PNS SCHOOL OF ENGINEERING & TECHNOLOGY

Nishamani Vihar, Marshaghai, Kendrapara

LECTURE NOTES ON

COMPUTER SYSTEM ARCHITECTURE

DEPARTMENT OF COMPUTER SCIENCE & ENGG.

3RD SEMESTER

PREPARED BY

Er.JYOTSNAMAYEE BISWAL

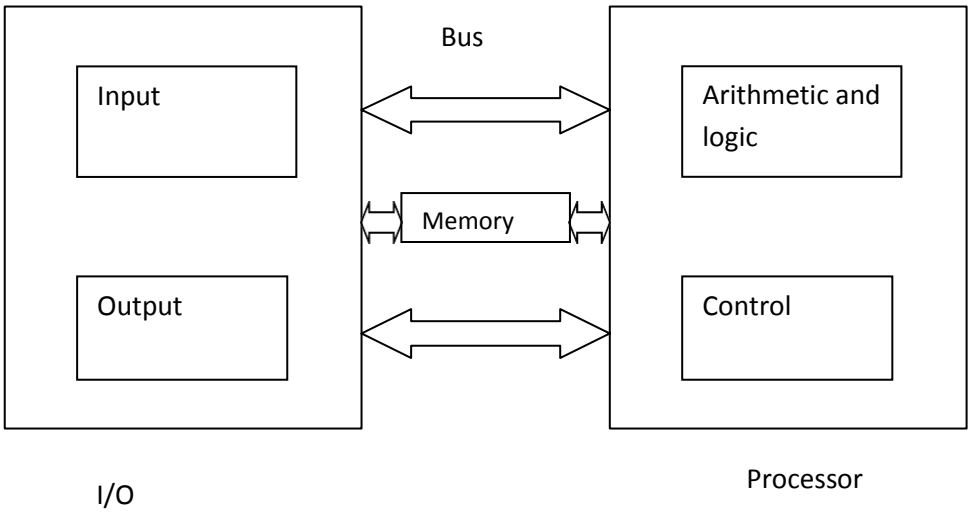
LECTURER IN COMPUTER SCIENCE & ENGG.

CONTENTS

CHAPTER-1

BASIC STRUCTURE OF COMPUTER HARDWARE

Computer is a machine which accepts input information in the digitized form, process the input according to a set of stored instruction and gives an output in form understandable by humans.



Function unit : computer consist of

1. Memory unit
2. General add special purpose registers.
3. Control circuits consisting of flip flops, decoders
4. Common data and address pulse.

Input unit:- Computer accept coded information through input unit and delivers the output through output unit.

Memory unit:-memory unit used to store programs as well as data. It may be classified into primary storage and secondary storage.

Primary storage:- It is a memory mode of up semi conductor storage cells. The these cells are grouped together in a fixed size called word. The word may be 16-bit to 64-bits.

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

CPU:-

The ALU and CU together form the control processing unit. It is known as processor (The computer performs the following functions.

⇒ Accepts program and data through input unit and store them in memory.

⇒ The store data are processed by the ALU under program control.

⇒ The process information is delivered through the logic and output unit.

⇒ The program that is to be executed is stored in memory the CPU then fetches the instructions from the memory one after another and program the required operation given by the instructions.

The data are number characters used as operands by the instructions.

Pulse:- Pulse is a single which carry some information CMOS- Complementary metal oxide semiconductor.

ASCII-American standard code for information and interchange.

Bit		X	Y	X+Y
Byte	A.D.	T	1	1
KB	Log	0	1	1
MB	Rel	1	0	1
GB	Boolen	0	0	0
TB				

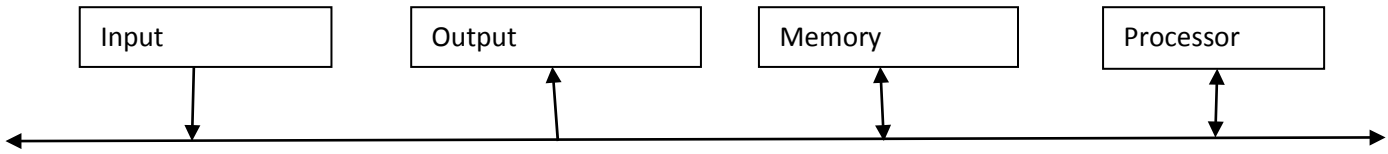
CSA

Computer components:-

Computer consists of central processing unit, main memory and the input output components. Those computer are inter connected in order axid the basic function of the computer.

The main memory stores both data and instruction in binary format. Every memory location can be individuals address instructions are educated one after another in a pre defined sequence.

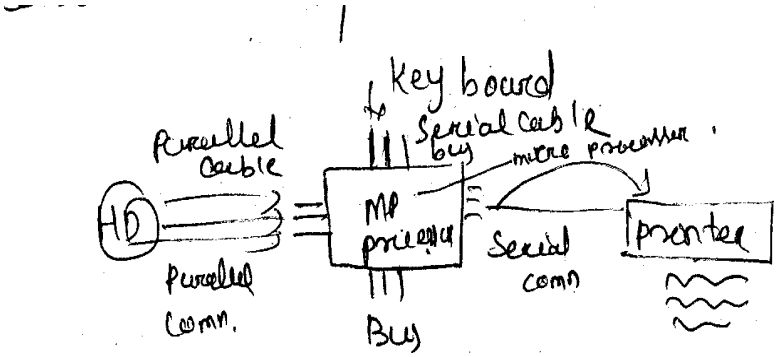
Bus structure:



All the components of the computer are interlive are inter connected so that word of data is transfer between the unit using parallel set up lines called Bus.

The lines that carry data most have the lines per address and controls purpose devices connected to the bus bary widely is their speed operations.

While some device are slow other like hard disk drive are faster memory and processor unit operate at electronic speed.



Performance:- performance is the ability of the computer and to quickly exact the program. the speed at which the computer execute the program OS design by its hardware and machine language instruction.

Basic performance mesure:-

The speed of operator of an com is generally decided by

i) Response time
ii) Through put.

1) **Response time:-**Response time is the time spent to complete and even or ad operation (Execution time).
2) **Through put:**- Through put is the amount of work done per unit of time that is called through put.
   ⇒ The amount of processing a accomplished during giving interval time (Band width).
   ⇒ Time spent from the start of execution of a program to its completion as called elapsed.

A = 100
G = 200
C = a+b
} ↓ time el time.

Performance parameter:-

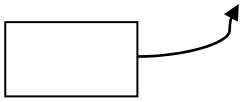
The basic performance parameter equation

$T = (NxS)/R$

Where T performance parameter of an application program N number of machine language instruction required to complete the execution required to complete the execution if an program S. Average num of basic steps required are clock rate of the processor in cycle per second.

Pipe lining and parallel processing:-

C:\\ DIR:→ abc



In case of pipe lining of the instruction to a computer have executed one after another. But in case of parallel processing multiple functional unit are used to create a parallel path. True which different on completed instructions can be execution. Thus due to state of execution several instruction in every where clock cycle.

Measuring performance:-

Response time and true pull are independent entities. When measuring the performance of a system.

Performance = 1/execution time

CPU performance equation:-

The CPU are construct using the clock running at a constant time.

CPU time = CPU clock cycle for program x clock cycle time

Cycles per instructions:-

Program consists of a number of CPU instruction be represented by instruction count.

CPI = instruction Count X Clock Cycle time / clock rate

Memory location and address:-

15	A	100
	100	101
110	111	1000

r index

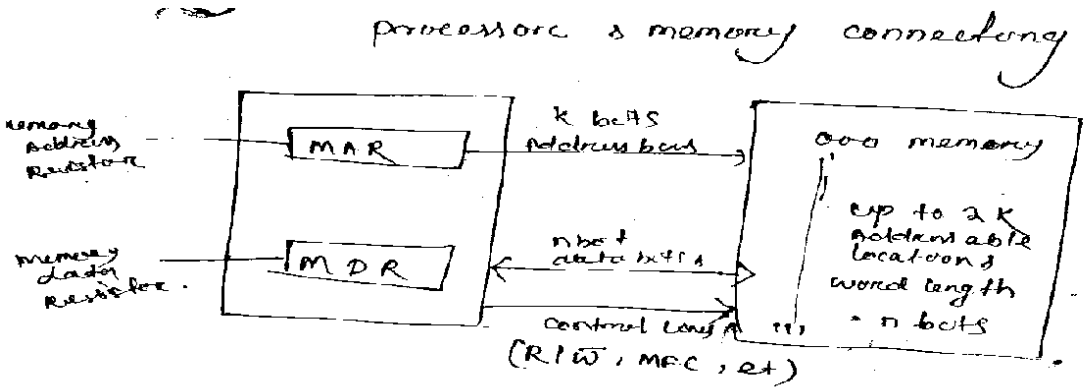
r (10)

r (1) ↓ 10	r (2) ↓ 21	r (3) ↓ 50	---	r (10) 100

The maximum size of the memory that can be used in nay computer are determine by the addressing skim. A 16 bit computer can generate a 16 bit address capable of addressing upto 2^{16} K (kilobit) memory locations

2^{32} = 4 GB

Memory operations:-



⇒ Memory is usually dissguing to store and retrieve data in word length quantities.

⇒ Data transfer between the memory and the processor takes place through the use of 2 processor register called.

1) MAR – Memory Address register
2) MDR – Memory data register.

If Mar is K bits long, it can address up to 2 K address location. It MDR is n bit longs during a memory cycle, n both data are transfer between processing and memory 2 controls line read, write bar and memory function complete.

Co-ordinate the data transfer

Resister

To Hold an information

Address MAR DATA MDR.

Instruction and instruction sequencing:-

The instruction said define many functions perform by the CPU. Information need to be provided on the various types of the data and operations to be perform on them.

This include the length instruction in bits, no of address to be used and the size of the each field.

The number of CP register that can be accessed by instruction for storage of data and operand.

Performance measures:-

Performance is the ability of the computer to quickly execute a program.

⇒ The speed at which the computer executes a program is decided by the design of its hardware and machine language instruction.

⇒ Computer performance measures is of very big term when used in context of the computer system.

⇒ System that execute program in less time are called to have higher performance.

Basic performance measures:-

The speed of operation of a system is generally decided by two fractions.

i) Response time
ii) Throughput.

Response Time:-

⇒ Response time is the time spend to complete an event or an operation.

⇒ It is also called as execution time or latency.

Throughput:-

Throughput is the amount of work done per unit of time. i.e. the amount of processing that can be accomplished during a given interval of time.

⇒ It is also called as bandwidth of the system.
⇒ In general, faster response time leads to better throughput.

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.

MIPS

A nearly measure of computer performance has the rate at which a given machine executed instruction.
⇒ This is calculated by dividing the no. of instruction and the time required to run the program

CPI/IPC

CPI – Clock cycle per Instruction

IPC – Instruction per cycle.

It is another measuring that which is calculated as the number of clock cycle required to execute one instruction (cycle per instruction) by the instruction executed per cycle.

Speed up:-

Computer architecture use the speed up to describe the performance of architectural charges as different improvement are made to the system.

It is defined as ratio of execution time before to the execution time after the charge.

$$\text{Speed up} = \frac{\text{execution time before}}{\text{Execution time after}}$$
$$= \frac{\text{performance of system after}}{\text{Execution time before}}$$

Amdahl's law:-

This law states that "performance improvement to be gained by using a faster mode of execution is limited by the fraction of time the faster made can be used".

Amdahl's law defines the term speed up.

Speed up = performance of entire task using enhancement
Performance of in time task without using enhancement

Performance = $\frac{1}{\text{Execution time}}$.

Speed up = execution time without using enhancement
Execution time with using enhancement

Factors affecting speedup are as follows:

i) The fraction of computation time in the original machine can be modified to use the advantage of the enhancement.

ii) This is called fraction enhanced which is always less than or equal to one.
Fraction enhanced ≤ 1
Ex: if a program that usually it will take 30 seconds for execution using the enhancement
Fraction enhanced = 30/100

2) Improvement granted by the enhanced execution made is the speed with which the taks could run faster using the enhancement.

Speed up > 1

Speed up enhanced = time in original mode
Time in enhance mode

Ex. Let us a program takes 5 second in enhanced mode while it takes 10 second earlier.

So, speedup enhanced = 10/8 = 2

The new execution time can be calculated as follow:

Execution time new =

Execution time original X ((1-fraction enhanced) + fraction enhanced/speedup enhanced)

The speedup overall = execution time original/execution time new

Speedup overall = 1/(1-fraction enhancement)+fraction enhances/Speedup enhance

Performance parameter

The basic performance equation is given by T = NxS/R

Where Q.t. – Performance parameter of an application program.

NS – No. of instruction required to complete the exe[2] of a program.

R-Clock rate of the processor in cycles per second.

S-Avg. no. of basic step required to execute one machine instruction.

Clock rate:-

Clock rate is one of the important performance measures by improving in the clock rate. There are two ways in which clock rate may be increased.

1) Improving IC technology which makes logic circuits faster thus reading time taken to complete a basic step.
2) By reducing the processing amount in one basic step which by reduces the clock period as R = 1/T.

CPU performance Equation:-

Normally the CPUs are constructed by using a clock running at a constant rate. This discrete time events is known as a clock cycle.

CPU time = the time of a program may be represented as

= CPU clock cycle for a program x clock cycle time

= CPU clock cycle/ clock rate

= instruction count X CPU

IC X CPI X C

Questions:-

1. **Define computer architecture.**
2. **Explain the basic functional units of a computer.**
3. **What is non-neuman architecture?**
4. **Explain virtual memory.**

CHAPTER-2

INSTRUCTION AND INSTRUCTION SEQUENCING

The types of addressing modes are available with instruction set.

1) functionality complex minimal instruction set.
2) Instruction set based on speed of execution 'RISC' reduce Instruction set computer.
3) A more elaborate instruction set that into frequently used sequence including a single processor operation.

Types of operands:-

Machine operations depends on the types of data being process . operand can be one of the following.

1. Address
2. Numbers
3. Characters
4. Logical data

Address:- Address are the for of number that represent specific location in memory.

Number data type:- Number data type are used by all machine language data type fix point floating point and decimal.

Character:- character are entered using ASCIIC. Encoding also another including character i.e. EBCDIC (Extended binary coded decimal inter change code) used for character.

Logical date:-

The Boolean data can be stored using 1(True) 0(false).

OPcode types:-

1. Arithmetic – add supply multiply divide
2. Logical – AND, OR, NOT
3. Data transfer – move, store, load, push, pop
4. Conversion – translate, convert
5. system control – reserve for operating system
6. Input output control – Input-readout – white.
7. Control transfer – Jump, returned, skip, halt.

Instruction format:- Opcode 4 bit Op1 Op2
6 bit 6 bit

An instruction is read into the instruction register (IR).

OP Code

Operation code	Address of Operand

→ Format of instruction

Operation Code (OP Code)

This specifies the operation to be performed.

e.g. :- Add, sub & Load

Address of Operand:-

The operand specified by the OP code may involve one or more sources for operands. These operands are the inputs for operations.

Classification of Instruction:-

Instruction are classified depending upon the number of operand address they contain such as classification is follows:-

1. **0-Address instruction**:-
   The 0-address type instruction do not contain any operand address. The operand address are implied.
   e.g. :- ADD TO S ← (A+B)
2. **1-Address Instruction:-**
   In 1-adress instruction only one operand address is specified in the instruction. The other operand is in accumulator.
   e.g.:- LOAD A AC ← M [A]
   ADD B AC ← AC+M[B]
3. **2-Address instruction:-**
   In 2-address instruction both operand address are specified. The result is placed in one of the specified address.
   e.g.:-MOV R_1 , A R_1 ← M[A]
   ADD R_1, B R_2 ← R_1 + M[B]
   MOV R_2, C R_2 ← M[C]
4. **3-Address instruction:-**
   In 3-address instruction two address are specified for two operands & one address for the result
   e.g.:-ADD R_1 , A, B R_1 ← M[A] + M[B]
   ADD R_2, C, D R_2 ← M[C] + M[D]
   MUL *, R_1, R_2 M[X] ← R_1*R_2

Addressing technique:-

To specify a memory address in an instruction word, the most obvious technique is simply to give the address in binary form. This called direct addressing, although direct addressing provides the most straight forward way to give a memory address, several other techniques are also used. The use of the these techniques is generally motivated by one of the following considerations.

(1) Desire to shorten address section
(2) Programmer convenience.
(3) System operation facilities.

Addressing mode:-

Each instruction needs data on which it has to perform the specified operation. The data (operand) may be in 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. We will discuss here only six type of addressing modes.

(1) Direct addressing
(2) Register addressing
(3) Register indirect addressing
(4) Immediate addressing
(5) Base register addressing
(6) Indirect addressing

(1) Direct addressing

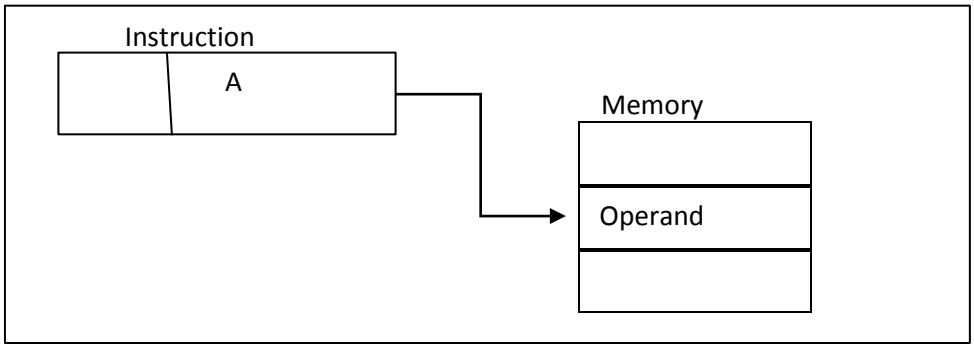
In direct addressing the address of the data (Operand) is specified within the instruction itself.

e.g. :-

STA 2500H - Store the contents of accumulator in the memory location 2500 H. Here 2500H is the memory address. Where a data are to be stored.

