

Computer Architecture

Instruction format

Addressing mode

Data transfer & manipulation

Status bit condition

Conditional branch Instruction

Subroutine call & return

Program interrupt

RISC - Reduced instruction set Computer

CISC - Complex instruction " "

Differences b/w them or advantage & Disadvantages

Data Path

① One Bus

② Two Bus

Memory organization

RAM & ROM chips

Memory address Map

Memory Connection to CPU

Auxiliary Memory

Associative Memory

Cache memory

Associative mapping

Dired mapping

Set - Associative mapping

Virtual memory

Address mapping using pages

Page replacement

Memory management H/w

Segmented page mapping

~~PLP~~ short note otherwise

~~PLP~~ Input-output Interface

~~PLP~~ Input-output Bus and Interface Modules

~~PLP~~ Input-output versus Memory bus

Isolated v/s Memory mapped I/O

Asynchronous Data transfer

≡ —

Aynchronous Serial Transfer

Mode of transfer

Interrupt Initiated I/O

~~vector interrupt & non vector interrupt diff~~

~~Input / output~~

Priority Interrupt

DMA Direct memory access

Bus Arbitration

Direct Controller
memory access

Direct memory access Transfer

~~DMA all diagram important~~

Input output processor



Instruction formats & the physical and logical structure of computer is normally described in some manuals provided by the computer system. and these manuals are internal construction of C.P.U, these manuals are list of all hardware-implemented instruction. specify their binary code format, which provide definition of each instruction code and provide the necessary control function.

→ Common field of instruction formats :-

- ① operation code field that specify operation to be performed .
- ② An Address field that designates a memory address .
- ③ A mode field that specify which effective address is determined .

① Operation code :- example of some operation are ADD, multiply, subtract, increment

the instruction format in this type of computer uses one address field.

(ADD X) For example this instruction specify an arithmetic addition.

if store
the addition
of

Result $AC \leftarrow AC + M[X]$

↑ previous data ↑ given current data.

Operation code "ADD" pop two number in the stack and added them and push into the stack.

~~register organization~~ (*)
 evaluate this expression $x = (A+B) * (C+D)$
 using Address instruction,

\Rightarrow By three Address instruction / by three register

$$R_1 \leftarrow M[A] + M[B]$$

$$R_2 \leftarrow M[C] + M[D]$$

$$M[X] \leftarrow R_1 * R_2$$

\Rightarrow By two address instruction

$$R_1 \leftarrow M[A]$$

$$R_1 \leftarrow R_1 + M[B]$$

$$R_2 \leftarrow M[C]$$

$$R_2 \leftarrow R_2 + M[D]$$

$$R_1 \leftarrow R_1 * R_2$$

$$M[X] \leftarrow R_1$$

\Rightarrow By one address instruction

$$AC \leftarrow M[E]$$

$$AC \leftarrow AC + M[B]$$

$$M[C] \leftarrow AC$$

$$AC \leftarrow M[C]$$

temporarily
register

$$AC \leftarrow AC + M[D]$$

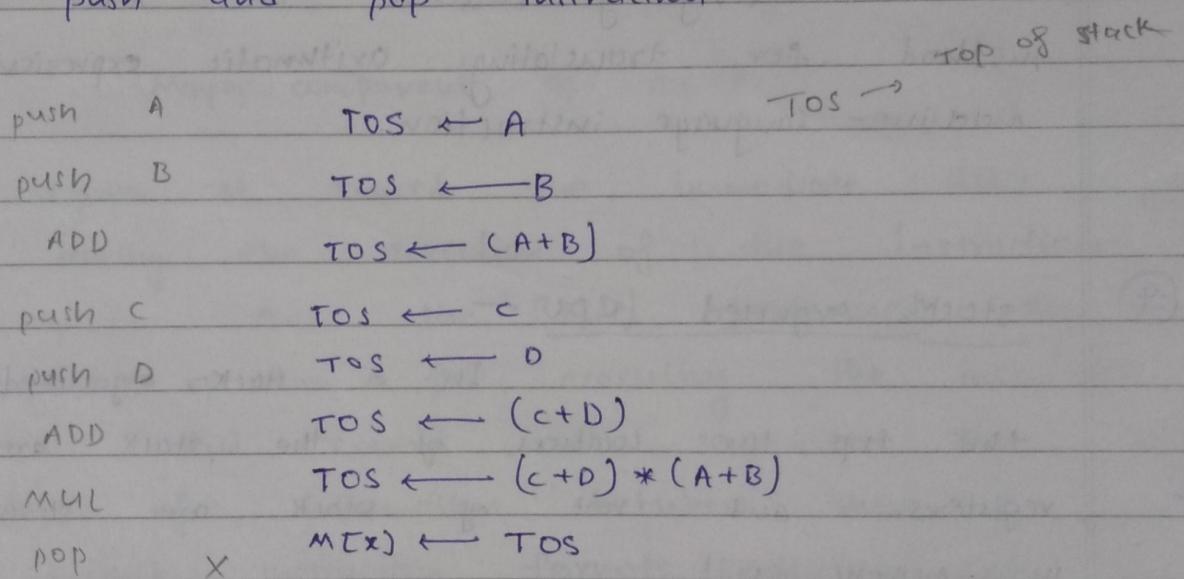
$$AC \leftarrow AC * M[T]$$

$$M[X] \leftarrow AC$$

⇒ zero - Address Instruction / stack organized computer

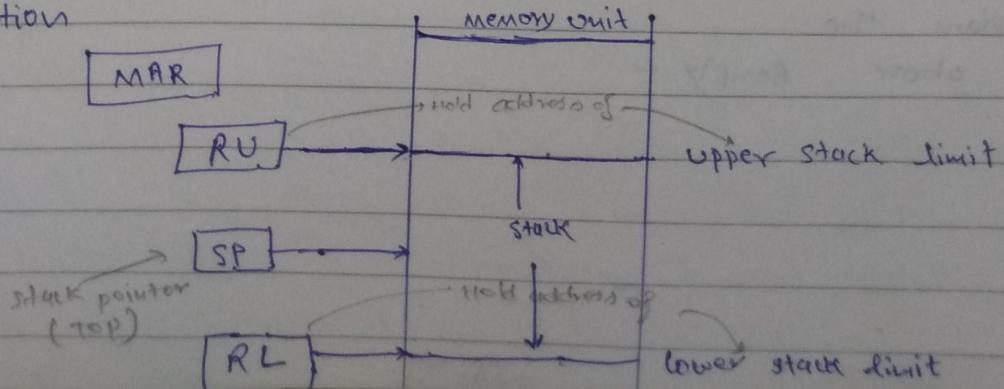
it doesn't use an address field

for the instruction ADD and MUL. it
use push and pop instruction



Stack organisation :-

the implementation of a stack
in the CPU is done by assigning a
portion of the computer memory to a stack
operation



polish Notation:-

$A + B$ → infix

$+ A B$ → prefix / polish notation

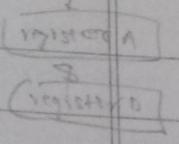
$A B +$ → postfix / reverse polish notation

Any expression should convert into polish notation any way because the most efficient method for translating arithmetic expression into machine-language instruction.



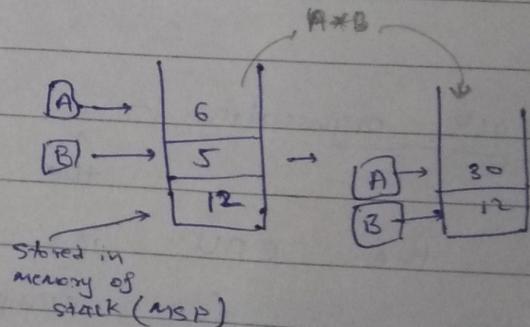
stack organized CPU :-

In a stack-organized CPU the top two location of the stack are processor registers, and rest of stack are stored in memory unit.



like

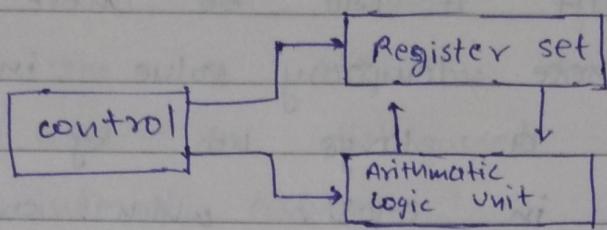
$$4 * 5 + 12$$



explain the

above briefly :-

CPU organisation:- the part of the computer that performs the bulk of data processing operation called central processing unit



Major component of CPU.

