

LEARNING MATERIAL
On
COMPUTER SYSTEM ARCHITECTURE
(For 3rd semester CSE)

PREPARED BY:
GIRIJA PRASAD SWAIN
GOVT. POLYTECHNIC, DHENKANAL

01/2020

Monday	-	6	13	20	27
Tuesday	-	7	14	21	28
Wednesday	1	8	15	22	29
Thursday	2	9	16	23	30
Friday	3	10	17	24	31
Saturday	4	11	18	25	-
Sunday	5	12	19	26	-

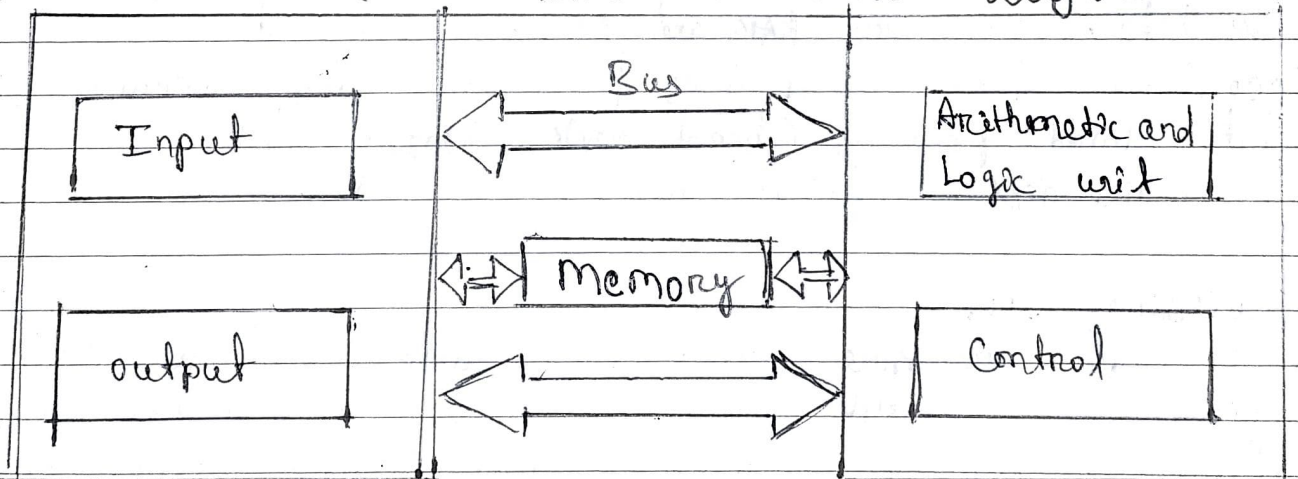
Chap-1

1. Basic structure of Computer H/W
- * Basic structure of Computer hardware
- * Functional units
- * Computer components
- * Performance measures
- * Memory addressing and operations.

11 Basic structure of computer hardware :

12 Computer :

13 Computer is a electronic machine which accepts input
 14 input information in the digitized form through
 15 input devices, process the input according to a
 16 set of stored information in memory through CPU
 and gives an output in form understandable
 by humans.

Functional unit - Structure / Block diagramI/P unit :

The computer accept coded information through
 input unit by the user. It is a device that is
 used to give required information to the
 computer.

notes

Phone

email

website

It accept coded information through input unit and delivers the output through output unit.
ex- Keyboard, mouse, joystick etc.

Memory unit :-

Memory unit used to store programs as well as data. It may be classified into primary storage and secondary storage in terms of Bytes. A memory unit is a collection of storage cells together with associated circuits needed to transfer information in and out of storage.

The memory unit consist of :

- (i) primary memory < RAM
- (ii) secondary memory < ROM

Primary storage : It is a memory made of up semiconductor storage cells. These cells are grouped together in a fixed size called word. The word may be 16-bit to 64 bits. units are RAM and ROM.

Secondary storage : It is used to store large amount of data and program (hard disk, floppy disk, magnetic disk, optical disk).

Processing unit :

A central processing unit (CPU), also called a central process or main processor, CPU performs basic arithmetic, logic, controlling and input/output operations specified by the instructions in the program. It runs the operating system and applications by receiving input from the user, it then processes the data and produces output. Stores data are processed by the ALU under program control.

notes

Phone

email

website

*

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

The ALU and CU together form the control processing unit. It is also known as brain of computer.

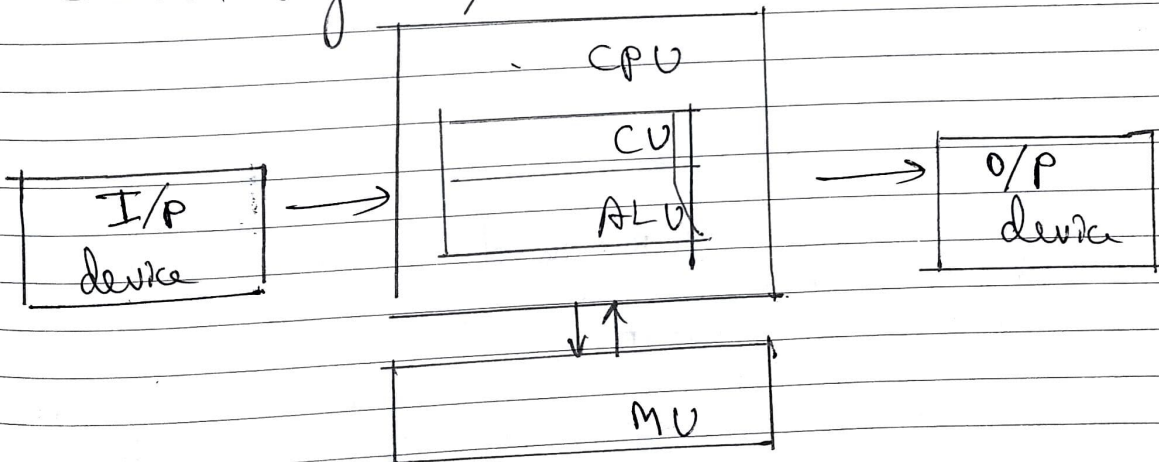
→ CPU consist of 3 units / subsystems:
(i) Arithmetic & logic unit
(ii) Control unit
(iii) Memory unit / registers.

Output

Any information that is processed by and sent out from a computer is called as output. It is a device can receive data from another device and generate output with that data, but it cannot send data to another device.

ex- Monitor, projector, speakers etc.

Block diagram / computer architecture:



Von neumann architecture

notes

Phone

email

website

*

Functional units:

- A Computer architecture describes the functions and design of the various units of a digital system.
- Computer architecture deals with the specification of the instruction set and the hardware units that implement the instructions.
- Computer hardware consists of electronic circuits, displays, magnetic and optic storage media and also communication facilities.
- Functional units are a part of CPU that performs the operations and calculations.
- Functional units of a computer system are parts of the CPU that performs the operations and calculations called by the computer program.
- A Computer consists of five main components: CPU, memory unit, ALU, Control unit (CU) and output unit.

