Unit 1

Computer Organization

Computer Organization refers to the Operational units and their interconnections that realizes the architectural specification.

Examples:

→ Control signals
→ Interfaces between computer and peripherals
→ Memory technology being used.

Basic structure of Computers:

Computer types:

1. Digital computers
2. personal Computers
3. Notebook Computers
4. Workstations
5. Enterprise system
6. Server
7. Super computer.

Digital Computer:

→ It is a fast electronic calculating machine which accepts digitilized input information, process it to list of internally stored instruction and produces the resulting output information.

→ The list of instruction is called as computes program and internal storage is called computer memory.

→ The Computer may widely vary in size, cost and power.

Personal Computer:

→ The most Common Computer is personal Computer which can be used in homes, ~~schools~~, schools, buriness, office etc.

→ It is the most common form of desktop Computer that have processing and storage unit, visual display, audio output unit and a keyboard.

→ In this storage media includes harddisc, CDROM and diskettes

Notebook Computers:

→ It is another version of personal Computer where all the components are packed into a single unit

→ It is portable and size is of thin brief case.

Ex: Laptops, Tablet PC

Workstations:

→ Workstations are of with high resolution graphics and I/o capability that performs more computational power than personal Computers

→. It can be used in engineering application and interactive design work.

Enterprise System: (or) mainframes System:

→ It is a large and powerful system that are und for busines data processing.

→ It computes at high powes and has vay large storage.

Servers:

→ It contains very large size database, storage unit and capable of handling large volumes of regnest to accers the data

→ Servers are widely used in education, business and personel user Communities.

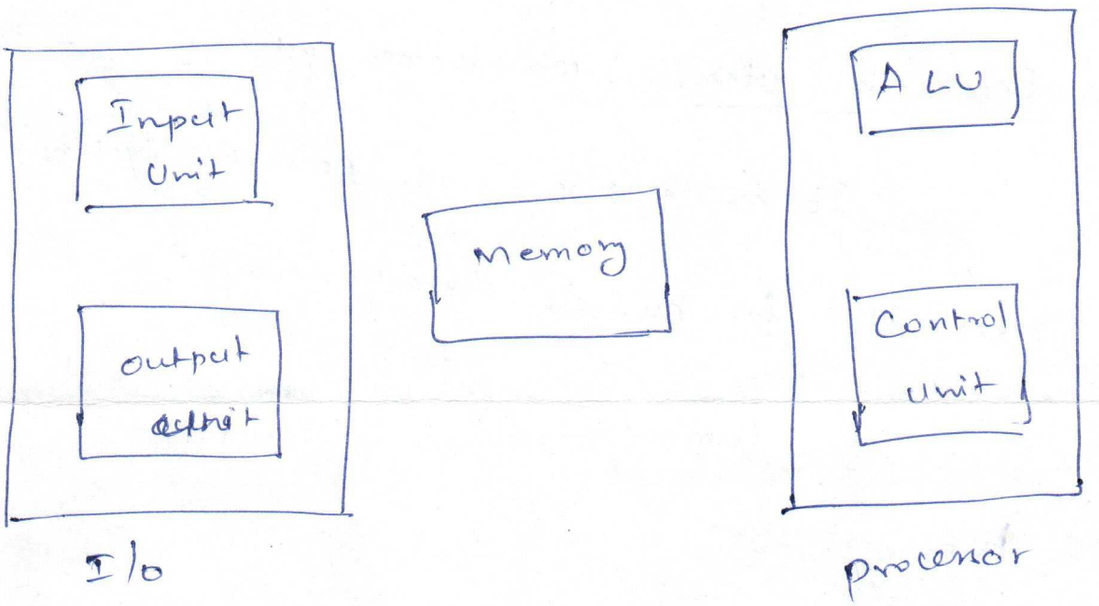
Super Computers:

→ It is und for large scale numerical calculation required in applications such as weather forecasting, aircraft design and Simulation.

Functional units:

A computer consist of five functional units

They are
1. Input unit
2. Memory unit
3. ALU
4. output unit
5. Control unit.



Basic functional units of Computer

→ The input unit accepts the coded information from the user through the electromechanical device such as keyboard. or from other computers.

→ The information received is either stored in memory or immediately used by ALU to perform desired operation.

→ The processing steps are determined by a program stored in the memory.

→ finally the results are sent back to the user through the output unit.

→ The input & output unit is often collectively called Input output unit.

