Computer Organization (CO)

UNIT-I

JNTUK
R16
2-2-CSE

Syllabus: Basic structure of Computers:

- Functional Units
- Basic Operational concepts
- Bus structures
- System Software
- Performance
- The history of computer development

TEXT-BOOK:

1. Computer Organization, Carl Hamacher, Zvonks Vranesic, Safea Zaky, 5th Edition, McGraw Hill

(1) Computer types:

→ Computer is a fast electronic calculating machine that accepts digitized input information, processes it according to a list of internally stored instructions and produces the resulting output information.

→ The list of instructions is called a computer program, and the internal storage is called computer memory.

→ The different computer types are,

(i) Desktop computers
  - personal Computer

(ii) notebook Computers

(iii) WorkStations

(iv) Enterprise Systems

(v) Servers

(vi) Super Computers.

→ Desktop Computers have processing and storage units, visual display and audio output units, and a keyboard that can all be located easily on a home (or) office desk.

→ The ~~most~~ storage media includes,
- hard disks
- CD-ROMs etc.

→ The most common form of desktop computers is <u>personal computer</u>, which has found wide use in homes, schools, and business offices.

→ Portable <u>notebook computers</u> are a compact version of the personal computer with all of these components packaged into a single unit the size of a thin briefcase.

→ <u>Workstations</u> have more computational power than personal computers, used in engineering applications, especially for interactive design work.

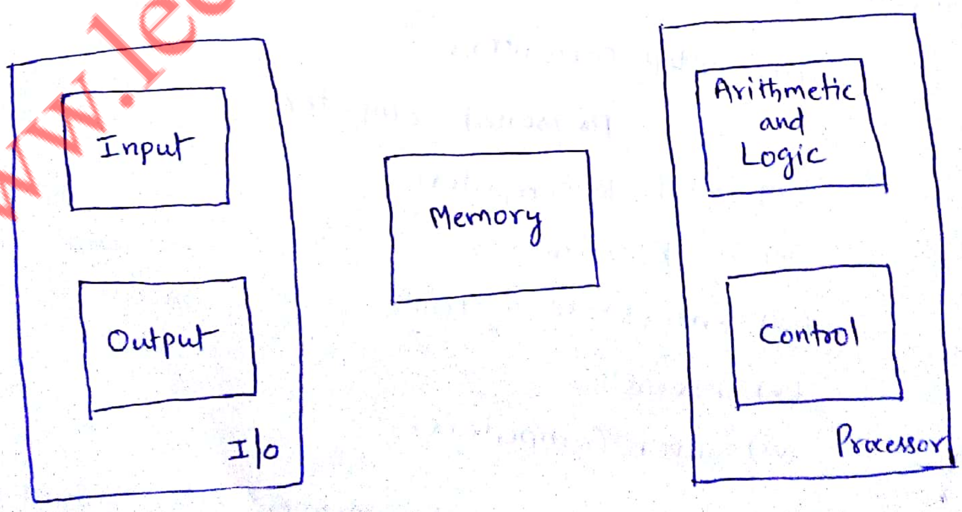
→ <u>Enterprise Systems</u> (or) <u>Mainframes</u>, are used for business data processing in medium to large corporations that require much more computing power and storage capacity than workstations can provide.

→ <u>Servers</u> contain sizable database storage units and are capable of handling large volumes of requests to access the data.

→ <u>Super Computers</u> are used for the large-scale numerical calculations required in applications such as weather forecasting and aircraft design and simulation.

② Functional Units:

→ <u>Basic Functional Units of a Computer</u> is depicted as,



→ A computer consists of five functionally independent main parts
- input
- memory
- Arithmetic and Logic
- output and
- control unit.

→ Instructions (or) Machine Instructions, are explicit commands that
- govern the transfer of information within a computer as well as between the computer and its I/O devices.
- specify the arithmetic and logic operations to be performed.

(i) Input Unit:

→ Computers accept coded information through input units, which read the data.
→ The most well-known input device is the Keyboard.
→ The other Input devices are,
- Mouse
- Joystick
- Scanner
- Touch Screen
- Light Pen, .. etc.

(ii) Memory Unit:

→ The function of the memory unit is to store programs and data.
→ There are two classes of storage
- Primary
- Secondary

→ Primary Storage is a fast memory that operates at electronic speeds.
→ Programs must be stored in the memory while they are being executed.
→ The Primary memory of a Computer is RAM (Random Access Memory)
→ The Secondary Storage is used when large amounts of data and many programs have to be stored, particularly for information that is accessed infrequently
→ The different Secondary storage devices are,
- HDD (Hard disk drive)
- FDD (Floppy disk drive)
- Magnetic disks and taps
- Optical disks (CD-ROM) - compact disk Read Only Memory

(iii) Arithmetic and Logic Unit :

→ Most computer operations are executed in the arithmetic and Logic Unit (ALU) of the Processor.

