

21/02/25

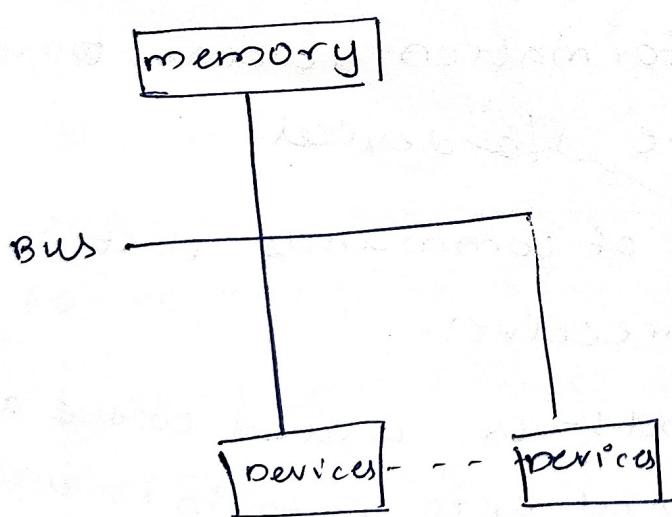
## Module-5

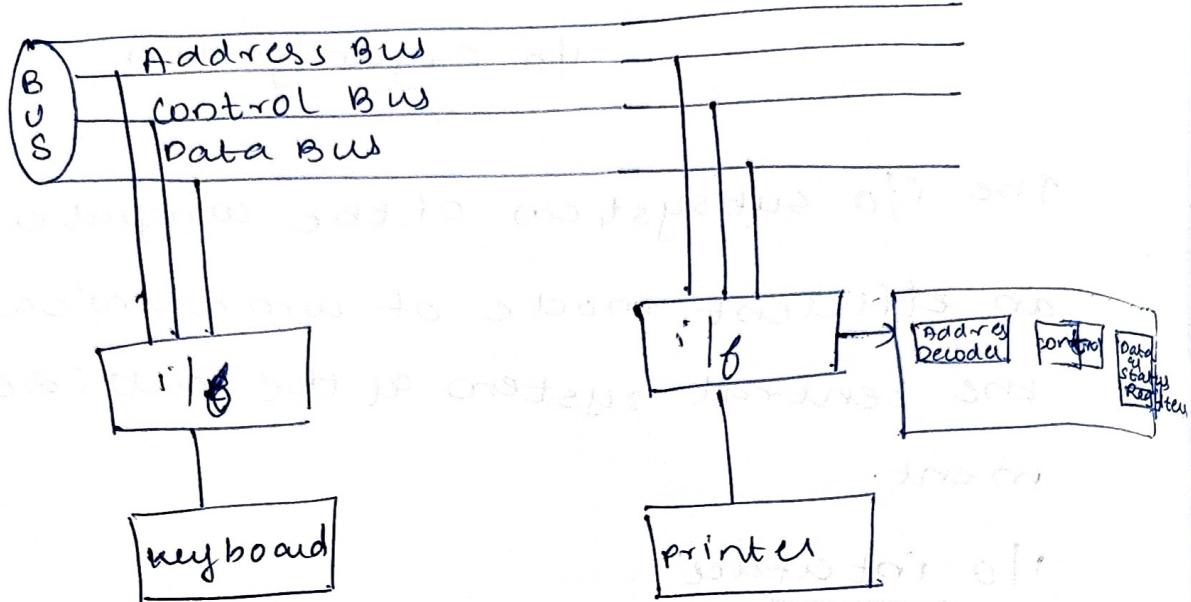
### I/O Organization

The I/O subsystem of the computer provides an efficient mode of communication b/w the central system & the outside environment.

#### I/O interface

This provides a method of transferring information b/w internal storage & external I/O devices. Peripherals connected to a computer need special communications links for interfacing them with the CPU. The purpose of the communications link is to resolve the differences that exist b/w the central computer & each peripheral.





Address Decoder is located in the peripheral.

It enable the device to recognize the address when the address appear on the address line.

Data & Status Register

It hold the data being transfer to or from the processor.

It contains information relevant to the operation of the I/O device.

There are 4 types of commands that an interface may receive:-

- (i) control command :- is issued to ~~the~~ activate the peripheral to inform it what to do.

- (ii) status command (rin & sout): is used to test various status cond' in the interface & the peripheral.
- (iii) data output command (write or DATA OUT): it causes the interface to respond by transferring data from the bus into one of its registers.
- (iv) data input command (Read or DATA IN): it allows the interface to receive an item of data from the peripheral & places it in the buffer register.