Here register set stored the immediate data used during the execution of the instruction.

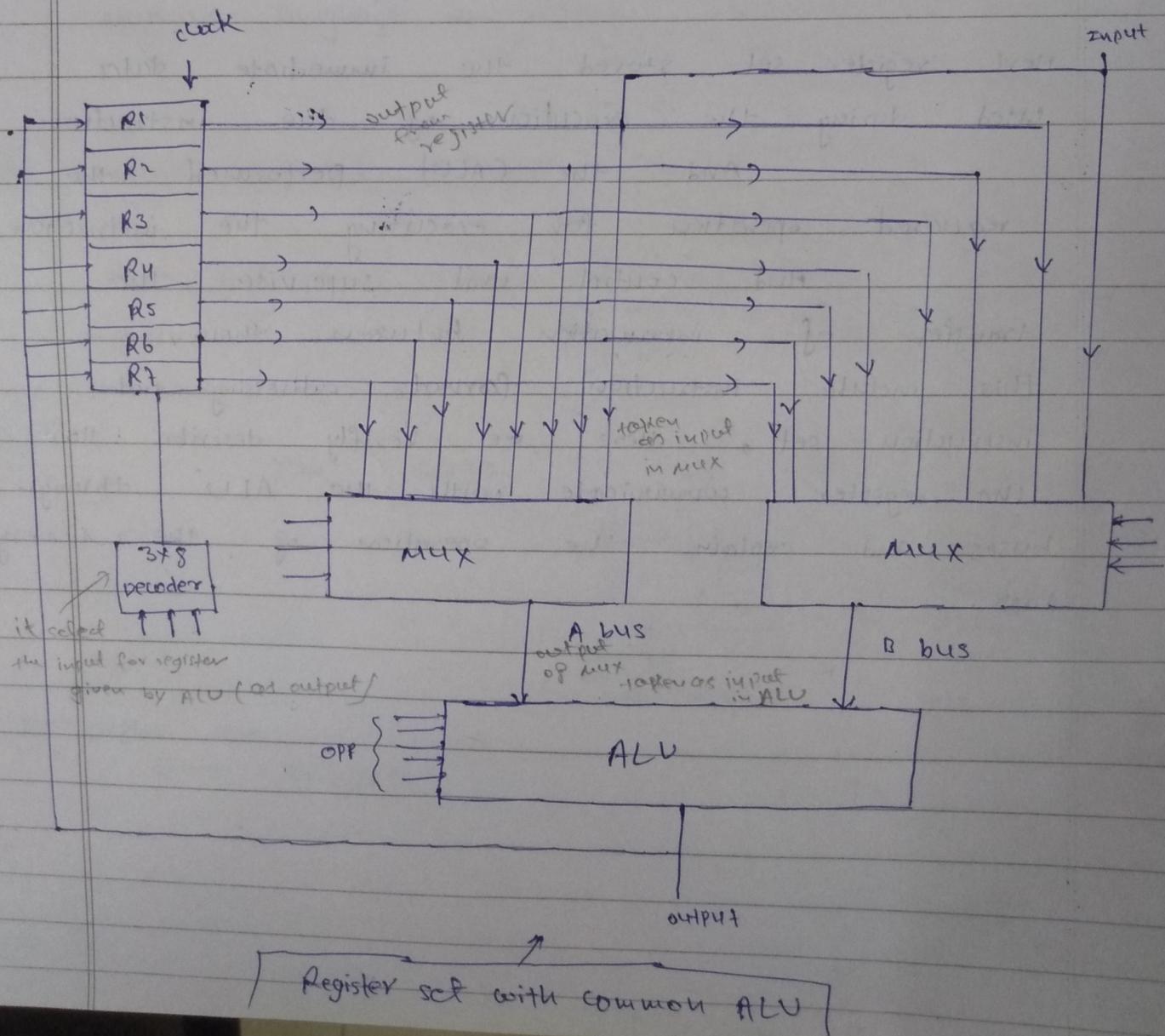
And the (ALU) performed the required operation for executing the instruction.

And control unit supervise the transfer of information between them.

This include instruction formats, addressing mode, instruction set, Here we briefly describe how the register communicate with the ALU through buses and explain the operation of the memory task.

General Register Organization :— generally we have studied that memory location are needed for storing pointer, counter, return address, temporary result.

But it will be more relevant / efficient if we store these temporary value in processor register when a large no. of register are included in C.P.U. which are the seven register used in common bus system.



Addressing mode :- the Addressing mode specifies a rule for interpreting the address field of the instruction before the operand is actually referenced.

An example of an address mode field is the I bit in the basic computer, which is used to distinguish between a direct and an indirect address.

Opcode Address

Reduced Instruction set Computer (RISC) :-

It is a instruction set which determined the way that machine language programs are constructed. Many computer have instruction set that include more than 100 times these instruction are used to translate higher level language to machine language.

A computer with a large number of instruction is classified as a complex instruction set computer (CISC).

RISC characteristics:-

RISC is a concept to reduce execution time by simplifying the instruction set of the computer.

- ① Relatively few instruction
- ② Relatively few addressing modes.
- ③ memory access limited to load and store instruction.
- ④ all instruction done within the register's of CPU
- ⑤ single - cycle instruction. execution.

CISC

- ① CISC architecture gives more importance to hardware

- ② complex instruction

- ③ It access memory directly

- ④ coding in CISC processor is simple

- ⑤ As it consists of complex instruction, it takes multiple cycles to execute

RISC

- ① RISC architecture gives more importance to software

- ② Reduced instruction

- ③ It required register.

- ④ Coding in RISC processor requires more number of lines.

- ⑤ It consist of simple instruction that take single cycle to execute.

6

Complexity lies in micro-program.

7

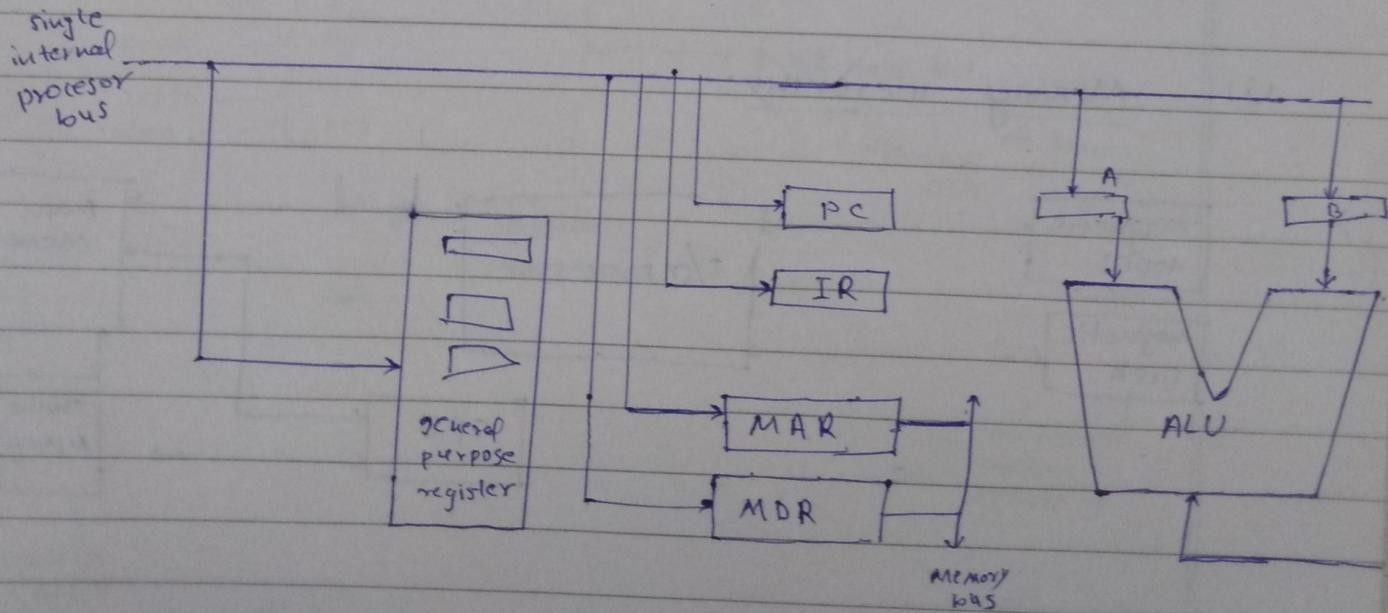
Complexity lies in compiler.