LDA 2500H - Load the accumulator with the contents of the memory location 2500 H.

In 01 - Read the data from input device whose address is 01, here 01 is the address of an input device where data are to be read.



(2) Register addressing

In register addressing the operands are located in general purpose registers.

In other words the contents of a register is the operand.

Therefore only the names of the registers are to be specified in the instruction.

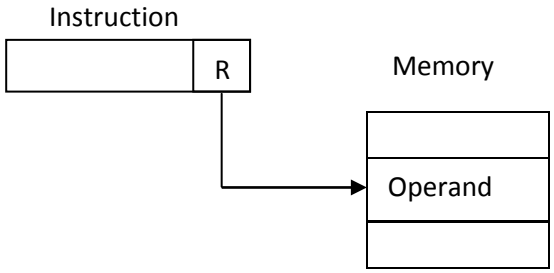
e.g. :-

MOV A, B = Transfer the contents of register B to register A. the Opcode of this instruction is 78H in binary form is 01111000.

The1st 2 bit 01 denote move operation, the next three bit 111 are binary code for register A & last three bit 000 are binary code of register B of Intel 8085.

ADDB = Add the contents of the register B to the content of the accumulator. The OP code of this instruction 804 in binary from 10000000. The 1st five bits 10000 specify the odd operation to be performed the last three bits 000 are for the

binary code of register B.

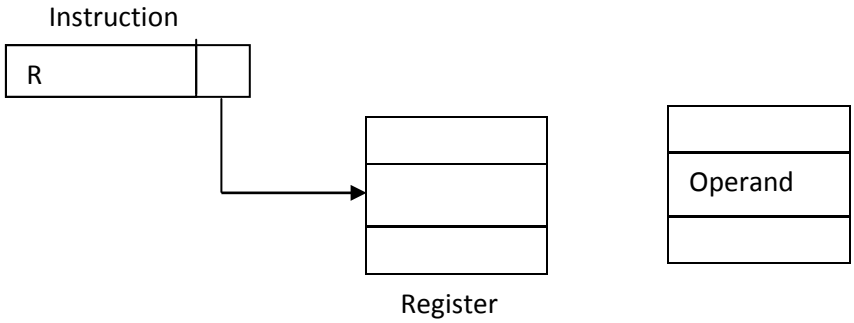


(3) In register indirect addressing address of the operand is given indirectly. The contents of the register or register pair are the address of operand.

LXIH, 2400 MOV A, M = Load H-L pair with 2400 H move the content of the memory location (M) whose address is in H-L pair)1, 0, 2400 H), to the contents of the accumulation.

In this example MOV A, M is an example of register indirect addressing. The address of the operand is not directly given but the address of the memory location is stored in H-L pair, which has been specified by the earlier instruction LXIH, 2400 H.

LXIH, 2200 H ADDM - Load H-L pair with 2200H add the content the memory location (M) whose address is in H-L pair to the contents of accumulator. Here ADDM is an example of indirect addressing.



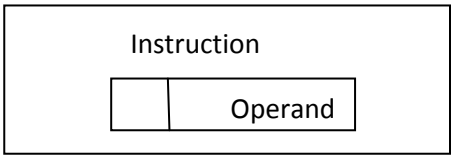
(4) <u>Immediate addressing:-</u>

In immediate addressing the operand is given in the instruction it self.

e.g. - # MV1, 06 move 06 to accumulator

ADI 05 ADD immediate 05 to the contents of a accumulator

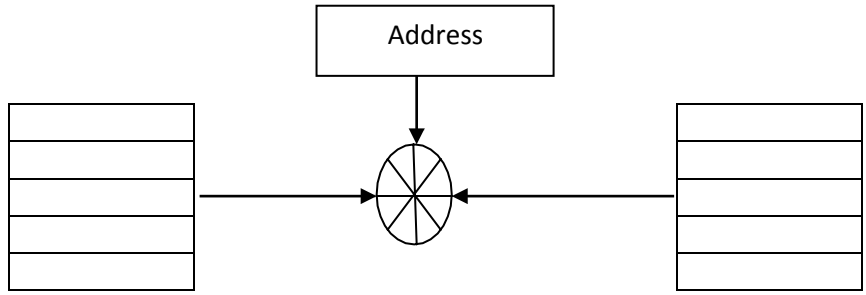
LXIH, 2500 H

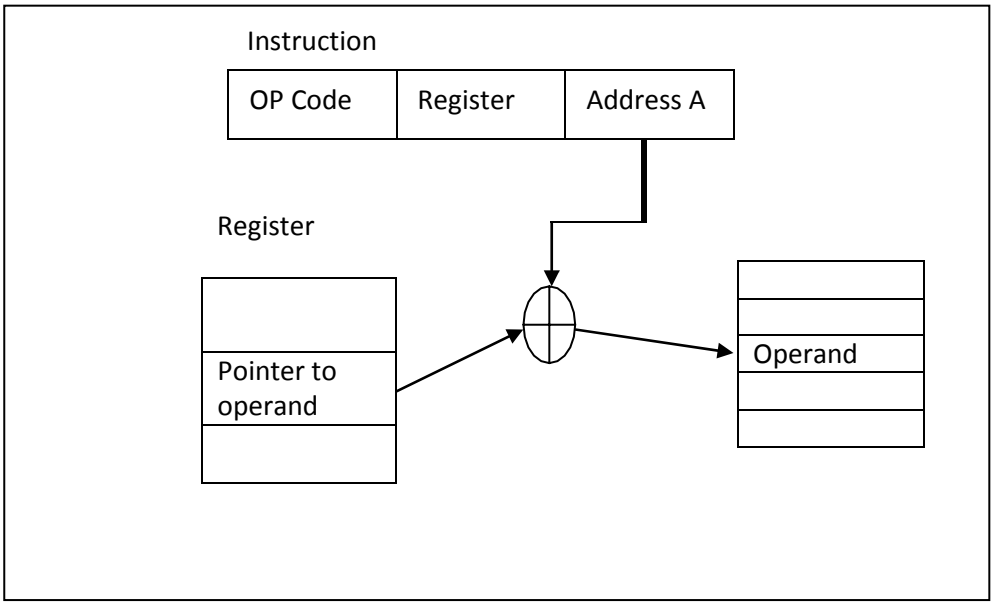


(5) <u>Base register addressing:-</u>

In many computer system it becomes necessary to move programme from one place to another, in the memory. To solve this problem many computer use base register. The addressing mode employing a base register is known as Base register. In this mode of addressing an offset is added to the contents of the base register to obtain the effective address.

The instruction contain an offset. When a program is moved into memory. The starting address is placed in the base register.

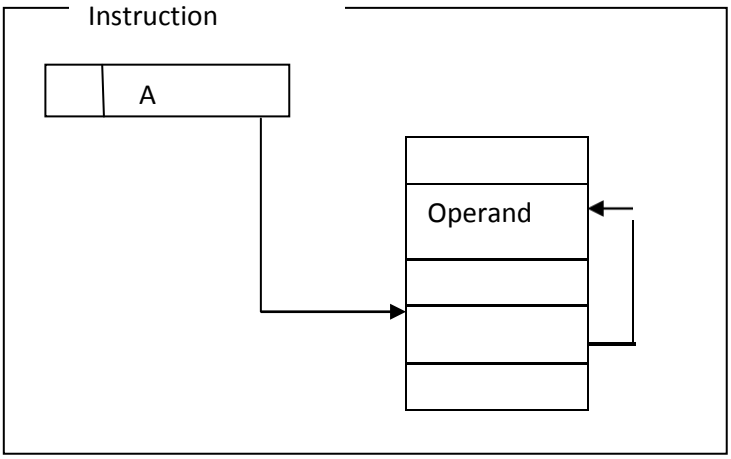




(6) <u>Indirect addressing:-</u>

In indirect addressing the address field of the instruction gives the address of the instruction gives the address of the operand. Control fetches the instruction from memory and uses its address port to access memory register to read the effective address.

e.g. - # IAD 302 – Indirect add the number whose address is stored at the address 302.



Instruction set:- Instruction set based on number of address.

1) Pre address instruction
   Ex:- x=(A+B)X(C+D)

   3Addness instruction:-
   ADD = R1 A, B, R1← M[A]+M[B] – add B with a
   Add = R2, C,D,R2 ← M[C]+M[D]
   NUL = X1R1,R2, M[X] ← R1*R2

Computer using 3 address instruction have an address field to specify either process register or a memory address from where operands can be paste.

Addressing technique:-

To specify a memory address in a instruction words the most obvious technique is simply to give the address in binary form.

This called direct addressing

Addressing mode:-

The operand field of instruction specified to address from where the data has to be fetched. The addressing mode specifies a rule for interpreting or translating of the instructor into an effective address from where the operand is actually reference. The control unit of computer should go through an instruction cycle to execute an instruction.

A computer has a resister called the program counter which keeps the stack program instruction that was store memory. It holds the address of next sequential instruction to be fetched. The decoding determines the operation to be perform.

The operands are fetch either form memory or the register.

The instruction is executed and the result will be put pack into the operand address. The various addressing model are.

1) **Immediate addressing:-** In the mode of addressing the operand in the part of instruction and it specified in address field. This called immediate addressing.
   Add-5 Operand
2) **Direct Addressing:-** The effective address is equal to the address part of the instruction. The operand register in the memory and its address is gen directly by the address field of instruction.
3) **In directing Addressing:**- the addressing field of the instructed uses the address filed of the operand stored in the memory.
4) **Resister addressing:**- here the addressing field register to register.

Register in directing addressing:- The address field refers to a resister and the effective address is the address given by the register given by of the address field.

Stack addressing:- Stack addressing in form of addressing.

Questions:

1. **Explain addressing mode.**
2. **Explain the various types of instructions.**

CHAPTER-3

PROCESSOR SYSTEM

Processor unit:-

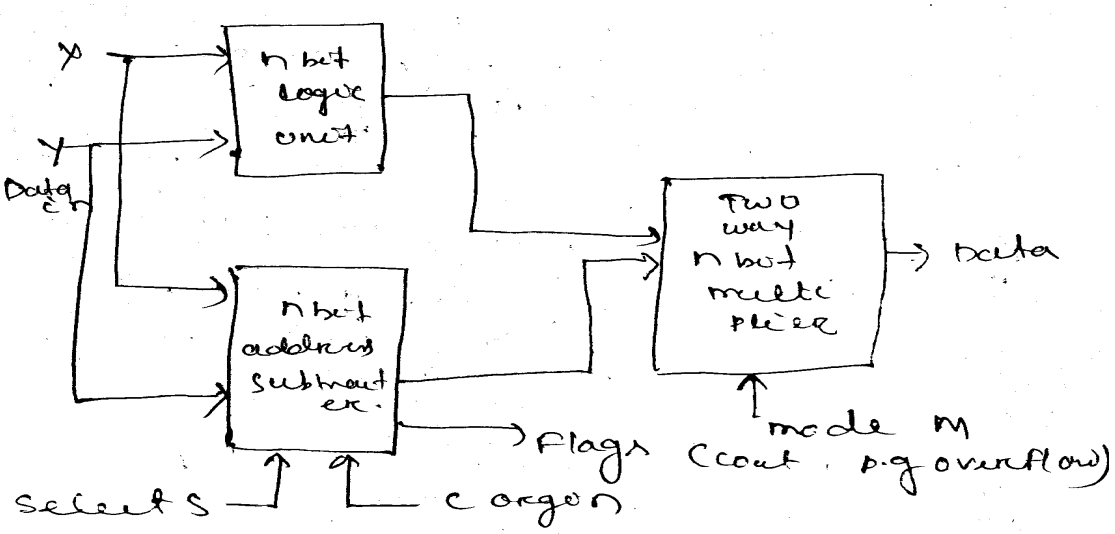
Design of ALU:- The ckt that carryout the data processor instruction. The arithmetic logic unit. The ALU using combinational ckt that performs, subtraction ward base logical operation multiplication and division. This can be complemented some time using auxiliary arithmetic unit called coprocessor.

80286 – 80287

80386 → 80387

Combinational ALU:-

Basic ALU n a bit:-

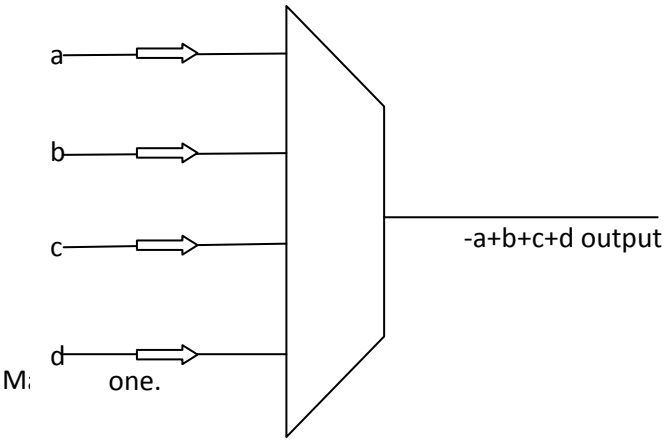


The fig shows the simple ALU combine with adder, subtractor with word base logic functions such as AND, Ex-OR, NOT for fixed point instruction.

Then the carry in select way.

The mode control line m attach to the two way n bit multiplexer determined the type of operation that is arithmetic or logical to be perform by the ckt.

The result in channel out thought the output bus Z. the select control line determines the specific operation to be perform by the sub unit. The multiplexer.



M one.

The ALU performs the bit close operations for input data lines. The maximum number of distance.
Logical operation is $16\text{-}2^4 = 16$.
So the select bus needs to be size of 4, $2^2 = 4$.
M1 = xy
M1 = xy
M1 = xy
M1 = xy
Sum of the product.
P(x,y) = xys3+
Dy so
Sy sit
Sy sn+

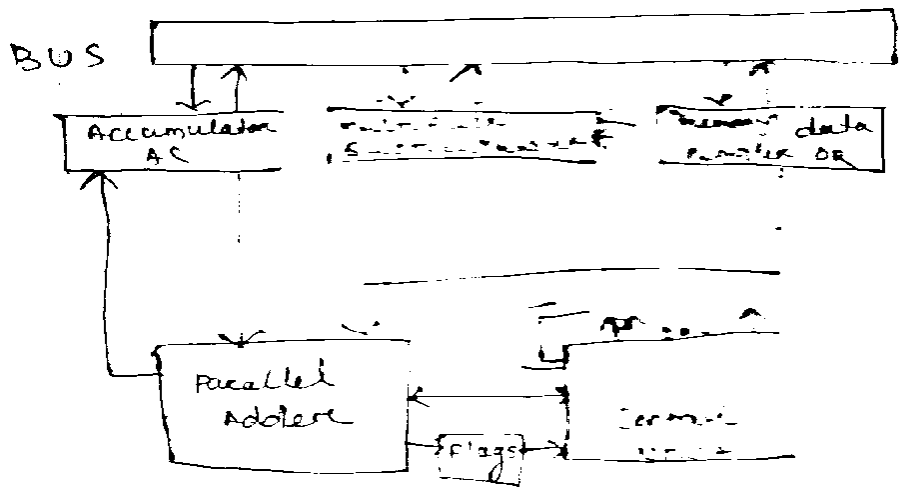
Arithmetic

Operations	Resister usage
Addition	AC = AC+DR
Subtract	AC = AC-DR
Multiply	AC-MQ = DRxMQ
Division	Ac-MR = MQ/DR

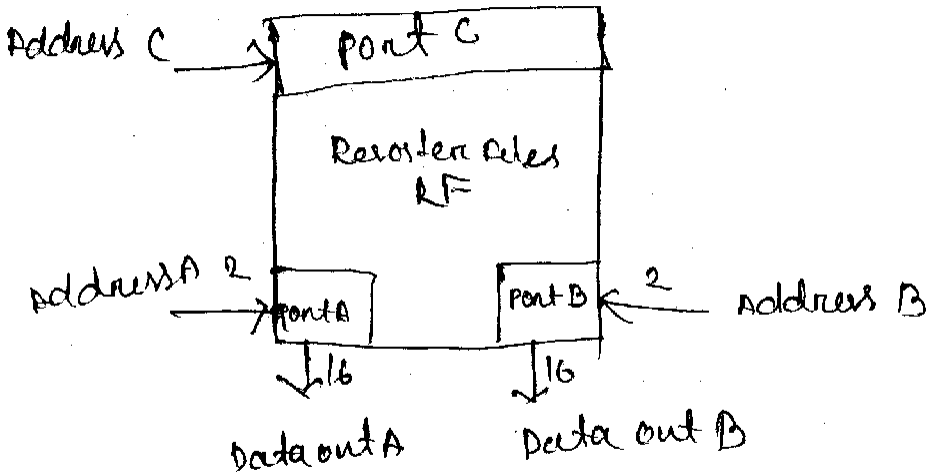
The multiplication and division is done using one sequential digit by digit sift and add or subtract algorithm. The one word resisters namely accumulator, multiplier/quotient (MQ) and the data resister (DR) are use for operand storage.

