CHAPTER 2

COMPUTER ARCHITECTURE

OVERVIEW:

In 1951 John Von Neumann and his team proposed a design of stored program computer. According to the design:

- A sequence of instructions is called a program.
- All the data and the programs are stored in memory.
- The machine reads all the instructions one by one and executes these instructions sequentially.

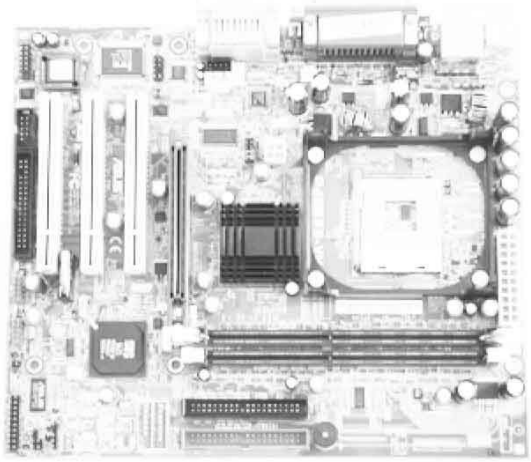
Q1. What is Computer Architecture? Describe the components of Architecture.

COMPUTER ARCHITECTURE:

- The organization and interconnection of various computer components is called computer architecture.
- The important components are: I/O devices, CPU, Primary or Main Memory, Secondary Memory or Storage devices, I/O unit, Registers and Bus Interconnection

I/O devices:

- It stands for input and output devices.
- Input devices: User can communicate with the computer through input devices e.g. Keyboard, Mouse, joystick, track-ball, Microphone, camera, scanner, bar code reader etc.
- Output devices are used to communicate with the user e.g. Monitor, Speakers, Printers and Projectors etc.



CPU:

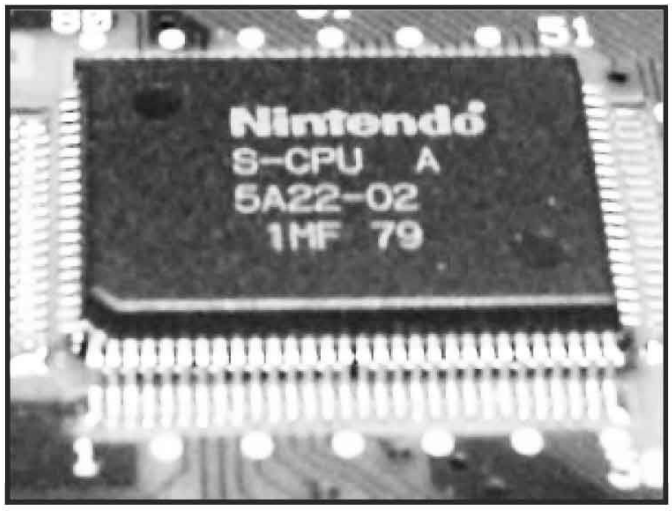
- It stands for "central processing unit".
- It is a small chip mounted on the circuit board (mother board).
- CPU is the brain of computer. It controls all the functions of computer system.
- CPU also performs data processing.
- There are two parts of a CPU, CU and ALU

CU:

- It stands for "Control Unit". The CU reads the instruction from memory, decodes and executes these instructions.
- It acts as supervisor of the computer. It controls all the parts and activities of the computer.
- CU generates electronic signals that direct the computer to execute the programs it doesn't itself execute the instructions rather it directs the other parts to do so.
- It also controls the flow of information and co-ordinates the activities of other devices.

ALU:

- It stands for "Arithmetic and Logical Unit".
- It performs two types of operations on data.
- Arithmetic operations (addition, subtraction, multiplication and division)
- Logical operations (the operations produces true or false results e.g. less than, greater than, equal to not equal to etc.)



Main Memory or Primary Storage:

- All the program instructions and data are loaded in memory for the execution.
- This unit temporarily stores the data and instructions.
- It is also called working area of the computer. A computer cannot work without memory.
- There are two types of memory: RAM (Random Access Memory) and ROM (Read Only Memory)

Secondary Storage devices:

- These devices are used to store data permanently in the computer for later use.
- These are CD's, Hard Disk, and Floppy Disk etc.

I/O unit:

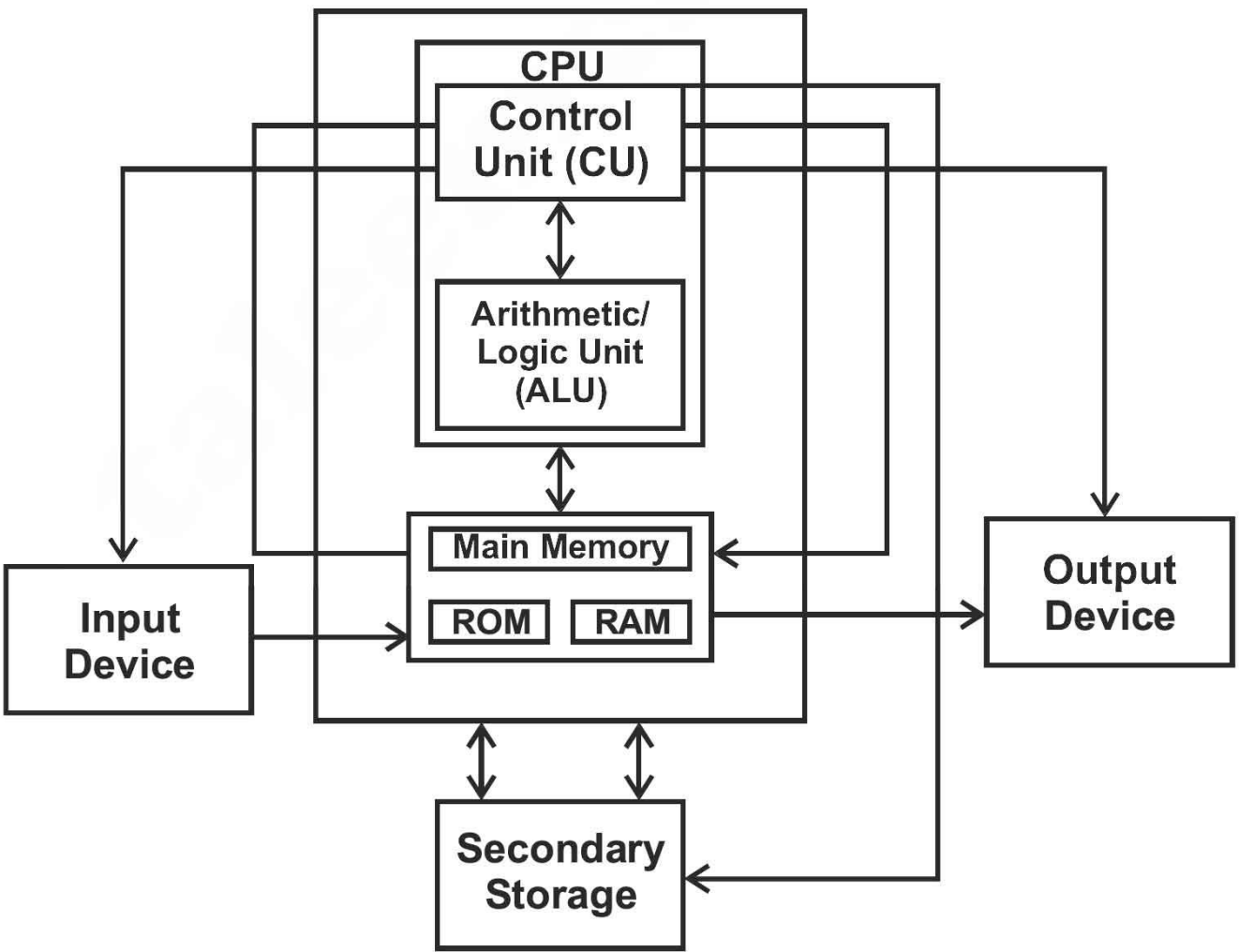
- It helps the processor to communicate with its peripheral devices.
- These are disk drives, monitor, printer etc.
- I/O unit controls different devices attached with computers.

Register:

- These are the high-speed memory locations built into the microprocessor. The CPU uses these locations to store data and instructions temporarily to perform certain operations.
- Data is processed and transferred from one component to another with the help of registers.

Bus Interconnection:

- It is used to connect different parts of the computer. It is a communication channel all the data instructions and commands flow through the bus interconnection.



Q2. What is Computer Memory? Describe its types.

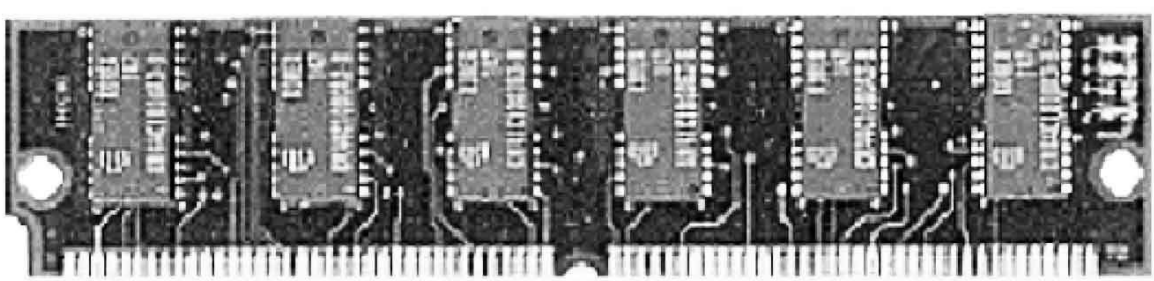
MAIN MEMORY:

- It is called working area of the computer. All the programs and data are stored in memory.
- A computer cannot work without main memory.
- It is a chip of semi-conductor material.
- Memory consists of identical cells. These cells are logically organized into a group of 8-bits (1-Byte). Each memory cell has a unique address.
- Data and instructions are stored in these cells in the form of bits.
- It is possible to randomly select any memory cell directly to store and retrieve data and instructions because each location of memory is accessible in equal time.
- It is faster than other storage devices.
- **Types of Memory:**

  RAM (Random Access Memory) and ROM (Read Only Memory)

RAM (Random Access Memory)

- Primary storage is also called RAM (**Random Access Memory**).
- It is used as read/write memory. It is also called volatile and user memory.
- The contents of RAM are lost if the electric supply is switched-off.
- Random Access Memory or RAM is the memory that the computer uses to temporarily store the information as it is being processed. The more information being processed the more RAM computer needs.
- RAM memory chips are available in different sizes and speeds, which are from 64MB to 2GB.
- As computer technology changes the type of memory changes and making old memory chips obsolete.
- It improves the data processing speed of the computer.



Types of RAM:

SRAM, DRAM

SRAM:

- It stands for static random access memory.
- No refreshing is required in SRAM.
- These chips are faster than DRAM.
- SRAM uses less power than DRAM.
- Its design is more complex than DRAM.
- SRAM is more expensive than DRAM.

DRAM:

- It stands for dynamic random access memory.
- It is slower than SRAM
- It requires refreshing.
- Its design is simple and is not expensive.
- **EDO-DRAM:** EDO-DRAM is an example of DRAM. It stands for Enhanced Data output Dynamic Random Access Memory. It is 50% faster than ordinary DRAM

CACHE MEMORY:

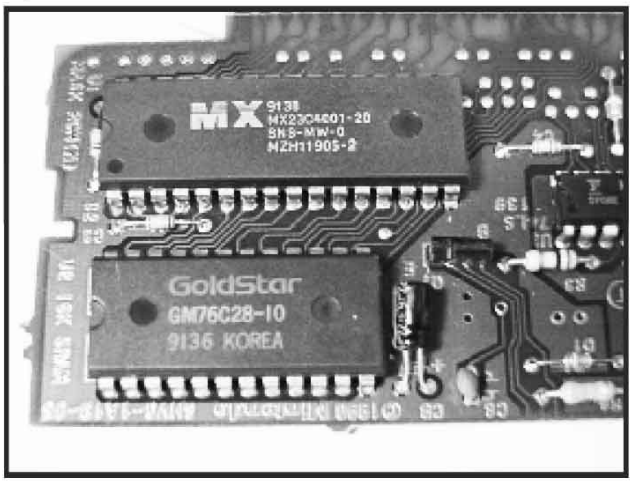
- It is very high speed memory. It is used to increase the speed of processing
- It is small memory between CPU and main memory.
- Its access time is close to the processing speed of CPU.
- It is used to store the frequently used programs and data.
- By making active programs and data available to the CPU at a rapid rate, it is possible to increase the performance of a CPU.

ROM (Read Only Memory):

- The contents of ROM are read only.
- New data cannot be written on it.
- Its contents are permanent.
- The system manufacture writes the ROM instructions and data at the time of manufacturing and the data and instructions cannot be changed afterwards.
- It is also called non-volatile, or system memory.
- When the power is switched off, the contents of ROM do not lost.
- As ROM is supplied by the manufacturer therefore it is impossible for the users to modify the stored programs in ROM.

- ROM chips are used to store frequently used programs like operating system routines and data which is not supposed to be change for a long time.
- **Types of ROM:**

  PROM, EPROM, EEPROM



PROM:

- It stands for programmable read only memory.
- It is initially blank and the manufacturer writes instructions and data with some special devices.
- Once the information is stored onto PROM, then it cannot be changed.

EPROM:

- It stands for erasable programmable read only memory.
- It is another type of ROM chip and it is possible to erase the stored information and the chip can be reprogrammed using special devices.
- Information can be erased using ultraviolet rays.

EEPROM:

- It stands for electrically erasable programmable read only memory.
- It is a kind of ROM chip that can be re-written using electrical devices.
- The stored information on EEPROM can be erased, modified or reprogrammed easily using special devices.

Q3. What is Computer Bus? Describe its types.

COMPUTER BUS:

- It is a set of parallel lines used to connect two or more devices of a digital computer.
- It is the most important component of computer architecture. A computer has more than one bus interconnection.

- All the components of a computer are connected with a set of parallel lines. All these lines are used to transfer data in the form of bits from one component to another component. These lines are called BUS.
- There are two types of buses: System Bus and Expansion Bus

SYSTEM BUS:

- System bus is used to connect main components of the computer.
- Generally there are 70-100 parallel lines in system bus.
- It is divided into three main categories.
  1. Control Bus
  2. Address Bus
  3. Data Bus

Control Bus:

- These lines are used to transfer control signals from one component to another.
- It specifies the type of operation that is to be performed.
- It also transmits the control signals like ACKS (acknowledgement signal). When a CPU gives command to the memory for writing data, then the memory sends an acknowledgement signal to the CPU after successful writing of data.
- A few commonly control signals are:

Command/ Control Signal	Meaning
MEMORY WRITE	It is used to write data to a given memory location
MEMORY READ	It is used to read data from a given memory location
I/O WRITE	It is used to write some data on output device.
I/O READ	It is used to read some data from input device.
BUS REQUEST	It is used to request a control on the bus so that the requesting device can use it to transmit data.
BUS GRANT	It is used by the bus controller to indicate the grant of the bus to a device.
TRANSFER ASK	This command is used to request for data transfer.