+

Data path:-

A data path is a collection of functional unit such as Arithmetic logic unit (ALU) or multipliers that perform data processing operations, registers and buses. Along with the control unit

simply it is a ALU, the set of registers, and the CPU's internal buses that allow data to flow between them.



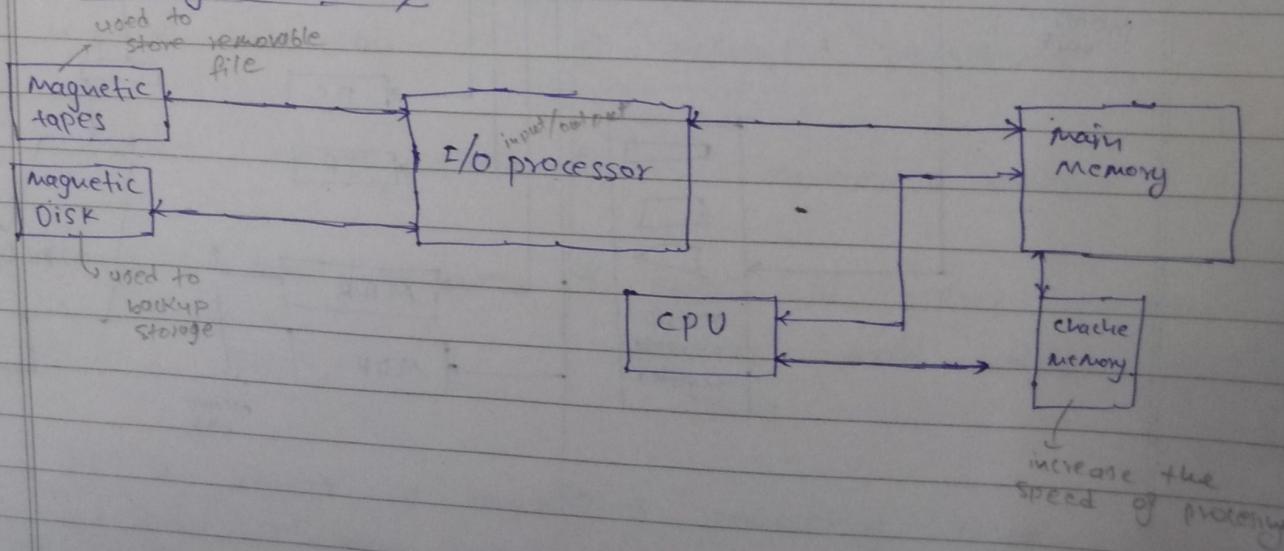
(+) one bus:- CPU register ALU used a single bus to move outgoing and incoming data used by one clock cycle.

(+) two bus:- faster solution general purpose register connected both buses and data can be transfer from two different register

Similarly three bus

Memory organisation:-

(i) Memory Hierarchy:-



Main Memory:- it is the central storage unit in a computer system it is a relatively large and fast memory used to store program and data during the computer operation.

RAM

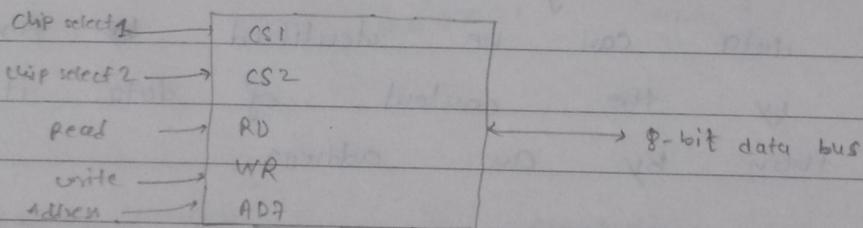
SRAM
→ store binary information
→ flip-flop (uses)
volatile

DRAM
→ store binary information
→ used electric charge

ROM

non volatile
secondary storage

Typical RAM chip



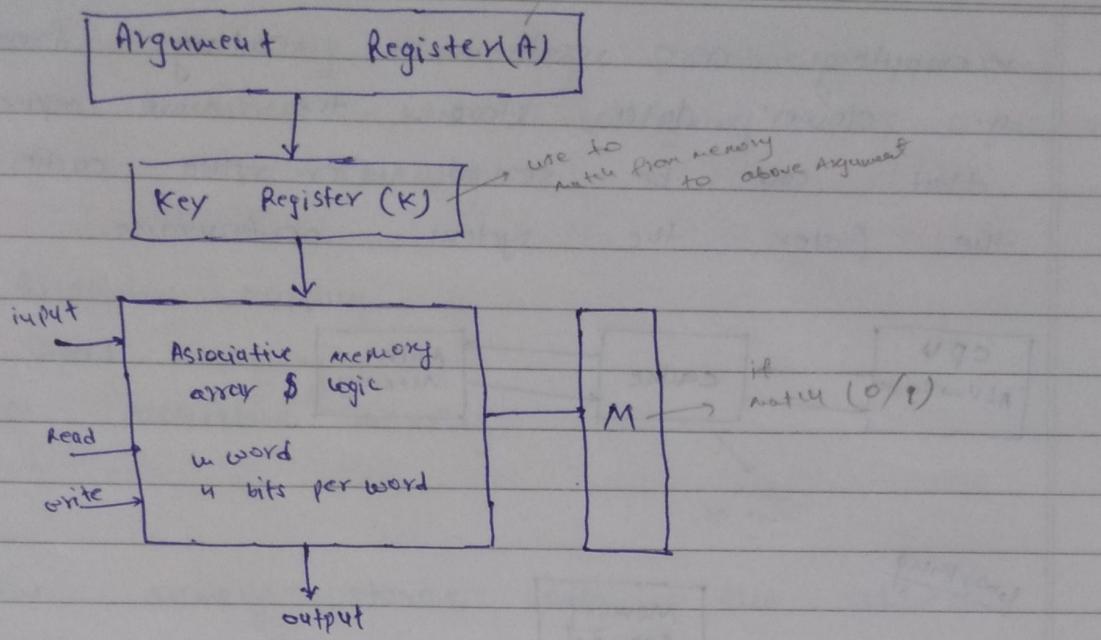
Block diagram,

Associative Memory + or [Content Addressable memory]

Arred the search procedure is a strategy for choosing sequence of address reading the content of memory of each address and comparing the information read with the item being searched until a match occurs.

the Number of memory access depends on location of item and and the efficiency of the search algo.

- Defn
- the time required to find an item stored in memory can be reduced, if stored data can be identified for access by the content of data itself rather than by an address.
 - A memory unit accessed by content is called an associative memory or content addressable memory.
 - Accessed simultaneously and in parallel on the miss of data content rather than by specific address.



