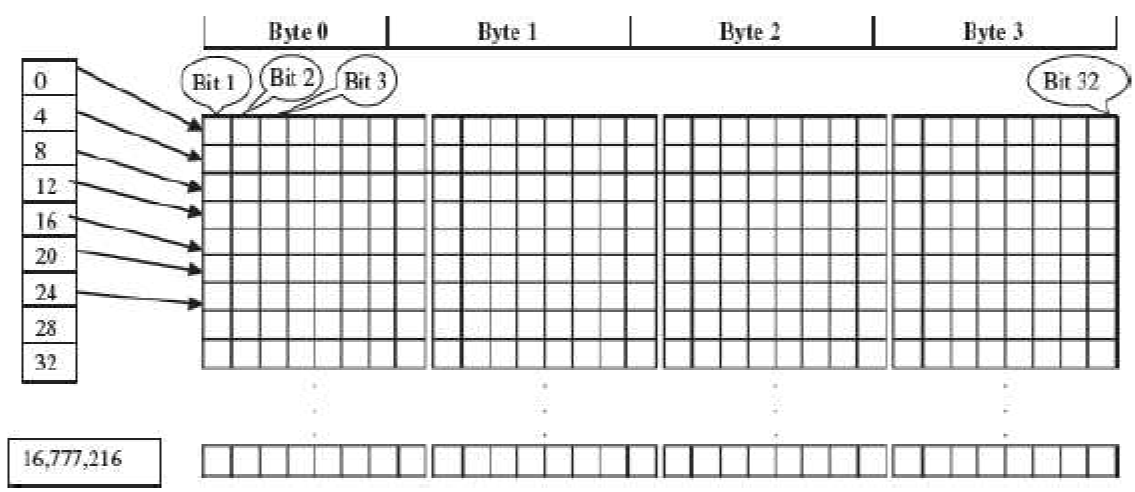
1 Kilobyte (KB) = 210 = 1024 bytes

1 Megabyte (MB) = 220 = 1024KB

1 Gigabyte (GB) = 230 = 1024 MB = 1024 *1024 KB

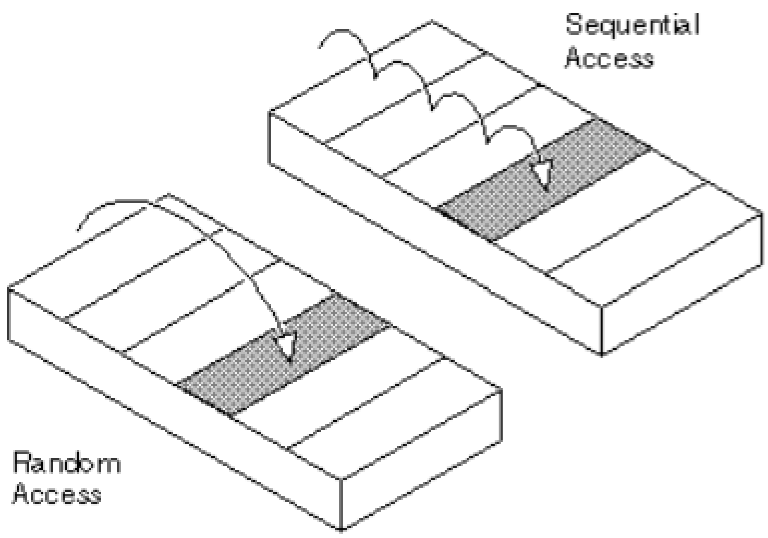
1 Terabyte (TB) = 240= 1024 GB = 1024 * 1024 *1024 KB

Memory is logically organized as a linear array of locations. For a processor, the range of the memory addresses is 0 to the maximum size of memory. Figure below shows the organization of a 16 MB block of memory for a processor with a 32-bit word length.



Access types of storage devices:

Access type of any storage device depends on how data can be read and written into it. Generally, the information stored in storage devices can be accessed in two ways: Sequential Access and Random Access (Direct Access).