ADDRESS BUS:

- It is a part of system bus. It is used to carry address signals to read and write data in the memory.
- Address bus is uni-directional.
- An address is a unique ID of each component connected to the system bus. It is called address of the component.
- When a component of a computer wants to communicate with another, it uses a few system bus lines to specify the address of destination; these lines are called address bus.

Data Bus:

- Data bus is used to transfer the data from one component to another.
- There are 32 or 64 parallel lines of data bus.
- The amount of data that a bus can transmit is called **bus-width.** A 64-line data bus can transmits 64 bits (8-bytes) at a time.
- If more lines are present in the bus, it can carry more data.
- Width of data bus has direct impact on the performance of the computer.
- Data bus is bi-directional.

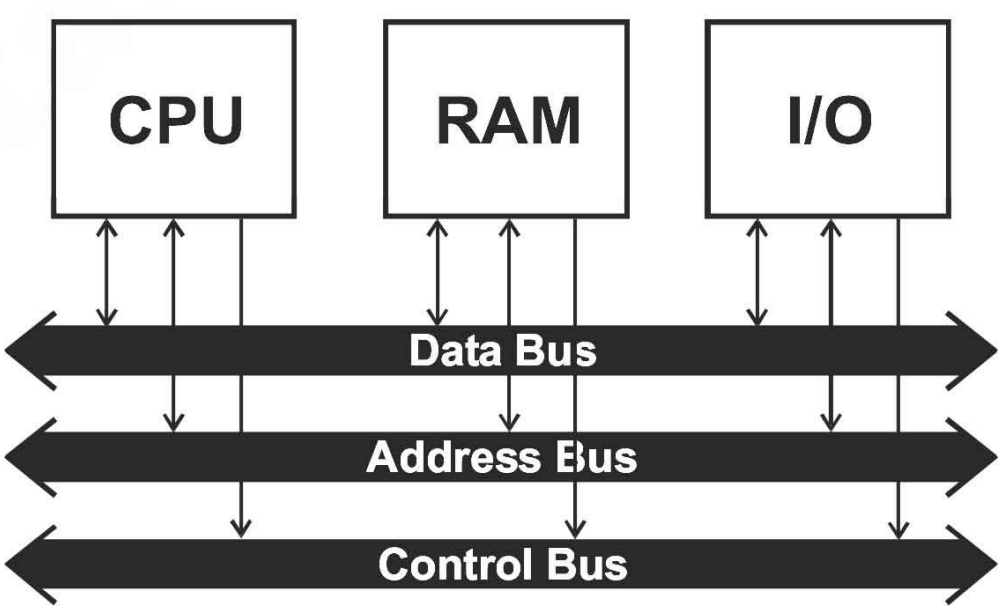
Expansion Bus:

- The major components are connected to the system bus and the remaining components are connected to expansion bus and the expansion bus is connected to the system bus.
- If all the components are attached with system bus then it will slow down the computer. All the components will have to wait longer to get access to the bus. We use the expansion bus to solve this problem.

Q4. What is I/O Unit? Describe Interrupts and DMA.

I/O UNIT:

- It handles the processor's communication with its peripheral devices.
- It is the part of a computer that is used to control the I/O devices.
- The I/O unit is responsible for controlling the different devices attached with it.
- It is also responsible for the compensation of speed difference between processor and I/O devices.
- Only the I/O unit is connected to the system bus and the remaining components are attached with I/O unit.
- There are two ways to transfer data from peripheral devices to the CPU. Interrupts and DMA



INTERRUPTS:

- Interrupts are the signals generated by I/O devices. These signals inform the CPU about the occurrence of certain events such as completion of an I/O operation. The CPU stops all operation and control is transferred to interrupting device.

- The disadvantage of this scheme is that it reduces the overall performance of the processor.

DMA:

- It stands for direct memory access. It is a technique of performing an I/O operation in which I/O operation is completed without involving the CPU.

- The CPU just issues an I/O command and the rest of the work is completed with the help of a special DMA module (DMA controller).

- CPU then continuous to do the other task when I/O operation is completed the DMA module sends a interrupt signal to CPU. Thus the CPU is involved in the beginning and at the end of I/O operation and the performance of the CPU is increased.

- The disadvantage of DMA scheme is that it is more complex and more hardware is required.

Q5. What are Registers? Describe register types.

CPU REGISTERS:

- These are the high speed memory locations built into the microprocessor. The CPU uses these locations to store data and instructions temporarily for processing.