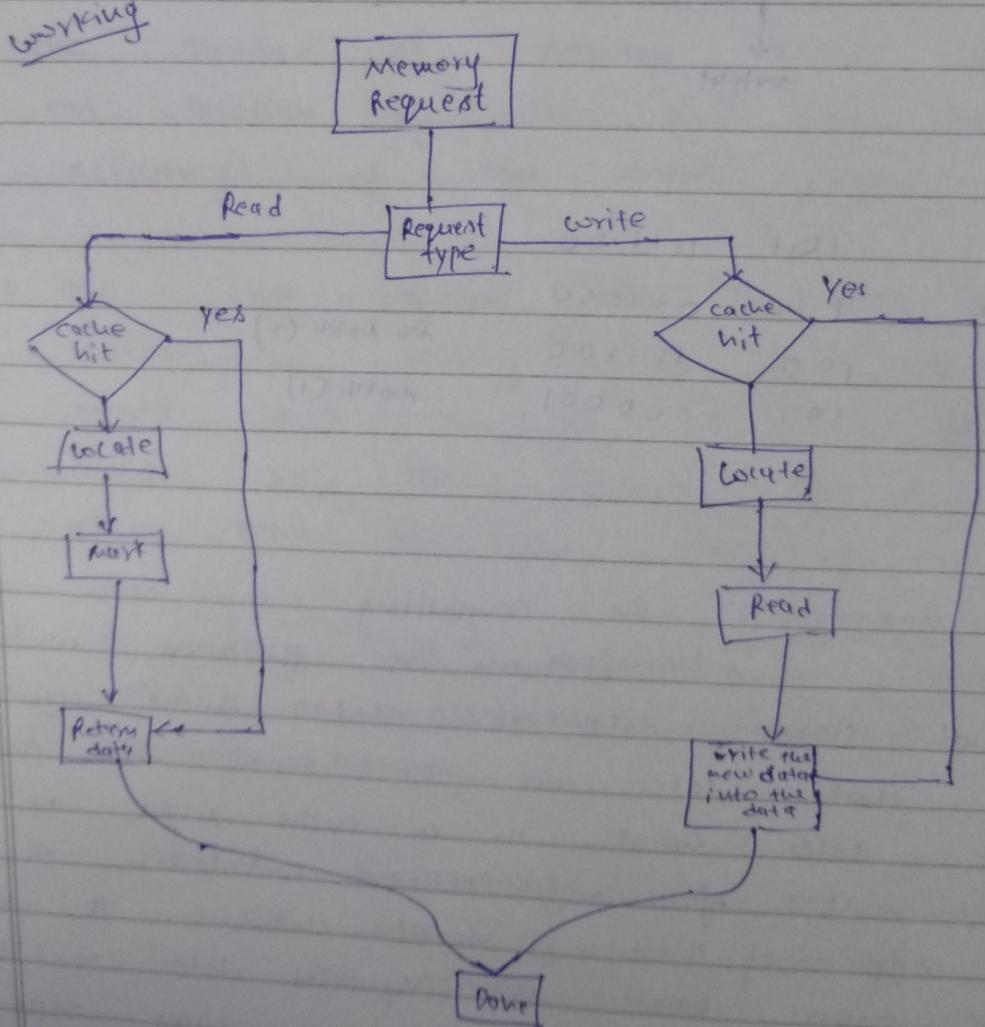
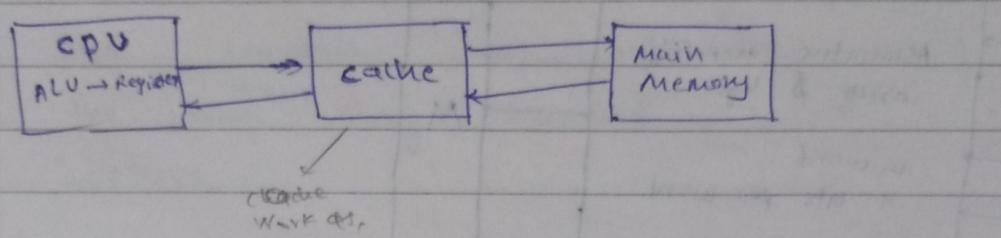
A	101	11100	
(mask)	K	no match 11	000000
word 1		100	111100
word 2		101	000001

no match (0)
match (1)

cache memory :- in computing, a cache is a hardware or software component that stores data so that future requests for that data can be served data faster. the data stored in a cache might be the result of an earlier computation or a copy of data stored elsewhere. A cache hit occurs when the requested data can be found in a cache, while a cache miss occurs when it cannot.

this is faster than

recomputing a result or reading from a slower data store, the more request that can be served from the cache, the faster the system performance.



* three type of mapping procedures are of practical interest when considering the organisation of cache memory.

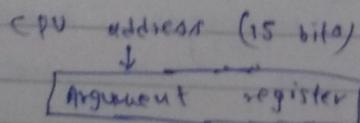
- (1) Associative mapping
- (2) Direct mapping
- (3) set associative mapping.

(1) the memory stores both the address and data of the memory word. A CPU address of 15 bits is placed in the argument register and the associative memory is searched for a matching/mapping address.

If the address is found the corresponding 12 bit data is read and sent to the CPU.

Address	Data
01000	3450
02777	6410
22345	1234

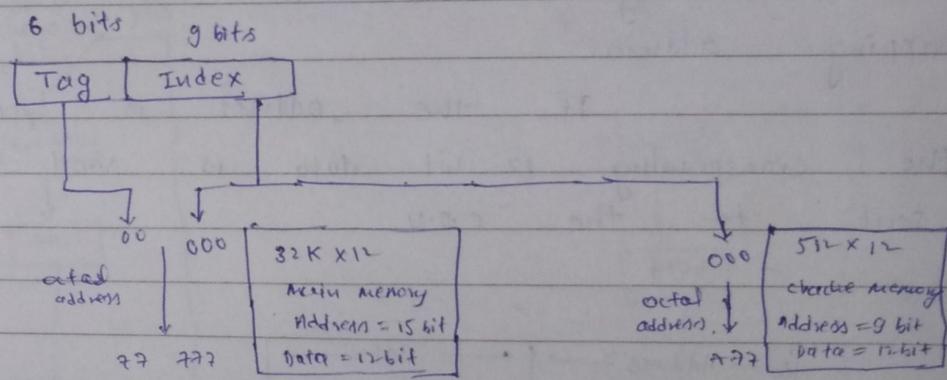
Associative mapping cache



(2)

Direct mapping :-

15 bit is divided into two field. The nine least significant bit constitute the index field and remaining six bit from the tag field. CPU generate a memory request the index field is used for the address to access the cache. The tag field of the CPU address is compared with the tag in the word read from the cache. If there is no match there is miss and the required word is read from main memory.



Direct mapping cache organisation

disadvantage of direct mapping is that two words with the same index in their address but with different tag value cannot reside in cache memory.

(7)

Set - Associative Mapping :- A third type of cache organisation set.

Associative mapping is an improvement over a direct mapping in that each word of cache can store two or more words of memory under the same index address.

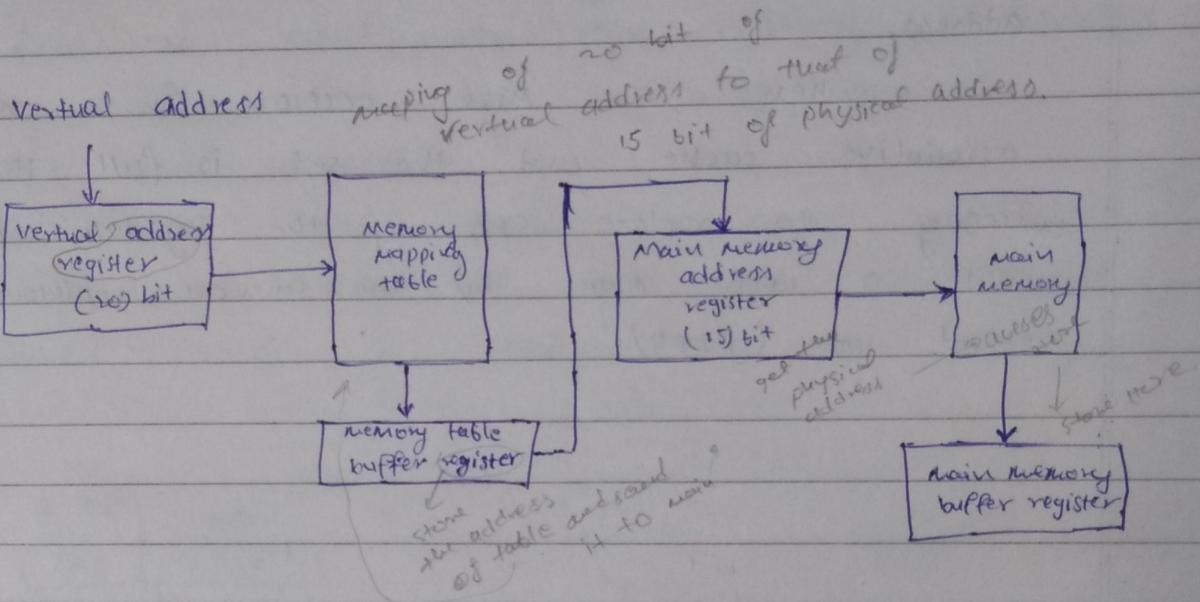
When a miss occurs in a set-associative cache and the set is full it is necessary to replace one of the tag-data item with a new value by the common replacement method is (FIFO)

#

Virtual Memory :- Virtual memory is a concept used in some large computer systems that permit the user to construct programs as though a large memory space were available, equal to the totality of auxiliary memory.