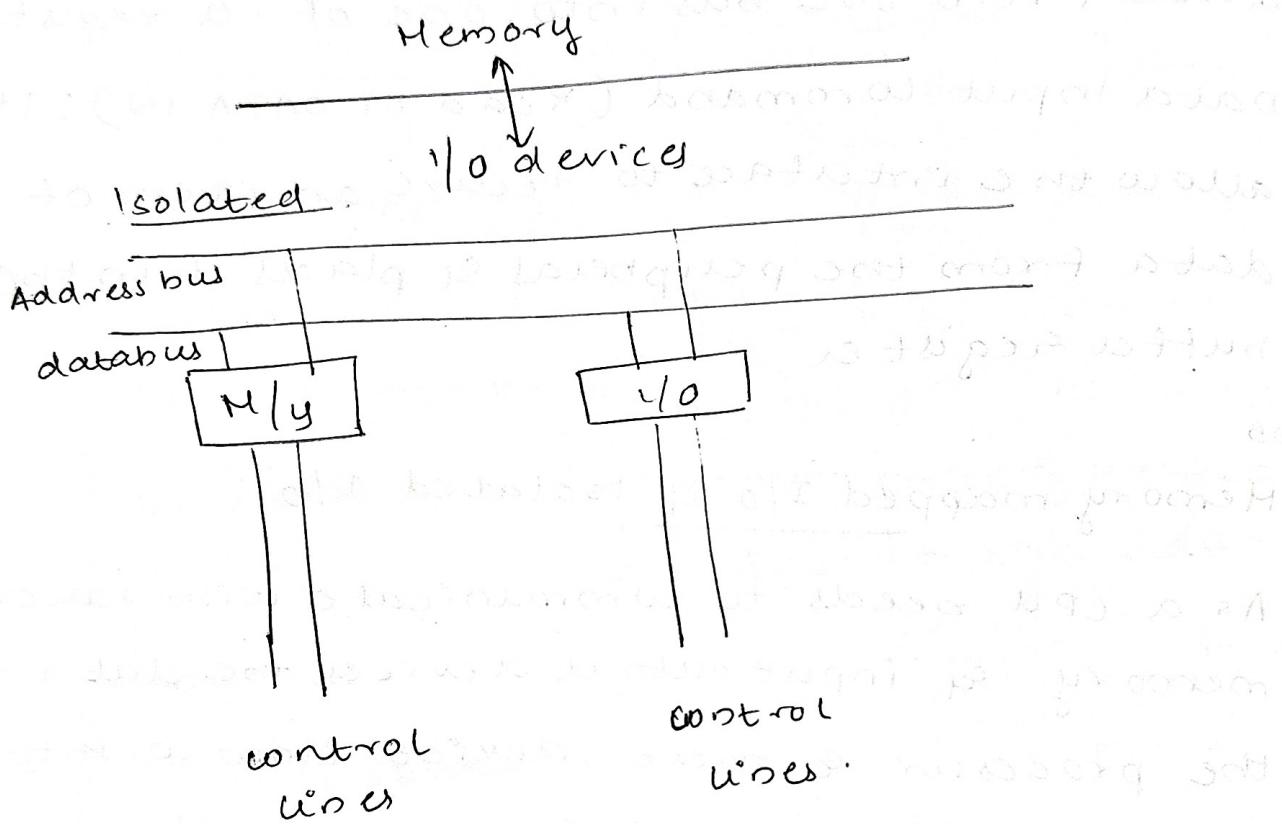
2/02/23

### Memory mapped I/O & Isolated I/O

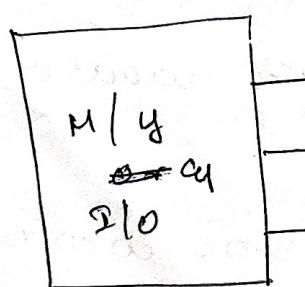
As a CPU needs to communicate with various memory & input output devices, the data b/w the processor & these devices flow with the help of a system bus. There are 3 ways in which the system bus can communicate:

- (i) use two separate buses, one for memory & other for I/O. In this method, all data, address & control lines would be separate for memory & I/O.
- (ii) Isolated I/O: use one common bus for both memory & I/O but have separate control lines. It has its own address space. (i.e. separate memory & I/O address space)

(iii) Memory Mapped I/O: A single set of instructions is used by both memory & I/O. Memory & I/O addresses share the common address space.



### Memory mapped



## Modes of transfer

Data transfer between the central computer & I/O devices must be handled in 3 modes.

- 1) programmed I/O
- 2) interrupt driven I/O
- 3) DMA(Direct Memory Access)

### Programmed I/O

In programmed I/O each data transfer is initiated by ~~by~~ the instruction in CPU.

- (i) Input instruction: Used to transfer data from I/O devices to CPU.
- (ii) Stored instruction: Used to transfer data from CPU to memory.
- (iii) Output instruction: Used to transfer data from CPU to I/O device.

Programmed I/O is used in very slow speed computer as it is not an efficient method.