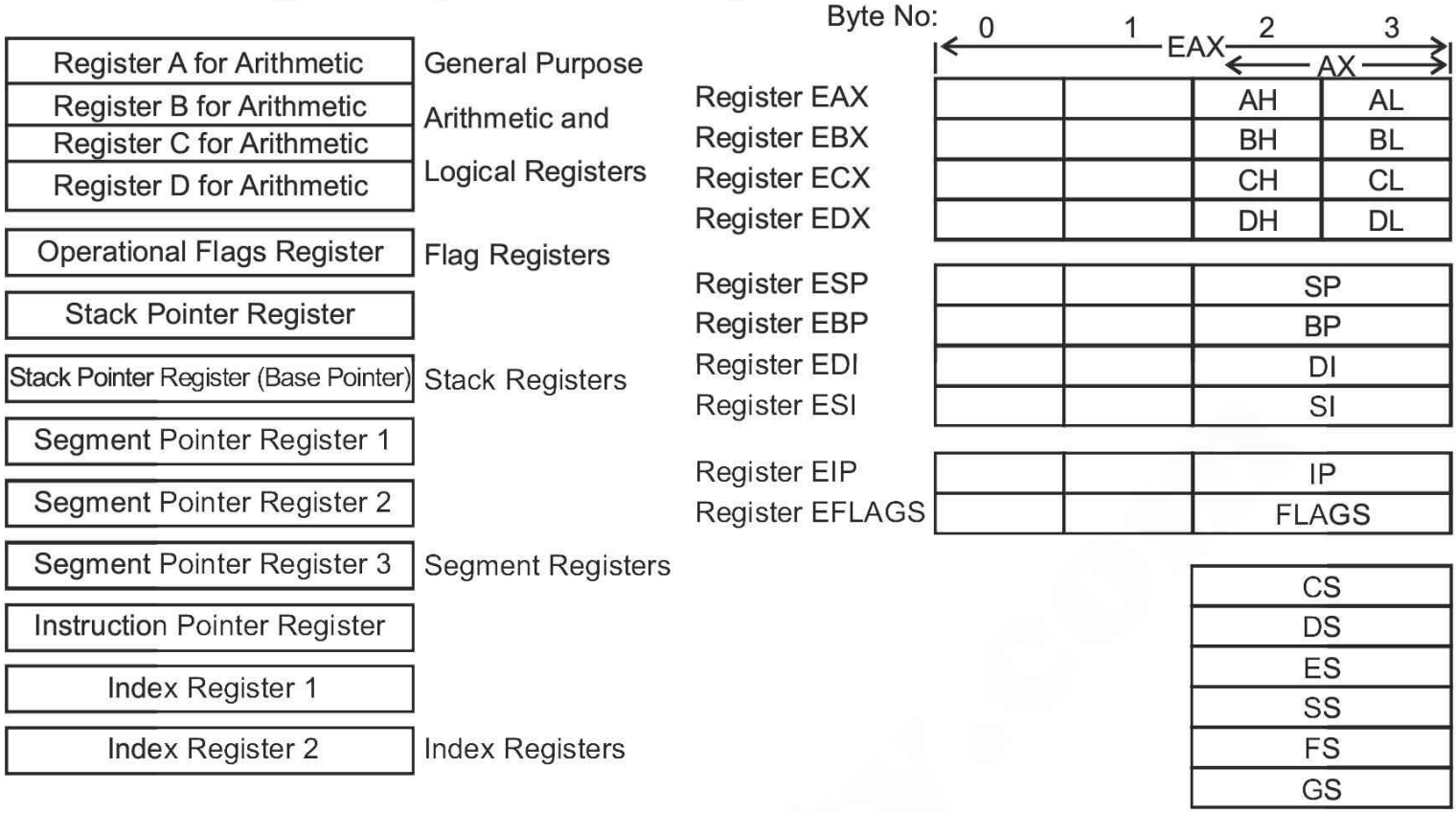
- CPU processes, stores and transfers data from one component to another with the help of registers.

- The number of register varies among computers.

- It increases the performance of CPU.

- Each register receives the information, hold it temporarily and pass it on, as directed by CU.

- The size of registers depends on the computer architecture.



- **Types of Registers:**
  - PC (Program Counter)
  - IR (Instruction Register)
  - MAR (Memory Address Register)
  - MBR (Memory Buffer Register)
  - IP (Instruction Pointer)
  - SP (Stack Pointer Register)
  - GPR (General Purpose Registers A, B, C, D)
  - Address or segment Registers (CS, DS, ES, SS)
  - Index Registers (DI, SI)
  - flag register

PC (Program Counter):

- It holds the address of the next instruction to be fetched and executed.
- When instruction is fetched, the value of PC is automatically incremented and it points to the address of next instruction.

IR (Instruction Register):

- It holds the current instruction that is being executed.