I/P:

CPU:

MU:

ALU:

- Arithmetic and logical operations of a computer are executed in the ALU of the processor. It performs arithmetic operations like addition, subtraction, multiplication, division and also the logical operations like AND, OR, NOT operation.

CU

The Control unit is a component of a computer's central processing unit that coordinates the operation of the processor.

*

notes

Phone

email

website

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

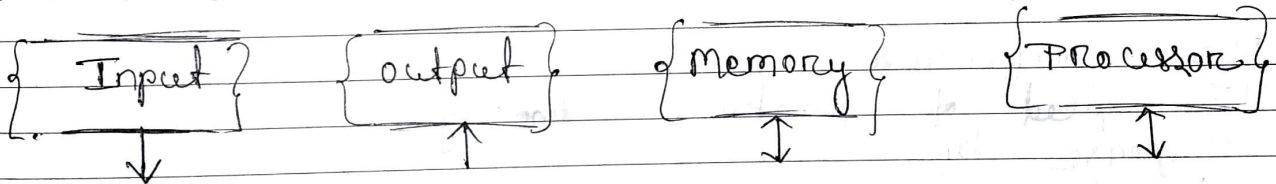
- It tells the Computer's memory, arithmetic/logic unit and output devices how to respond to a program's instructions.
- The control unit is also known as the nerve center of a computer system.
 - CU is also known as central nervous system of the computer. It controls the CPU and all the peripheral devices connected to it.

Output :

Components of a Computer :

- I/O unit
- CPU
- MU

Bus structure :



- All the components of the computer are interlinked and connected so that word of data is transfer between the unit using parallel set up lines called BUS.
- The lines that carry data must have the lines per address and controls purpose devices connected to the bus widely is their speed operation.
- To achieve a reasonable speed of operation, a computer must be organized.

notes

Phone

email

website

*

Single bus structure is

→ Low Cost

→ Very flexible for attaching peripheral devices.

Multiple Bus Structure Certainly increases, the performance
But also increases the cost significantly.

Performance Measure:

→ Performance is the ability of the computer and to quickly exact the program. The speed at which the computer execute the program or design by its hardware and machine language.

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

→ The total time required to execute the program is elapsed time (amount of time that passes from the beginning of an event to its end) is a measure of the performance of entire computer system.

Basic performance measure:

The speed of operation is two types:

- (i) Response time
- (ii) Through put.

Response time: It is also called latency.

Response time is the time spent to complete and even or an operation (time from start to completion of a task) includes:

- operating system overhead
- waiting for I/O and other processes

- Accessing disk and memory
- Time spent executing on the CPU or execution time.

※

notes

Phone

email

website

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

Through put:

- 8 The total amount of work done in a given time on per unit of time is called through put.
- 9 → The amount of processing accomplished during giving interval time (Band width).
- 10 → Time spent from the start of execution of a program to its completion as called elapsed.

CPI :

- 12 → Cycle per instructions is clock per instruction
- Program consist of a number of CPU instruction i.e represented by instruction count.

$$\text{Clock rate} = 1 / \text{clock cycle}$$

- work according to clock cycle / signal.

- Clock cycle is measured in (10^{-9} second) nano sec.

- Instruction Count (IC)

* total number of instructions executed the program.

- CPI → cycle per instruction.

Elapsed time :

- Elapsed time is a time spent from the start of execution of the program to its completion is called elapsed time.

- This performance measure is affected by the clock speed of the processor and the concerned input output device.

- In other words it is the difference betⁿ the time at which a task finishes and the time at which the task started.

MIPS: "Million Instructions per second". It is a method of measuring the raw speed of a computer's processor.

→ Measure of computer performance has the rate at which given machine executed instruction.

→ This is calculated by dividing the no. of instruction and the time required to run the program.

Functional units of Computer:

Five functional units

→ Input

→ Memory

→ Arithmetic and logic unit

→ Output

→ Control unit.

Memory addressing and operations:

Memory addressing:

→ A memory address is a unique identifier used by a device or CPU for data tracking.

→ Address is identified by binary code. This binary address is defined by an ordered and binary sequence, allowing the CPU to track the location of each memory byte.

→ In computing, a memory address is a reference to a specific memory location used at various levels by software and hardware.

notes

Phone

email

website

Types of Memory addressing:-

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

Memory addressing is of two types :

- (i) physical address
- (ii) logical address

Physical address:

- A digital computer's main memory consist of many memory locations. Each memory location has a physical address, which is a code.
- The CPU (or other device) can use the code to access the corresponding memory location.
- Only system software, i.e. the BIOS, operating systems, and some specialized utility programs (e.g. memory testers), address physical memory using machine code operands or processor registers, instructing the CPU to direct a hardware device called memory controller.
- The memory controller bus consist of a number of parallel lines, each represented by a binary digit (bit).
- Fast is memory unit.

Logical address:

- A computer program used memory addresses to execute machine code and to store & retrieve data.
- Now a days due to the introduction of virtual memory computers are using logical address or virtual address, using the computer's memory management unit and operating system.
- It is generated by CPU.
- Logical address is used like a reference to access the physical address.

notes → Logical address is virtual address and * can be viewed by user -

Phone

email

website

10

Day 010-356 Wk-02
FRIDAYBasis for ComparisonLogical addressphysical address

Basic

It is the virtual address generated by CPU

The physical address is a location in a memory unit.

Address space

Set of all logical addresses generated by CPU in reference to a program is referred as logical address space.

Set of all physical addresses mapped to the corresponding logical address is referred as physical address.

Visibility

The user can view the logical address of a program

The user can never view physical address of program

Access

The user uses the logical address to access the physical address.

The user can not directly access physical address.

Generation

The logical address is generated by the CPU

physical address is computed by MU.

Memory operations :

→ Memory is usually designed to store and retrieve data in word length quantities.

→ Data transfer between the memory and the processor takes place through the use of

two processor registers called:

notes

Phone

email

website

**

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

(i) MAR - Memory address register.
(ii) MDR - Memory data register.

MAR:

It is the address register which is used to store the address of the memory location where the operation is being performed.

MDR:

It is the data register which is used to store the data on which the operation is being performed.

* Memory read operation:

Memory read operation transfers the desired word to address lines and activates the read control line.

* Memory write operation:

Memory write operation transfers the address of the desired word to the address lines, transfers the data bits to be stored in memory to the data input lines. Then it activates the write control line.

Amdahl's law:-

The law states that performance improvement to be gained by using a faster mode of execution is limited by the fraction of time the faster mode can be used?