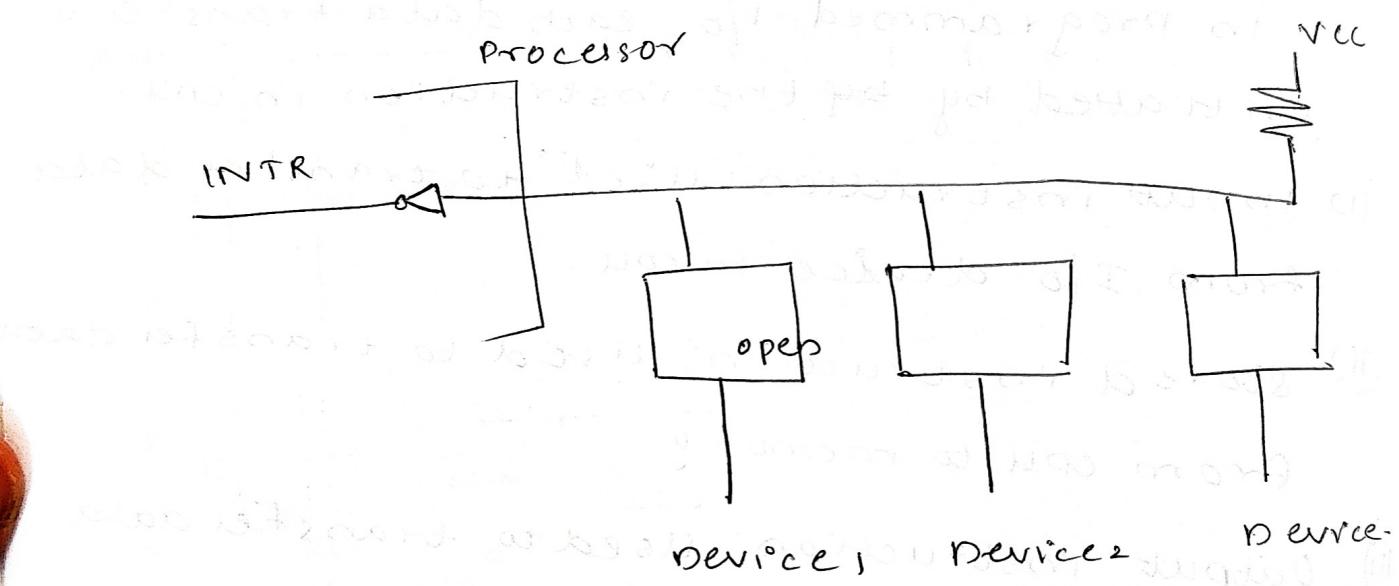
### Interrupt driven I/O

When a program enters a wait loop or an interrupt is encountered, it will check the device status. During this period, the processor will not perform any fn. There are many situations where other task can be performed.

ed while waiting for an I/O device to become ready. To allow this happen, we can arrange an INTR to alert the processor when it becomes ready i.e sending a hardware signal called an interrupt to the processor.

ISR - Interrupt service Routine. - The routine executed in response to an interrupt request.

### Interrupt Hardware



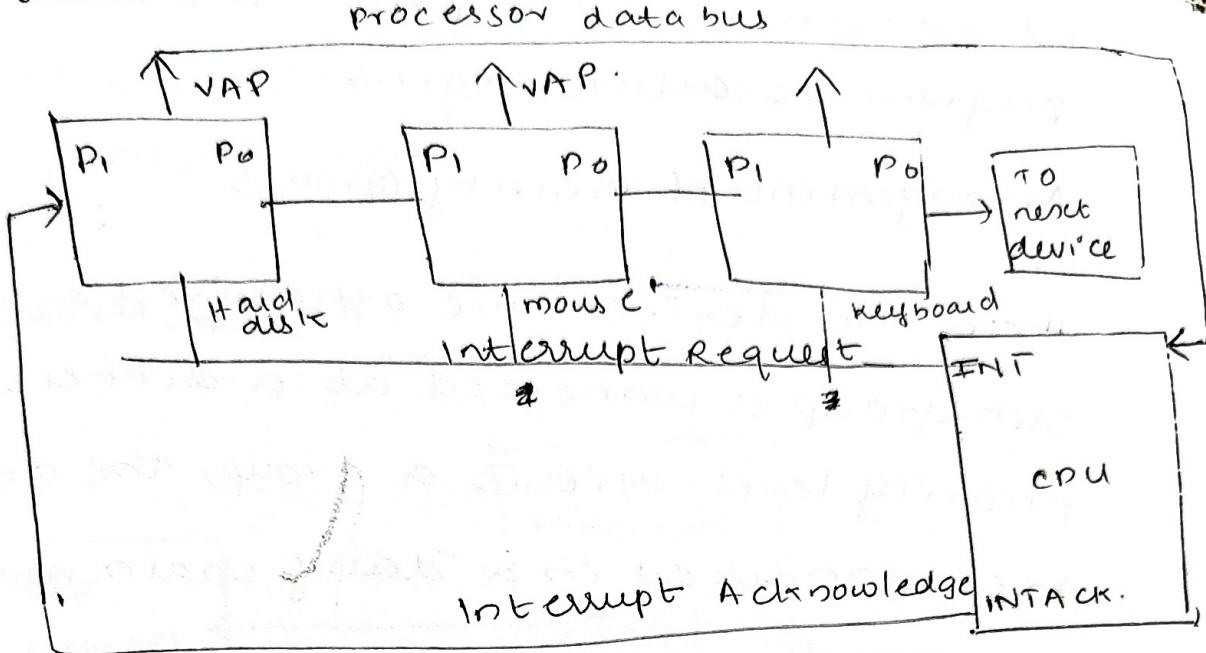
### Handling multiple interrupts

To handle multiple interrupt, two mechanism are used:-

① Daisy chaining

② Arrangement of priority groups.