→ Consider an example, suppose two numbers located in the memory are to be added.

→ They are brought into the processor, and the actual addition is carried out by the ALU.

→ The sum may then be stored in the memory (or) retained in the processor for immediate use.

(iv) Output Unit:

→ The output unit is the counter part of the input unit.

→ Its function is to send processed results to the outside world.

→ The different output devices are,

- Monitor
- Printer
- Plotter

(v) Control Unit:

→ All activities inside the machine are directed and controlled by the control unit.

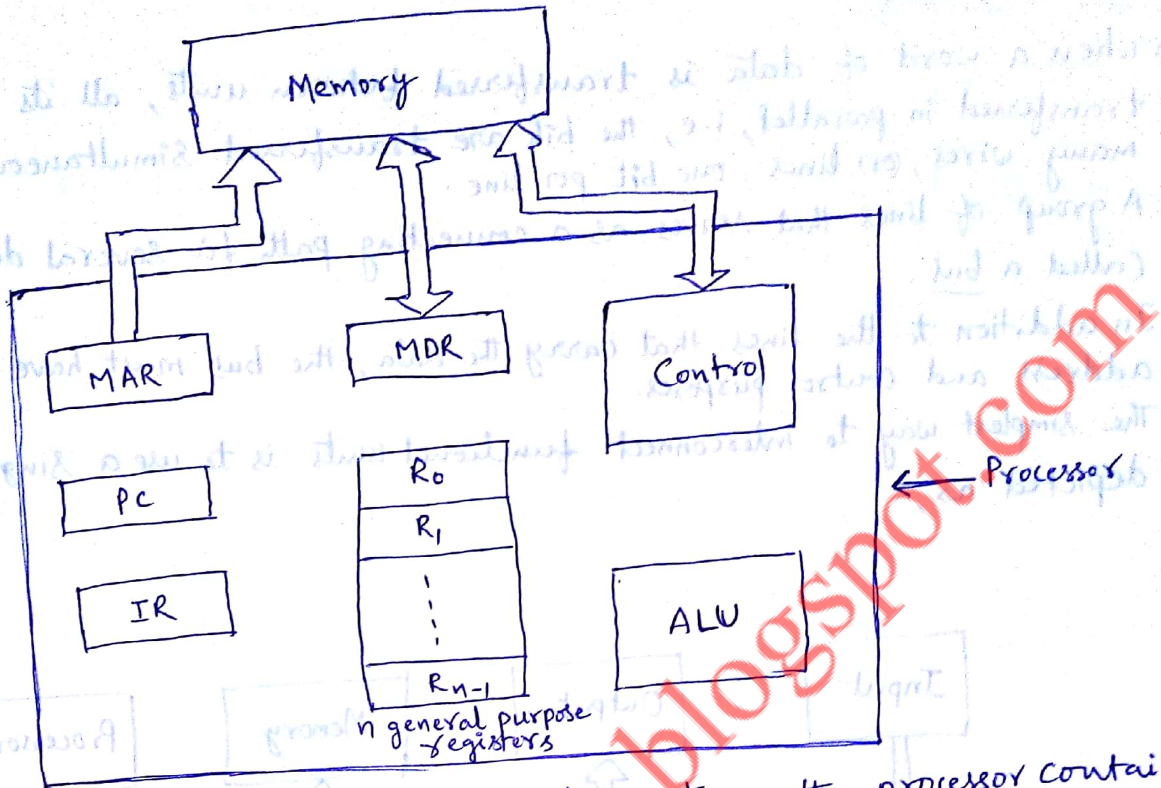
→ The control unit is effectively the nerve center that sends control signals to other units and senses their states.

③ Basic Operational Concepts

→ Transfers between the memory and the processor are started by sending the address of the memory location to be accessed to the memory unit and issuing the appropriate control signals.

→ The data are then transferred to (or) from the memory.

→ The connections between the Processor and the memory is depicted as,



→ In addition to the ALU and the control circuitry, the processor contains a number of registers used for several different purposes.

→ The Instruction Register (IR) holds the instruction that is currently being executed.

→ Its output is available to the control circuits, which generates the timing signals that control the various processing elements involved in executing the instruction.

→ The Program Counter (PC) keeps track of the execution of a program and it contains the memory address of the next instruction to be fetched and executed.