AC and MQ are organize as a single register and at capable of sifting right or lect. The adder substractor unit and the derived input spore AC and DR placed its result in AC. The MQ resister stores multiplier during multiplication and quotient division.

Basic sequential ALU:



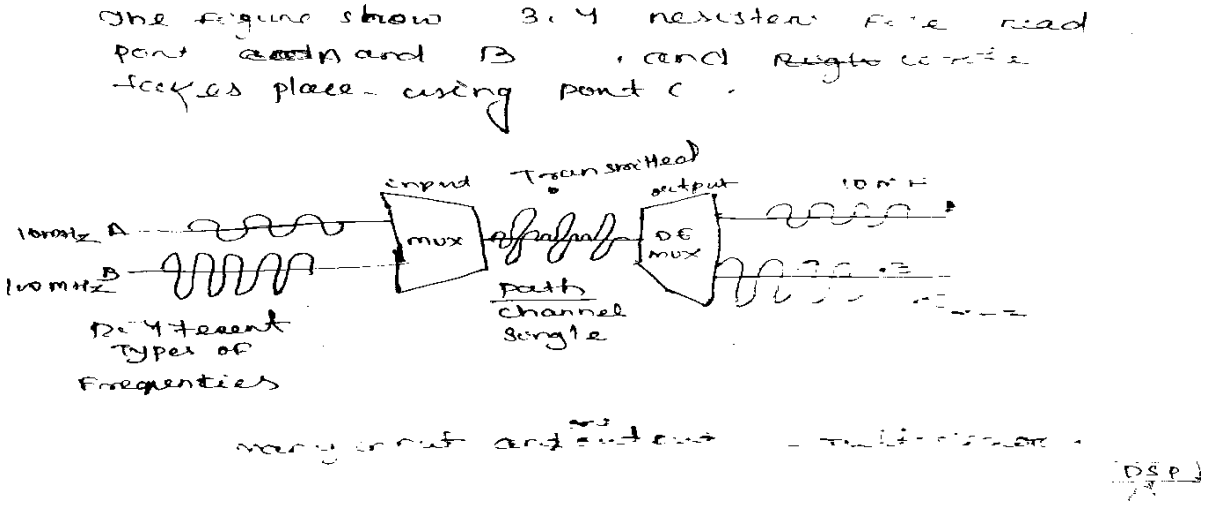
Register files:-



In modern CPU have a set up general purpose register R02, M-1 called resister file replacing AC-(Accumulator), DR-(Data Resister), AR – (Address Resister).

Each resister are subscript Rm. Rm register file is individually quick address subscript m0.
Sta R1, R2
Add R3, R1 R2
R3 = F(R1, R2)
R2 = F(R1, R2).

The register file performs its function it random access memory (RAM).
A multiport resister file is build a set up resister and multiplexor and A multiplexor combination. Allows the data to be operation transform any thus register to output port read and various input port to register.



Data path design:-

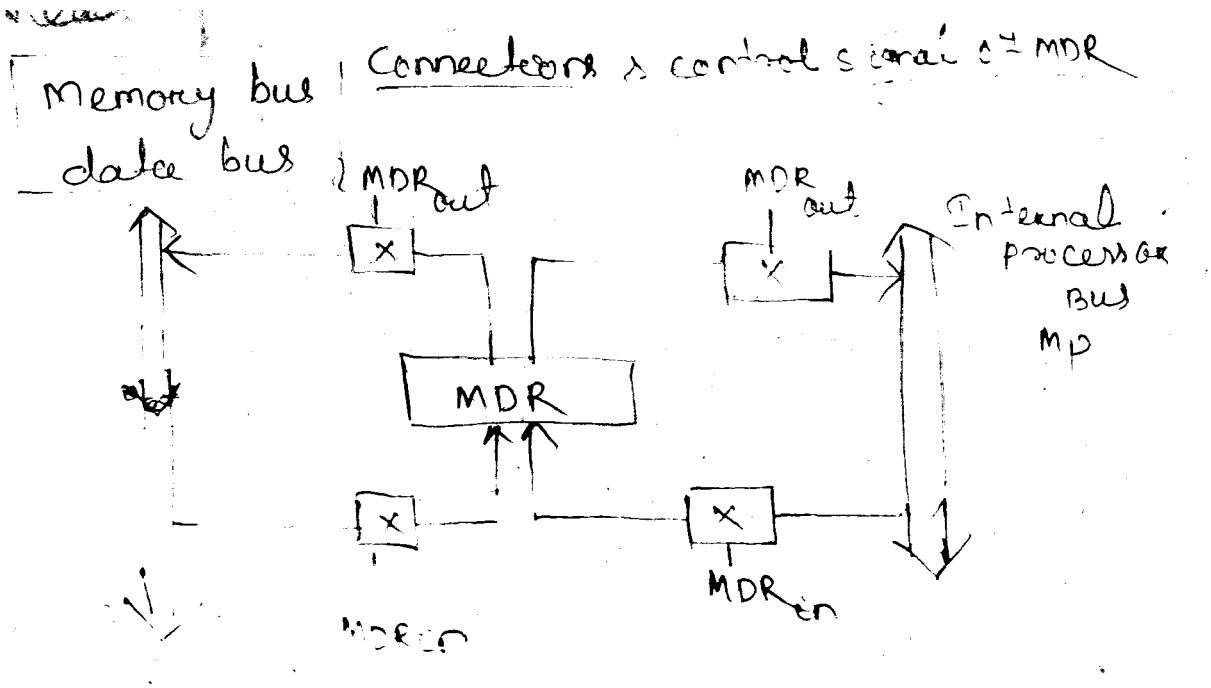
⇒ It is a unit for implementing logical and fixed point operation it consists of a register file and a combinational ALU capable of addition or subtraction. The entire sequential ALU for fixed point no. 5 can be built on a single IC. The ALU can be designed for expansion to handle operand of bigger size using (i) special expansion bit slice processor, (ii) Temporal expansion bit slice processor.

1) **Special expansion bit slice processor:-**
   This connects K-copies and M-bit ALU to form a single ALU of processing K-M-bit directly for addition the resulting array like circuit is said to be bit slice because each component ALU concern the processor a separate slice form each km bit operand.
2) **TEBSP:-**Temporal expansion bit slice processor:- this uses n copy of n bit ALU in the manner of serial adder to perform and operation on KM-bit word in k conjugative pluck cycles. Here the ALU process a separate m bit slice or each operand. This is called multi cycle processing. Here the data buses and register rules of individual slices are effective input to increase the size from 0 bit to 16 bit.

Basic memory operation:-

All data and instructions are stored in memory before and after their used. These data should be transfer back and fourth using read, write operations in memory.

Read operation:-



The information to be fetched from memory may represent on instruction in a program. The processor has to specify the address of the memory location where this information is stored on request read operation the processor transfer the address to MAR. The requested data are receive from memory and store in MDR (Memory data register).

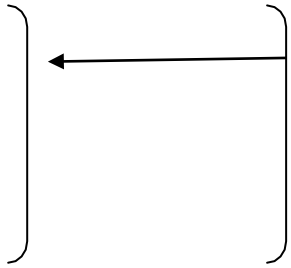
During the memory read and write operation the timing of internal processor operations most be coordinated with the response of the address device on memory bus.

The processor complete one internal data transfer in one clock cycle. The control signal called memory function completed is used for thus purpose.

Ex: MOV R3 R4
11 11
MAR←[R3]
R4←[MDR]

Start a read operation on the memory bus. Wait for MRC response from the memory. Load the MDR from the memory bus.

1. R3 out, MAR in, READ
2. MDR in, MFC
3. MD Rout, R4 in



Memory read operation using 3 steps:-
Memory write operation using 3 steps:-

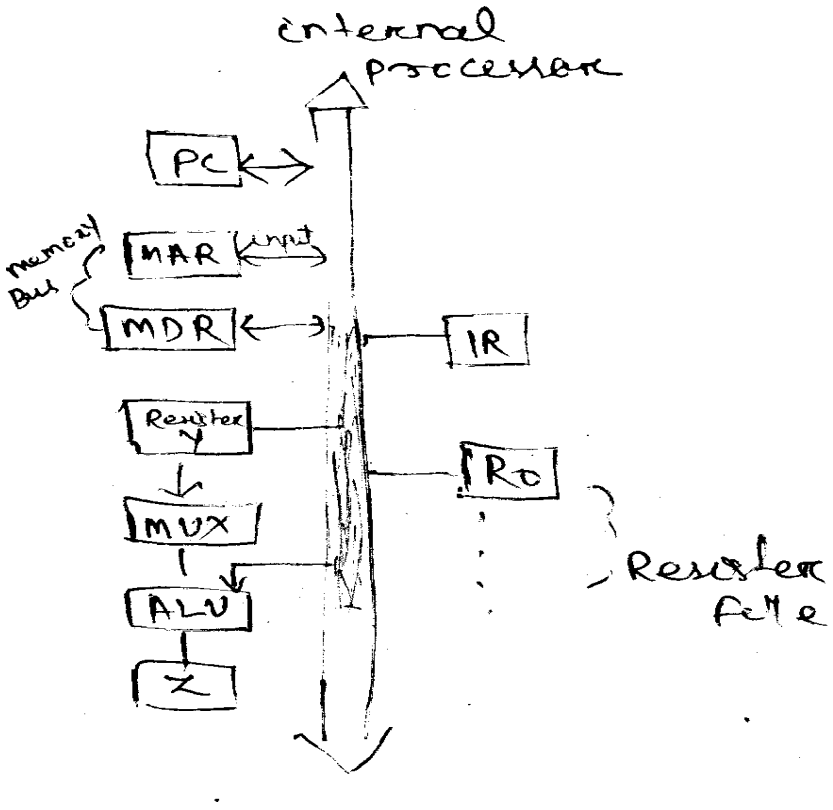
1. R3 out, MAR in
2. R4 out, MDR in, WRITE
3. MDR out, MFC

The memory address where the data is to be written is loaded into MAR. the data should be written are loaded into MDR and a write command issued until the memory operation is completed [MFC].

Complete instruction execution:-
ADD (R4), R2

Executing this instruction requires the following:
Steps:-

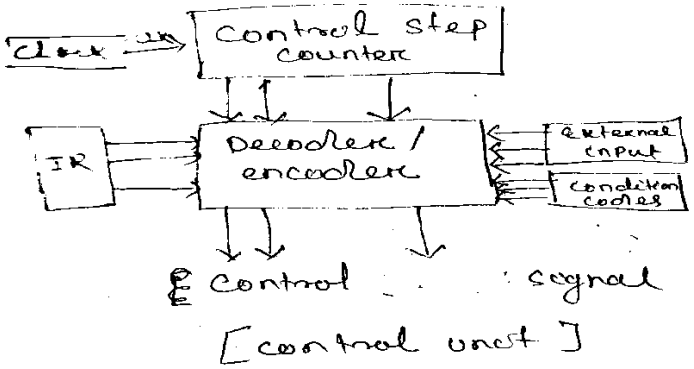
1. Fetch the instruction
2. Fetch the first operand
3. Perform the addition
4. Load the result into R2

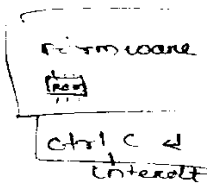


1. Pcout, MAR in, Read, Select-4, ADD, Zin
2. Zout, Pcon, Yin, Mfc
3. MDR out, IRin
4. R4, MAR in, READ
5. R2out, yin, MFC
6. Rout, Select Y, ADD, Zin
7. Zout, R2in, End

PC- Program counter points to the address of the next instruction to be executed.

Hardware Control:-





For an instruction to be executed the processor most generate the control signal in proper sequence. i.e. either in Hardware control on microware control.
In the control unit each step of the sequence is completed in one clock period. A counter is used to keep track of the control steps.

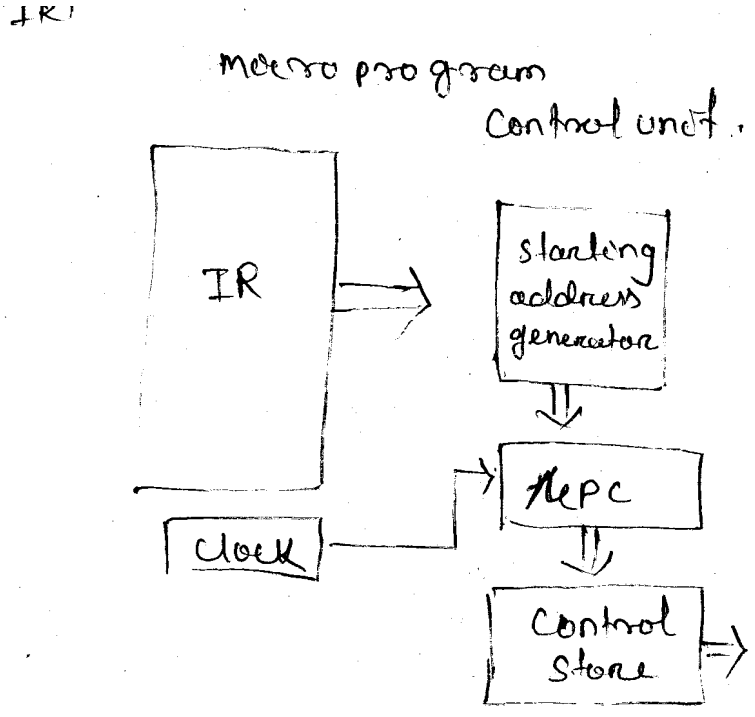
The control signals are determined by the instruction contend in the following

1) Step counter
2) Condition code flag
3) External input singles. Such as MFC inter request.

Microprogrammed Control:-

In microprogrammed control a program similar to the machine language program generates the control signal. A control word (CW) represents various control signals. Each of the control steps in the control sequence of an instruction detain a unique combination of 1s, 0s in the control word. The individual control word in the micro routine is refer as micro instructions. The micro routin for all instructions of a computer are stored in a special memory called the control store. The control unit can generate the control signal for any instructions.

A microprogram counter is used to read control words from the control stored. The starting address generator is loaded to micro program counter every time a new instruction is loaded into instruction register (IR).



The control signal are delivered to various pants of the processor in the current sequent.

```
MOV     R1, R2 ……
ADD     R1, R2
SUB
HLT
```

The MPC (micro Program counter) is coded with.

1. Starting address of the micro instructions when a new instruction is loaded.
2. Branch address when a branch micro instruction is encountered and the condition is satisfied.
3. The address of the first control load of the next instructions when end is encountered.

```
001     MOV   R1 R2
        ADD   R1 R2
 00     SUB
        JMP
010     END
```

CONTROL SIGNALS

To execute instructions, the CPU must have some means of generating control signals in the proper sequence computer designers have used a wide variety of techniques to solve this problem. Most of these techniques, however, fall into one of two categories.

1. Hardwired control
2. Micro programmed control

Hardwired control:-

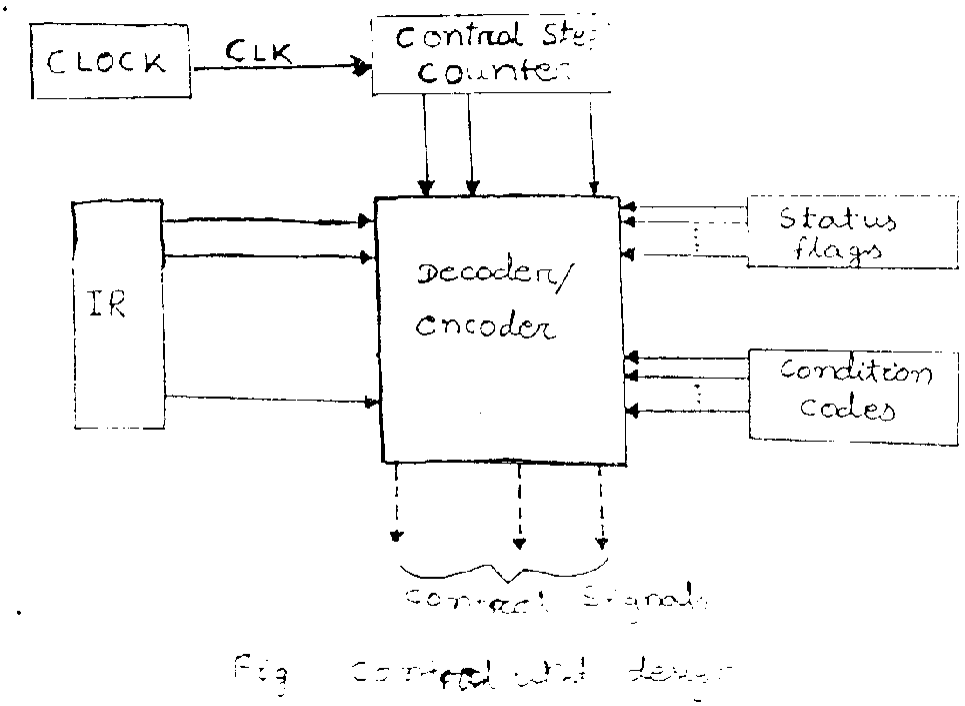


Fig Control unit design

The control unit based on the use of a counter device by a clock signal, CLK, as shown in the above figure. Each state or count of this counter corresponds to one of the control signal. The control signals are uniquely determined by the following information:

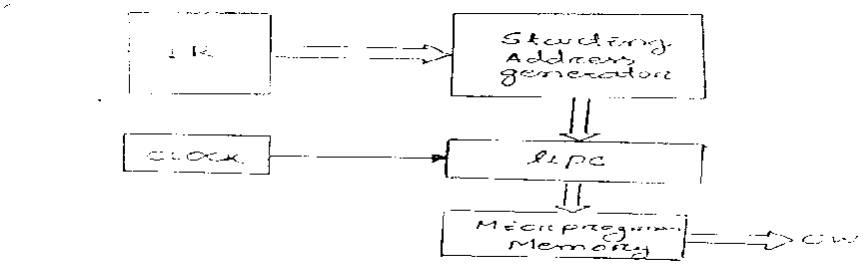
contents of the control counter

contents of the instruction Register

contents of the condition code or other status flags.