→ It defines speed up by using multiple processors.

notes

Phone

email

website

*

12

Day 012-354 Wk-02
SUNDAY 144CSA (Computer system architecture):

- The computer architecture deals with the functional behaviour of a computer system as viewed by programmer.
- It is also described as, the logical structure of the system unit that is used electronic components.
- The computer architecture borrows the lessons from building a successful computer system.

Ex- vonneumann architecture

- Computer s/w and h/w are joined together and interact to make a computer work.
- There are three categories:
 - system design
 - Instruction set architecture
 - microarchitecture:

notes

Phone

email

website

*

Chap - 2

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

Instruction and instruction sequencing

- 1) Fundamentals to instructions
- 2) operands
- 3) OP codes
- 4) Instruction formats
- 5) Addressing Modes.

Instruction :

- 1) An instruction is a single operation of a processor defined by the processor instruction set.
- 2) The size or length of an instruction is 4 bit to multiples bytes.
- 3) These have two sources : operand & opcode.
- 4) An instruction is an order given to a computer processor by a computer program.
- 5) Each instruction is a sequence of 0s and 1s.

Operands

- 1) In computers an operand is the part of a computer instruction that specifies data. That is to be operating on manipulated and by extension, data itself.
- 2) operand on which operation is to be performed and computer instruction describes an operation (add, sub, mult, div).
- 3) An operand is the object of a mathematical operation.

ex - "1+2", 1 and 2 are operands and

plus symbol is operator.
operand is address location where instruction can meet.

*

OpCodes :

→ An opcode is the portion of a machine language instruction that specifies the operation to be performed.

→ It is a single instruction that can be executed by the CPU.

→ It is in binary form, also known as instruction machine code.

→ opcode means operation code.

Ex - "add memory location A to memory location B".

So, add is the opcode.

→ opcode is a part of the instruction that tells the processor what should be done.

Types of operands :

- (i) Address
- (ii) Numbers
- (iii) Characters
- (iv) Logical data

address :

It is the form of data. Here some calculations must be performed on the operand reference in an instruction, which is to determine the physical address of an instruction. Address are the number that represent specific location in memory.

Numbers :

All computers will support numeric data types. There are

some commonly used numeric data types available, they are:

*:

		3	10	17	24
Monday	-	4	11	18	25
Tuesday	-	5	12	19	26
Wednesday	-	6	13	20	27
Thursday	-	7	14	21	28
Friday	1	8	15	22	29
Saturday	2	9	16	23	-
Sunday					

- Integer
 → point
 → decimal
- { number data types are used by all machine language data type like point floating point and decimal.

characters:

- To provide documentations, the commonly used data is text or some character.
- in form of codes characters are designed for binary data and represented by sequence of bits.
- The most commonly used character code is in (ASCII - American standard code for information interchange) form - each character is represented by 7-bit and 8th bit set to 0 (Zero).
- EBCDIC (extended binary coded decimal interchange code) is used to encode the characters.

Logical data:

- The data is stored using 1 (True) or 0 (false).
- Each word or other addressable unit is treated as a single unit of data.

Opcode types:

- Arithmetic - add, sub, mult, div
- Logical - AND, OR, NOT
- Data transfer - move, store, load, push, pop etc
- Conversion - translate, convert
- system control - reserve for operating system
- Input output control - input - readout - write
- Control transfer - jump, returned, skip, halt.

notes

Phone

email

website

*

Instruction format:
 It describes the internal structures (lay out design) of the bits of an instruction. An instruction format includes an opcode, and address is dependent on availability of particular operands.

Classification of instruction:-

(i) 0- Address instruction:

The 0- address type instruction do not contain any operand address. Stack is included in the CPU for performing ALU instructions with no address, it does not use an address field.
 e.g - $\text{ADD } \overset{\text{Top of Stack}}{\text{to S}} \leftarrow (A+B)$ (Tos is the top of stack)

for the instructions ADD 4 MUL .

(ii) 1- address instruction:

In 1- address instruction only one operand address is specified in the instruction. The other operand is in accumulator, uses Accumulator register for all data operations.

e.g: $\text{LOAD A } AC \leftarrow M[A]$
 $\text{ADD B } AC \leftarrow AC + M[B]$

All operations are done between the AC register and a memory operand.

(iii) 2- address instruction:-

In 2- address instruction both operand address are specified. The result is placed in one of the specified address in which each address field can specify either a processor register or a memory word.

e.g - $\text{MOV R}_1, R_1 \leftarrow M[A]$
 $\text{ADD R}_1, R_1 + M[B]$
 $\text{MOV R}_2, M[C]$

notes

Phone MOV instruction moves/transfers the operands to and from memory and processor registers.

email

website

*

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	1	8	15	22	29
Saturday	2	9	16	23	-
Sunday					

(iv) 3-address instructions -

In 3-address instruction two addresses are specified for two operands and one address for the result. It is used to specify using address field to specify either a processor register or a memory operand.

E.g : $ADD R_1, A, B \quad R_1 \leftarrow M[A] + M[B]$
 $ADD R_2, C, D \quad R_2 \leftarrow M[C] + M[D]$
 $MUL \infty, R_1, R_2 \quad M[X] \leftarrow R_1 \times R_2$

Here R_1 & R_2 are two processor registers and $M[A]$ denotes the operand at memory address specified by A.

Addressing mode :-

Each instruction needs data on which it has to perform the specified operation. The data (operand) may be in a accumulator, general purpose register, therefore there are various ways to specify data. The techniques of specifying the address of the data are known as addressing modes.

Six type of addressing modes:

- (i) Direct addressing
- (ii) Indirect addressing
- (iii) Immediate addressing
- (iv) Register addressing
- (v) Register indirect addressing
- (vi) Base register addressing

Direct addressing :-

In direct addressing the address of the

*

Monday	-	6	13	20	27
Tuesday	-	7	14	21	28
Wednesday	1	8	15	22	29
Thursday	2	9	16	23	30
Friday	3	10	17	24	31
Saturday	4	11	18	25	
Sunday	5	12	19	26	

data (operand) is specified within the instruction itself.

- Single memory reference to access data
- No additional calculations to find the effective address of the operand.

Indirect addressing mode:

In this, the address field of instruction gives the address where the effective address is stored in memory. This slows down the execution, as this includes multiple memory lookups to find the operand.

Immediate addressing mode:

In this mode, the operand is specified in the instruction itself. An immediate mode instruction has an operand field rather than the address field.

Register addressing mode:

In this mode the operand is stored in the register and this register is present in CPU. The instruction has the address of the register where the operand is stored. The operand is contained in a register set. No reference to memory is required to fetch the operand.

Register indirect addressing mode:

Means that the location of an operand is held in a register. It is also called indexed ^{indirect} addressing mode. To load a value from memory into a register

using register - indirect addressing, we use second register known as base register.

notes

Phone

email

website

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

This base register holds the actual memory address that the program is interested in.