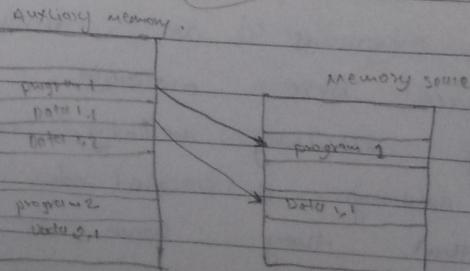
Virtual memory is used to give programmers the illusion that they have a very large

memory at their disposal. A virtual memory system provides a mechanism for translating program-generated address into correct main memory locations. This is done dynamically which can be seen by a mapping table.



* Address space and Memory space \div

An address used by a programmer will be called a **virtual address** and the set of such addresses the **address space**. An address in main memory is called a **location** or **physical address**. The set of such locations is called the **memory space**. The memory space is divided into equal size blocks.

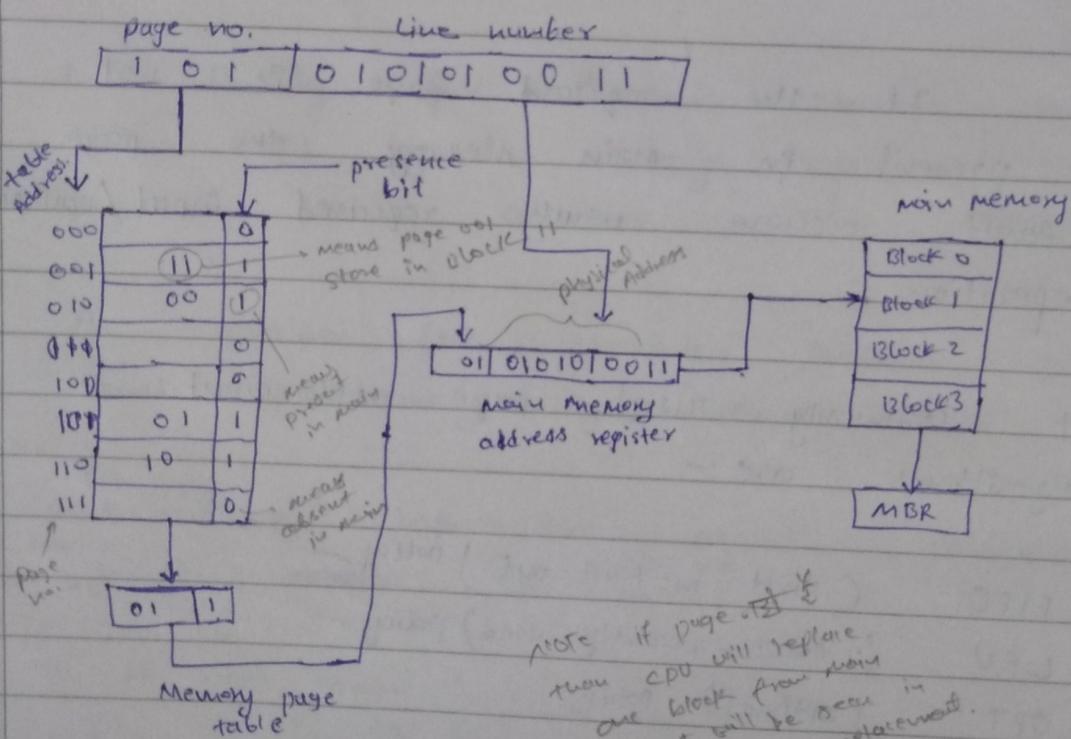


Address mapping using pages:-

the table implementation

of the address mapping is simplified if the information in the address space and memory space are each divided into groups of fixed size.

physical memory is broken down into group of equal size called block.



Note: if page 011 is replaced then CPU will replace one block from main memory that will be seen in page replacement.

(+)

Page replacement :-

In a computer architecture system that uses paging for virtual memory management, page replacement decision within memory page to page out when a program starts execution page of memory needs to be collected within memory page to page out when a program starts execution one or more pages are transferred into main memory and the page table is used to indicate their position if the required page is not present in main memory the page fault operation is performed. If the page is found in main memory then the page fault occurs with required input / output result of the operation.

Most commonly used page replacement algorithms are :-

- ① FIFO (First - in - first - out) policy → Here older page are被淘汰 (evicted) from memory to make room for new arrival of pages.
- ② LRU (Least - recently - used) policy → Within the interval of time the most used pages are被淘汰 (evicted) from memory to make room for new arrival of pages.
- ③ OPT (Optimal) policy → It is never淘汰 (evicted) from memory to make room for new arrival of pages.

Memory Management Hierarchies :-

Memory management hierarchies are a memory management technique used in a part of memory residing in various programs and procedures for hardware and software systems as a collection of a collection of hardware and software procedures for memory management.

(1) A facility for dynamic storage relocation :-

The basic components of a memory management unit are :-

It is a part of an overall operating system in many computers, it is a part of memory management system / software.

(2) A facility for dynamic storage relocation :-

A fault maps logical memory reference into physical memory addresses.

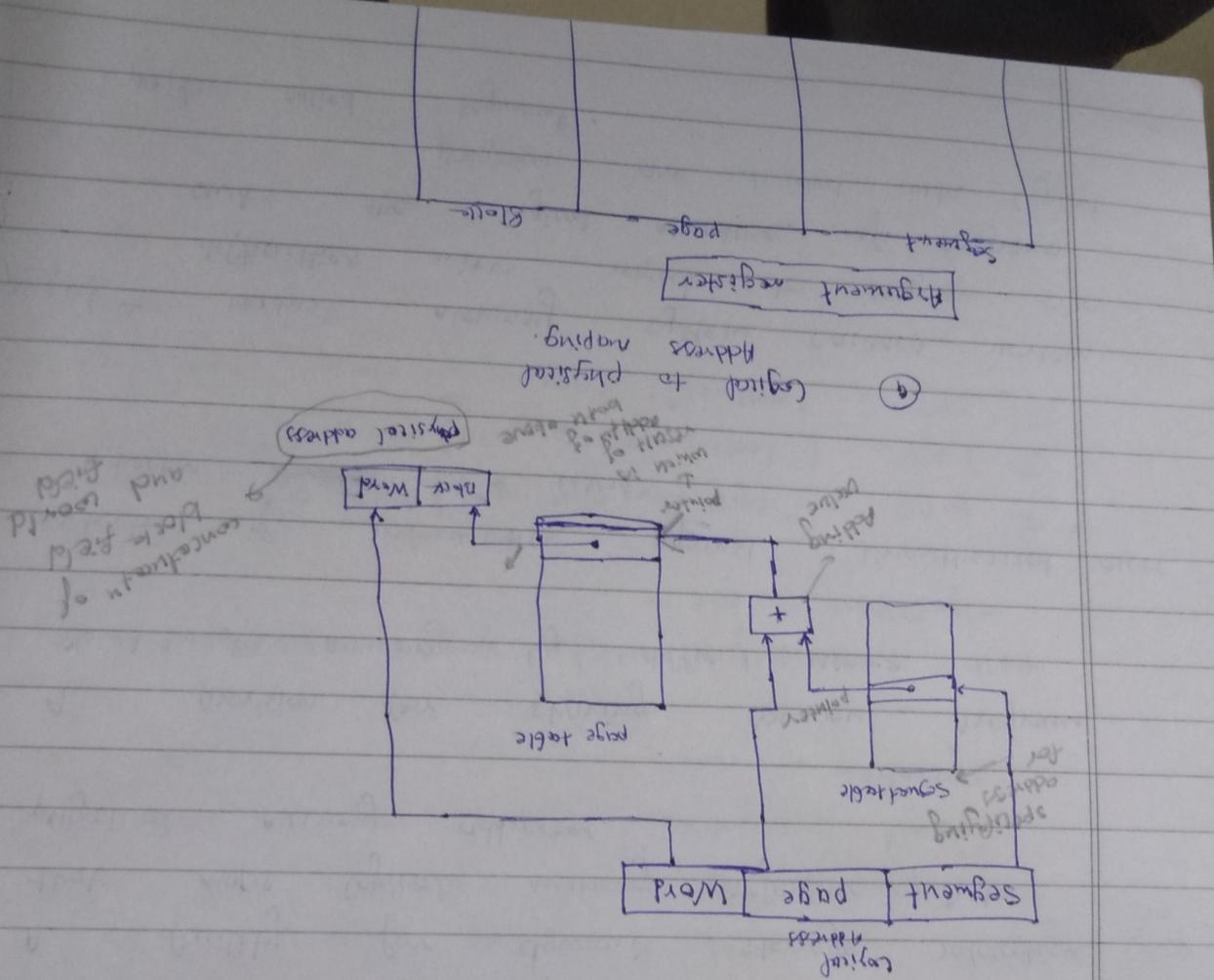
(3) A provision for sharing common program :-

Shared in memory by different users.