→ The processor fetches the instruction from the memory and performs the desired operation.

→ The computer is completely controlled by the stored program, except for possible external interruption by an operator or by I/o devices connected to machine.

→ Data are the numbers or encoded characters that are used as operands by the instruction.

→ As computer knows only machine language, the source program (HLL) is translated into machine language program by the compiler

→ Each number, character or instruction is encoded as a binary digits called bits, each have one of two possible values 0 or 1

→ Two standard binary code formats are ASCII (American standard code for Information Interchange) where each character is represented as a 7 bit code and EBCDIC (Extended binary coded decimal Interchange code) which is represented as a 8 bit code.

Input unit:

→ Computers accepts coded information through Input units that reads the data.

→ The most well known input device is the keyboard where a key is pressed, the corresponding letter(or) digits is automatically translated into its corresponding binary code and the same is ~~stored on~~ transferred to either memory or the processor.

Memory unit:

→ The function of the memory unit is to store the programs and data

→ There are two classes of storage called primary and secondary storage.

→ Primary memory is the fast memory where the programs are stored while they are executed.

→ The memory contains a large number of storage cells, each capable of storing one bit of information

→ Each cells are used as a fixed size cell called as word. The memory is organized so that the content of one word containing 'n' bits can be stored or retrieved in one basic operation

→ Each cell in the memory is given a distinct address to provide easy access to the word.

→ The given word is accessed by specifying its address and issuing a control command signal that starts storing or retrieval process.

→ The no. of bits in each word is referred to word length. that ranges from 16 to 64 bits

→ Memory in which any location can be reached in short and fixed amount of time after specifying its address is called as RAM.

→ The time required to access one word is called as memory access time.

→ Secondary storage is used to store large amount of data and many programs

→ The typical secondary storage devices are magnetic disk, magnetic tapes, and optical disks.

Arithmetic and logic unit:

→ The basic arithmetic operations are performed in ALU. Eg: two numbers to be added

→ The two numbers are brought into the processor and actual addition is carried out by ALU

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

→ When operands are brought into processer they are stored in high speed storage element called registers.

→ Access time to register is faster than the access time to memory.

Output unit:

→ Its function is to send processors results to the outside world (or) user.

Ex: monitor, printer etc.

→ Some units such as graphic displays provide both input function and output function.

Control unit:

→ The memory, ALU, i/p and o/p unit store and process information and perform i/o operations The Control unit sends the control signals to other units and senses their states.

→ The actual transfers are generated in the Control circuits by timing signals that determines when a given action is to take place.

The operation of computer can be summarized as follows:

→ The computer accepts information in the form of programs and data through i/p unit and stores it in memory

→ Information stored in the memory is fetched under program control, into arithmetic & logic unit where it is processed.

→ Processed information leaves the computer through output unit

→ All the activities inside the computer are directed by the Control unit.

Basic Operational Concepts:

Consider the following instruction

Add LocA, R0

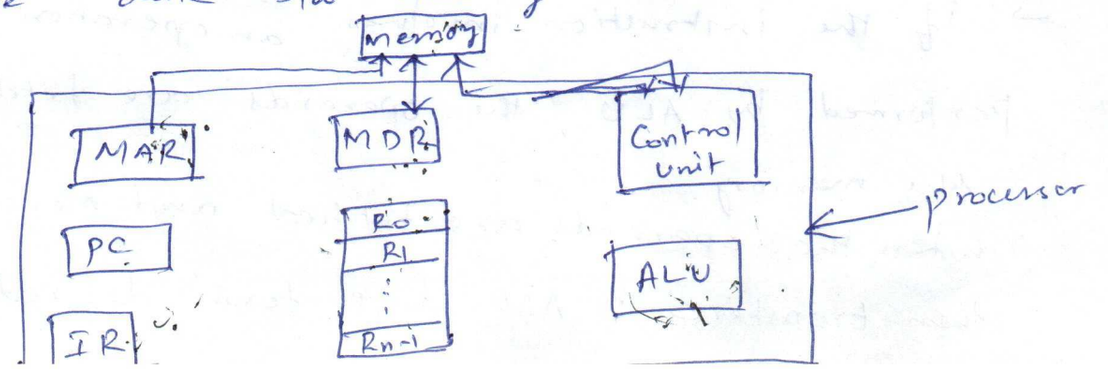
→ This instruction adds the operand at memory Loc to the operand in a register R0, which is present in processor and their sum is stored in R0 which is overwritten

→ The above instruction performs the following steps

1. The instruction is fetched from the memory into processor
2. The operand at LocA is fetched and added to the contents of R0
3. The resulting sum is stored in Register R0

→ Transfer between memory and processor are started by sending the address of memory location to be accessed to the memory unit and issuing appropriate Control signal

→ The below figure shows how the transfers can be done b/w memory and processor.