Base register:

Base register is the register, acts as a address holder of the base storage location from where the data was stored continuously. Assembler uses the base register value to find the data that is required. Address serving as a reference point ("base") for other addresses.

Instruction code:- Instructions are encoded as binary instruction codes. Each instruction code contains of a operation code, or opcode, designated by add, sub, Move, input etc

Mnemonic code: A device such as pattern of letters, ideas, or associations that assists in remembering something. so its usually used by assembly language programmers do, like "ADD" & "MUL" and "MOV" etc. This is a assembler specific.

Processor system

Monday	-	6	13	20	27
Tuesday	-	7	14	21	28
Wednesday	1	8	15	22	29
Thursday	2	9	16	23	30
Friday	3	10	17	24	31
Saturday	4	11	18	25	-
Sunday	5	12	19	26	-

- 1 > Register files
2 > Complete instruction execution

- Fetch
- Decode
- Execution

- 3 > Hardware control
4 > Micro program control

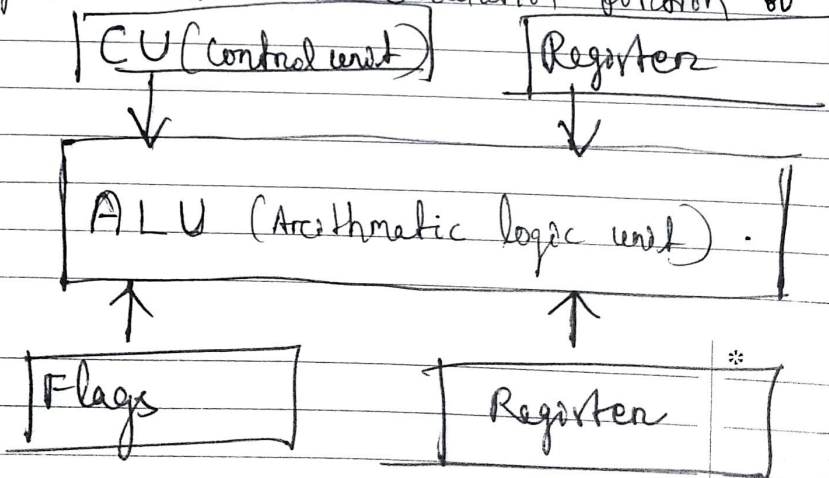
Register files :

A register file is an array of process registers in a CPU. Modern integrated circuit (IC) based register files are usually implemented using static RAM with multiple ports.

The instruction set architecture of a CPU will almost always define a set of registers which are used to move data between memory and functional unit of chip. Register is a small amount of storage available as a part of CPU.

Processing unit / ALU : On design of ALU :

The circuit carryout the data processor instruction. The ALU. The ALU performs sub, Add, mult, div and logical operation (AND, OR, NOT) etc. It represents the fundamental building block of the CPU of computer. It is the calculation portion of computer.



notes

Phone

email

website

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

An arithmetic logic unit (ALU) is a major component of the central processing unit of the computer system. It does all processes related to arithmetic and logic operations that need to be done on instruction words. ALU is divided into the arithmetic unit (AU) and the logic unit (LU).

ALU is also known as an Integer unit (IU). The arithmetic logic unit is that part of the CPU that handles all the calculations the CPU may need.

Different operation as carried out by ALU :

→ Logical operations: These include operations like AND, OR, NOT, XOR, NOR, NAND, etc.

→ Bit-shifting operations: This pertains to shifting the positions of the bits by a certain number of places either towards the right or left, which is considered a multiplication or division operations.

→ Arithmetic operations: This refers to bit addition and subtraction. Although multiplication and division are sometimes used, these operations are more expensive to make. Multiplication and subtraction can also be done by repetitive additions and subtractions respectively.

ALU performs basic arithmetic and logic operations. All information in a computer is stored and manipulated in the form of binary numbers, i.e. 0 and 1. Transistor switches are used to manipulate binary numbers in two

possible states of switch: open or closed. There is no current represents 0. A closed where there is current represent 1.

* when which transistor

Monday	6	13	20	27
Tuesday	7	14	21	28
Wednesday	1	8	15	22
Thursday	2	9	16	23
Friday	3	10	17	24
Saturday	4	11	18	25
Sunday	5	12	19	26

Complete instruction and execution:

The basic function performed by a computer is execution of a program, which consists of a set of instructions stored in memory. Instruction processing consists of two steps:

The processor reads (fetches) instructions from memory one at a time and executes each instruction.

Six steps are involved in execution of complete instruction:

- (i) Fetch instruction
- (ii) Decode instruction
- (iii) perform ALU operation
- (iv) Access memory
- (v) update register file
- (vi) update the PC (program counter).

Fetch:

Execution cycle starts with fetching instructions from main memory. The instruction at the current program counter (PC) will be fetched and will be stored in Instruction register (IR). This is getting the instruction from memory into the processor.

→ The program counter (PC) stores the address of the next instruction to be fetched. This is copied to the memory address register (MAR), so, MAR is loaded with the address of the instruction that is being fetched.

→ MAR places the address to be used in the "Address bus" (locating the memory location of the address).

Monday	1	10	17	24
Tuesday	4	11	18	25
Wednesday	5	12	19	26
Thursday	6	13	20	27
Friday	7	14	21	28
Saturday	1	8	15	22
Sunday	2	9	16	23

- The CPU sends a signal over the Control bus to read the memory. So, the memory chip is mean memory places the contents of the address in to the data bus.
- The data from the address is transferred back to the CPU over data bus where it is loaded in to the MDR (memory data register).
- The contents of the MDR is copied into the CIR (Current instruction register). This consists of the instruction (opcode) and the data part (operand).

Decode:

During this cycle the encoded instruction present in the IR (instruction register) is interpreted by the decoder. Next step for the CPU to interpret the data just fetched is to decode it.

- The decode unit of the CPU analyses the instruction (opcode) in the CIR and decodes it. The decode unit has an instruction set that defines the opcode.

→ A series of micro-signals are sent to different areas to prepare the CPU for reading for next step by CPU.

- The data to be operated on (operand) may be passed on to the ALU or the address of the data may be used with the operation which is then copied to the MAR or the data to be operated on is copied to the MDR.

Execute:

The appropriate instruction is carried out* upon the data (operand) in the MDR by the execution unit (containing an arithmetic logic unit that carries out

Calculations on the data). This is the executing of the instruction.

- The results of the arithmetic logic (ALU) are output to another register called the accumulator.
- The control unit assigns new signals to reset the CPU to prepare for the next cycle.