## Daisy Chaining



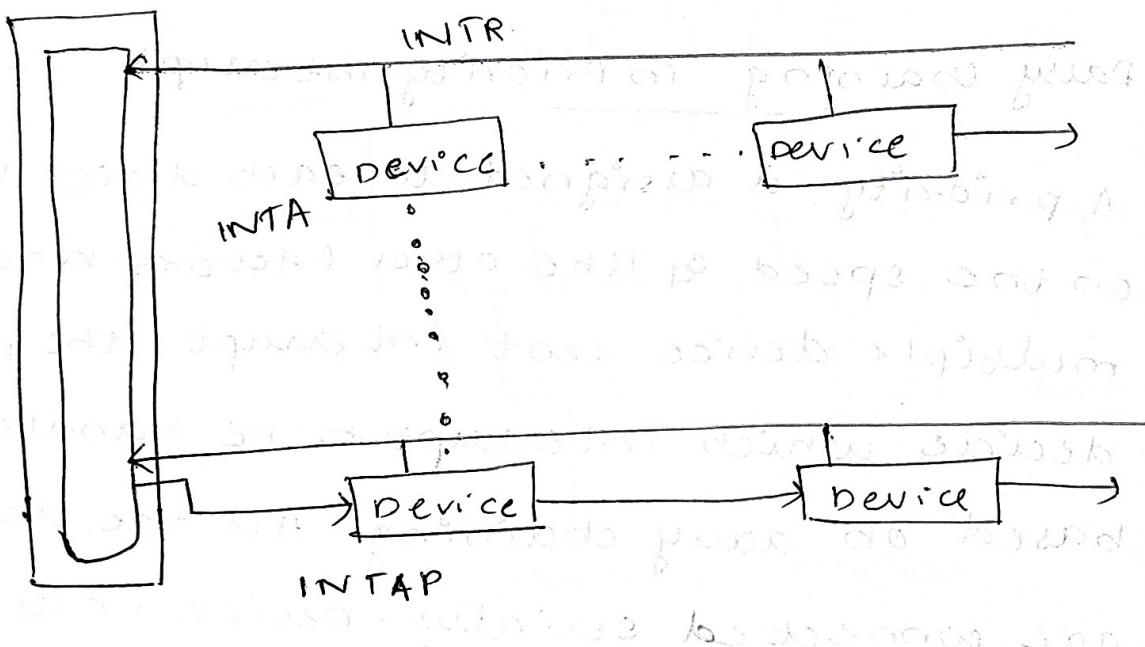
## Daisy Chaining in Priority Interrupt

A priority is assigned to each device based on the speed & like other factors. When multiple device sent interrupt, the processor decide which interrupt to be handled is based on daisy chaining. All the devices are connected serially. Device 1 has ~~the~~ highest priority than device 2 & so on. If any interrupt occurred, the interrupt request line will be set to 1. Then the interrupt acknowledgement also becomes 1. If the first device generate the interrupt means, it consumes 1 & P<sub>O</sub> will be 0 & it goes to the next device. Otherwise, the 1 is passed to next device through ~~P<sub>O</sub>~~ of operation goes on. When the device consumes 1, vector

ed address of ISR is passed to the CPU & program execution begins.

## Arrangement of Priority Groups

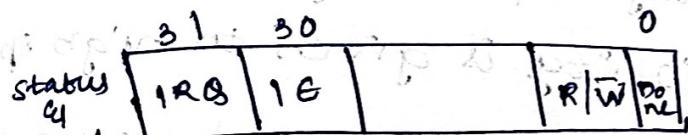
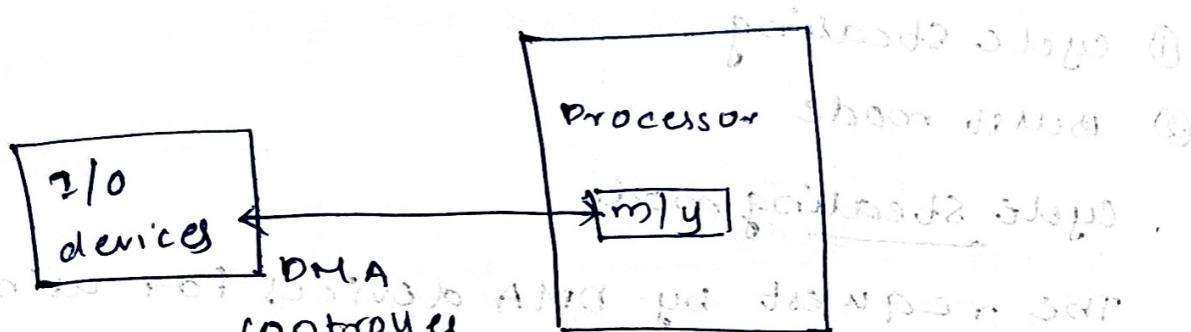
Here the devices are organized as groups & each group is connected at a different priority level. Within a group, the devices are connected in a daisy chain manner.



## DMA Design Methodology Description

The process of transferring data directly b/w memory & an external device without continuous intervention of the CPU or the processor. It is a parallel technique used for high speed I/O device. DMA transfers are performed by a control circuit called DMA controller. To initiate the data transfer, the processor sends

- (i) Starting address or word to which data is transferred
  - (ii) No. of words in the block which is stored in memory
  - (iii) Direction of transfer: Byte bit word double word
- 00100100 registers in DMA



static address in the descriptor passing to processor

$R = m/y$  to I/O device

$m =$  from device to  $m/y$

word count:  $m/y$  processor will keep at time

completed transmission to processor

R/W

when Read=1, the data is transferred from memory

to I/O device. Write is enabled when the

bit at 0 of the data is transferred from

device to memory. Done=1, the controller has

completed the transferring of data & is

ready to receive another command. Then  $IE$  = 1, it causes the controller to raise an

interrupt signal to the processor.

interrupt after it has completed transferring the block of data.

IRQ=1 :- It indicate that the controller has requested an interrupt.

### Mode of Transfer

① cycle stealing

② burst mode

#### Cycle Stealing Mode

The request by DMA devices for using the bus are having higher priority than processor request. Top priority is given to high speed peripherals like disc, high speed network interface, graphic display devices.

In such case, if the processor originates most memory access cycles, the DMA controller can be

said to steal the memory cycles from the processor. This interviewing techniques

is called cycle stealing.