In order to gain some insight into the structure of the control unit we will start by giving a simplified view of the hardwired involved. The decoder-encoder block is a combinational circuit that generates the required control outputs depending on the state of all its inputs. The instruction register contains the opcode of the current instruction which is used to determine which micro operation to perform during the execution cycle. The status flags are needed by the control unit to determine the status of the CPU and the outcome of the previous ALU operations.

Micro programmed control:-



let us start by defining a control word (CW) as a word whose individual bits represents various control signals. The micro program corresponding to the instruction set of computer are stored in a special memory called micro program memory. The control unit can generate the control signals for any instruction by sequentially reading the CWs of the corresponding microroutine form the micro program memory. To read the control words sequentially from the micro program memory a microprogram counter (MPU) is used. Every time a new instruction is loaded into the IR, the output of the block labled "starting odder generator" is loaded into the MPU. The MPU is then automatically incremented by the clock, causing successive micro instructions to be read from the memory. Hence the control signals will be delivered to various parts of the CPU in the correct sequence.

Questions:

1. **What is sequential ALU?**
2. **What is Micro Programming?**
3. **Write notes on Micro Programmed control unit.**
4. **List the advantages and disadvantages of Micro Programmed control.**

CHAPTER-4

MEMORY SYSTEM

Characteristics of memory:-

The characteristics of the memory is its capacity for internal memory, it is expressed in terms of byte or words. Word lengths are 8,16,32,64 bits. External memory capacity is topically expressed in terms of bytes.

Word:- it is the natural unit organization of memory. The length of word is topically equal to the number of bits used to represent a number and to the instruction length (normal comput 32 bit).

Addressable Unit:- The addressable unit is the word. 2A=N many system allow necessary in byte legal.

Unit of transfer:- This is the number of bit read out or written onto memory at a time.

Sequential:- Access most be made in a specific linear sequence.

Direct Access:- The individual blocks or records have a unique address based on physical location. Access is accomplished by direct access to reach to the final location.

1	Name	ADD	Sal	Deign
2				
3				
4				
5000, 5				
100,000				

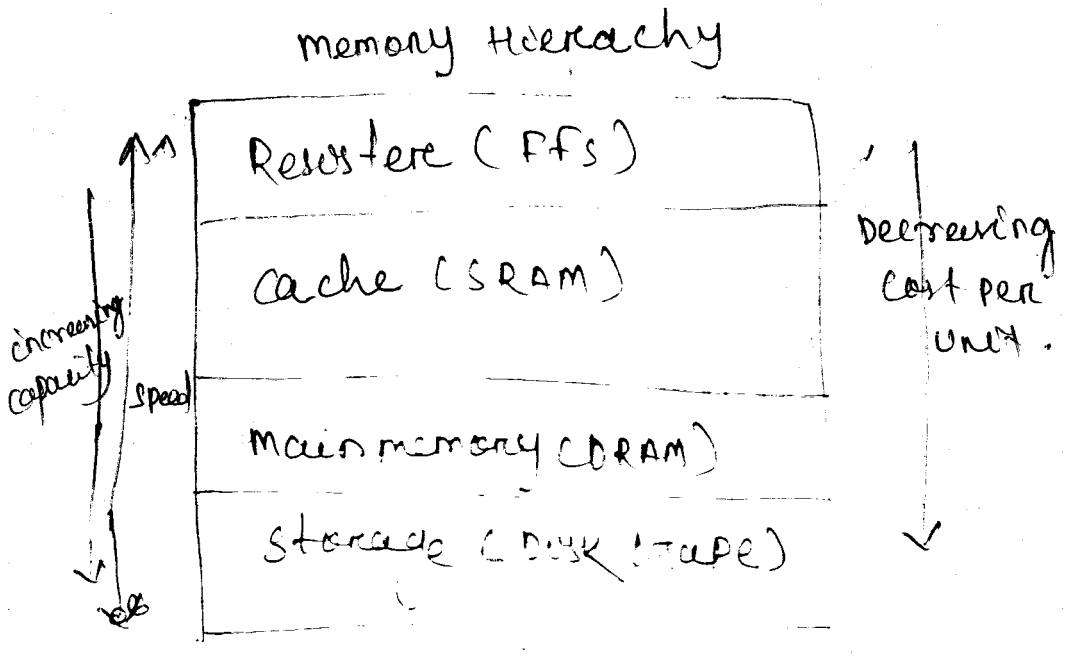
Random Access:- Each addressable location in memory has a unique, physically addressable mechanism any location can be selected at random and directly accessed. The main memory and cache memory used random access.

Three performance parameters are used:-

1. **Access time (latency):-** For random Access memory this is the time of takes to perform read or wrote operations. For Non RAM, access time is the time taken to position the read/write mechanism at the designed location.
2. **Memory cycle time:-** It is applied to random access memory and consist of access time and any addition time required before a second access can start. It is concerned with system bus.
3. **Transferred:-** This is the rate at which date can be transfer in and out of a memory unit. For RAM it is equal to 1 for non RAM it is
   TN = TA+N/R
   TN – Average time to read or write n bits.
   TA – Average Access time
   N – Number of bits.
   R – Transferees in Bits for second.

Memory Hierarchy:-

The principle of locality (temporal locality and special locality) states that programs do not access code and data uniformly.



Basic concept of memory:-

→ byte address

Word address

0	1	2	3
4	5	6	7
2^k-4			2k-1

2^k-4

Big-endian assignment

→ byte address

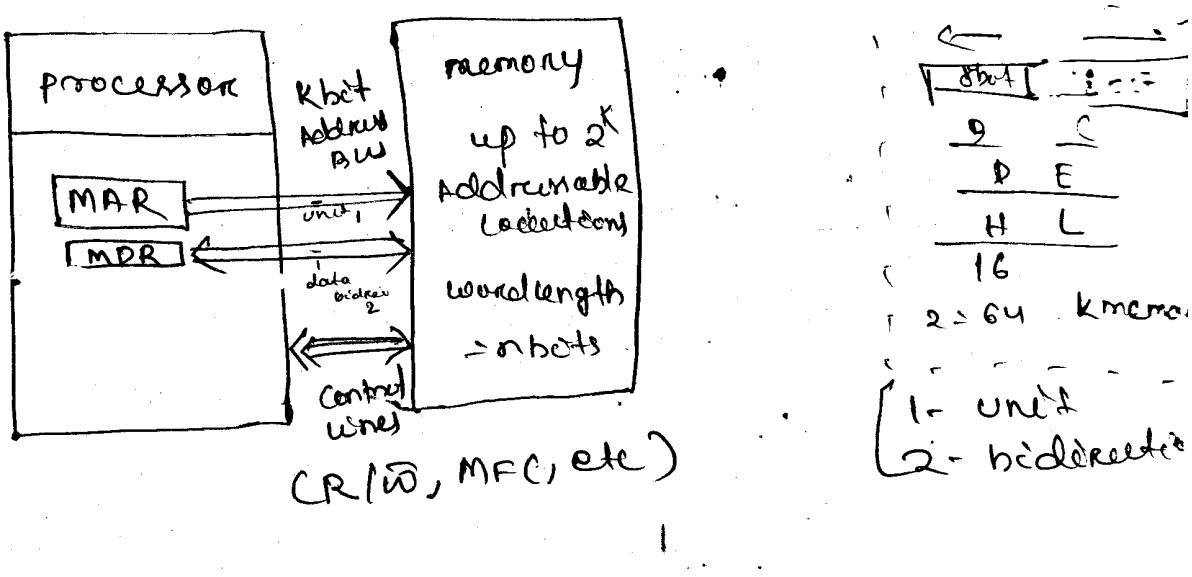
0	3	2	1	0
4	7	6	5	4
	2^k-1			2k-4

2^k-4

Little-endian assignment

T e memc

address is capable of addressing upto 2 = 64 K (kilo bite) memory location. When lower byte address is use for more significant bytes(left) of the work it is known as big-endian assignment. Otherwise the right most assignment is called little-endian. The memory is normally design to store and retrieve data in word length quantities. The data transfer between the memory and the processor takes place through the use of two processor registers. Namely 1) memory address register (MAR), 2) Memory data register (MDR).

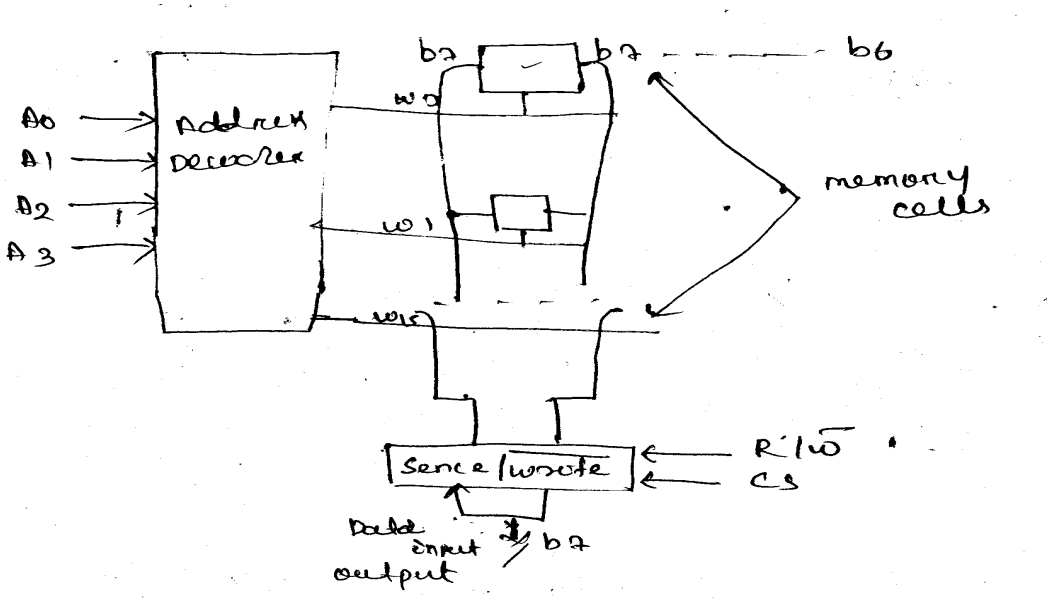


If MAR is k bit long, MDR is 8 bits long, than the memory context upto 2^k address locations and n bit of data transfer between the memory and the processor in a memory cycle. The read write bar (R/W) and memory function completed co-ordinates the data transfer. To read the data from memory by loading the address of the required memory location into MAR and setting R/W line to high.

The memory responds by putting the data from the address location on the data line with MFC signal. Similarly for write operation set R/W line to low. The processor writes the data into memory location by loading the address into MAR. the access are synchronies using a clock.

SEMICONDUCTOR RAM

Organization of memory chip:-

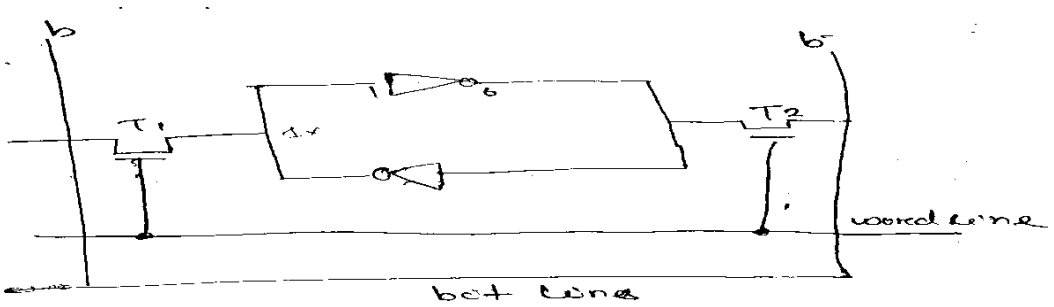


Memory cells are organized in the form of an arrays. Each cell is capable of storing 1 bit information each row of cells constitute a memory word of 8 bit. And the cells of a rows are connected to a common signal line called word line. Which is driven by address decoder. The sence/ Write ckt are connected data I/O lines. During read operations the ckt reads information the store in the cell and transmit to output lines. During write operations the sence write ckt receive the input information and store of in. the selected cell. The R/W signal satisfies the required operation. The CS signal line is selct a given chip in multi chip memory system.

Static RAM:-

The memory i.e. capable of retaining of current states is long as the power is applied is known as static RAM.

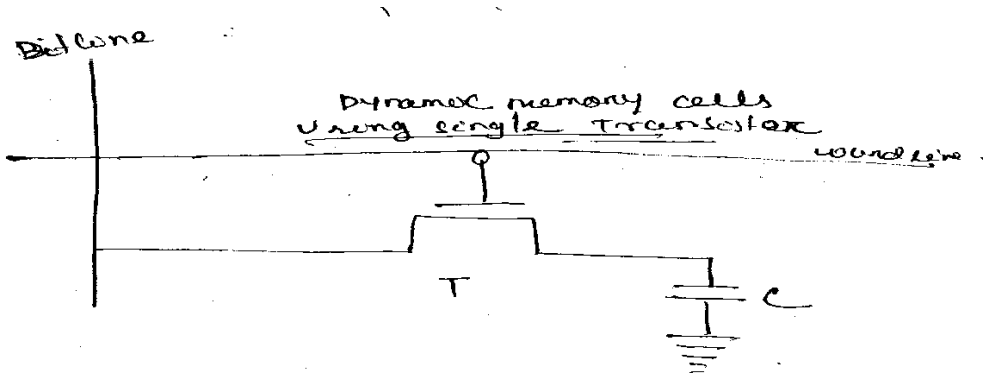
Static RAM



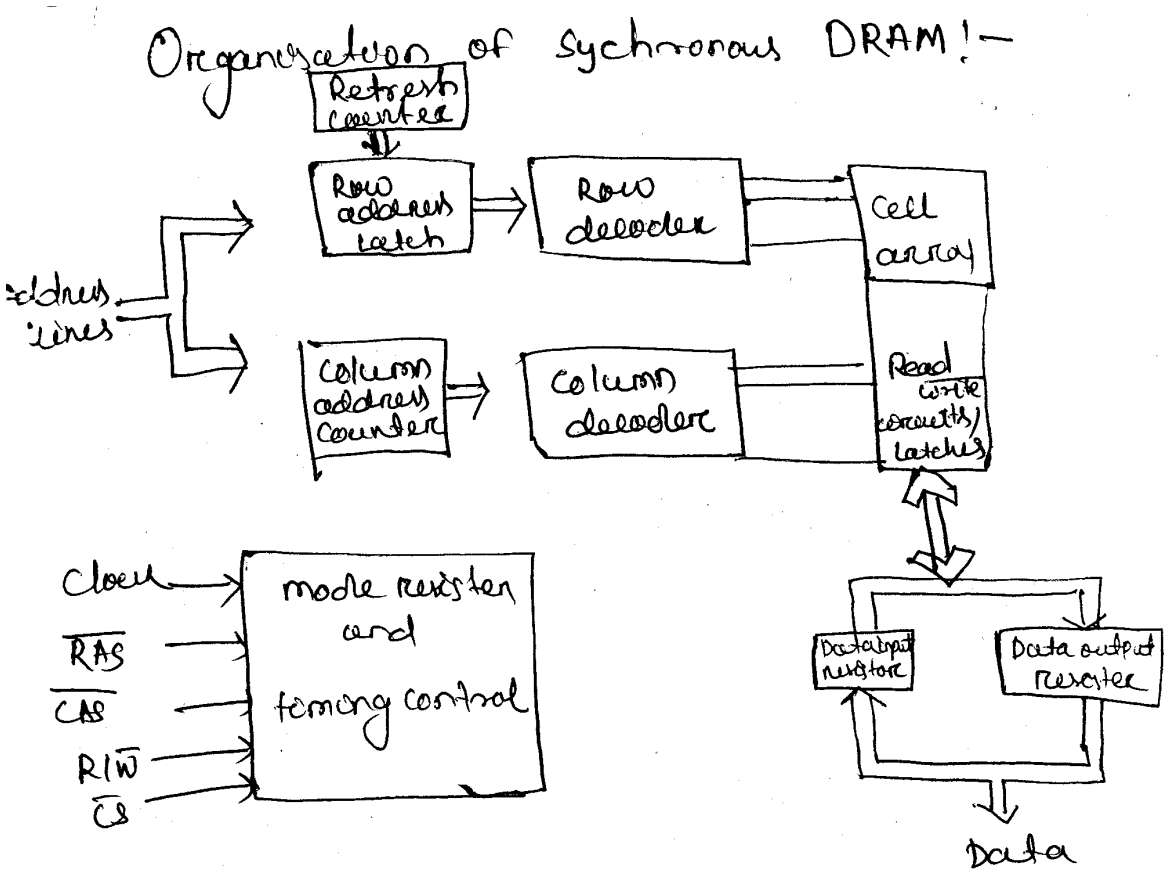
This latch is connected to 2 bit lines by transistors T1 and T2. These transistor can be open or closed. Under the control of word line. When the word line is at ground label, the transistor are turn off and latches returutten their original state. When the cell is in state 1 if the logic value at point x is one and at point y is zero. T0 read SRAM (Static RAM cell) the word is activated to close switches T1 and T2. If the cell is state 1 the signal on bit line is high i.e. b=1, b=0. If the cell is to be state zero, the switches T1 and T2 should be open.