Hardware Control:

It is implemented as logical circuit (gates, half-blops, decoders etc.) in the hardware. This organization is very complicated if we have a large control unit. In architecture, if the design has to be modified or changed, requires changes in the wiring among the various components. Thus the modification of all the combinational circuits may be in different ways -

Microprogrammed Control unit:

A microprogrammed control unit is implemented using programming approach. Micro-program, consisting of micro instructions is stored in the control memory of the control unit. Execution of a micro-instruction is responsible for generation of a set of control signals.

Attributes

Hardware Control unit

Microprogrammed control unit

- | | | | | |
|-----------------------------|---|---|---|---|
| (i) Speed | → | speed is fast | → | speed is slow |
| (ii) Cost of implementation | → | more costlier | → | cheaper |
| (iii) Flexibility | → | not flexible to accommodate new system. | → | more flexible to accommodate new system or new set. |

Monday - 3 10 17 24
 Tuesday - 4 11 18 25
 Wednesday - 5 12 19 26
 Thursday - 6 13 20 27
 Friday - 7 14 21 28
 Saturday - 8 15 22 29
 Sunday - 9 16 23 30

Wk-04 Day 025-341

▶▶

SATURDAY

25

Hardware

(i) Instruction set of size

small

large

(ii) Control memory

As sent

present

(iii) Chip area required

less

More

(iv) occurrence

occurrence of error is more

occurrence of error is less

(v) Decoding

Complex decoding and sequencing logic.

Easier decoding sequencing logic.

(vi) Ability to handle complex instructions

Difficult to handle complex instruction sets.

Easier to handle complex instruction sets.

(vii) Ability to handle

difficult to handle

easier to handle.

notes

Phone

email

website

Memory system

- 1) Memory characteristics
- 2) Memory hierarchy
- 3) RAM and ROM
- 4) Interleaved memory
- 5) Cache memory
- 6) Virtual memory

Memory characteristics:

(i) Location:

- * CPU registers
- * Cache
- * Internal memory (main)
- * External (secondary)

(ii) Capacity:

Word size: Typically equal to the number of bits used to represent a number and to the instruction length. For address of length A (in bits), the number of addressable units is 2^A . (Instruction length is 32 bits) normally.

(iii) Unit of transfer:

- * word
- * block

(iv) Access method:

* Sequential access:

Access must be made in a specific linear sequence. Time to access an arbitrary record is highly variable.

* Direct access:

Individual blocks or record have an address based on physical location. Access is by indirect access to general vicinity of desired information, then ~~some~~ search. Access time is still variable, but not as much as sequential access.

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

* Random access:

Each addressable location has a unique, physical location. Access is by direct access to desired location. Access time is constant and independent of prior accesses.

* Associative:

Desired units of information are retrieved by comparing a sub-part of unit. Location is needed. Most useful for searching.

* Performance:

* Access time (Latency):

For random access memory, latency is the time it takes to perform. A read or write operation that is the time from the instant that the address is presented to the memory to the instant that the data have been stored available for use.

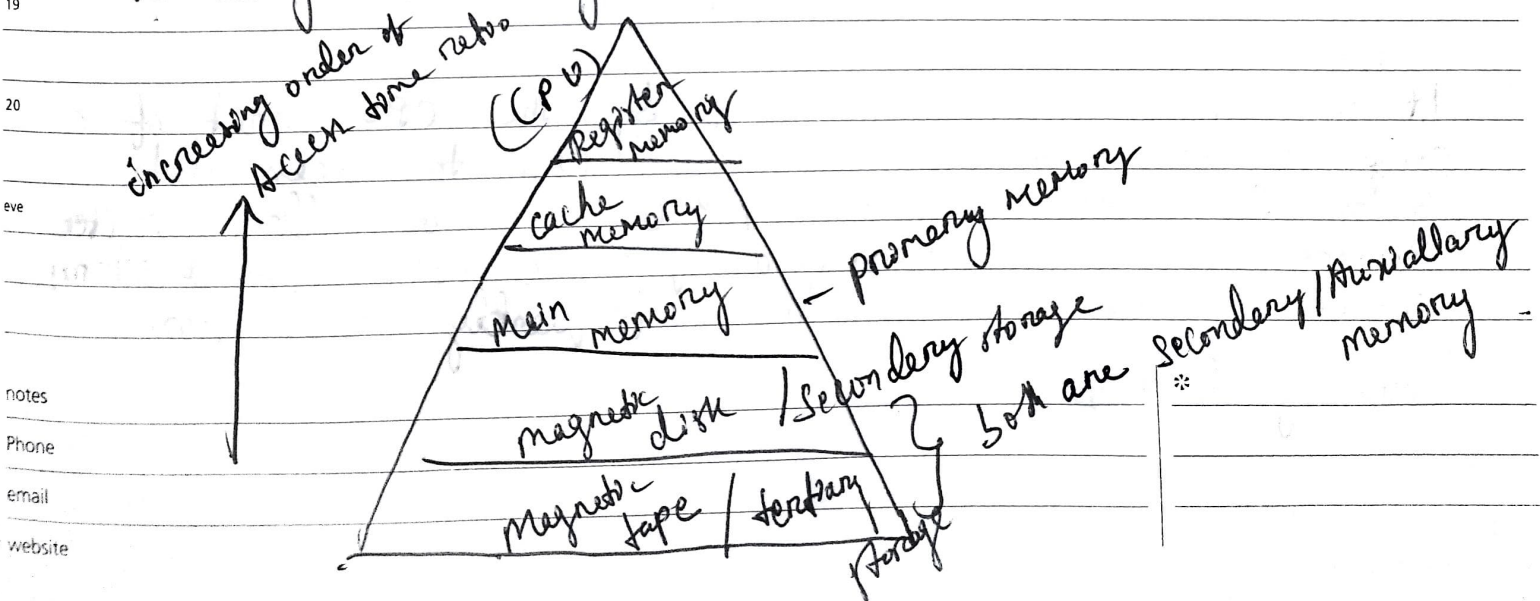
* Memory cycle time:

Access time + additional time required before a second access can begin (refresh time).

* Transfer rate:

Generally measured in bit / second.

Memory hierarchy:



Monday	-	6	13	20	27
Tuesday	-	7	14	21	28
Wednesday	1	8	15	22	29
Thursday	2	9	16	23	30
Friday	3	10	17	24	31
Saturday	4	11	18	25	-
Sunday	5	12	19	26	-

Memory hierarchy:

The memory hierarchy design is a computer system mainly includes different storage devices. Most of computers were built with extra storage to run more powerfully beyond the main memory capacity. Memory hierarchy diagram is a hierarchical pyramid for computer memory. The designing of the memory hierarchy is divided into two types such as primary (internal) memory and secondary (external) memory.

Register memory:

The register is a static RAM in the processor of the computer which is used for holding the data word which is 64 or 128 bits. It is the smallest and fastest memory in a computer. It is located in the CPU in the form of registers, which are the smallest data holding elements. A register holds frequently used data, instructions, and memory address that are to be used by CPU. Registers hold a small amount of data around 32 bits to 64 bits. The speed of a CPU depends on the number and size (no. of bits) of registers that are built in the CPU. Register include accumulator, data register, the Address Register, Program Counter (PC) and more.