Sequential Access Devices:

- Sequential access means that computer must run through the data in sequence, starting from the beginning, in order to locate a particular piece of data.
- Magnetic tape is an example of sequential access device.
- Let us suppose that magnetic tape consists of 80 records. To access the 25th record, the computer starts from first record, then reaches second, third etc. until it reaches the 25th record. Sequential access devices are generally slow devices.

Direct Access Devices:

- Direct access devices are the ones in which any piece of data can be retrieved in a non-sequential manner by locating it using the data's address. It accesses the data directly, from a desired location.
- Magnetic disks and optical disks are examples of direct access devices. There is no predefined order in which one can read and write data from a direct access device.
- In a magnetic disk consisting of 80 records, to access the 25th record, the computer can directly access the 25th record, without going past the first 24 records.
- Based on access, magnetic tapes are sequential access devices, and, magnetic disks, optical disk and magneto-optical disks are direct access devices.

Using the Computer Memory:

The computer starts using the memory from the moment the computer is switched on, till the time it is switched off. The list of steps that the computer performs from the time it is switched on are:

- Turn the computer on.
- The computer loads data from ROM. It makes sure that all the major components of the computer are functioning properly.
- The computer loads the OS from the hard drive into the system's RAM. CPU has immediate access to the OS as the critical parts of the OS are maintained in RAM as long as the computer is on. This enhances the performance and functionality of the overall system.
- Now the system is ready for use.
- When we load or open an application it is loaded in the RAM. Since the CPU looks for information in the RAM, any data and instructions that are required for processing (read, write or update) is brought into RAM.
- To conserve RAM usage, many applications load only the essential parts of the program initially and then load other pieces as needed. Any files that are opened for use in that application are also loaded into RAM.
- The CPU requests the data it needs from RAM, processes it and writes new data back to RAM in a continuous cycle. The shuffling of data between the CPU and RAM happens millions of times every second.
- When you save a file and close the application, the file is written to the secondary memory as specified by you. The application and any accompanying files usually get deleted from RAM to make space for new data.
- If the files are not saved to a storage device before being closed, they are lost.

<u>Note:</u> Sometimes, when you write a program and the power goes off, your program is lost if you have not saved it. This is because your program was in the RAM and was not saved on the secondary memory; the content of the RAM gets erased when the power is switched off.

Input-Output Unit:

An I/O unit is a component of computer. The I/O unit is composed of two parts—input unit and output unit. The input unit is responsible for providing input to the computer and the output unit is for receiving output from the computer.

Input Devices:

Input devices allow users and other applications to input data into the computer, for processing. The data input to a computer can be in the form of text, audio, video, etc. The data is entered manually by the user or with minimal user intervention. Input devices are classified as follows:

<u>Human data entry devices</u>

- Keyboard
- Pointing devices—mouse, trackball, joystick, digitizing tablet
- Pick devices—light pen, touch screen

Source data entry devices

- Audio input—speech recognition
- Video input—digital camera
- Scanner—hand-held scanner, flat-bed scanner
- Optical Scanner—Optical Character Recognition (OCR), Optical Mark Recognition (OMR), barcode reader

Output Devices

Output devices provide output to the user, which is generated after processing the input data. The processed data, presented to the user via the output devices could be text, graphics, audio or video. The output could be on a paper or on a film in a tangible form, or, in an intangible form as audio, video and electronic form. Output devices are classified as follows:

Hard Copy Devices

- Printer
- Plotter
- Computer Output on Microfilm

Soft Copy Devices

- Monitor
- Visual Display Terminal
- Video Output
- Audio Response

I/O Port

The peripheral devices can be connected to computer in several ways. Devices such as network adapters and sound cards are connected to expansion slots inside the computer. Printers and scanners are connected to ports on the backside of the computer. Also, in a portable computer, the PC Card connects to the PC Card slot on it.

The I/O ports are the external interfaces that are used to connect input and output devices like printer, modem and joystick to the computer. The I/O devices are connected to the computer via the serial and parallel ports, Universal Serial Bus (USB) port, etc.