Synchronous dynamic RAM (SDRAM)

SRAM are faster and their cells requires several transistor. Such cells don't written their state indefinitely and hence are called (DRAM) Dynamic RAM. Information is store in a dynamic memory cell that consist of a capacitor and a transistor. As the cell is required to store the information for longer period its contains are periodically refreshed.



Synchronous DRAM



Hence, the memory operation is directly synchronies by a clock. The address and data connection are buffer by means of register and the output is connected a latch.

In road operation the containt of all the cells in a selected row are loaded into these latches. Data held in the latch that corresponds to the selected column are transfer to the data output register.

1) Basic memory operation
   Read write operation
   Static RAM
   Dynamic RAM
2) Describe fix point addition, subtraction and floating point multiplication.
3) Write short notes (any two).
   i) Inter leaved memory.
   ii) Floating point arithmetic operations
   iii) Design of ALU
4) Indentify the advantages of parallel processing, what is a pipeline? Draw a space time diagram to show how an instruction is executed.

5) Identify the techniques/modes which makes data transfer to and from peripherals. Explain DMA method of data transfer by a suitable diagram.
6) Define the function of control unit. What do you means by micro program control? How it is different from hardware control?
7) What is the function of cache memory? What is memory mapping process 1 explain the structure the mapping process on cache memory.
8) Answer any five questions:
    a) Define virtual memory. How the memory table is used for mapping a virtual address?
    b) What are the basic memory operations? Explain how an instruction gets executed in a computer.
    c) Describe the architecture, function and working of I/W channels.
    d) What is the function of bus? Explain bus architecture.
    e) What do you mean by addressing mode? Explain different addressing mode.
    f) Describe Flynn's classification.
    g) Explain interrupt handling tech.

9) Answer all question:
    i. Name two replacement techniques in cache memory.
    ii. Define hit ratio.
    iii. What is the function of an interface?
    iv. What do you mean by interrupt?
    v. Define speed up.
    vi. What do you mean by performance measure?
    vii. Define instruction format.
    viii. What is strobe?
    ix. Define handshaking.
    x. What is register memory?

Virtual memory:-

Virtual memory is the most fundamental memory management concept to be implemented for memory management functions such as space allocation, program relocation, code sharing and protection.

The key idea is to allow or user program more memory locations than those available in a physical memory. This concept work and as follow:-

A virtual address is generated by a user program and the set of virtual address constitutes the virtual address space.

Hence it was necessary to divide the program into small portion called overlay, to fit them into the primary memory. A programmer has to design overlays to make them independent to each other with this kind of provision, one can successively bring each overlay into the main memory and execute them

in a sequence. In the a system using virtual memory there of virtual address. Space is much larger than the physical address space.

A virtual memory system performs a series of mapping operations that mechanizes the process of overlay generation.

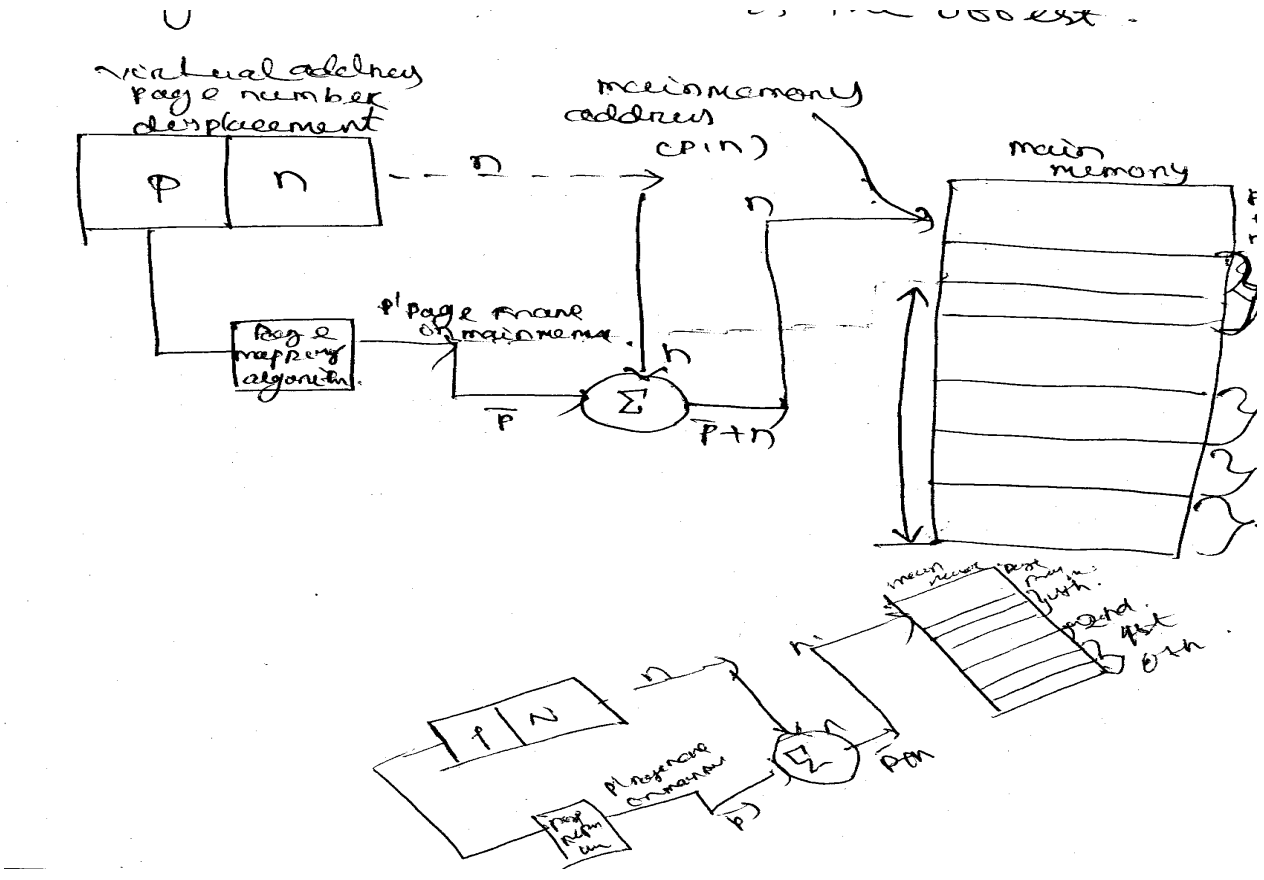
A virtual memory system can be configured as follows:-

1. Memory paging
2. Memory segmentation.

Memory paging:-

In memory paging the virtual memory is divided into blocks of equal size. There are called pages. The physical memory is also divided into frames in the same way. A page has the same size as a frame and may be blocks of 512, 1027 or 2048 words.

In paging system each virtual address can be considered to be an ordered power of <p,n>, where p is the page number and n is the offset.



Mapping scheme for virtual address:-

Virtual address (K)	Page number
0-2	0
2-4	1
4-6	2
6-8	3
8-10	4
10-12	5
12-14	6
14-16	7
16-18	8
18-20	9
20-22	10
22-24	11
24-25	12
26-28	13
28-30	14
30-32	15

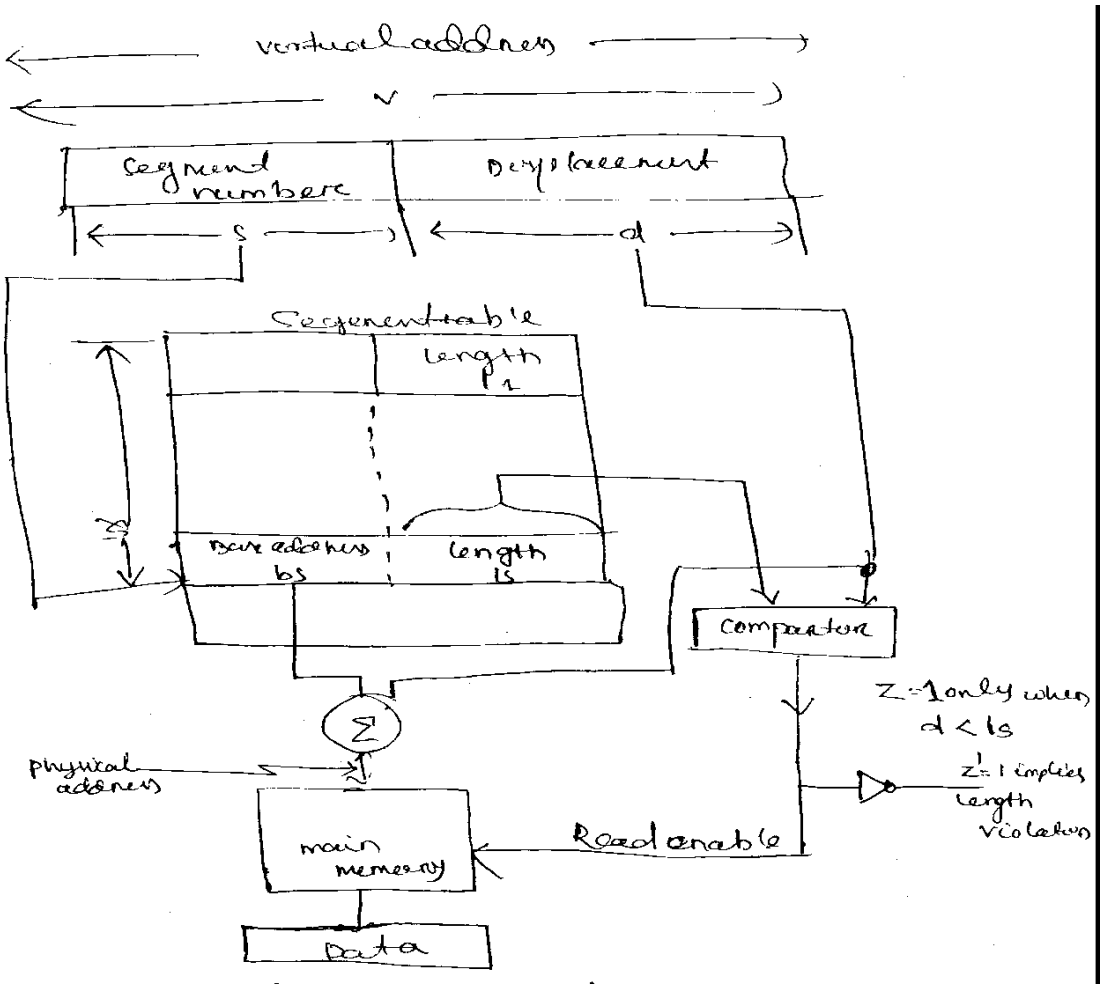
Frame		Physical Address (K)
Binary	Decimal	
06	6	0-2
01	1	2-4
10	2	4-6
11	3	6-8

Memory segmentation:-

The paging concept is viewed as one dimensional technique as virtual addresses generated by a program increase linearly from 0 to same maximum value. In a segmentation system, each logical entity like stack, array subrountine etc. have a separate virtual address space. The virtual address is called a segment and each segment can grow from zero to a maximum value. As each segment refers to a separate virtual address space, it can grow or shrink independently without affecting other segment.

Typically, a segment descriptor consists of the following information.

1. Segment base address b
2. Segment length 1
3. Segment presence bit
4. Protection bits.



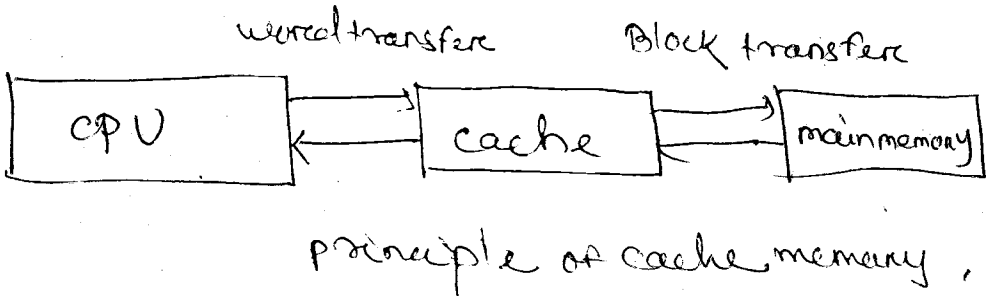
The following common protection protocols are used on a segmentation system:-

1. Read only
2. Execute only
3. Read and execute only
4. Unlimited access
5. No access.

Cache memory:-

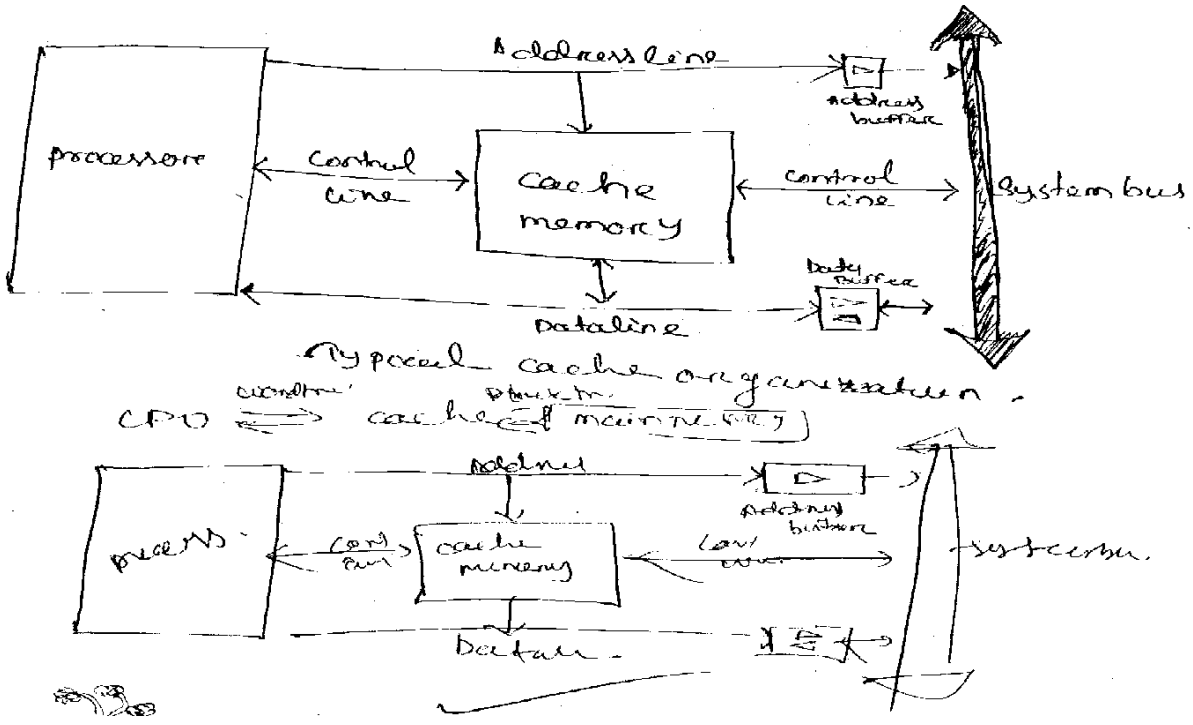
Cache memory provides a memory with the speed approaching that of the fastest available memories and a large memory size at the price of less expensive semiconductor memories.

A smaller and faster cache memory is connected to the large and slow main memory. A copy of portions of the main memory is contained in the cache. When the processor attempts to read a word word of memory, a check is made to find if the word is available in the cache. If available the word is delivered to the processor, other wise a block of main memory, consisting of some fixed number of words, is read in to the cache and then the required word is delivered to the processor.



principle of cache memory.

If the word is found in the cache, it is delivered to the processor. Otherwise, the block containing that word is loaded into the cache from main memory and then the word is delivered to the processor. On one side, the cache connects to the processor via data, control and address buffers are disabled and communication is only between processor and cache, with no system bus traffic. When a cache does not contain the required data, the desired address is loaded on to the system bus and the data are returned through the data buffer to both the cache and the processor. The desired word is first read into the cache and then transferred from cache to processor.



Questions:

1. **What is deadlock?**
2. **What are deadlock prevention?**
3. **What is deadlock avoidance?**
4. **What are the condition of deadlock?**

CHAPTER-5

INPUT–OUTPUT SYSTEM

I/O channel Architecture:-

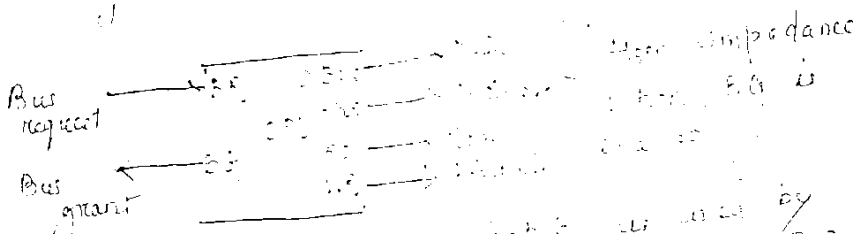
An increasing complexity and sophistication of individual components were brought in as computer system and evolved. This is more evident in the development of I/O function and the evolutionary steps can be summarized as below.