Cache memory:

It is a hardware cache used by the CPU unit of a computer to reduce the average cost to access data from the main memory. A cache is a smaller, faster copy of the data from frequently used main memory locations.

notes

Phone

email

website

*

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

Cache memory is important because it improves the efficiency of data retrieval. It stores program instructions and data that are used repeatedly in the operation of programs or information that the CPU is likely to need next.

Main memory:

The main storage region of a computer, to which the central processor has immediate or direct access. The main memory is defined as the central storage in a computer. Main memory is intimately connected to the processor, so, moving instructions and data into and out of processor is very fast. Main memory is sometimes called RAM. The memory where programs and data are kept when the processor is actively using them. Δ (ex)

Magnetic disk:

The magnetic disks in the computer are concular plates fabricated of plastic otherwise metal by magnetized material. All the disks in computer turn jointly at high speed. The tracks are lots which are stored within the magnetized plane in spots. There are separated into sections which are named as sectors.

Magnetic tape:

This tape is a normal magnetic recording which is designed with a slender magnetizable covering on an extended, plastic film of the thin strip.

This is mainly used to backup huge data*.

RAM and ROM:

RAM

1. Temporary storage
 2. Store data in MBs
 3. RAM chip is a volatile memory.
 4. Writing data is faster
 5. It loses the data stored in it when the power is turned off.
 6. It is costlier than ROM
 7. It can hold a large amount of data as compared to ROM
 8. The data in RAM can be modified easily.
 9. It is used in the normal operations of a computer.
 10. Stores during and after processing.
 11. Generate less heat
 12. Used as Main Memory
 13. Uses less power
- Ex - Static and Dynamic RAM

ROM

1. permanent storage
 2. Store data in GBs.
 3. ROM chip is a non-volatile memory.
 4. Writing data is slower
 5. The data inside it retains even if the power of the CPU is switched off.
 6. It is cheaper
 7. It can only store small amount of data.
 8. ROM can be hardly or never be modified.
 9. It is used primarily in the startup process of a computer.
 10. stored by manufacturer.
 11. generate more heat
 12. Used as cache
 13. Uses more power.
- Ex - PROM, EPROM and EEPROM are types of ROM.

notes

Phone

email

website

*

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

RAM:

Random access memory (RAM) is the hardware in a computing device where the operating system (OS), application programs and data in current use are kept so they can be quickly reached by the device's processor.

RAM is the main memory in a computer, and it is much faster to read from and write to than other kinds of storage, such as a hard disk drive (HDD) etc.

ROM:

It is a type of non-volatile memory used in computers and other electronic devices. Data stored in ROM cannot be electronically modified after the manufacture of the memory device. It is a type of storage medium that permanently stores data on personal computers (PCs) and other electronic devices. It contains the programming needed to start a PC, which is essential for boot up.

Interleaved memory:

It is a design made to compensate for the relatively slow speed of dynamic RAM by spreading memory address evenly across memory banks. Contiguous memory reads and writes are using each memory bank in turn, resulting in higher memory throughput due to reduced waiting for memory banks to become ready for desired operations.

→ It is a technique for increasing speed.

→ It is a process that makes the system more efficient, fast and reliable.

notes

Phone

email

website

It is two types : 2 way & 4 way

two memory blocks are accessed at same time for writing and reading operation.

four memory blocks are accessed at the same time.

Cache memory :

— do —

Virtual memory :

Virtual memory is a memory management capability of an operating system (OS) that uses hardware and software to allow a computer to compensate for physical memory shortages by temporarily transferring data from random access memory (RAM) to disk storage (ex- windows). The purpose of virtual memory is to enlarge the address space, the set of addresses a program can utilize. A computer can address more memory than the amount physically installed on the system. This extra memory is actually called virtual memory and also it is a section of a hard disk. This process is done temporarily and is designed to work as a combination of RAM and space on the hard disk. When the RAM runs low, virtual memory will move the data out of RAM then transfer them in to a space called paging file.

Monday	30	2	9	16	23
Tuesday	31	3	10	17	24
Wednesday		4	11	18	25
Thursday		5	12	19	26
Friday		6	13	20	27
Saturday		7	14	21	28
Sunday	1	8	15	22	29

Input - output system

- 1> Input - output interface
- 2> Mode of data transfer
- 3> Interrupt driven I/O
- 4> DMA
- 5> I/O processor

Input - output interface:

provides a method for transferring information between internal storage and external I/O devices. Peripherals connected to a computer need special communication links for interfacing them with the CPU.

- decodes the device address (device code).
- decodes the commands (operation).
- provides signals for the peripheral controller.

It is required because there are many differences between the central computer and each peripheral while transferring information.

Mode of data transfer:

- programmed I/O
- I/O interface
- Direct access memory

Programmed I/O:

It is a method of transferring data between the CPU and a peripheral. Each data item transfer is initiated by an instruction in the program, involving the CPU for every transaction. It is the exchange of data on any types of communication between the processor and external devices. Each data item

Common transfer is initiated by an instruction in the program, involving the CPU for every transaction. PMIO refers to transfers using a special address space outside the normal memory.

Direct access memory :

DMA is a method of transferring data from the computer's RAM ^{main memory} to another part of the computer without processing it using the CPU to speed up memory operations. The process is managed by a chip known as DMA Controller. It is also used for memory to memory "copying or moving" of data within memory. It is a mode of data transfer between the memory and I/O devices. DMA Controller is a hardware unit that allows I/O to be accessed directly. It contains several processor registers that can be written and read by CPU. DMA also used for intra-chip data transfer in multi-core processors. DMA can also be used for "memory to memory" copying or moving of data within memory.

notes

Phone

email

website

*

03/2020					
Monday	30	2	9	16	23
Tuesday	31	3	10	17	24
Wednesday	-	4	11	18	25
Thursday	-	5	12	19	26
Friday	-	6	13	20	27
Saturday	-	7	14	21	28
Sunday	1	8	15	22	29

Chapter - 6

Wk-06 Day 035-331
TUESDAY

04

I/O interface and Bus Structure:

1) BUS and System Bus:

2) Types of system Bus

→ Data

→ Address

→ Control

3) Bus Architecture

4) SCSI

5) USB

6) Basic parameters of Bus design.

BUS: It is a high speed internal connection.

In computer architecture a bus is a communication system that transfers data between components inside a computer or between computers. This covers all related hardware components (wire, optical fiber, etc). and software including communication protocols. It is a subsystem that is used to connect computer components and transfer data between them.

Ex- An internal bus connects computer internals to the motherboard. A bus may be parallel or serial.