Parallel Port

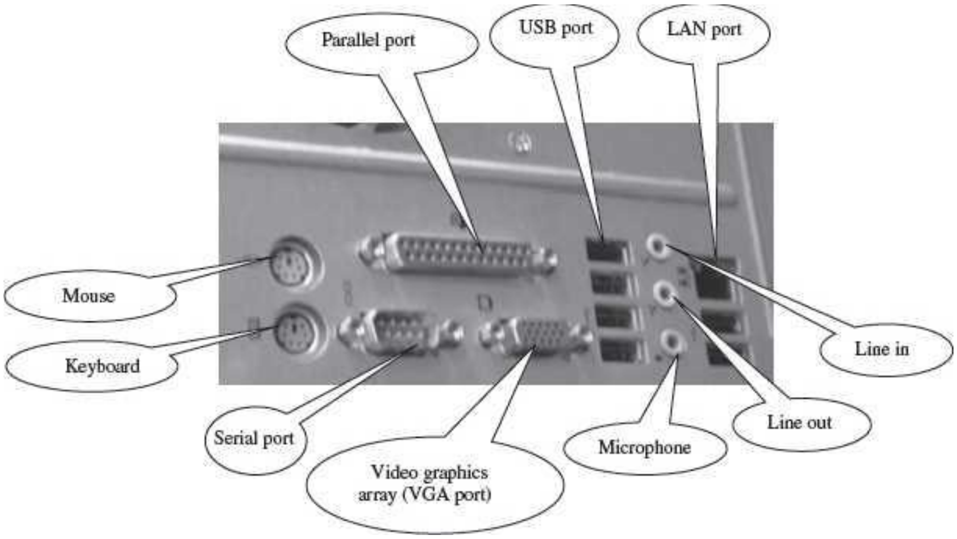
A parallel port is an interface for connecting eight or more data wires. The data flows through the eight wires simultaneously. They can transmit eight bits of data in parallel. As a result, parallel ports provide high speed data transmission. Parallel port is used to connect printer to the computer.

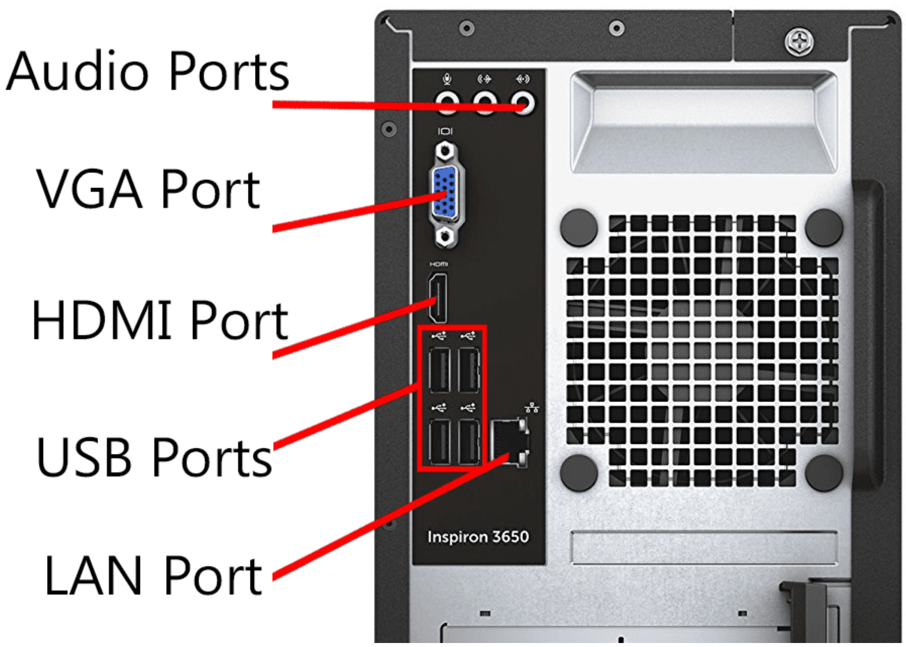
Serial Port

A serial port transmits one bit of data through a single wire. Since data is transmitted serially as single bits, serial ports provide slow speed data transmission. Serial port is used to connect external modems, plotters, barcode reader, etc.