1. As seen in a simple microprocessor controlled device, the CPU directly control is a peripheral device.
2. A controller or an I/O module is added between the CPU and I/O. the CPU uses programmed I/O without interrupts and the CPU becomes some what removed from the specific details of the external device interface.
3. The same configuration in step 2 is improved to use interrupts. The CPU spends no time waiting for an I/O operation, thus improving efficiency.
4. The I/O module is given direct access to memory using DMA. Now, a block of data to or from the memory can be moved without involving the CPU, except at the beginning and end of the transfer.
5. The I/O module becomes of processor of its own right, with a specialized instruction set centre on I/O. the CPU allows the I/O processor to execute an I/O program on memory, as it fetches and executes these instructions without CPU interrupted only when the entire sequence has been completed. Thus type of I/O modules as referred to as an I/O channel.
6. A local memory is added to the I/O module., making it a computer in its own night with this architecture, a large number of I/O devices can be added and controlled with minimal CPU involvement. A common use for such architect are called I/O processor is to control communication with interactive terminates and the I/O processor takes care of most of the taks5.

   As one travels along the evolutionary path, an increasing number of the I/O functions are performed without CPU involvement. The CPU is increasingly relived of I/O related tasks, thus improving performance and efficiency.

Direct memory access (DMA)

⇒ In DMA transfer, the CPU is edle and has no control of the memory busses. The transfer of data takes place directly between the fast storage device and memory controlled by DMA controller which mange the bus.



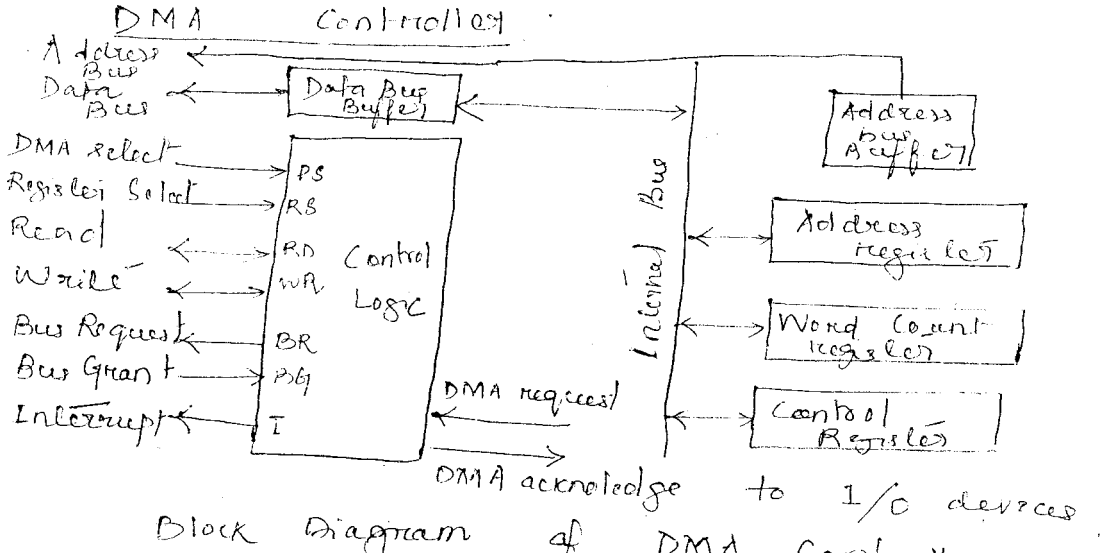
⇒ The Bus request input is used by the DMA controller to request the CPU to relinquish control of the buses.

⇒ The CPU activates the Bus Grant output to inform the external DMA that the buses are now rol under the control of CPU and can be used for memory transfer without the processer intervention.

⇒ Transfer of data can be done in many ways.

⇒ In burst transfer, a block sequence consisting of a number of memory words is transferred in the continuous hugest while the DMA controller is master of the memory.

⇒ In cycle stealing the DMA controlled transfer are data word at a time, after which it returns the control of busses to the CPU. The CPU uses the bus for one memory cycle.



Block Diagram of DMA

⇒ The unit communicates with the CPU via data bus and control lines.

⇒ The register in the DMA are selected by the (in through the address bus by enabling the DS (DMA select) and RS (register select) inputs.

⇒ The RD and WR inputs are bialirection.

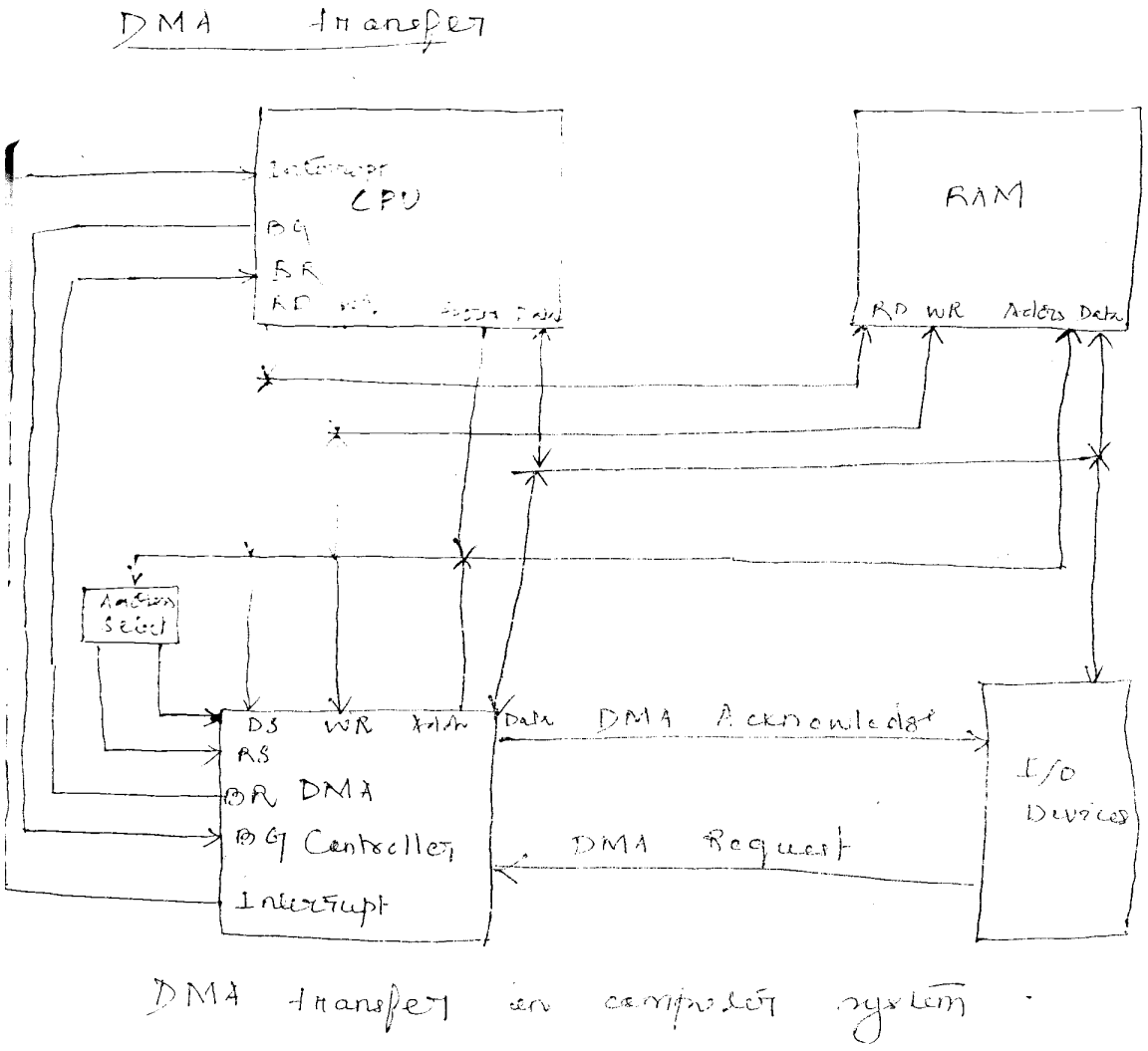
⇒ When the BG=0, the CPU can communicate with the DMA resisters to read framer write to the DMA register.

⇒ When the BG=1, the CPU has relinquished the buses and the DMA can communicate directly with the memory specifying the address in the address bus and activating the RD at WR control.

⇒ The address register control an address to specify the desired location in memory.

⇒ The work count register holds the number of words to be transferred.

⇒ The control register specifies the mode of transfer.



DMA transfer in computer system

⇒ In DMA has its own address which activates the DS and RS lines.

⇒ When the peripheral device send a DMA request, the DMA controller activates the BR ins, in framing the CM to relinquish the buses.

⇒ The CM responds with its B4 low, informing the DMA that its buses are disabled.

⇒ The DMA plots the current value of its address register into its address bus. Initiates the FR or WR signal and sends a DMA acknowledge to the peripheral device.

⇒ When BG=0 the CPU communicates with the internal DMA register.

⇒ When BG=1, the RD and WR ace O/P line from the DMA controller to KAM.

⇒ When the peripheral device receives a DMA acknowledge, it puts a word in it data bus (for write) or receives a word from the data bus (for read).

⇒ For each word that is transferred the DMA increments the address register and decrements its word count register.

⇒ Till the word count is not zero the DMA checks the DMA request line.

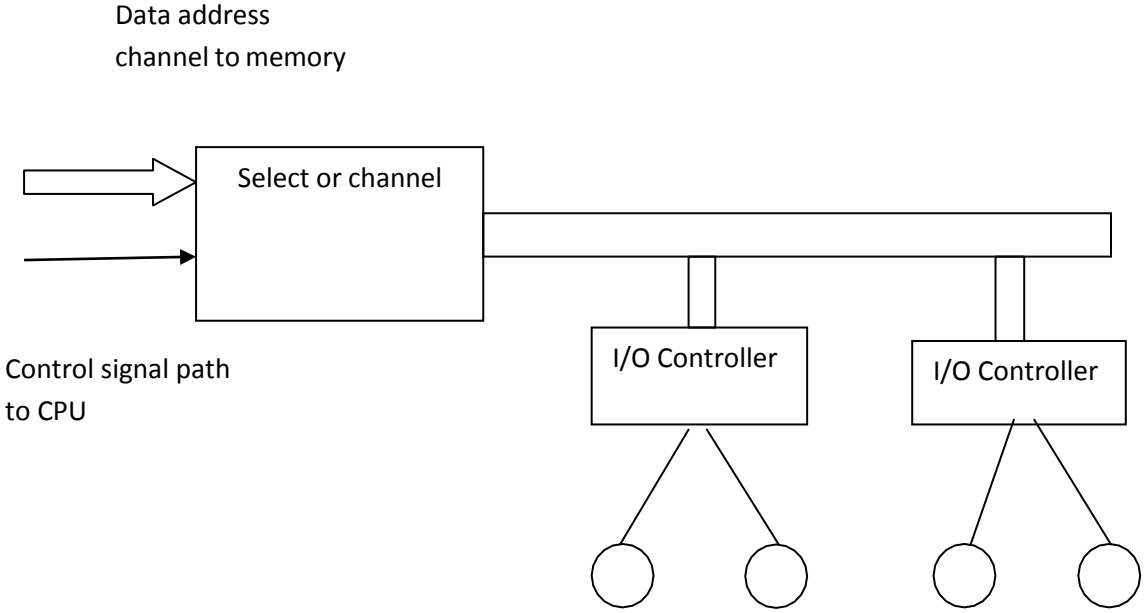
I/O channel architecture:-

The I/O channels can be considered as an extension of DMA module. Here another component used called I/O processor. The I/O channels has the ability to execute I/O instruction because it gets complete control over I/O operation. In such a computer system with I/o channels, the I/o instruction are not executed by the CPU but are stored in main memory to be executed by a special purpose I/o processor. Thus the CPU initiate an I/o transfer by instructing the I/o channels to execute a program in memory. The I/o processor execute an I/o program in memory without CPU intervention. The CPU is interrupted only when the entire sequence has been completed.

Two major type of I/o channels are commonly used namely.

a) Selector
b) Multiplexor

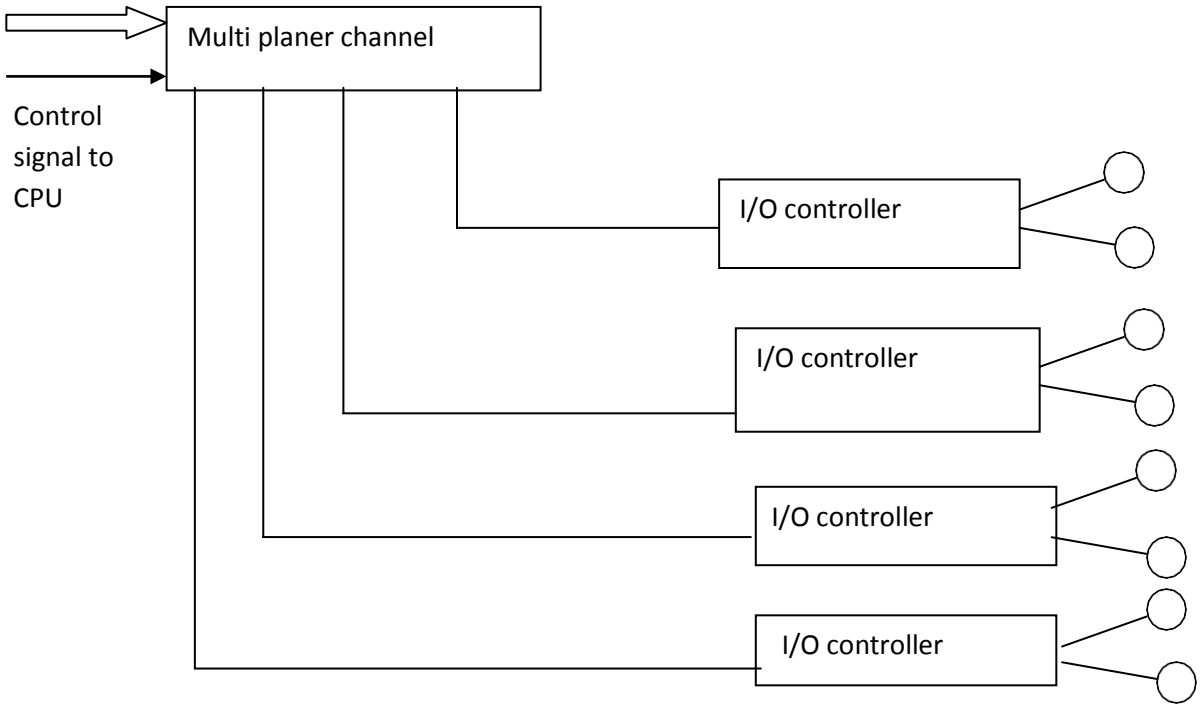
a) Selector:-



A selector channel controls multiple high speed I/O device and is dedicated to the transfer of data with one of those of any point of time.

b) Multiplexor:-

Data address Channel
to main Memory



A multiplexer channel can handle multiple I/O device at the same time.

Questions:

1. **Difference between compiler, interpreter.**
2. **Different phases of compiler.**
3. **Describe disk scheduling algorithm.**

CHAPTER-6

I/O INTERFACE & BUS ARCHITECTURE

BUS INTERCONNECTION:-

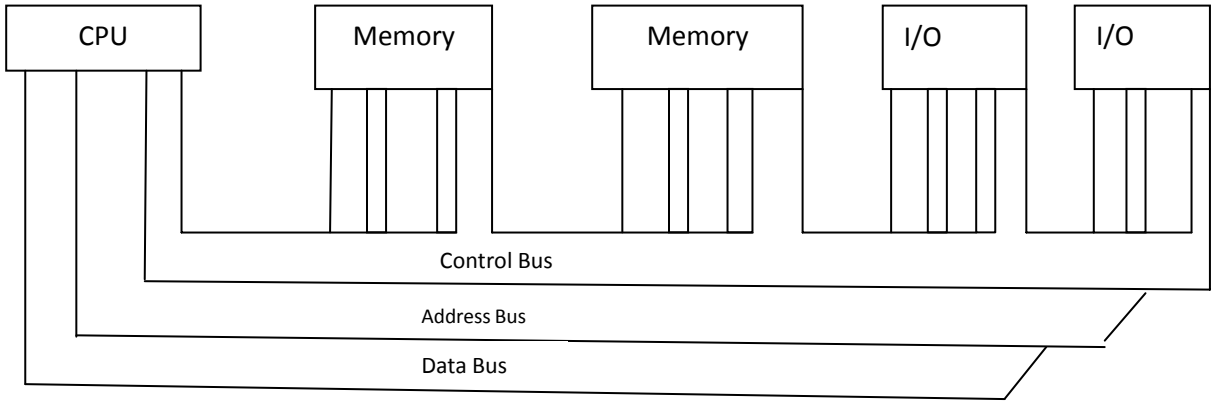
Bus consists of a number of communication lines and each lines can carry a signal representing a binary '0' or '1'. Bus is an inter connecting pathway between two or more device. It is a shared communication medium connected to multiple device.