(4) Protection of information against unauthorized access between ~~user~~ users :-

Scheme :-

Virtual memory system causes certain difficulties with respect to program size and the logical structure of programs. Programs are divided into logical segments and called segments.

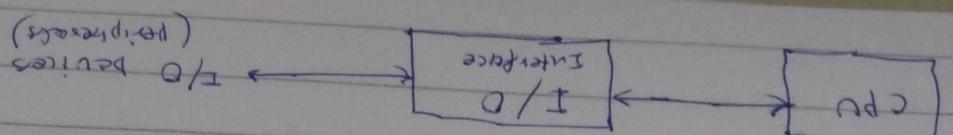


the segment number given in the logical address
two page table base is added to a
address for a page table base. The
segment for page mapping :- the entry in the
segment table if a filter

program is controlled at logical address.
The addresses generated by a segmented

⊕

program is controlled at logical address.
The addresses generated by a segmented



at effective and efficient processing,

Data transfer rate and peripheral are different from the word format.

Data transfer rate are usually slower. So a synchronized interface may be needed.

To resolve these difference, computer system included special hard ware component better than the CPU and peripheral to supervise all input and output and output transfer to these components are called interface units.

peripheral device Input - output device connected to the computer are also called peripherals.

Input - output operations of the computer if modules all the system

and external devices for transferring between method provide a method

Input - Output Interface =

③

④

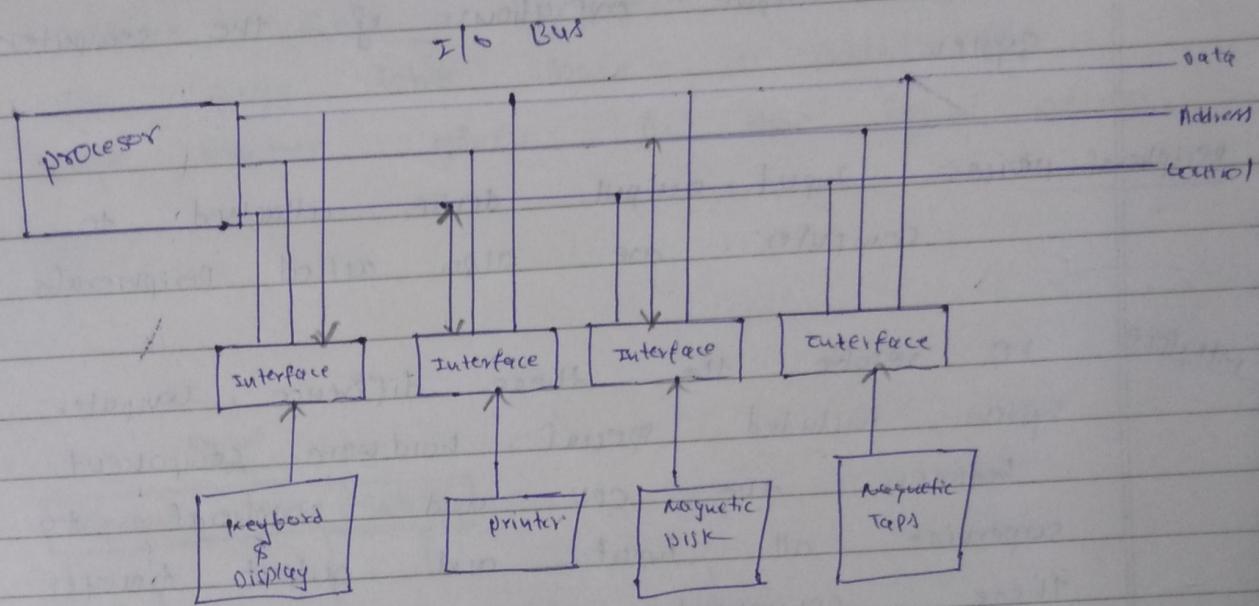
⑤

Interface

peripheral

#

④ I/O bus and interface modules :-
 the above
 difference can be synchronized by
 this device



I/O command :- instruction executed in the interface
 and if attached peripheral device.

control command :- used to activate the command
 peripherals and to tell it what
 to do.

status command :- used to test various status
 condition in the interface &
 the peripherals.

Output data :- cause the interface to respond by transferring data from bus into one of its register.

Input data :- interface receive an item of data from the peripherals and place it in its buffer register.

or

causes the I/O modules to obtain an item of data from the peripheral.

Asynchronous Data Transfer :-

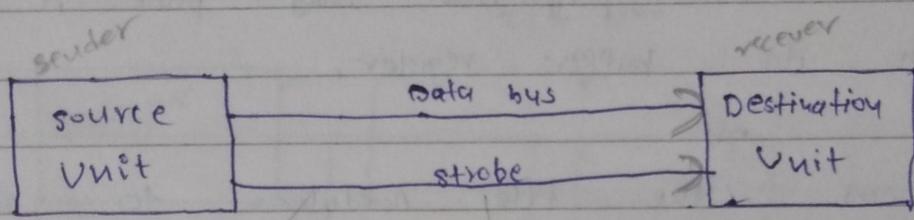
the internal operation in a digital system are asynchronized by means of clock pulse, supplied by a common pulse generator.

clock pulse are applied to all registers within a unit and all data transfer among internal register occur simultaneously during the occurrence of a clock pulse.

if the registers in the interface share a common clock with the CPU register, the transfer between the two unit is said to be synchronized.

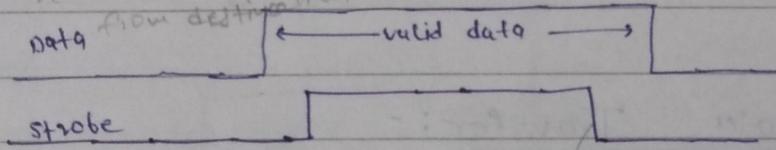
a strobe pulse is supplied by one of the unit to indicate to the other unit when the transfer was to occur.

Strobe:- One way of achieving this is by means of a strobe pulse applied by one of its unit to indicate to the other unit when the transfer has to occur,



① Block Diagram "to receive data,"

also use to receive in source file
from destination



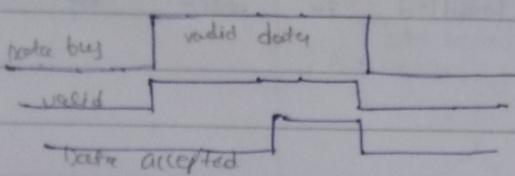
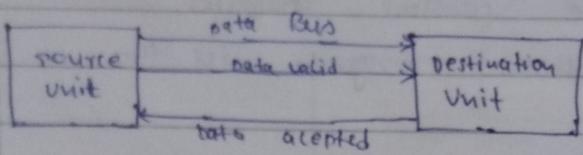
② Timing diagram "destination Receive until the strobe pulse is on"

figure
Source - initiated strobe for data transfer.

Handshaking:- the disadvantage of the ~~strobe~~ strobe method is that the source unit that initiates the transfer has no way of knowing whether the destination unit actually received the data or not.

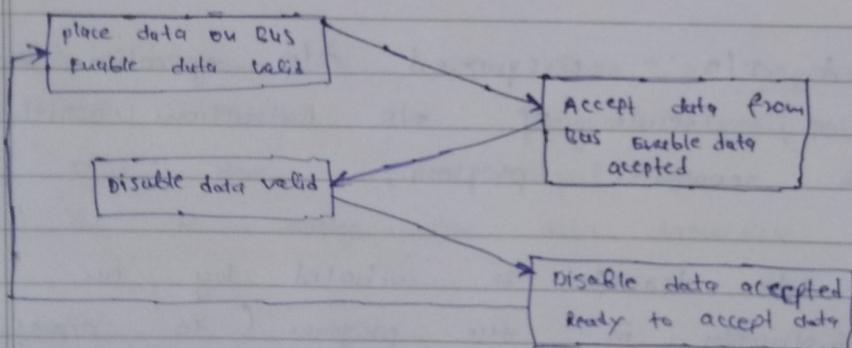
→ A control signal is accompanied with each data item being transmitted to indicate the presence of data.