The speed of bus is measured in MHz (megahertz).

→ It is a set of physical connections (cables, printer circuit etc), which can be shared by multiple hardware components in order to communicate with one another.

System Bus:

The system bus is a pathway composed of cables and connectors used to carry data between a computer microprocessor and the main memory.

The system bus is a parallel collection of

conductors (circuit has two or more paths for current flow and voltage remain same) that carry data and control signals from one component to other.

Types of system bus:

There are three major system bus types:

➤ Address bus:

It is a unidirectional pathway, which means that information can only flow one way. Its function is to transfer data between devices. A collection of wires connecting the CPU with main memory that is used to identify particular locations (addresses) in main memory. The width of the address bus (that is, the number of wires) determines unique memory locations can be addressed. The address is stored in the form of binary numbers to enable the data bus to access memory storage, that are identified by the hardware address of the physical memory (physical address).

- ➔ The address bus consists of 16, 24, or 32 signal lines.
- ➔ CPU reads/writes data from the memory by addressing a unique location; o/p of location of data by the address bus. memory uses this address to access the proper data.
- ➔ The address bus is unidirectional.

➤ Data bus:

It is also called a processor bus which is governed by a bus controller that regulates the speed of information between components. The data bus allows data to travel between microprocessor (CPU) and memory RAM. It is a system within a computer or device,

consisting of a set of wires, that provides *transportation for data - It also transfer information between two computers.

Monday	30	2	9	16	23
Tuesday	31	3	10	17	24
Wednesday	-	4	11	18	25
Thursday	-	5	12	19	26
Friday	-	6	13	20	27
Saturday	-	7	14	21	28
Sunday	1	8	15	22	29

The width and clock rate of the data bus determine its data rate (the number of bits per second it can carry). Common data bus widths include 1, 4, 8, 16, 32 and 64 bit. The data bus is bidirectional.

Control bus:-

Carries control signals from the processor to other components. The control bus also carries the clock's pulses. The control bus is unidirectional. Control bus is used by CPU to communicate with devices that are contained within the computer. This occurs through physical connections such as cables or printed circuits. Control bus minimize the lines that are needed for communication. It determines data processing. Without control bus the CPU cannot determine whether the system is receiving or sending data. The control bus is bidirectional. Control bus is used to transmit command and timing information. Command indicates operations and timing signal indicates duration of validity of data & address information.

BUS structure:

Bus: used to connect diff parts of system (comp).

Def: A group of lines that serves as a connecting path for several devices is called BUS.

A bus is a subsystem that transfers data between computers containing components inside it or betⁿ computers. Early computer buses were literally parallel electrical wires with

07

Day 038-328 Wk-06

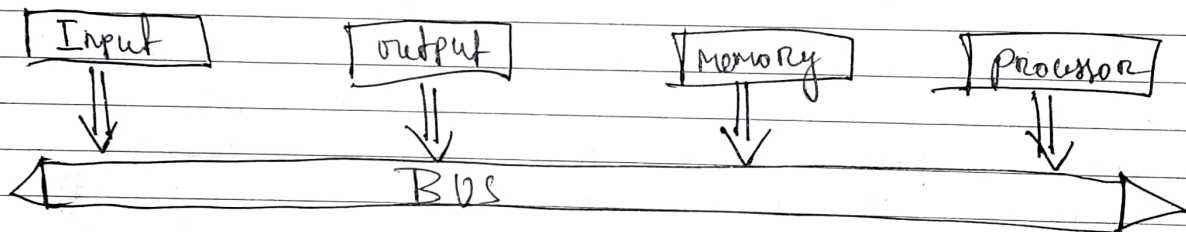
FRIDAY

Monday	-	3	10	17	24
Tuesday	-	4	11	18	25
Wednesday	-	5	12	19	26
Thursday	-	6	13	20	27
Friday	-	7	14	21	28
Saturday	1	8	15	22	29
Sunday	2	9	16	23	-

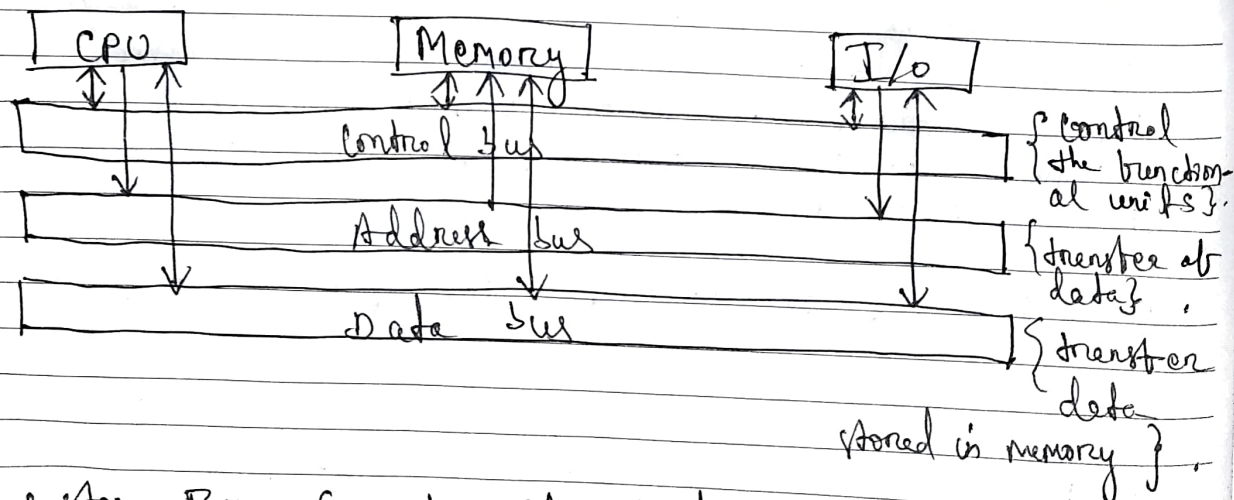
Multiple Connections, but Modern Computer Buses can use both parallel and both serial connections.

Lines: → carry data
→ Address
→ Control purposes.

→ simplest way to connect functional units to a single bus.



Architecture



The system Bus can be classified into three groups:

→ Data Bus

→ Address Bus

→ Control Bus

} Define or explain all again

notes

Phone

email

website

			03/2020	
Monday	30	2	9	16 23
Tuesday	31	3	10	17 24
Wednesday	-	4	11	18 25
Thursday	-	5	12	19 26
Friday	-	6	13	20 27
Saturday	-	7	14	21 28
Sunday	1	8	15	22 29

SCSI (Small computer system interface) :-

SCSI refers to a standard bus defined by American national standard institute (ANSI). A SCSI bus may have 8 data line which is called narrow bus and transfers data one byte at a time. A wide SCSI bus 16 data lines and transfers data 2 bytes at a time. The SCSI bus is connected to the processor bus through an SCSI controller which uses DMA to transfer data from main memory to the device or vice-versa. SCSI is the standard interface for connecting peripheral devices to a PC. SCSI is used to increase/enhance performance, deliver faster data transfer transmission and provide larger expansion for devices such as CD-ROM, Scanners, DVD drives and CD writers etc. SCSI has a controller in charge of transferring data between the devices and SCSI bus.