Bus structure:-

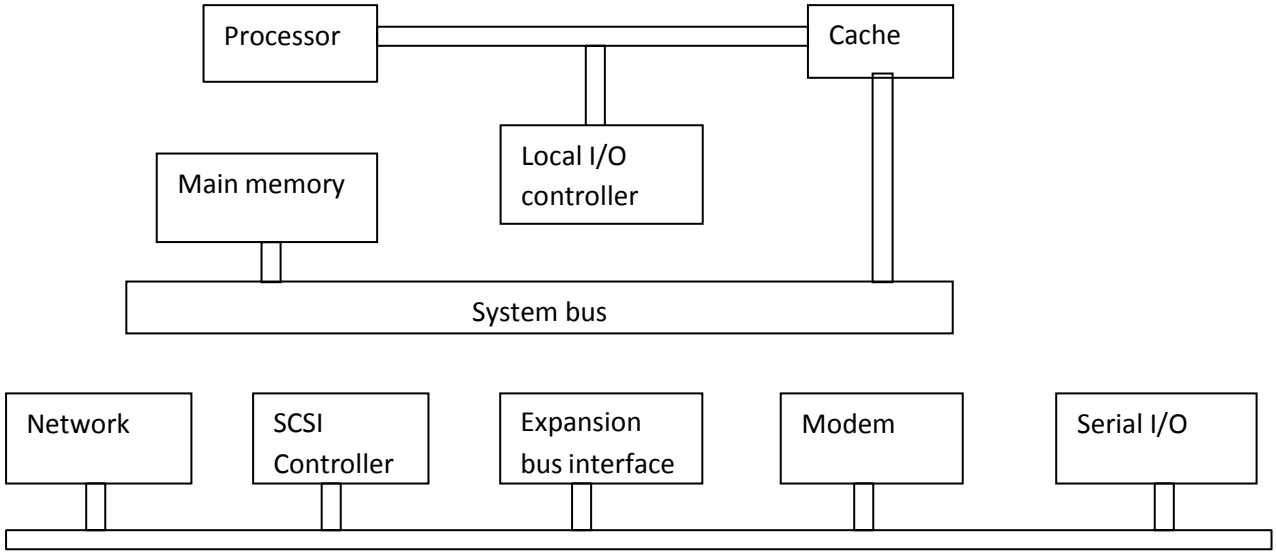
System bus consists of 50 to 100 communication line and each line assigned a function. The system bus can be classified into three basic groups i.e. data bus, address bus, control bus.

1. **Data Bus:-**
   The data bus provide a path for data transfer between system module. The data bus will consist of 16,32 or 64 lines referred as width of data bus. The width of data bus is equal to the memory word length.

2. **Address bus:-**
   The address lines are used to locate the destination of data on the data bus. The width of the address bus determines. The maximum memory capacity of the system. If 16 bit address bus means memory capacity is 2^{16} location.



3. **Control bus:-**
   Control Bus is used to transmit command and timing information. The command indicates the operation to be performed and timing signal indicates the duration of validity of data and address information.



Basic parameter of Bus design:-

The following are the Basic parameters to be considered while designing a bus.

i) Type of Bus:-
   - Dedicated
   - Multiplexed

ii) Method of timing:-
   - ⊗ Asynchronies
   - ⊗ Synchronous

iii) Method of Arbitration:-
   - ⊗ Centralized
   - ⊗ Distributed

iv) Width of bus:-
   - ⊗ Data
   - ⊗ Address
   - ⊗ Control

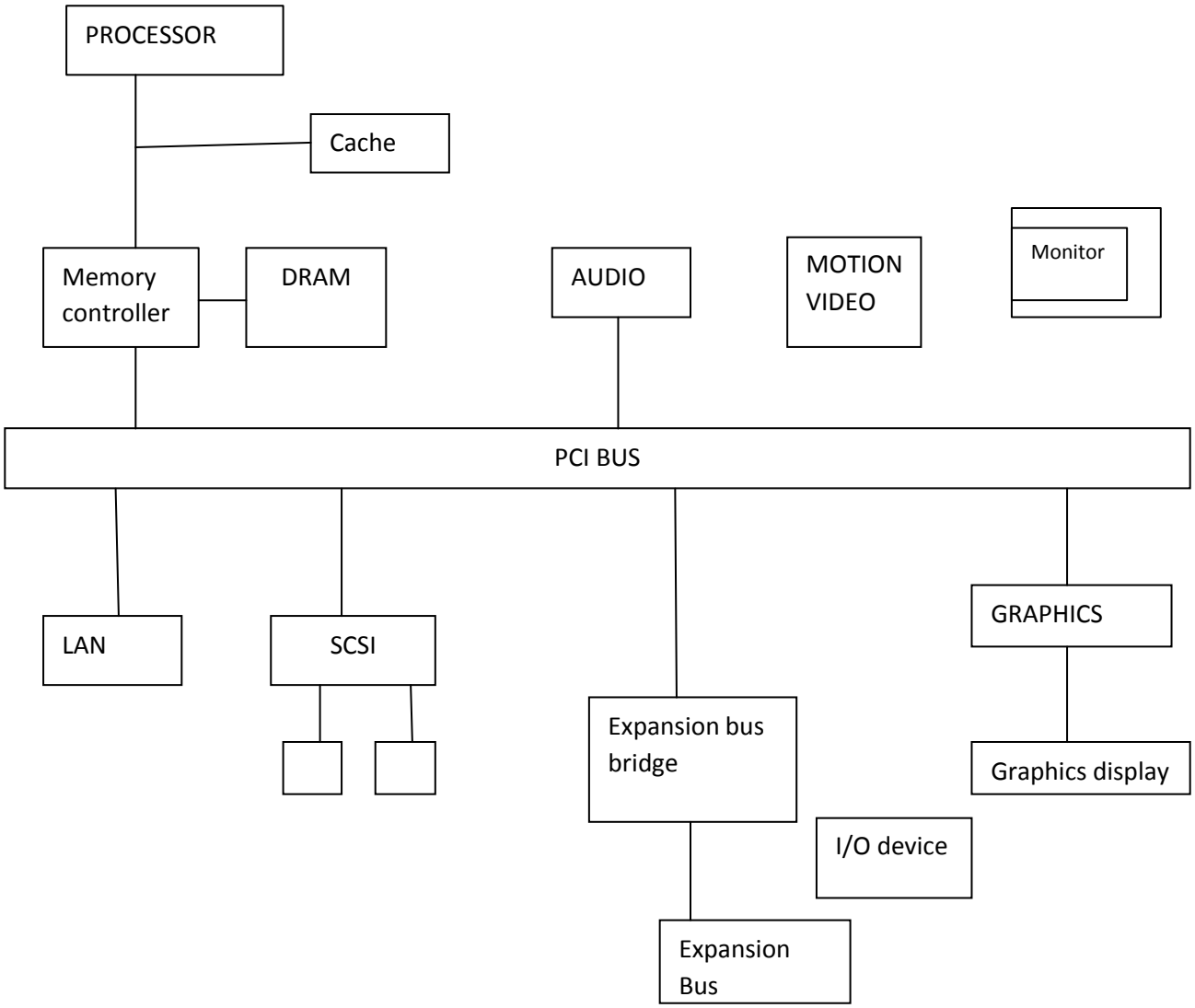
v) Type of data transfer:-
   - ⊗ Read
   - ⊗ Read-modify-write
   - ⊗ Read-after-write
   - ⊗ Block transfer.

Peripheral component inter connect BUS (PCI)

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

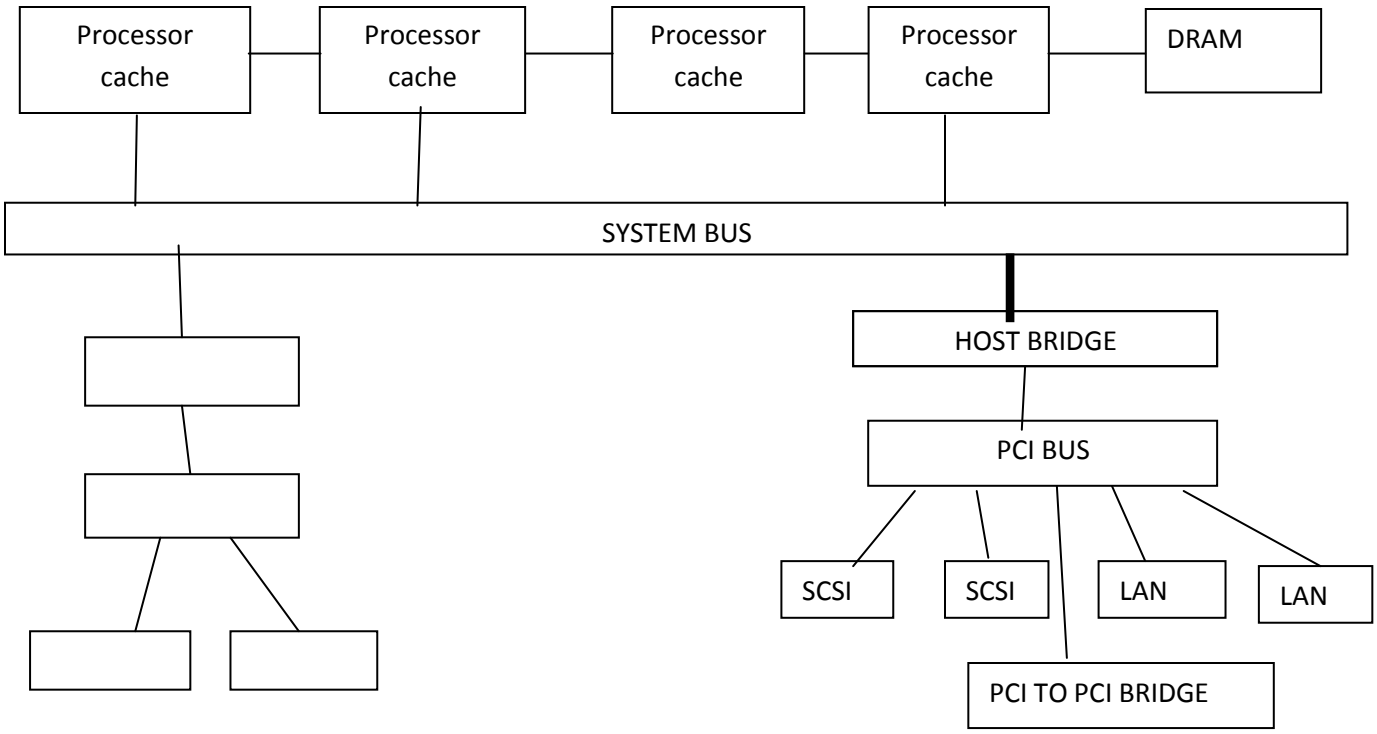
PCI bus use upto 64 data lines at 66 MHZ speed. The data transfer rate is 528 MBPS to 4.224 GBPS also economical for most of the modern I/O system requirements. It requires only a few chips to implement and support other buses.

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



PCI bus makes use of synchronous timing and a centralized arbitration scheme.

In a multiprocessor system as shown below more than one processor with multiple PCI configuration is connected by bridges to the processors system bus. The system bus supports only processor/ can be main memory and PLI bridge.



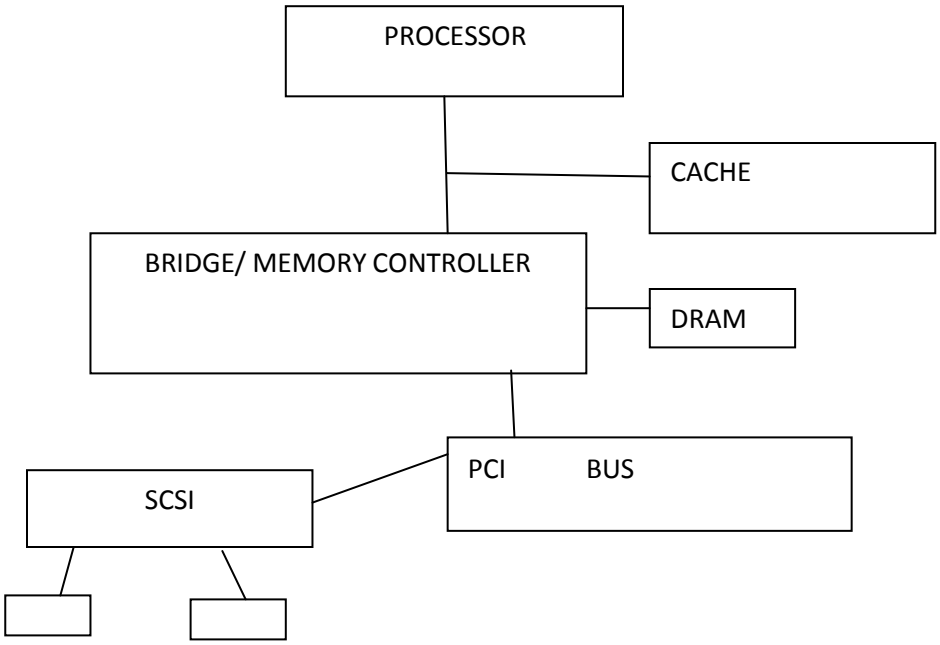
Small computer system interface (SCSI):-

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.

The operation of SCSI bus is explained using a disk drive. Communication with a disk drive differs from main memory. Data are stored on disk sectors and not in contiguous sector. Hence for a read or write operation, need to access several disk sector which may not be continuous. Due to the constraints of the mechanical motion of the disk. There is a delay of several millisecond before the 1^{st} sector from which data are to be transferred is reached. A block of data is transferred at high speed and after some delay another set of data my follow.

$\Rightarrow$ The SCSI connector may have 50,68 or 80 pins and maximum transfer rate 5 to 16 Mbps.



Universal serial Bus (USB):-

A modern computer system uses variety of devices such as key board micro phone, camera, speaker and display device.

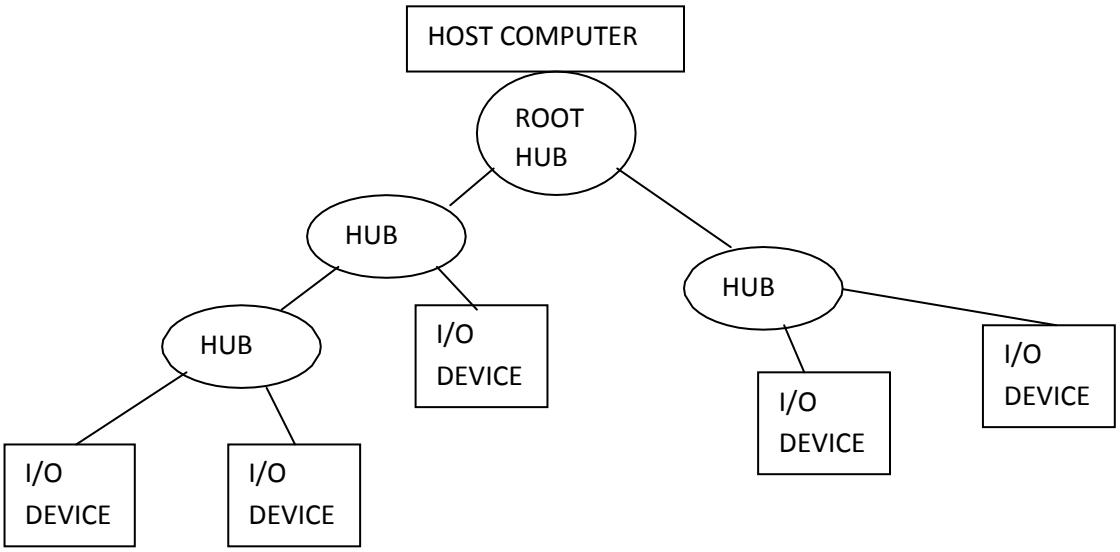
⇒ The simple, low-cost mechanism to connect these device to computer is the universal Serial Bus(USB)

⇒ USB has been designed to meet several objectives mainly are

- Plug & play mode of operation
- To connect variety of device including telephone and internet.
- Simply and low cost and easy to use
- To add many device to computer at any time.

⇒ The USB supports two speed operations one called low speed at 1.5 Mbps and the other full speed at 12 mbps. A recent version introduced a third speed operation, called high speed at 480 Mbps.

USB architecture:-



A large no of devices can be added or removed at any time, the USB uses the tree structure as shown in the fig.

⇒ Each hub has a no. of ports where device may be connected of which one device can be HUB. A message sent by the host computer is board cost to all the I/O devices and only the addressed device respond to that message.

⇒ The message sent from an I/O devices sent only Up stream towards the root of the tree and is not seen by other devices. Hence the USB enable the host to communicate with the I/O device but it doesn't enable these device to communicate with each other.

Questions:

1. **Describe a Bus structure.**
2. **What is USB and explain?**

CHAPTER-7

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.

Register transfer language:-

A micro operation is an elementary operation performed on the information store in one or more register. The result of the operation may replace the previous binary information of a register or may be transferred to another register.

e.g. shift, count, clear, load.

The internal h/w/ organization of a digital computer is best defined by specifying.