#### Burst mode

In this mode, the DMA controller has exclusively access to the main memory to transfer a block of data without interruption.

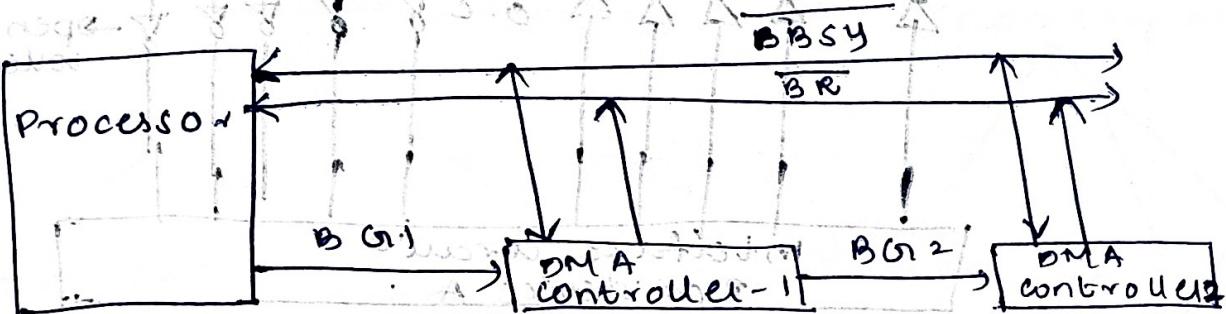
## Bus Master (Processor or I/O device)

The device which initiates the data transfer on the bus at a given time.

### Bus Arbitration

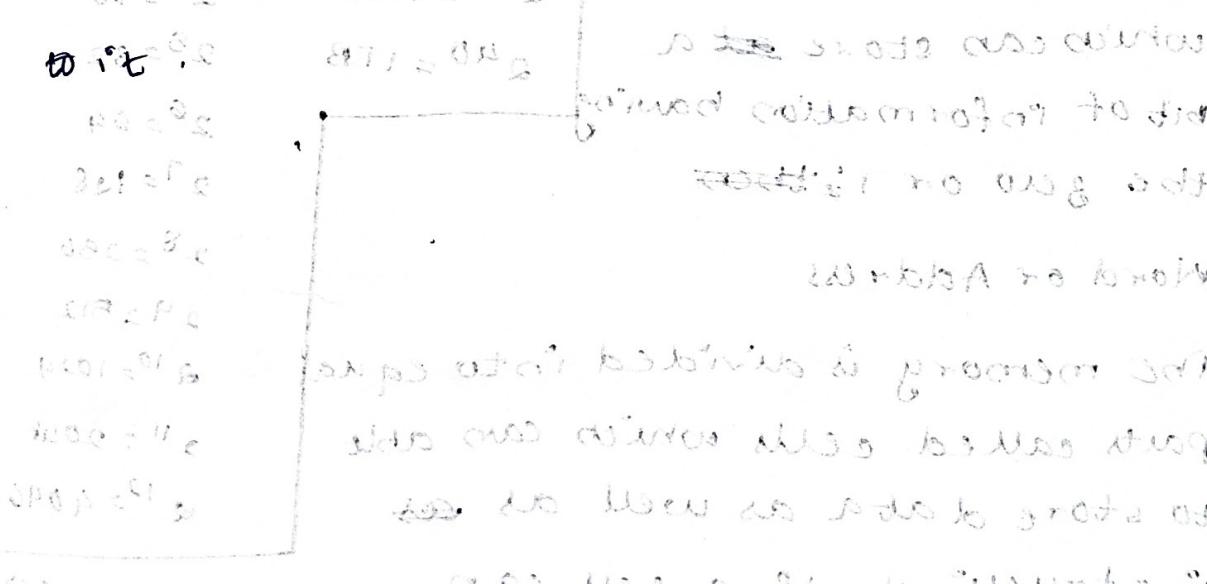
- ① centralized Arbitration
- ② distributed Arbitration.