MAR (Memory Address Register):

- It holds the address of active memory location.
- When CPU wants to stores or read data from the memory, CPU stores the required address of memory location in MAR.

MBR (Memory Buffer Register):

- It holds the contents of the memory location read from or written in the memory.

SP (Stack Pointer Register):

- Arrangement of data is called **stack** and data is stored in or retrieved from stack using LIFO (last in first out).
- LIFO is also known as FILO (first in last out). PUSH and POP instructions are usually used in stack to store or retrieve the values.
- SP register is used to manage the stack and it stores the value of top of stack.

GPR (General Purpose Registers A, B, C, D):

- These registers are used to perform arithmetic and logical operations.
- EAX, EBX, ECX, EDX are 4 byte registers. AX, BX, CX, and DX are 2 byte registers. AH, AL, BH, BL, CH, CL, DH and DL are 1 byte registers.
- There are four general purpose registers.

AX stands for Accumulator Register:

- It is used for arithmetic and logical operations.
- It holds the initial data to be operated upon, and the final results of processing.
- The results of arithmetic operations are returned to accumulator for transfer to the main memory through buffer register.

BX stands for Base Register:

- It is used for arithmetic and data movement operations.

CX stands for Counter Register:

- It is used for counting purpose.
- It acts as counter in loops.

DX stands for Data Register:

- It has special role in division and multiplication.
- The size of these registers is from 1 to 4 bytes.

Address or Segment Registers (CS, DS, ES, SS) :

- It is a group of 4 registers CS, DS, ES and SS.
- Size of each segment register is 2 bytes.

CS (Code Segment) Register:

- It holds the base location of all executable instructions in the program.
- CS along with IP register fetches the next instruction.

DS (Data Segment) Register:

- It is the default base location for memory variables.
- DS points to data in memory using DI or SI registers.

- CPU calculates the offset of variables using the current value of DS.

ES (Extra Segment) Register:

- It is an additional base location for memory variables.

SS (Stack segment) Register:

- It contains the base location of the current program stack.

IP (Instruction Pointer) Register:

- CS along with IP register fetches the next instruction.

Index Registers (DI, SI):

- DS points to data in memory using DI (Destination Index) or SI (Source Index) registers.

Q6. What is an Instruction set and explain types of Instructions?

TYPES OF INSTRUCTIONS:

- Each CPU provides a number of instructions to perform different operations.
- The set of all the instructions provided by CPU is called **Instruction Set.**
- The instruction set increases the performance of a CPU.
- There are four types of instructions provided by a CPU.

  Data transfer instructions

  Arithmetic and logical instructions

  I/O instructions

  Control Transfer instructions

Data Transfer Instructions:

- These instructions are used to transfer data from or to the memory.
- The professionals use these instructions to bring the data into CPU and copy data from CPU to the memory.

Arithmetic and Logical Instructions:

- CPU can perform arithmetic and logical operation with the help of arithmetic and logical instructions.
- Arithmetic instructions are addition, subtraction, multiplication and division.
- These operations are performed on signed, unsigned and floating point numbers.
- Logical operation is like comparison of two numbers. XOR shifting and rotating of numbers are also logical operations and logical instructions perform logical operation.

I/O Instructions:

- Each CPU provides instructions to read data from peripheral devices and writing data to peripheral devices.
- These instructions are used to read and write data to and from the Input/Output devices e.g. a programmer uses input and output commands e.g. input and print commands are used for these operations

Control Transfer Instructions:

- Each CPU provides instructions to control the flow of operations.
- These instructions are used to execute the instructions repeatedly and the programmer uses these instructions to transfer the control from one instruction to another e.g. jump, jumpz (jump if zero).

Q7. What is an Instruction format? Explain different types of Instructions format.

INSTRUCTION FORMAT:

- The CPU fetches the instruction from memory and decodes it and executes the instruction according to the requirement of the instructions.
- The structure of instruction is called instruction format. Each instruction has a specific format and generally it consists of:

opcode: A code for the instruction indicates the type of action and is 4-bit code

Address of the operands: These are the addresses of data locations on which CPU performs an action and each address is 6-bits.

opcode(4-bits) operand referce1(6-bits) operand reference2(6-bits)

The different address formats are:

1. Zero Address format
2. One address format
3. Two address format
4. Three address format

Zero Address format:

- It does not use any address field in the computational instruction.
- ADD and MUL instructions don't use any address field of the stack (set of memory location); however, address field is necessary for PUSH and POP instructions.

One address format:

- The instruction with one address format uses only one address field. It uses accumulator AC register for all types of data manipulation.

Two address format:

- The instruction with two address format uses two address fields. Each address field can specify a register or a memory location.
- Examples are: ADD, MOV, CMP, BIS etc.

Three address format:

- The instruction with three address format uses three address fields. Each address field can specify a register or a memory location.
- A program with arithmetic expressions produces fast and better results.
- A binary coded instruction requires too many bits to specify three addresses.
- These instruction formats consists of either three register address fields or two registers and one memory address field.