Instruction Register (IR):

It holds the instruction that is currently being executed

Program Counter (PC):

It contains the address of the next instruction that is to be fetched.

Memory address Register (MAR):

It holds the address which is to be accessed.

Memory Data register (MDR)

It contains the data to be written into or read out of the addressed location

→ Execution of the program starts when the PC is set to point to the first instruction of the program
→ The Content of pc are transferred to the MAR and 'read' control signal is sent to the memory
→ Then the word is read out of the memory and loaded into the MDR
→ The Contents of MDR are transferred to the IR.
→ If the instruction involves an operation to be performed by ALU, the operands are fetched from the memory.
→ When the operands are fetched and moved to MDR then transfered to ALU to perform desired operation.

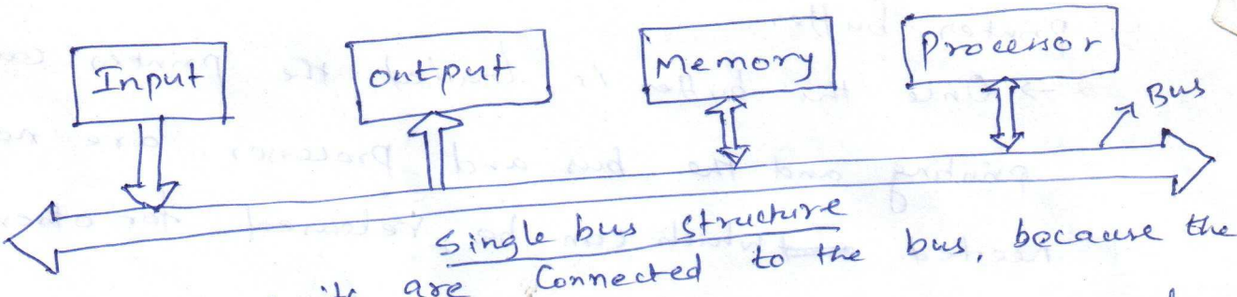
Bus structure:

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

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

→ The lines can carry the data, address and control signals

→ The simplest way to interconnect functional units is to use single bus which is shown below.



Single bus structure

→ All the units are Connected to the bus, because the bus can be used for only one transfer at a time, only two units can actively use the bus at any given time

→ The main thing of single bus structure is its low cost and its flexibility for attaching peripheral devices

→ Systems that contains multiple buses achieve more Concurrency in operations by allowing two or more transfers to be carried out at the same time. which leads to better performance but at an increased cost.

→ Memory and processors operate at electronic speed making them the fastest part of a Computer.

→ All the devices must Communicate with each other over a bus, so an efficient transfer mechanism must be used to smooth out the differences in timing among processors, memories and external devices

→ A common approach is to include buffer registers with the devices to hold the information during transfer.

Ex:
→ Consider the tranfer of an encoded character from the processor to a character printer. The Processor sends the character over the bus to the printer buffer.
→ Once the buffer is loaded, the printer can start printing and the bus and processor are no longer needed ~~and~~ which can be released for other activity.
→ The printer Continues printing the character in its buffer and it is not available for further transfer until this process is completed.
→ Thus buffer registers Carry timing differences among processors, memory and I/o devices

Bus:

Bus Connects various Components in a Computer system

Bus types:

1. Internal bus (or) System bus
2. External bus
3. Control bus
4. Synchronous bus
5. Asynchronous bus
6. Data bus
7. address bus

Internal bus:

→ This bus is used inside the processor System

→ It is also refered as System bus, which is used to transfer data, address and control signals

External bus:

→ This bus is used to interface with the devices outside the processor system.

→ It is typically used to Connect I/o devices.

Control bus:

→ The Control bus carries the transaction specific control information

→ Some typical control signals are

1. Memory Read and memory write
2. I/o Read and I/o write
3. Ready
4. Bus request and Bus grant
5. Interrupt and interrupt acknowledgement.
6. DMA request and DMA acknowledgement
7. clock
8. Reset

Data bus:

→ A data bus allows the computer subsystem for the transfer of data from one unit to other unit.