#### centralized

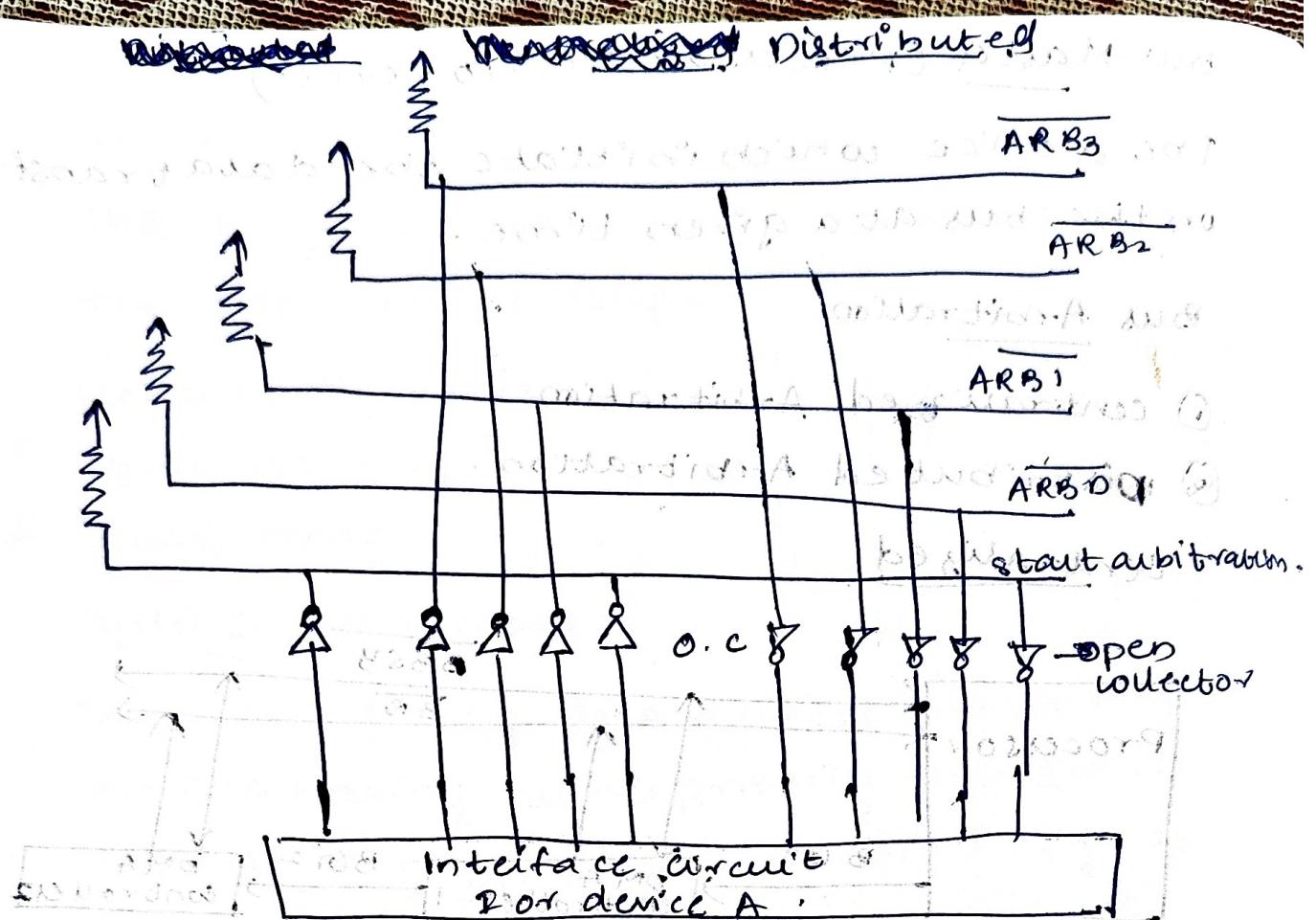


### Bus Arbitration

Bus Arbitration:- It is the process by which the next device to be come the bus master is selected and the bus mastership is transferred to it.



The diagram illustrates the bus arbitration process. The central bus (BBSY) is shown with multiple segments, each labeled "BBSY". Arrows point from the central bus to various peripheral components like Processor, RAM, ROM, Cache, and interface chips. This visualizes how arbitration is performed across the entire bus system to determine the next bus master.



## Memory Systems

The memory consists of many millions of storage cells each of which can store ~~at~~ a bit of information having the zero or 1's.