USB Port

Nowadays, USB is a common and popular external port available with computers. Normally, two to four USB ports are provided on a PC. USB allows different devices to be connected to the computer without requiring re-boot of the computer. USB also has the plug and play feature which allows devices ready to be run simply by plugging them to the USB port.





Working of I/O System:

The working of I/O system combines I/O hardware and I/O software. The I/O hardware includes ports, buses and device controllers for different devices, and I/O devices. The I/O software is the device driver software that may be embedded with operating system or comes with each device.

The working of I/O system is described as follows:

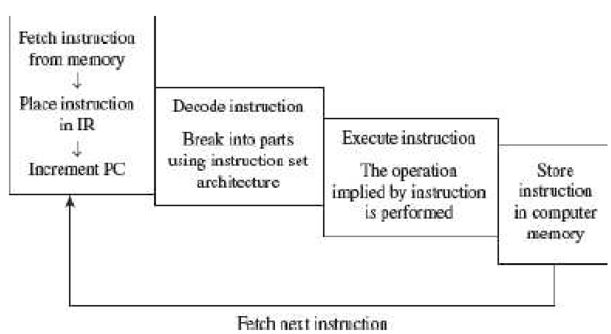
- I/O Devices are attached to computer via the ports of computer. There are many standard ports available on the backside of the computer case like serial port and parallel port. If one or more devices use a common set of wires, it is called a bus. For example, Peripheral Component Interconnect Bus (PCI bus), PCI Express bus, etc.
- Device Controller operates on a bus, a port or a device. It controls the signals on the wires of port or bus. The controllers have one or more registers for data and control signals. Controller may be simple like a serial port controller for a serial port, or, complex. Some devices have their own built-in controllers.
- Device Driver is software via which the operating system communicates with the device controllers. Each device has its own device driver, and a device controller which is specific to the device. The device drivers hide the differences among the different device controller and present a uniform interface to the operating system.
- Application programs use an I/O device by issuing commands and exchanging data with the device driver. The device driver provides correct commands to the controller, interprets the controller register, and transfers data to and from device controller registers as required for the correct device operation.
- Operating system ------ Device Drivers -------- Device Controllers -------- Devices

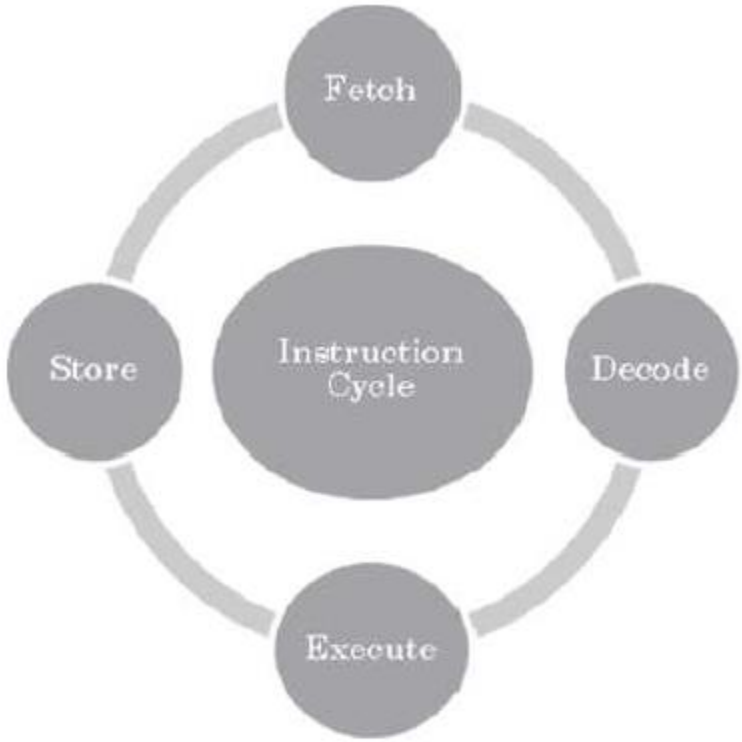
Instruction Cycle:

The primary responsibility of a computer processor is to execute a sequential set of instructions that constitute a program. CPU executes each instruction in a series of steps, called instruction cycle.