→ During the execution of an instruction, the contents of the PC are updated to correspond to the address of the next instruction to be executed.

→ The PC points to the next instruction that is to be fetched from memory.

→ It has also n-general purpose registers R_0 through R_{n-1}

→ Finally, two registers facilitate communication with the memory

→ These are the Memory Address Register (MAR) and Memory Data Register (MDR)

→ The MAR holds the address of the location to be accessed.

→ The MDR contains the data to be written into (or) read out of the addressed location.

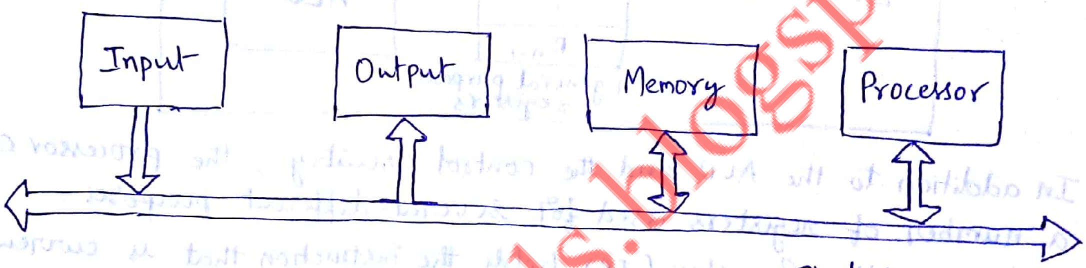
④ Bus Structures:

→ When a word of data is transferred between units, all its bits are transferred in parallel, i.e, the bits are transferred simultaneously over many wires, (or) lines, one bit per line.

→ A group of lines that serves as a connecting path for several devices is called a bus.

→ In addition to the lines that carry the data, the bus must have lines for address and control purposes.

→ The simplest way to interconnect functional units is to use a single bus is depicted as,



Single bus structure

→ The bus can be used for only one transfer at a time, only two units can actively use the bus at any given time.

→ Bus control lines are used to arbitrate multiple requests for use of the bus.

→ The main virtue of the single-bus structure is its low cost and its flexibility for attaching peripheral devices.

→ Systems that contain multiple buses achieve more concurrency in operations by allowing two (or) more transfers to be carried out at the same time

→ This leads to better performance but at an increased cost.

⑤ Software: (System software)

→ In order for a user to enter and run an application program, the computer must already contain some system software in its memory.

→ System software is a collection of programs that are executed as needed to perform functions such as,

- Receiving and interpreting user commands.
- Entering and Editing application programs and storing them as files in secondary storage devices.
- Managing the storage and retrieval of files in secondary storage devices.

- Running standard application programs such as processors, spreadsheets, (or) games, with data supplied by the user.
- Controlling I/o units to receive input information and produce output results.
- Translating programs from source form prepared by the user into object form consisting of machine instructions.
- Linking and running user-written application programs with existing standard library routines, such as numerical computation packages.

→ System Software is thus responsible for the coordination of all activities in a computing system.

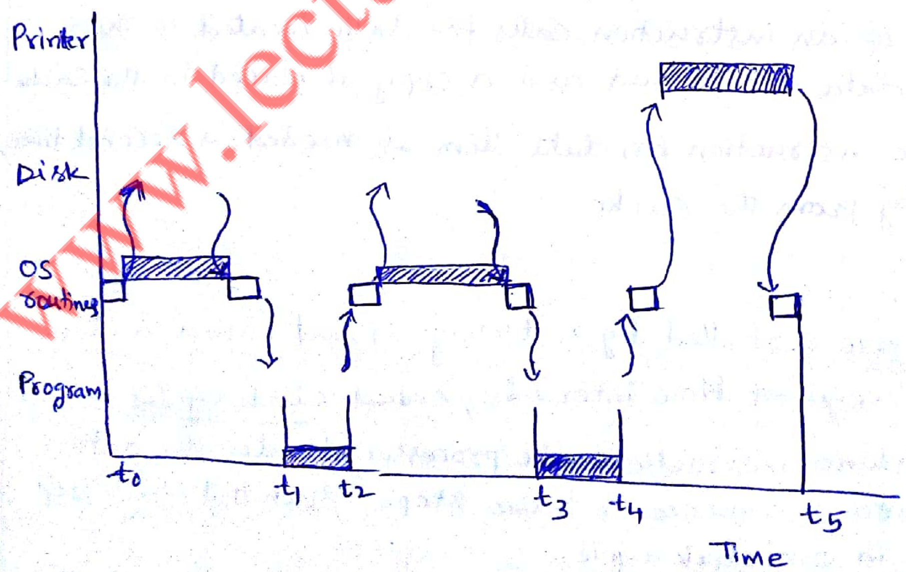
→ Application Programs are usually written in a high-level programming language, such as C, C++, Java, in which the programmer specifies mathematical (or) text-processing operations.