USB (Universal serial bus) :-

In modern computers USB was designed to standardize the connection of peripherals like printing devices, keyboards, digital still and video cameras etc. But soon devices such as printers, portable media players, disk drives etc use USB to communicate and to supply electric power. USB connectors have replaced* other types for battery charges, serial port and parallel ports etc.

notes

Phone

email

website

- 8) The simple, low cost mechanism to connect these device to computer is the universal serial BUS.
- 9) USB are designed to meet several objectives such as:
- plug and play mode of operation
 - To connect variety of device including telephone and internet.
 - simply and low cost and easy to use.
 - To add many devices to computer at any time.
- 10) The USB supports two speed operations one called low speed at 1.5 Mbps (Mega bytes per second) and another is full speed at 12 Mbps. A recent version introduced a third speed operation called high speed at 480 Mbps (Mega byte per second).

17) Basic parameter of BUS design:-

- 1) Bus Types
- 2) Method of Arbitration
- 3) Timing
- 4) Bus Width
- 5) Data Transfer Type
- 6) Block Data Transfer

1) Bus Types

A) **Dedicated**

A line is permanently assigned either to one function.

An example of functional dedication is the use of separate dedicated address and data line.

B) **Multiplexed**

Using the same lines for multiple purpose.

Eg:- Address and data information may be transmitted over the same set of lines.

At the beginning of the data transfer the address is placed on the bus and the address valid line is activated.



The address is then remove from the same bus line is used for data transfer.

C) Physical Dedication

The use of multiple buses,each of which connects to only a subset of modules.

2) Method of Arbitration

Determining who can use the bus at a particular time.

A) Centralized

A single hardware device called the bus controller or arbiter allocate time on the bus.

The device may be a separate or a part of a processor.

B) Distributed

There is no centralized controllers.

Each module contains assess control logic and the modules act together.

3) Timing

A) **Synchronous Timing**

Bus includes a clock line upon which a clock transmits a regular sequence of alternating 1's and 0's

A single 1-0 transition is referred to as a clock cycle or bus cycle.

All other devices on the bus can read the clock line.

All events start at the beginning of a clock cycle

B) **Asynchronous Timing**

The occurrence of one event on a bus follows and depends on the occurrence of a previous event.

Harder to implement and test than synchronous timing.



4) **Bus Width**

The width of data bus has an impact on the databus has an impact on the system performance.

The wider data bus, the greater number of bit transferred at one time.

The wider address bus, the greater range of location that can be referenced.

5) **Data Transfer Type**

Read-Modify-Write : A read followed immediately by a write to the same address.

Read-After-Write : Consisting of a write followed immediately by a read from the same address (for error checking purposes).

6) **Block Data Transfer**

One address cycle followed by n data cycles.

First data item to or from specified address.

Remaining data items to or from subsequent addresses.

Peripheral component inter connect BUS (PCI)

PCI bus is a well known, high bandwidth and processor independent bus. It gives better performance for high speed I/O like graphics adapter, network interface and DMA controllers for disk.

PCI bus uses up to 64 data lines at 66 MHz speed. The data transfer rate is 528 MBPS to 4.224 GBPS also economical for most of the modern I/O system requirements. It requires only a few chips to implement and support other buses.

⇒ It can support both single and multiprocessor system and provide a general purpose set of functions.

Flynn's classification

2012
OCTOBER

01

Monday

S M T W T F S							S	M	T	W	T	F	S
							1	2	3	4	5	6	7
9	10	11	12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30						

Sep 2012

- Flynn's uses the stream concept for describing a machine's structure.
- stream means a sequence of items (data or instructions).
- The classification of computer architectures based on the number of instruction streams and data streams (Flynn's taxonomy).

02

Tuesday

Flynn's taxonomy:

- SISD → single instruction single data
- classical von Neumann architecture.
- SIMD → single instruction multiple data.
- MISD → multiple instructions single data.
- non-existent, just listed for

Completeness

- MIMD:- Multiple instructions multiple data.
 → Most common and general parallel Machine.

SISD

- Corresponds to the traditional mono processor (von Neumann Computer).
 A single data stream is being processed by one instruction stream.
- A single-processor computer (uni-processor) in which a single stream of instructions is generated from the program.

Thursday

04

SIMD

- Each instruction is executed on a different set of data by different processors i.e. multiple processing units of the same type process on multiple data streams.
- executing the same instruction on all the CPUs but operating on different data streams.

2012
OCTOBER

05

Friday

Sep 2012
S M T W T F S S M T W T F S
1 2 3 4 5 6 7 8
9 10 11 12 13 14 15 16 17 18 19 20 21 22
23 24 25 26 27 28 29 30

MISD → Each processor executes a different sequence of instructions.

→ Multiple processing units operate on one single - data stream.

→ Multiprocessor machine capable of executing different instructions on different PEs ^(complexity doesn't) but all of them operating on the same dataset.

06

Saturday

MIMD :- → Each processor has a separate program.

→ An instruction stream is generated from each program.

→ Each instruction operates on different data.

→ Several processing units operates on multiple data streams.

Nov 2012

S	M	T	W	T	F	S	S	M	T	W	T	F	S
				1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20	21	22	23	24
25	26	27	28	29	30								

2012

OCTOBER

07

Sunday

SISO → Machine instructions are processed

in a sequential manner and computers adopting this Model are popularly called sequential computers.

- All the instructions and data to be processed have to be stored in primary memory.
- speed is limited by the rate at which the computer can transfer information internally.

Monday

08

PARALLEL PROCESSING

PARALLEL PROCESSING

Parallel processing is a technique used to provide simultaneous data processing task for the purpose of increasing the processing speed of a computer system. Instead of processing each instruction sequentially as in conventional computer, a parallel processing system is able to perform concurrent data processing to achieve faster execution time. The system may have two or more ALU's and be able to execute two or more instructions at the same time. Further more, the system may have two or more processors operating concurrently. The purpose of parallel processing is to speed up the computer processing capability and increase its throughput.

Parallel processing mechanism:-

A number of parallel processing mechanism have been developed in uniprocessor computers. We identify them in the following six categories.

- ❖ Multiplicity of functional units.
- ❖ Parallelism and pipeline within the CPU
- ❖ Overlapped CPU and I/O operations
- ❖ Use of hierarchical memory system
- ❖ Balancing of subsystem bandwidths.
- ❖ Multiprogramming and time sharing.