An instruction cycle involves four steps:

- Fetching The processor fetches the instruction from the memory. The fetched instruction is placed in the Instruction Register. Program Counter holds the address of next instruction to be fetched and is incremented after each fetch.
- Decoding The instruction that is fetched is broken down into parts or decoded. The instruction is translated into commands so that they correspond to those in the CPU’s instruction set. The instruction set architecture of the CPU defines the way in which an instruction is decoded.
- Executing The decoded instruction or the command is executed. CPU performs the operation implied by the program instruction. For example, if it is an ADD instruction, addition is performed.
- Storing CPU writes back the results of execution, to the computer’s memory.





Random Access Memory (RAM) types:

RAM has two primary forms:

- Static RAM (SRAM)
- Dynamic RAM (DRAM)

Dynamic RAM:

DRAM is widely used as a computer's main memory. Each DRAM memory cell is made up of a transistor and a capacitor within an integrated circuit, and a data bit is stored in the capacitor.

Since transistors always leak a small amount, the capacitors will slowly discharge, causing information stored in it to drain; hence, DRAM has to be refreshed (given a new electronic charge) every few milliseconds to retain data.

Static RAM:

SRAM is made up of four to six transistors. It keeps data in the memory as long as power is supplied to the system unlike DRAM, which has to be refreshed periodically.

As such, SRAM is faster but also more expensive, making DRAM the more prevalent memory in computer systems. SRAMs are used to build cache memory.

Basis for Comparison	Static RAM	Dynamic RAM
Speed (Access Time)	Faster	Slower
Physical Size	Small	Large
Storage	Smaller	Larger
Cost	Expensive	Cheap
Construction	Complex and uses large no. of transistors	Simple and uses a single transistor and capacitor
Charge leakage	Not present	Present and hence requires continuous power refresh
Power Consumption	Low	High
Application	Cache memory	Main memory

Read Only Memory (ROM) types:

The basic types of ROM are:

- Programmable Read Only Memory (PROM)
- Erasable Programmable Read Only Memory (EPROM)
- Electrically Erasable and Programmable Read Only Memory (EEPROM)

PROM - Programmable Read Only Memory

PROM is read-only memory chip that data can be written only once by a user. The difference between it and the read only memory is that PROM is manufactured as a blank memory, while the ROM is programmed during the manufacturing process.

The user buys a PROM, the user will need a special device called a PROM programmer or PROM burner to write the desired data onto the blank PROM chip. The process of programming a PROM is sometimes called burning the PROM. The memory can be programmed just once after manufacturing, which is an irreversible process.

EPROM - Erasable Programmable Read Only Memory

EPROM is a special kind of read only memory chip that has the opportunity to erase the programmed data, which the feature can be seen from its name. The programmable read-only memory can be programmed to write data with high voltage, and the data remains until it is exposed to ultraviolet light for lasting up to 10 minutes or longer.

Usually, an EPROM eraser can achieve this purpose, making it possible to reprogram the memory.

EEPROM - Electrically Erasable and Programmable Read Only Memory

EEPROM is also a kind of read only memory that the principle of operation is similar to EPROM, but the ways to program and erase are done by exposing it to an electrical charge.

It can be erased and reprogrammed about 10,000 times. Both erasing and programming take about 4 to 10 milliseconds. In the EEPROM, users can selectively erase and program any location and it can be erased one byte at a time instead of being erased the entire chip. Therefore, the process of reprogramming can be flexible but slow.