→ Operating System is a large program, (or) actually a collection of routines, that is used to control the sharing of and interaction among various computer units as they execute application programs.

→ The OS routines perform the tasks required to assign computer resources to individual application programs.

→ A system program that all programmers use is a text editor. It is used for entering and editing application programs.

→ User Program and OS routine sharing of the processor is depicted as,



→ The Operating System manages the concurrent execution of several application programs called as multiprogramming (or) multitasking

(6) Performance:

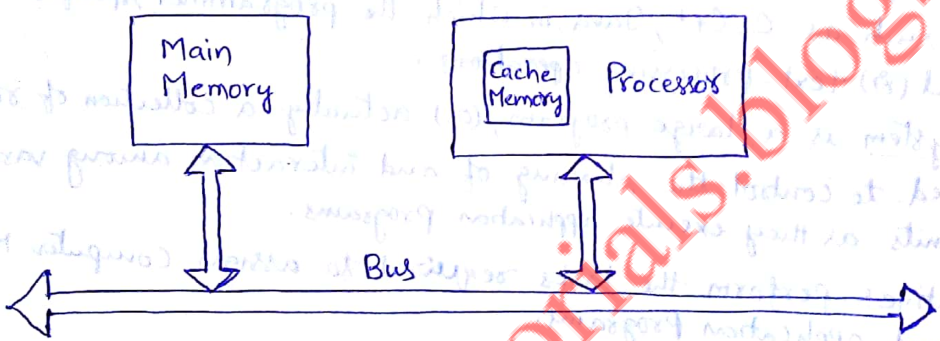
→ The most important measure of the performance of a computer is how quickly it can execute programs.

→ The speed with which a computer executes programs is affected by the design of its hardware and its machine language instructions.

→ To represent the performance of the processor, we should consider only the periods during which the processor is active.

→ These are the periods labeled Program and OS routines. The sum of these periods as the processor time needed to execute the program.

→ The processor cache is depicted as



→ At the start of execution, all program instructions and the required data are stored in the main memory.

→ As execution proceeds, instructions are fetched one by one over the bus into the processor, and a copy is placed in the cache.

→ When the execution of an instruction calls for data located in the main memory, the data are fetched and a copy is placed in the cache.

→ Later, if the same instruction (or) data item is needed a second time, it is read directly from the cache.

(i) Processor clock:

→ Processor circuits are controlled by a timing signal called a clock.

→ The clock defines regular time intervals, called clock cycles.

→ To execute a machine instruction, the processor divides the action to be performed into a sequence of basic steps, such that each step can be completed in one clock cycle.

→ The length P of one clock cycle is an important parameter that affects processor performance.

→ Its inverse is the clock Rate, $R = 1/P$, which is measured in cycles per second.

→ In standard electrical engineering terminology, the term cycles per second is called Hertz.

(ii) Basic Performance Equation:

→ Let T be the processor time required to execute a program

→ Assume that complete execution of the program requires the execution of N machine language instructions.

→ The number N is the actual number of instruction executions, and is not necessarily equal to the number of machine instructions in the object Program.

→ Suppose that the average number of basic steps needed to execute one machine instruction is S, where each basic step is completed in one clock cycle.

→ If the clock rate is R cycles per second, the program execution time is given by

$$T = \frac{N \times S}{R}$$

This is referred as Basic Performance Equation

(iii) Pipelining and SuperScalar Operation:

→ A substantial improvement in performance can be achieved by overlapping the execution of successive instructions, using a technique called pipelining.

→ Consider the instruction

Add R1, R2, R3

- which adds the contents of registers R1 and R2, and places the sum into R3.
- The contents of R1 and R2 are first transferred to the inputs of the ALU.
- After the add operation is performed, the sum is transferred to R3.
- The processor can read the next instruction from the memory while the addition operation is being performed.
- Then if that instruction also uses the ALU, its operands can be transferred to the ALU inputs at the same time that the result of the Add instruction is being transferred to R3.

→ A higher degree of concurrency can be achieved if multiple instruction pipelines are implemented in the processor.

→ This means that multiple functional units are used, creating parallel paths through which different instructions can be executed in parallel.

→ With such an arrangement, it becomes possible to start the execution of several instructions in every clock cycle. This mode of operation is called Superscalar execution.

(iv) Clock Rate:

→ There are two possibilities for increasing the clock rate, R.