Q8. What is Fetch-Decode and Execute Cycle? Explain different steps involved in this process.

FETCH-DECODE AND EXECUTE CYCLE:

- This cycle describes the process of execution of an instruction within the computer.
- One round of steps from getting an instruction back to getting the next instruction is called the **Machine Cycle**.

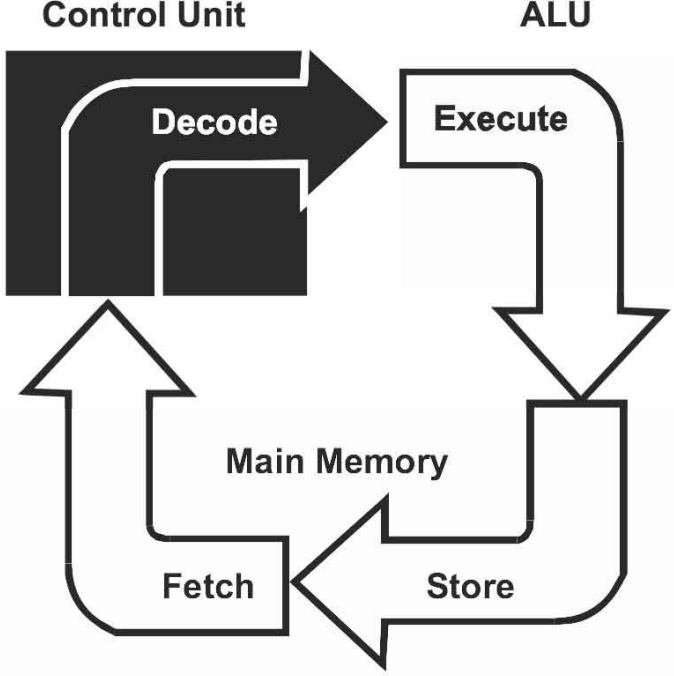
The Machine Cycle

Fetch - Get an instruction from Main Memory

Decode - Translate it into computer commands

Execute - Actually process the command

Store - Write the result to Main Memory



Fetch Instruction:

- The CPU reads the value of PC and the instruction is copied into the instruction register (IR). This process involves two following steps:
  1 Copy the contents of PC into MAR and the request for memory read.
  2. Copy the data from the memory to MBR and instruction is copied into IR.
- After completing the above steps the value of PC is incremented and it points to the address of next instruction.

Decode Instruction:

- Decoding means to activate the appropriate circuit to execute the instruction.
- After completing the fetch process the CU decodes the instruction by analyzing the opcode of the instruction.
- CU also reads the values of operands specified in the instruction.

Execute Instruction:

- The process of performing an action on the decoded instruction is called execution.
- After decoding the CPU executes the instruction by using the activated circuit.
- After execution the results are copied into the registers and memory.

Q9. What is an Operating system? Explain different functions of OS.

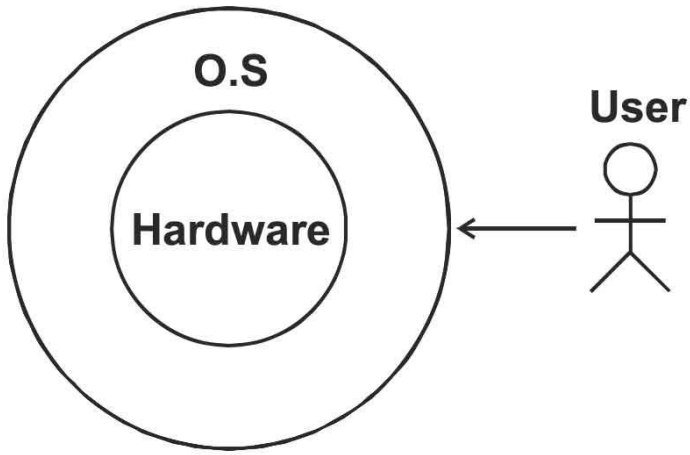
OPERATING SYSTEM:

- It is a collection of operative programs that provides an interface for the user to interact with the computer.

OR

- It is a set of programs running in the background on a computer system and providing an environment in which other programs can be executed and the computer system can be used efficiently.

- Important operating system are; windows, DOS, Unix, Linux, Solaris etc.



Functions of Operating System:

Manage Hardware Resources:

- OS must provide programs to manage the hardware resources of the computer like disk, memory and CPU.

Memory Management:

- Every program must be loaded in the computer memory during execution and if several programs are loaded simultaneously in the memory then it is the job of operating system to run each program individually.

  For multi-programming time-sharing technique is used.

  The operating system is responsible for the protection of data of one program from the action of other programs.

Load and Execute Programs:

- Operating system provides the facility of easily loading programs in the main memory and starts its execution.

Booting:

- Booting is a process of starting or restarting the computer. Operating system starts the compute to work. It checks the computer and makes it ready to work.

Data Security:

- OS must protect the user's data from an illegal access and modification.