### Word or Address

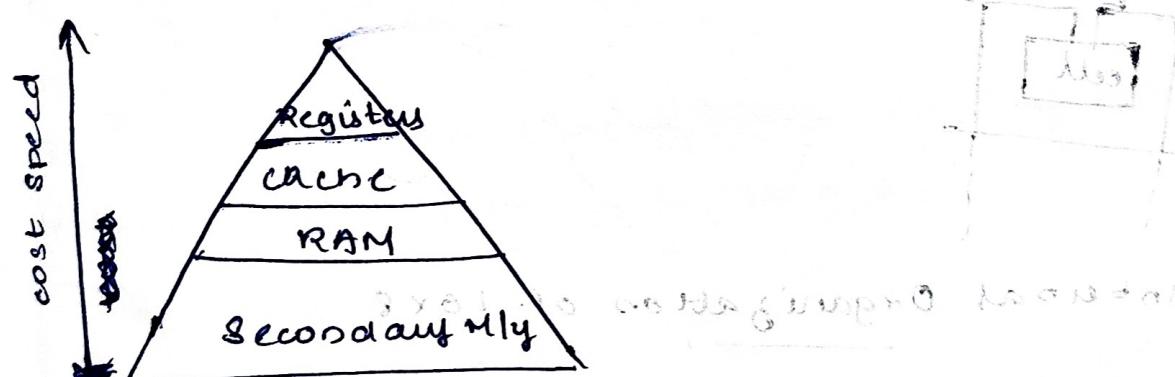
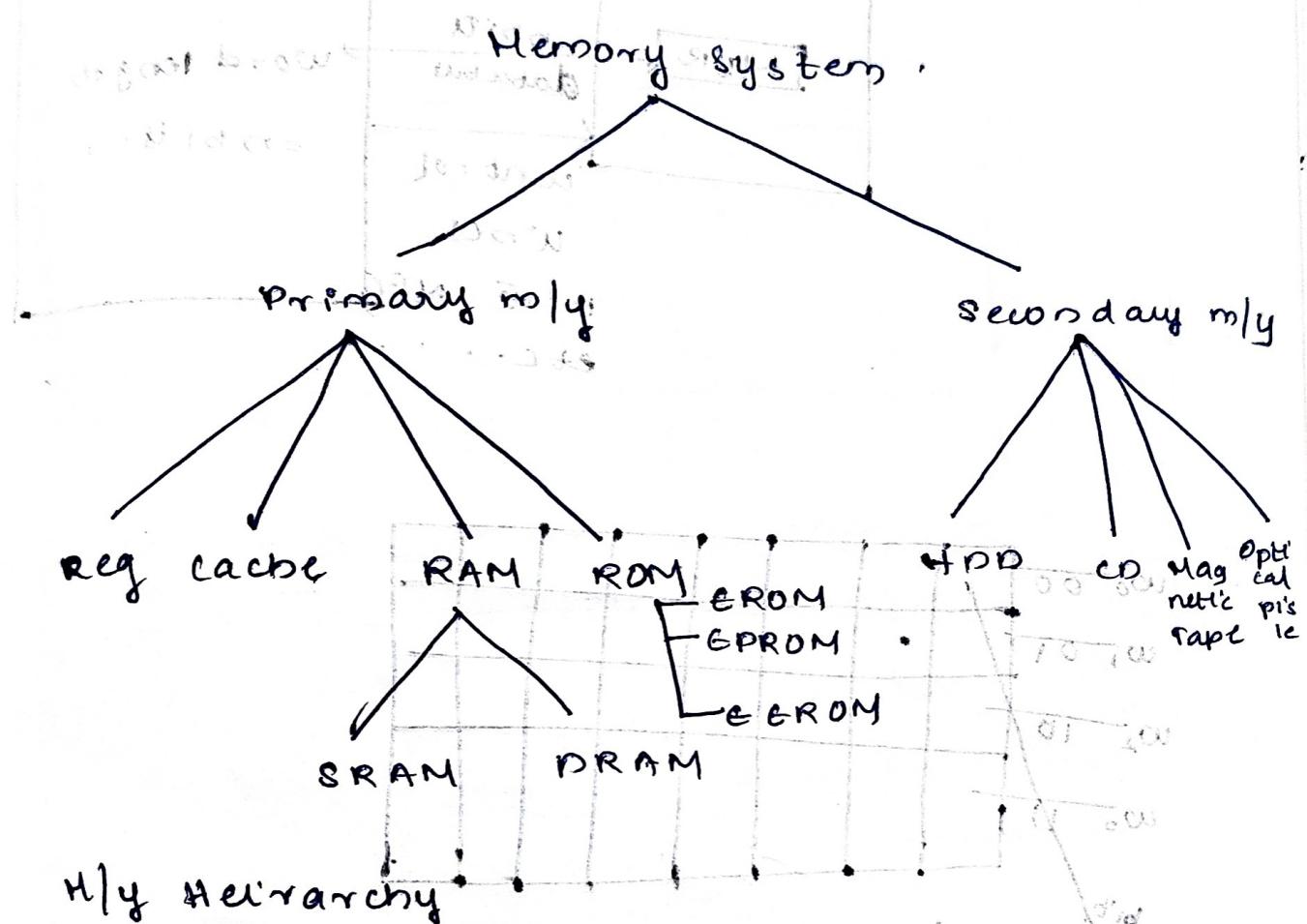
The memory is divided into equal parts called cells which can able to store data as well as ~~as~~ instructions. If a cell can store a maximum of 8 bits or one byte is called byte addressable. If a cell can store more than 1 bytes it is word