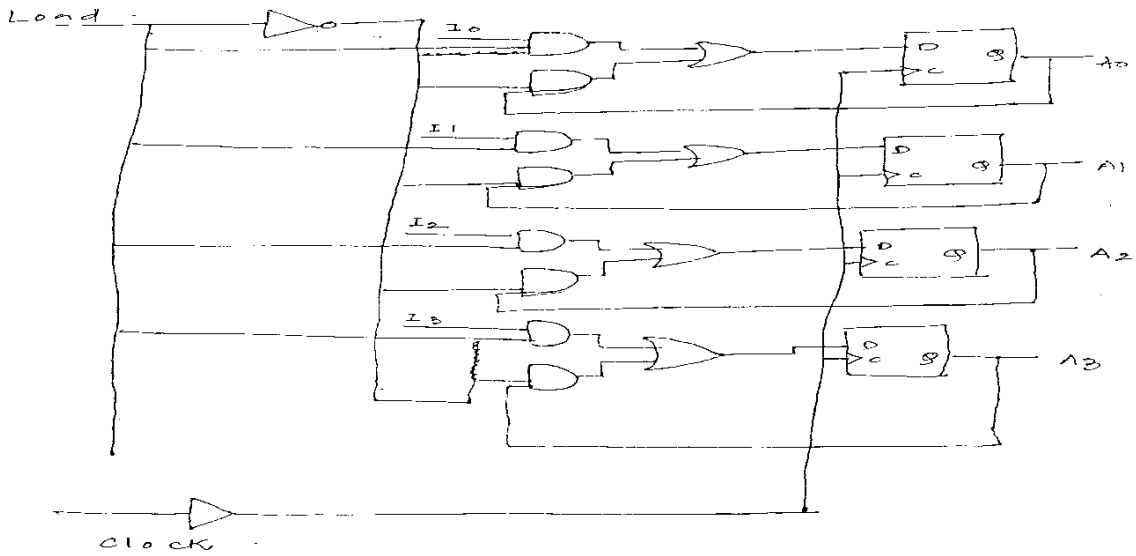
1) The set of register it contains and their function.
2) The sequence of micro operation performed in the binary information stored in the register.
3) The control that initiated the sequence of micro operation.

The symbolic notation used to describe the micro operation transfer among registers is called a register transfer language. The term register transfer implies the availability of h/w logic ckts that can perform a stated micro operation and transfer the result of the operation to the same or another register. The term language is borrowed from programming language.

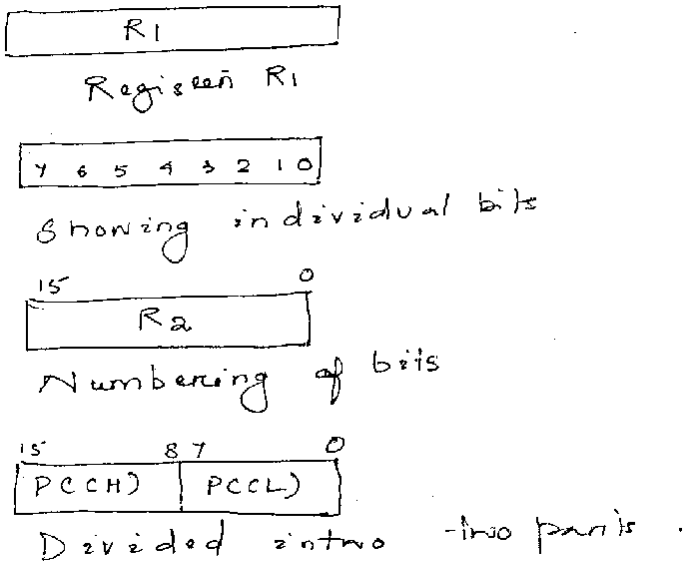
Register transfer:-

A resister consists of a group of flip flops that hold binary information and the gates that control when and how new information is transferred to the flip flop.

A 4 bit register with parallel load



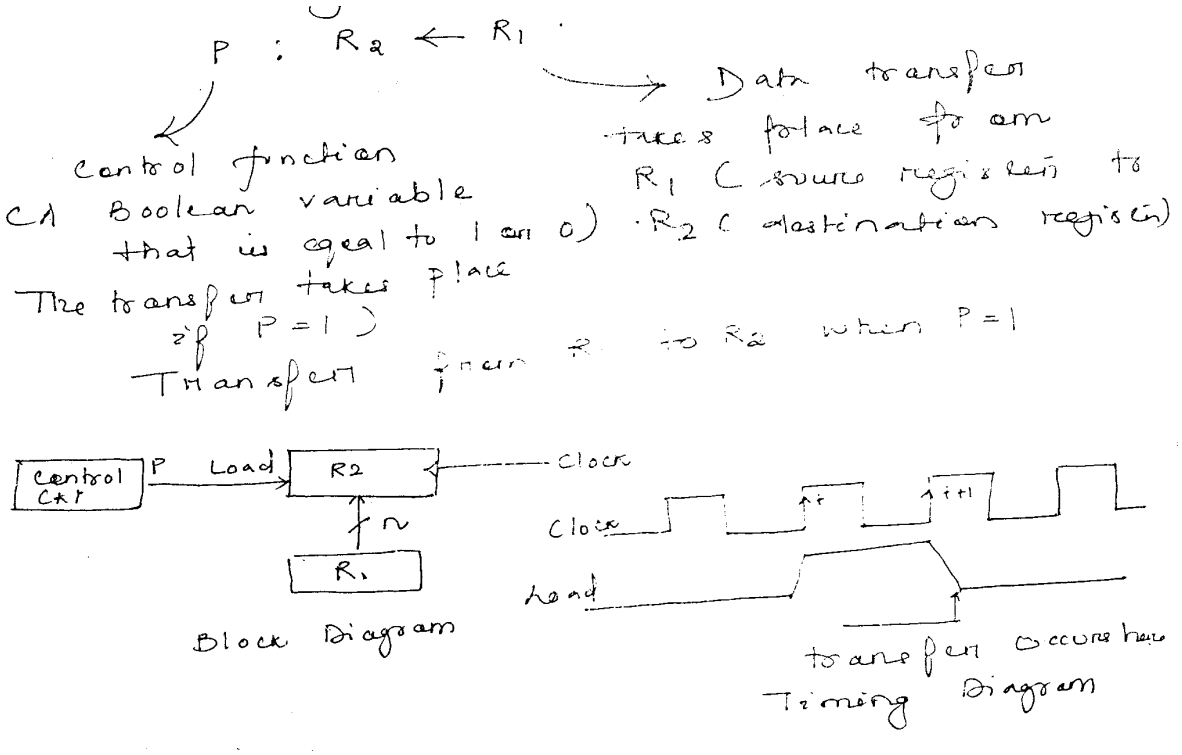
Representation of resisters an Block Diagram



Resister transfer

A register transfer is the transfer of data from one register to another or to the same.

A register transfer statement .



Flynn's Classification:-

Flynn's classic taxonomy is based on the number of control units and the number of processes available in a typical computer. Accordingly, the basic computers can be classified as follows:-

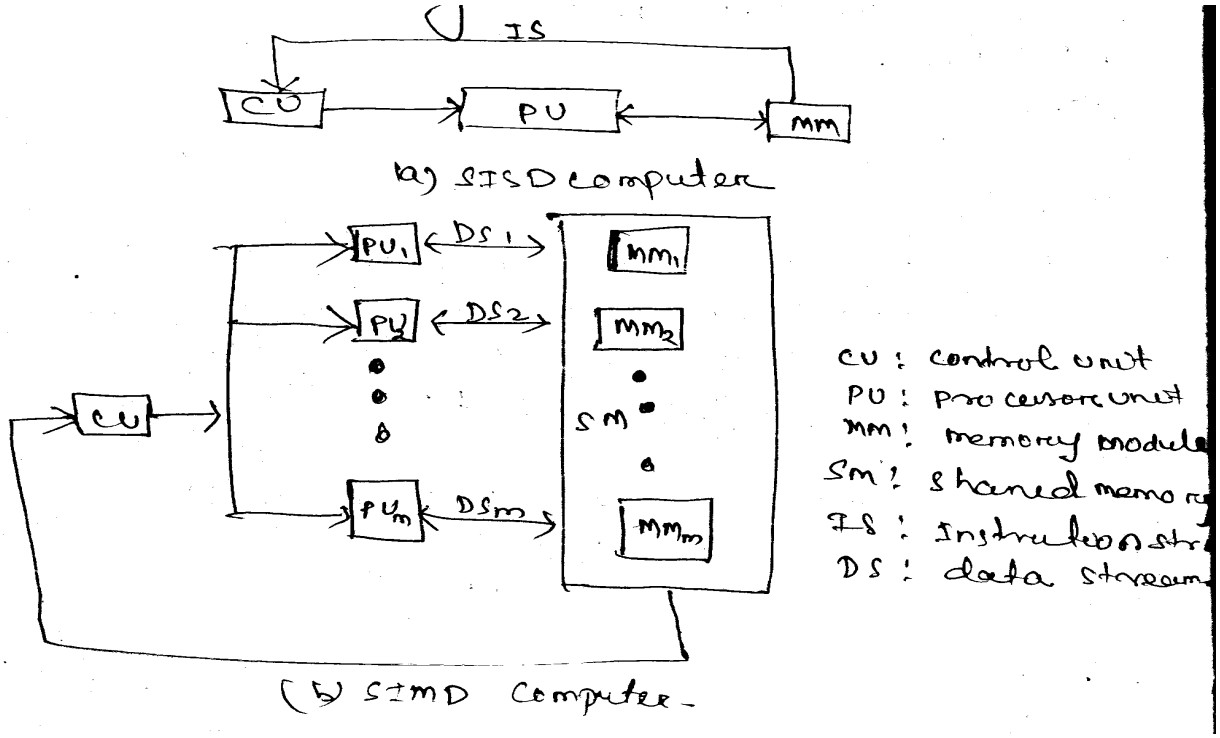
Single instruction:- Single data Stream (SISD) – This architecture has a single control unit producing a single stream of instructions. These instructions are executed sequentially but may be overlapped in their execution pipeline stages. An SISD computer may have more than one functional unit and all the functional units are under the supervision of one control unit.

Single instruction – Multiple Data Stream (SIMD):- in this implementation, a single machine instruction controls the simultaneously execution of a number of processing element and each processing element has its won associated data memory. This facilities each instruction to be executed on different sets of data by different processors.

Multiple instruction:- Single data stream (MISD):- A single processor is available which executes a single data stream of data using multiple instruction. This structure has never been implemented.

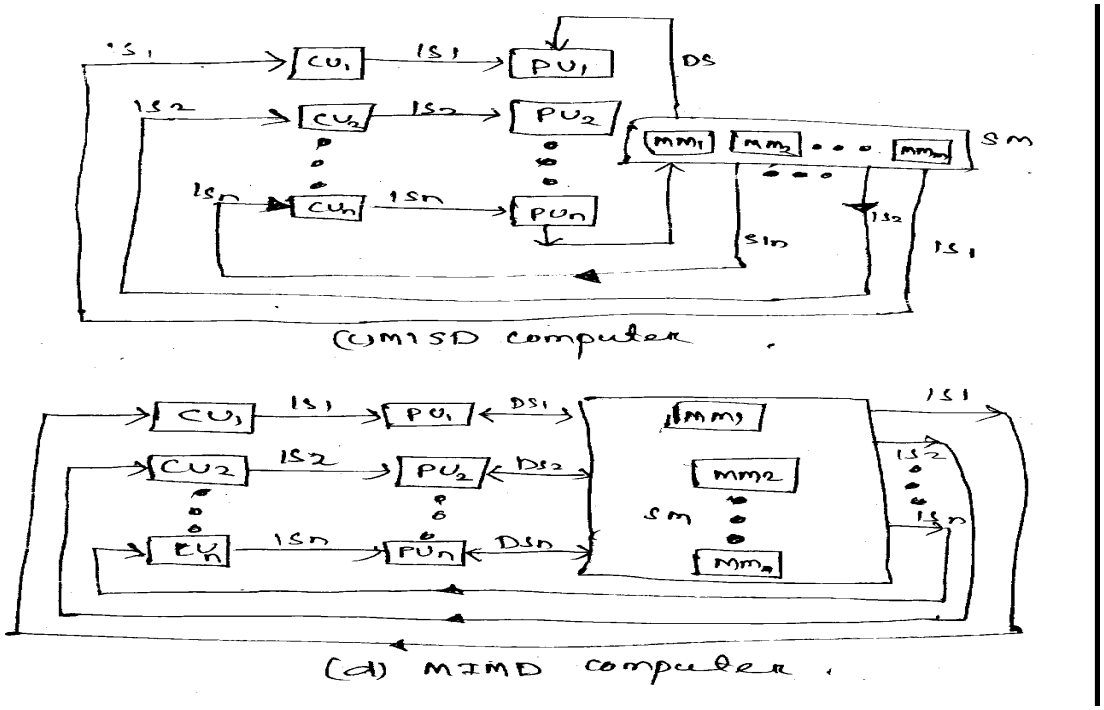
Multiple instruction:- Multiple data stream (MIMD):- A set of multiple processor is available which executing a different stream of data using different set of instruction sequences. Symmetric multiprocessor (SMP) and Non-Uniform Memory Access System (NUMA) are such kind of implementations.

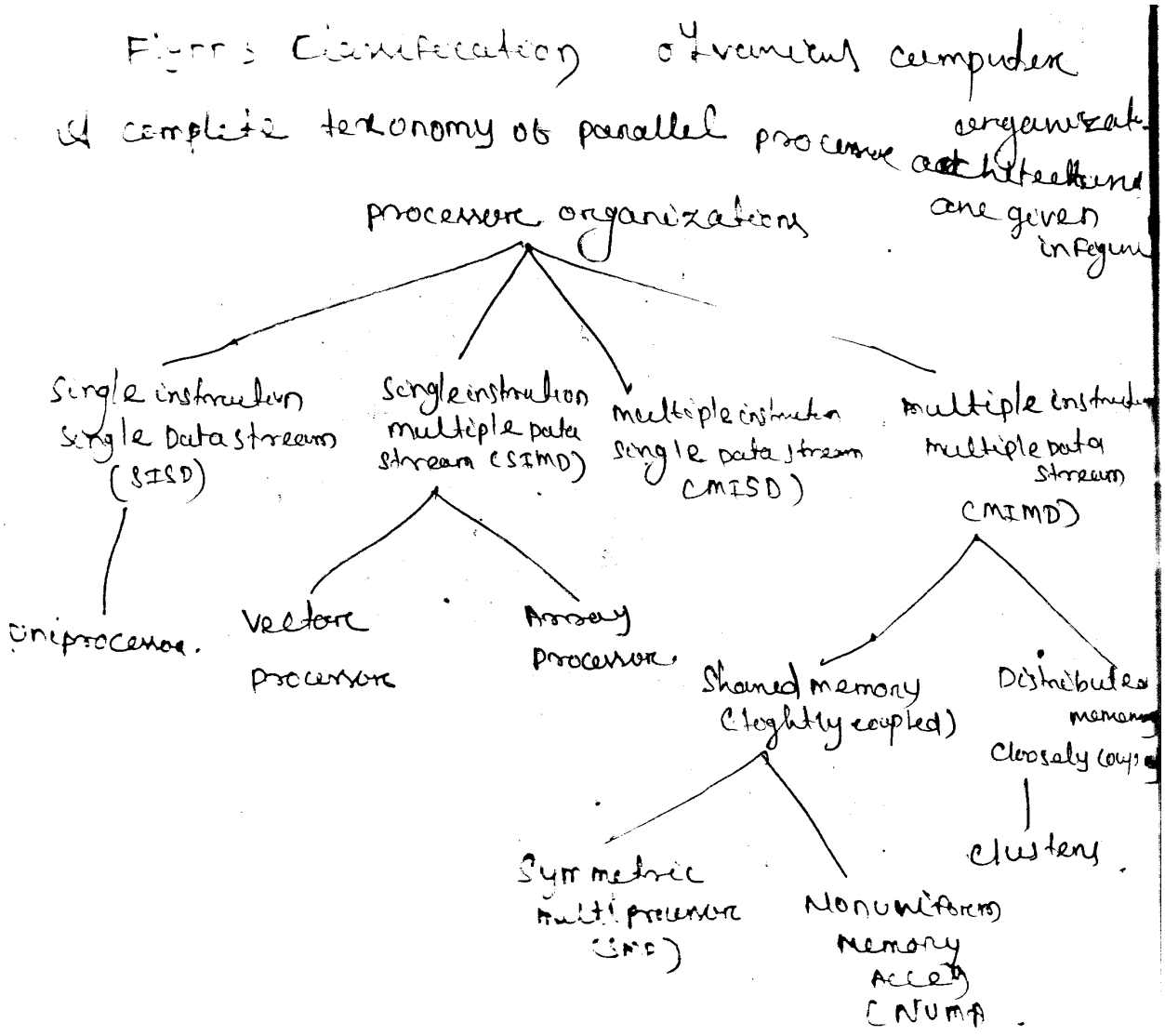
A comprehensive implementation of all the above types of organizations are shown in figure:-



Flynn's classification of various computer

A complete taxonomy of parallel processor arrangement are given in figure





In the parallel architecture taxonomy, the MIMD model has clearly emerged as the chore architecture on recent years. It offers the flexibility of operating as a single user machine providing high performance for one application and as a multi programmed machine running many task simultaneously. It can be built on the off the shelf microprocessor available in the market with a better cost/performance advantage.

The existing MIMD machines fall into two classes, viz centralized (Sym-metric) shared memory arithmetic tunes. (tightly coupled) and distributed shared memory architecture (loosely coupled).

Questions:

1. **What is parallel processing?**
2. **Explain Flynn's classification.**