Interface to the Users:

- It is an interface between user and computer and also between software and computer.
- There are two type of interfaces:
- Command-line interface
- Graphical user interface

Command-line interface (CLI):

- In command line interface users communicate with the system by typing commands using keyboard or
- It provides a command prompt to the users for typing commands to interact with the computer.
- Each command activates a program of the operating system.
- Examples are: MS-DOS, early versions of UNIX, Novell Netware etc.

Graphical user interface (GUI):

- A GUI operating system provides a graphical user interface to establish the user communication with the computer.
- It consists of windows, menus, icons and pointers.
- The user issues the commands by selecting different object with the help of pointing devices.
- MS-Windows 95, Windows 98, Windows XP are the examples of GUI operating system.

Q10. What are Programming Languages? Explain different Types of Programming Languages.

PROGRAMMING LANGUAGE:

- It is the way of communication between User and Computer.
- It is used to write computer programs.
- A language provides the rules to write an instruction is called its **syntax**. Each language has its own syntax.

Types of Programming Languages:

There are two types of computer languages.

1. Low Level Language
2. High Level Language

Low Level Language (LLL):

- The language which is close to machine language is called low level language.
- A Computer can easily understand low-level language.
- The low level language requires a deep understanding of the machine architecture.

There are two types of low-level language.

1. Machine language
2. Assembly language

Machine language:

- It is the native language of computer.
- The computer does not need any translator to understand it.
- Every machine language instruction consists of 0's and 1's.
- It is difficult for human beings to understand and learn it.
- It is difficult to locate and remove errors in the program.

- It is directly executed by the computer.
- The machine language programs are machine dependent.

Assembly language:

- Machine language instructions(0's and 1's) are replaced with English like words known as mnemonics (Ne-monics)
- It is also called symbolic language.
- An assembler is required to translate the assembly language programs into machine language.

High Level Language (HLL):

- It is closed to human language.
- It is user friendly language.
- User can learn and understand high level language easily.
- The instructions of HLL are written in English like statements.
- The programs of HLL are not directly executed on the computer.
- A language translator is required to translate the HLL into low level language.
- These translators are compilers and interpreters
- The programs written in HLL are machine independent.
- Each language has its own translator.
- HLL programs are easy to modify, debug and more reliable.
- The HLL does not require a deep understanding of the machine architecture.
- HLL describes a well defined way of writing programs.
- These languages are mostly used for writing application programs; Examples :
- C and C++ (used for writing System Software), Java (Java is equipped with strong Network features), COBOL (common business oriented language and used for business applications), PASCAL, FORTRAN (Formula Translation and used for mathematical problems),BASIC (Beginners all purposes symbolic instruction code)

Source Program:

- The program written in High Level Language is called source program.
- The computer does not understand the source code.
- The source code is converted into machine code and then it is directly executed on the computer.

Object Program:

- The program in a machine language is called object program.
- The computer understands the object code directly.

Q11. What are Language Translators? Discuss its different Types.

LANGUAGE PROCESSOR/TRANSLATOR:

- It is software that is used to translate the high level language programs (source code) into machine language (object code).
- Each language has its own translator. Only one type of translator is used in any language. There are three types of language processor.
  1. Compiler
  2. Interpreter
  3. Assembler

Compiler:

- The language translator translates the source code into object code and the whole program is translated at the same time.
- If a program contains errors then compiler can not convert the source code into machine code until all the errors are removed from the source program.

Interpreter:

- The language translator translates the source code into object code statement by statement.
- The working of interpreter is slower than the compiler.
- Any error in the program stops its execution. After removal of errors execution resumes.

Difference between compiler and interpreter:

- Compiler converts the whole program at a time and interpret / translate the source code instructions one by one.
- Compiler converts each instruction only once but the interpreter may translate an instruction several times.
- A compiled program runs faster than an interpreted program.

Assembler:

- The language translator translates the assembly language into machine code i.e. it produces the binary instructions for a given assembly language program.

Q.12 What is a port? Explain different types of ports.

Ans. Port:

- The peripheral devices are connected to the system unit through a special device called the port. It is an interface or connecting socket.
- It provides a standard way of communication between the computer and its input and output devices (or peripheral devices). For examples, keyboard, mouse, printer, monitor, and modem etc. are connected with the computer by plugging their connectors into ports.

Types of Ports:

There are three basic types of ports:

- Serial port.
- Parallel port.
- USB port.

Serial Ports:

- Serial ports provide connection for transmitting data one bit at a time. Serial ports are often referred to as communicating (COM) ports. The mouse, keyboard and modem are usually connected to the serial ports.

Parallel Port:

- A parallel port provides a connection for transmitting data 8-bits at a time. Therefore, it is eight times faster than serial port. The printers and scanners are connected to the parallel ports. Parallel ports are often referred to as Line Printer Ports (LPT).

USB:

- USB stands for Universal Serial Bus. It is a plug-and-play hardware interface. It is used to connect the peripherals such as keyboard, mouse, scanner, and modem etc. It allows up to 127 different peripheral devices to be connected to the computer.