→ the Receiving unit responds with another control signal to acknowledge the receipt of data.



Source

Destination



Modes of Transfer :-

an information received from external device is usually stored in memory for latter processing. Information transferred from the central computer into an external device originate in the memory unit.

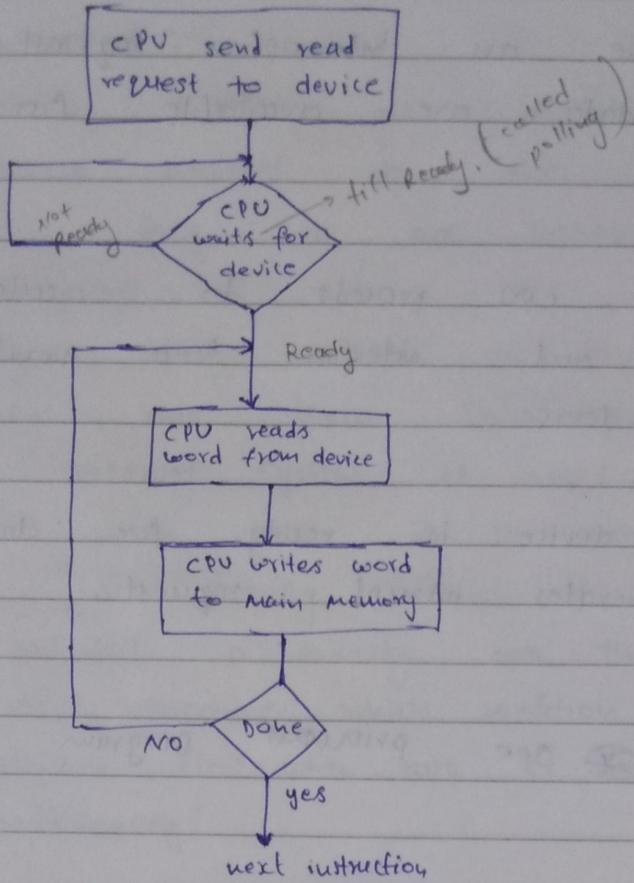
Data transfer to and from peripherals may be handled in one of three possible mode.

- 1) programmed I/O
- 2) Interrupt-initiated I/O
- 3) Direct memory access (DMA)

① programmed I/O :- programmed I/O operation are result of I/O instructions written in computer program.

→ each data transfer is initiated by an I/O instruction in the program (to access register memory on a device)

→ Transferring data under program control requires constant monitoring of the I/O devices by the CPU.



in program I/O CPU make a request and then CPU stays in program loop (polling) unit the input output device indicate that it is ready for data transfer.

Disadvantage

time consuming process since it keep CPU busy.

② interrupt initiated I/O :-

instead of continuous monitoring of CPU, interface will be informed

to issue an interrupt request signal
when data are available from the device.

- Meanwhile, CPU proceeds to execute another program and interface keep monitoring the device.
- When device is ready for data-transfer it generates interrupt request.

④ See previous diagram

#

Priority Interrupt :-

There are number of input / output device attached to the computer.

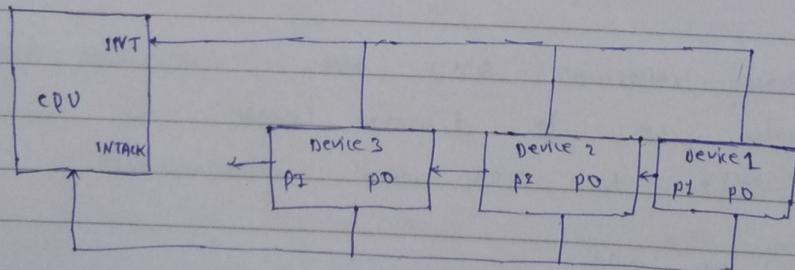
They all are capable of generating the interrupt.

When the interrupt is generated from more than one device simultaneously.

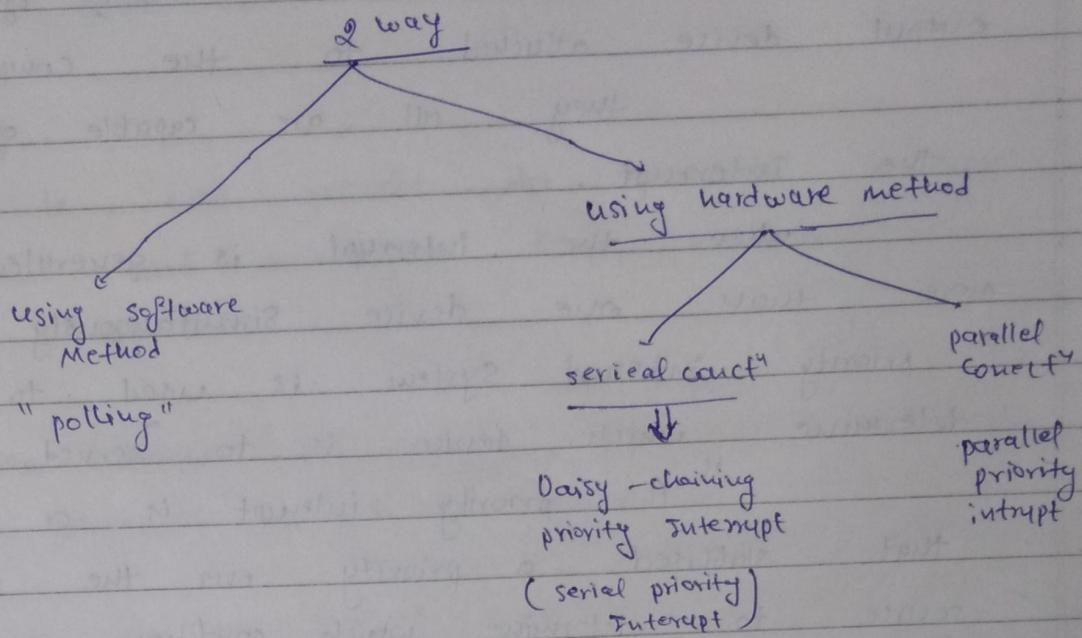
Priority interrupt system is used to determine which device is to be served first.

"A priority interrupt is a system that stabilised a priority over the various source to determine which condition is to be serviced first when two or more request arrive simultaneously."

→ Device with high speed priority transfers are given higher priority (eg: "Harddisk") and slow device are given lower priority, (Keyboard)



④ stabilising priority of simultaneous Interrupts :-



Direct Memory Access (DMA) :-

By the help of DMA
we can directly store the
data into memory without use
of CPU

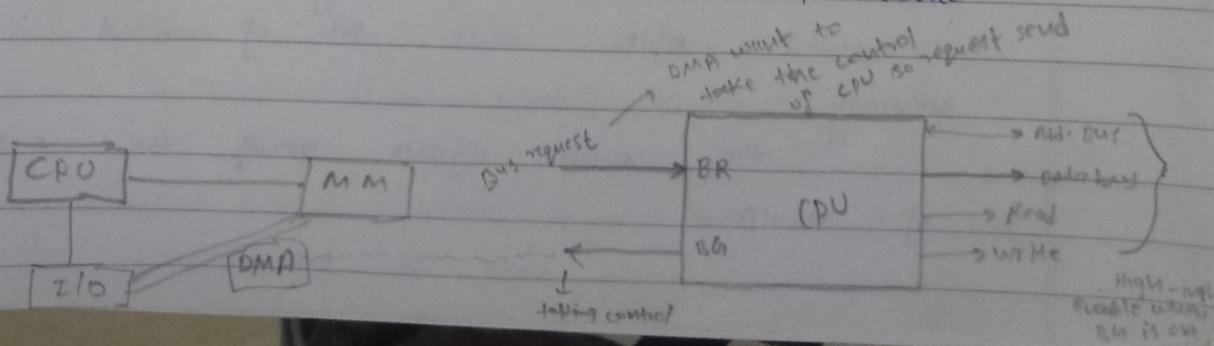
Removing the CPU from the path and letting the peripheral device manage the memory busses ~~will~~ directly would improve the speed of transfer.

bus request :- the bus request (BR) input is used by the DMA controller to request the CPU to leave control of the busses.

bus grant :- the CPU activate the bus grant (BG) to inform the external DMA that the busses are in the high-impedance state.

Cycle stealing :- cycle stealing allow the DMA controller to transfer one data word at a time after which it must return control of the busses to the CPU.