→. This includes transfering data to and from memory or from processor to other components.

→ The amount of data a data bus can handle is called as bandwidth.

→ A typical data bus is 32 bit wide means that 32 bit of data can travel through bus every second.

→ Newer computer data buses can handle 64 bit and even 96 bit data path.

Address bus:

→ A address is a computer bus that is used to specify a physical address

→ when a processor needs to read or write to a memory location, it specifies that memory location on the address bus

→ The width of address bus is 32 bits

Synchronous bus:

→ A bus used to interconnect devices that comprise a computer system where the timing of transactions between devices is under the control of synchronizing clock signal.

→ A device connected to a synchronous bus must guarantee to respond to a command within the period set of the clock signal or transmission error will occur

Asynchronous bus:

→ A bus used to interconnect devices ~~that~~ of the computer system where information transfer b/w devices are self timed rather than controlled by synchronizing clock signal.

→ A connected device indicates its readiness for transfer by activating a request signal

→ A responding device indicates the completion of transfer by activating an acknowledgement signal.

Software: → Computer s/w is the list of instruction that makes the hardware friendly to the user.

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

① Receiving and interpreting user commands

② Entering, editing application program and storing them as files in secondary storage device.

③ Managing the storage and retrieval of files in secondary storage device

④ Running the standard application programs such as word processor, Spread Sheets or games with data supplied by the user

⑤ Controlling I/o units to receive input information and produce output result.

⑥ Translating the program from source program to the object form

⑦ Linking and running user written application programs with existing standard library functions

→ Examples of System Software are Compiler, Operating System, Interpreter, editors and so on.

→ Application s/w are developed to assist the user to perform specific task.

→ In order to run an application program, the Computer must already contain these kinds of System softwares.

Performance:

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

→ Let us see some of the performance measures

(1) Processor clock:

→ Processor Circuits are Controlled by a timing signals called as a clock. The clock defines regular time intervals called clock cycle

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

(2) Basic performance equation:

Let T be the processors time required to execute a program that has been prepared in some high level language.

∴ The processor execution time for a program is given by $T = \frac{N \times S}{R}$.

where N = No. of actual machine instruction

S = avg No. of steps to be taken to execute one machine instruction.

R = clock rate per second.

Pipelining and Superscalar Operation:

→ The improvement of a performance can be ~~executed~~ achieved by overlapping the execution of successive instruction called Pipelining.

→ When multiple no. of instructions are to be executed by creating parallel paths ie the several instruction can be executed in every clock cycle. This mode of operation is called Superscalar operation

clock rate:

There are two possibilities for increasing the clock rate. They are

① Improving the integrated circuit technology to work in a fast way

② Reducing the amount of processing done in one basic step.

Instruction set: CISC and RISC:

→ Complex instruction set Computer has bigger instruction set ie it involves more number of steps for processing a single instruction

→ Reduced instruction set Computer has smaller instruction set ie it requires minimum number of steps for processing

Compiler:

→ A compiler translates a high level language program into sequence of machine instructions. So we need to have a suitable machine instruction set, then the compiler will provide good performance.

Performance Measurement:

→ The Computer Community adopted the idea of measuring Computer performance using benchmark programs.

→ A non profit organization called System Performance Evaluation Corporation (SPEC) selects and publishes the representative application programs for different domains.

Multiprocessors and Multicomputers:

→ Large Computer systems may contain number of processors units, in which they are called as multiprocessors System.

→ There System execute a number of diffrent task in parallel or they execute subtasks of a single large task in parallel.

→ All the processors have accers to all of the memory in such Systems. and they are termed as shared memory multiprocessor system.

→ The high performance of these System comes with much increased Complexity and cost.

→ In addition to multiple processors and memory unit the cost is increased because of the need for more complex interconnection Networks

→ In Contrast, it is also possible to use an interconnected group of Complete Computers to achieve high total Computational power.

→ when the tasks they are executing need to Communicate data, they do so by exchanging messages over a Communication Network. This is refered as message passing multicomputers.

Reference:

1. **Digital Design, 6th Edition, M. Morris Mano, Pearson Education, 2018.**
2. **Computer Organization, Carl Hamacher, ZvonkoVranesic, SafwatZaky, 6th edition, McGraw Hill, 2023.**

Prepared by:

Mr A Srinivasan,

Associate Professor,

Dept of CSE-DS