- First, improving the Integrated-Circuit (IC) technology makes logic circuits faster, which reduces the time needed to complete a basic step
  This allows the clock period, P, to be reduced and the clock rate R, to be increased.
- Second, reducing the amount of processing done in one basic step also makes it possible to reduce the clock period P.

(v) CISC and RISC:

→ Simple instructions require a small number of basic steps to execute

→ Complex instructions involve a large number of steps.

→ A key consideration in comparing the two choices is the use of pipelining.

→ CISC and RISC are used for complex instructions.

→ CISC – Complex Instruction Set Computers
   RISC – Reduced Instruction Set Computers.

(vi) Compilers:

→ A compiler translates a high-level language program into a sequence of machine instructions.

→ To reduce N, we need to have a suitable machine instruction set and a compiler that makes good use of it.

→ An optimizing compiler takes advantage of various features of the target processor to reduce the product N×S, which is the total number of clock cycles needed to execute a program.

(vii) Performance Measurement:

→ Computer designers use performance estimates to evaluate the effectiveness of new features.

→ Manufacturers use performance indicators in the marketing process.

→ Buyers use such data to choose among many available computer models.

→ The computer community adopted the idea of measuring computer performance using benchmark programs.

→ To make comparisons possible, standardized programs must be used.

→ The performance measure is the time it takes a computer to execute a given benchmark.

→ The accepted practice today is to use an agreed-upon selection of real application programs to evaluate performance.

→ A nonprofit organization called SPEC (System Performance Evaluation Corporation) selects and publishes representative application programs for different application domains, together with test results for many commercially available computers.

→ The SPEC rating is computed as,

$$\text{SPEC rating} = \frac{\text{Running time on the reference computer}}{\text{Running time on the computer under test}}.$$

→ The overall SPEC rating for the computer is given by

$$\text{SPEC rating} = \left(\prod_{i=1}^{n} \text{SPEC}_i \right)^{1/n}$$

where n is the number of programs in the suite.

(7) The History of Computer Development:

→ Computers as we know them today have been developed over the past 60 years

→ A long slow evolution of mechanical calculating devices proceded the development of computers.

→ Development of the technologies used to fabricate the processors, memories, and I/O units of computers has been divided into four generation.

(i) First Generation (1945-55)
(ii) Second Generation (1955-65)
(iii) Third generation (1965-75)
(iv) Fourth Generation (1975 - present)

(i) The First Generation:

- → The key concept of a stored program was introduced by John von Neumann
- → Programs and their data were located in the same memory, as they are today.
- → Assembly language was used to prepare programs and was translated into machine language for execution.
- → Basic arithmetic operations were performed in a few milliseconds using vacuum tube technology to implement logic functions.
- → Magnetic core memories and magnetic tape storage devices were also developed.

(ii) The Second Generation:

- → The transistor was invented at AT&T Bell Laboratories in the late 1940s and quickly replaced the vacuum tube.
- → This basic technology shift marked the start of the second generation.
- → Magnetic core memories and magnetic drum storage devices were more widely used in the second generation.
- → High-level Languages such as FORTRAN were developed.
- → Compilers were developed to translate these high-level language programs into a corresponding assembly language program.
- → Seperate I/O Processors were developed.
- → IBM became a major computer manufacturer during this time.

(iii) The Third Generation:

- → The ability to fabricate many transistors on a single silicon chip, called Integrated Circuit (IC) technology developed.
- → Integrated circuit memories began to replace magnetic core memories.
- → Other developments included the introduction of micsoprogramming, parallelism, and Pipelining.
- → Operating System software allowed efficient sharing of a Computer system by several user programs.
- → Cache and virtual memories were developed.
- → Cache memory makes the main memory appear faster than it really is, and virtual memory makes it appear larger.
- → System 360 mainframe computers from IBM and the line of PDP mini-computers from Digital Equipment Corporation were dominant commercial products of the third generation.

(iv) The Fourth Generation:

→ Tens of thousands of transistors could be placed on a single chip, and the name Very Large Scale Integration (VLSI) was coined to describe this technology.

→ VLSI technology allowed a complete processor to be fabricated on a single chip called Microprocessor.

→ Companies such as Intel, National Semiconductor, Motorola, Texas Instruments, and Advanced Micro devices, were the driving forces of this technology.

→ Organizational concepts such as concurrency, pipelining, caches, and virtual memories evolved to produce the high-performance computing systems of today. ~~as the~~

→ Portable Notebook computers, desktop PCs, and workstations, interconnected by local area networks (LAN), WAN (wide area networks), and the Internet.

→ Centralized computing on mainframes is now used primarily for business applications in large companies.