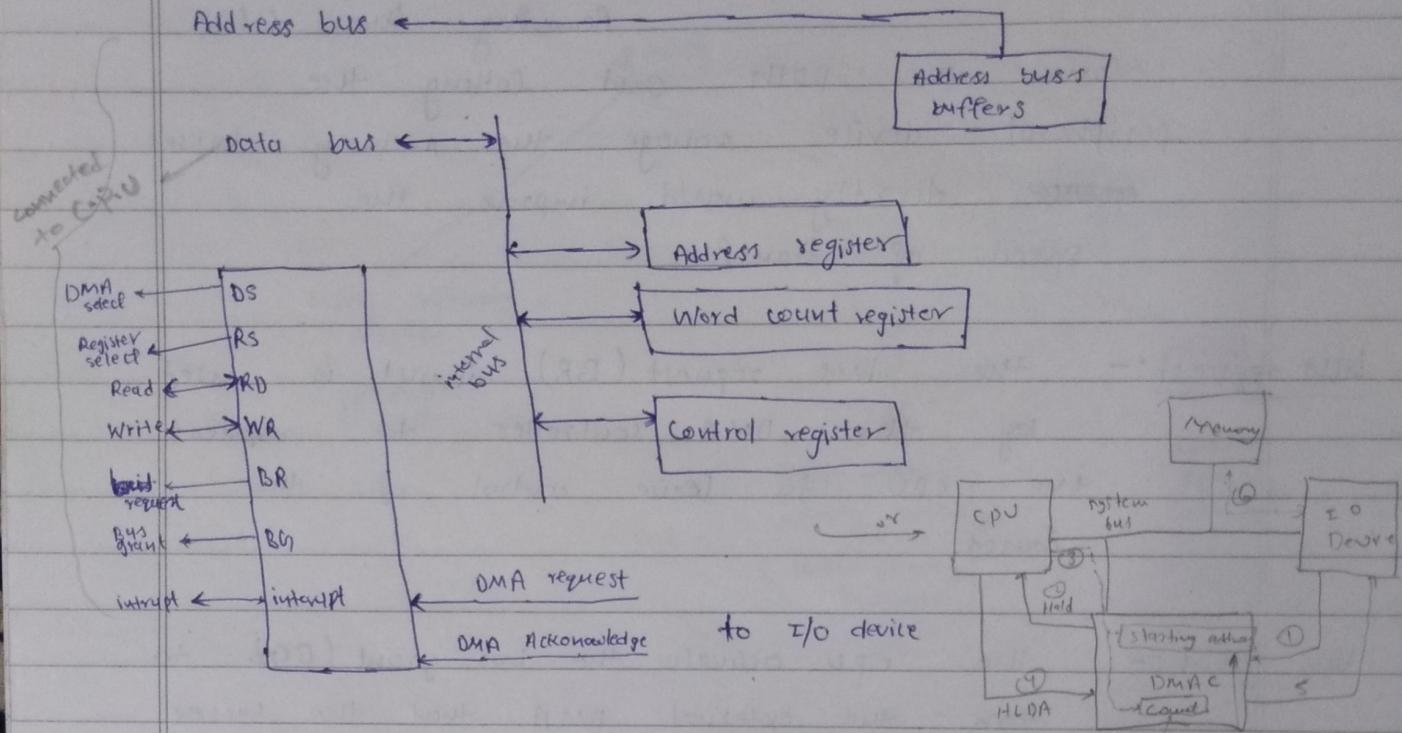
① DMA controller :- the DMA controller needs to usual circuit of an interface to communicate with CPU and I/O device



Block Diagram

DMA \times Controller

- There are three mode of transfer
- ① Programmed I/O
- ② Interrupt I/O
- ③ DMA Transfer



The registers in the DMA are selected by the CPU through the address bus, by enabling the (DMA select) and (register select). The read and write unit are bidirectional.

DMA controller has three register in address, a word count register and a control register.

Address register specify the desired location in memory, and it is incremented after each bit of word in memory.

The control register specify the mode of transfer.

Working

DMA is first ~~set~~ initialized by the CPU after that the DMA starts and continue to transfer data between memory and peripheral unit.

- 1) the starting address of the memory block where data are available (for read) and data stored (for write)
- 2) the word count, which is the number of words in memory block.
- 3) control → specify the mode of transfer.

(*)

#

DMA Transfer :-

During DMA transfer CPU can perform only those operation which do not require system bus.

Because bus has been accessed with I/O device for the shortest interval of time.

* it has three mode of transfer.

- ① Burst Mode
- ② cycle stealing
- ③ Interleaving DMA

① Burst Mode :- Burst of data is transferred before CPU takes the control of buses back.

② Cycle Stealing :-

→ Slow IO device takes some time to prepare the word.

→ During this type CPU keeps control of the buses.

→ Once word is ready CPU gives the control of the bus to DMA for 1 cycle. In which prepared word is transfer to memory.

③ Interleaving DMA :-

Here whenever CPU does not require system buses then only control of the buses will be given to DMA

→ CPU will not be blocked due to DMA

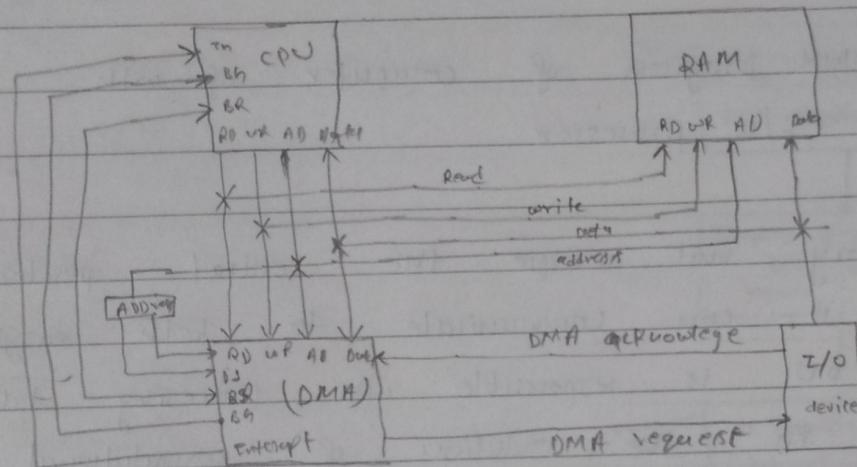
make it easy.

time required for data

transfer : Interleaving > cycle stealing > Burst mode

so. speed of

transfer: Burst mode > cycle stealing > Interleaving

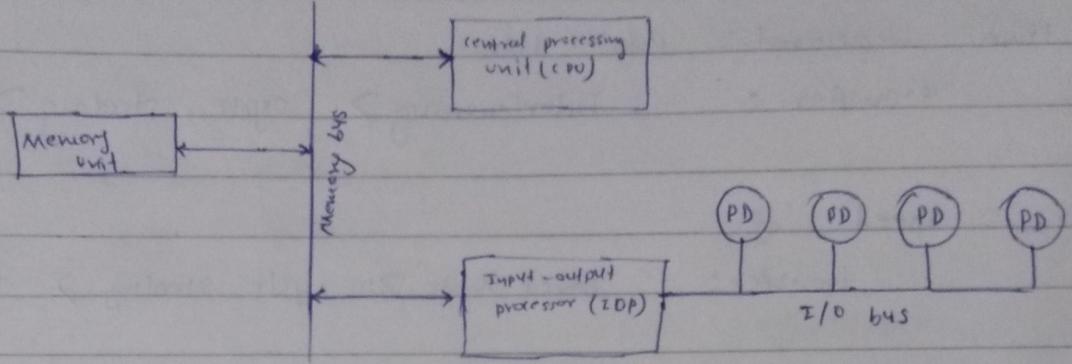


DMA Transfer in a computer system.

Input - Output Processor (IOP) :-

the IOP is similar to the CPU except that it is designed to handle the details of I/O processor. just like the DMA controller that must be set entirely by the CPU.

the input output processor can ~~not~~ fetch and execute ~~its~~ its own instruction without involvement of CPU.



Block Diagram of computer of with
I/O processor

Here memory unit occupy the central position so that it can communicate to both easily. CPU is responsible for processing data needed in the solution of computational task.

The communication between the IOP and the device attached to it is similar to the programmed controlled method of transfer. Communication with memory is similar to the DMA method.

② CPU - IOP communication:-