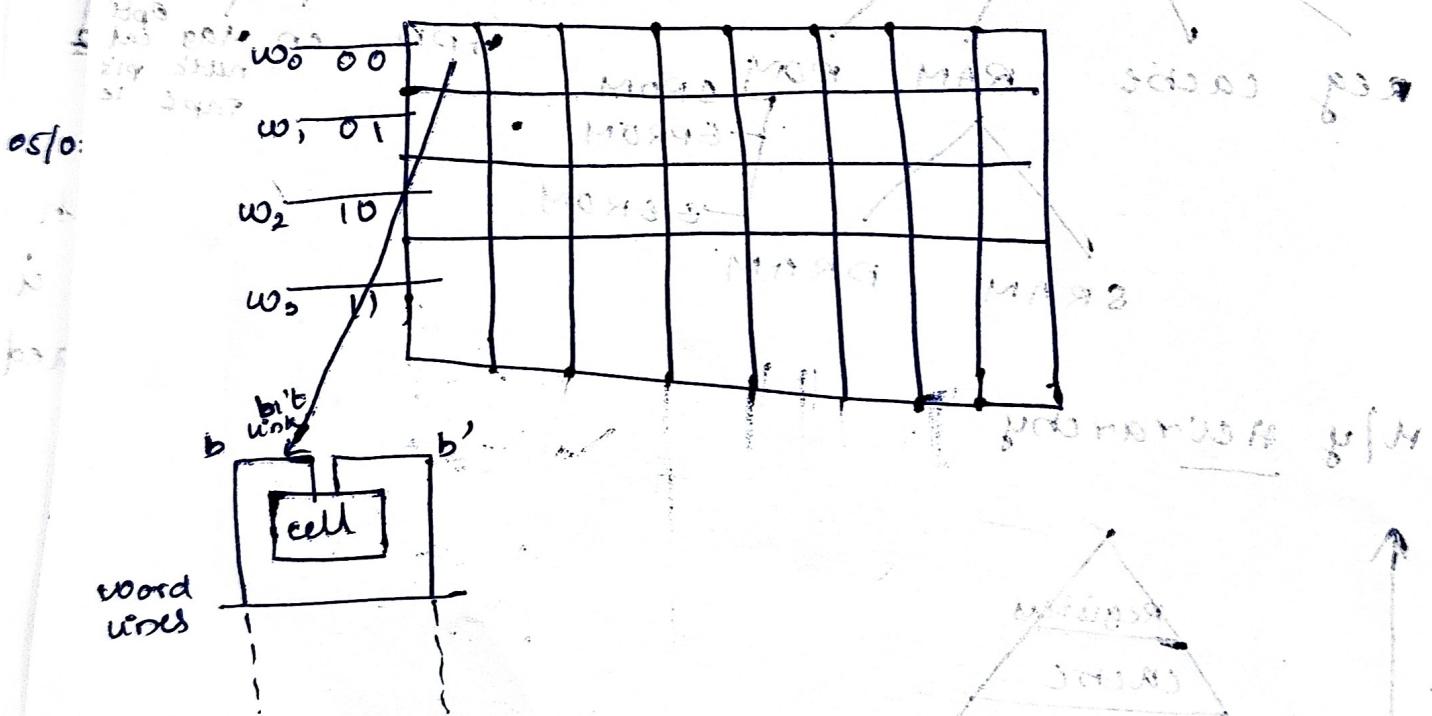
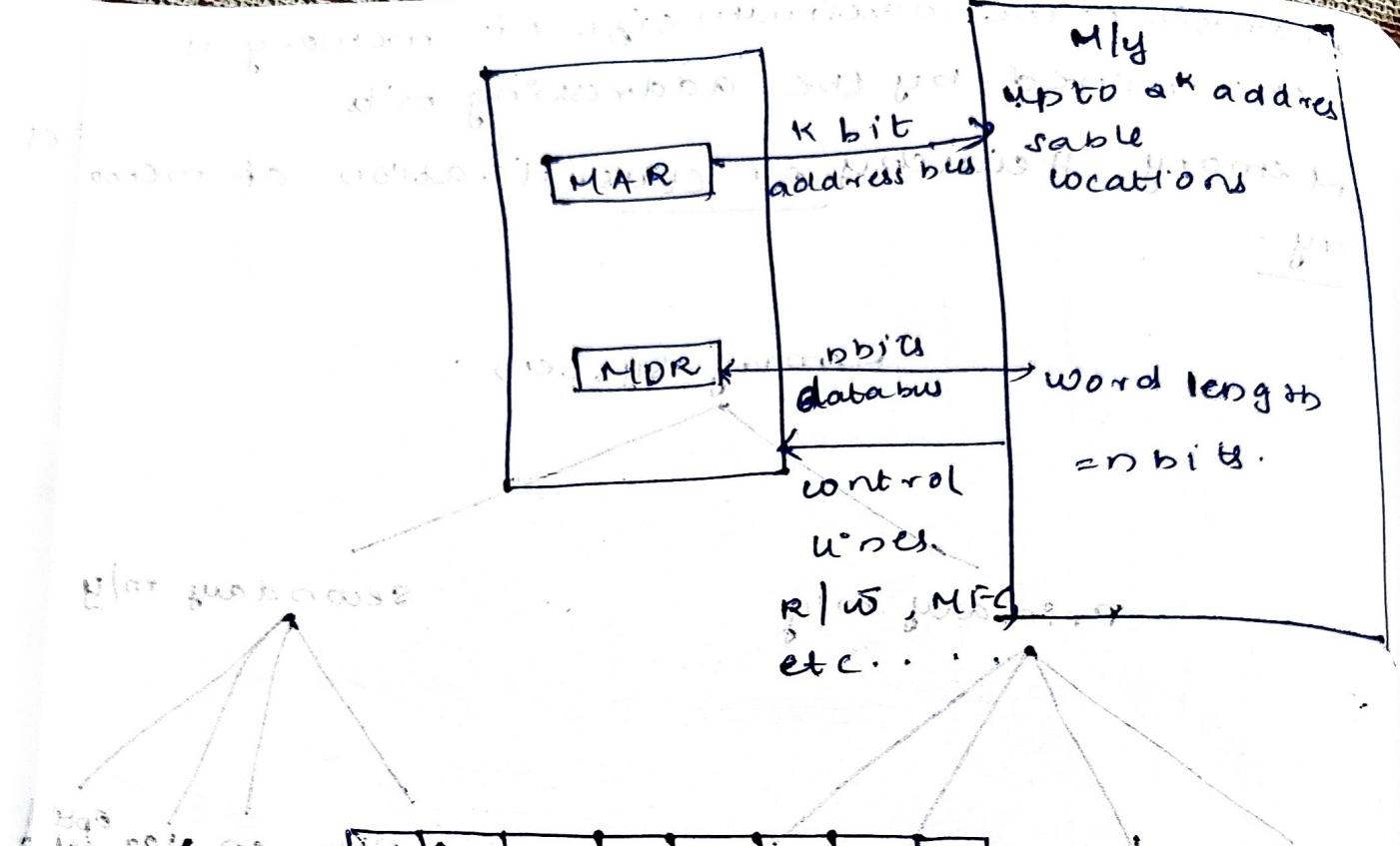
$2^{10} = 1024 = 1 \text{ KB}$	$2^2 = 4$
$2^{20} = 1 \text{ MB}$	$2^3 = 8$
$2^{30} = 1 \text{ GB}$	$2^4 = 16$
$2^{40} = 1 \text{ TB}$	$2^5 = 32$
	$2^6 = 64$
	$2^7 = 128$
	$2^8 = 256$
	$2^9 = 512$
	$2^{10} = 1024$
	$2^{11} = 2048$
	$2^{12} = 4096$

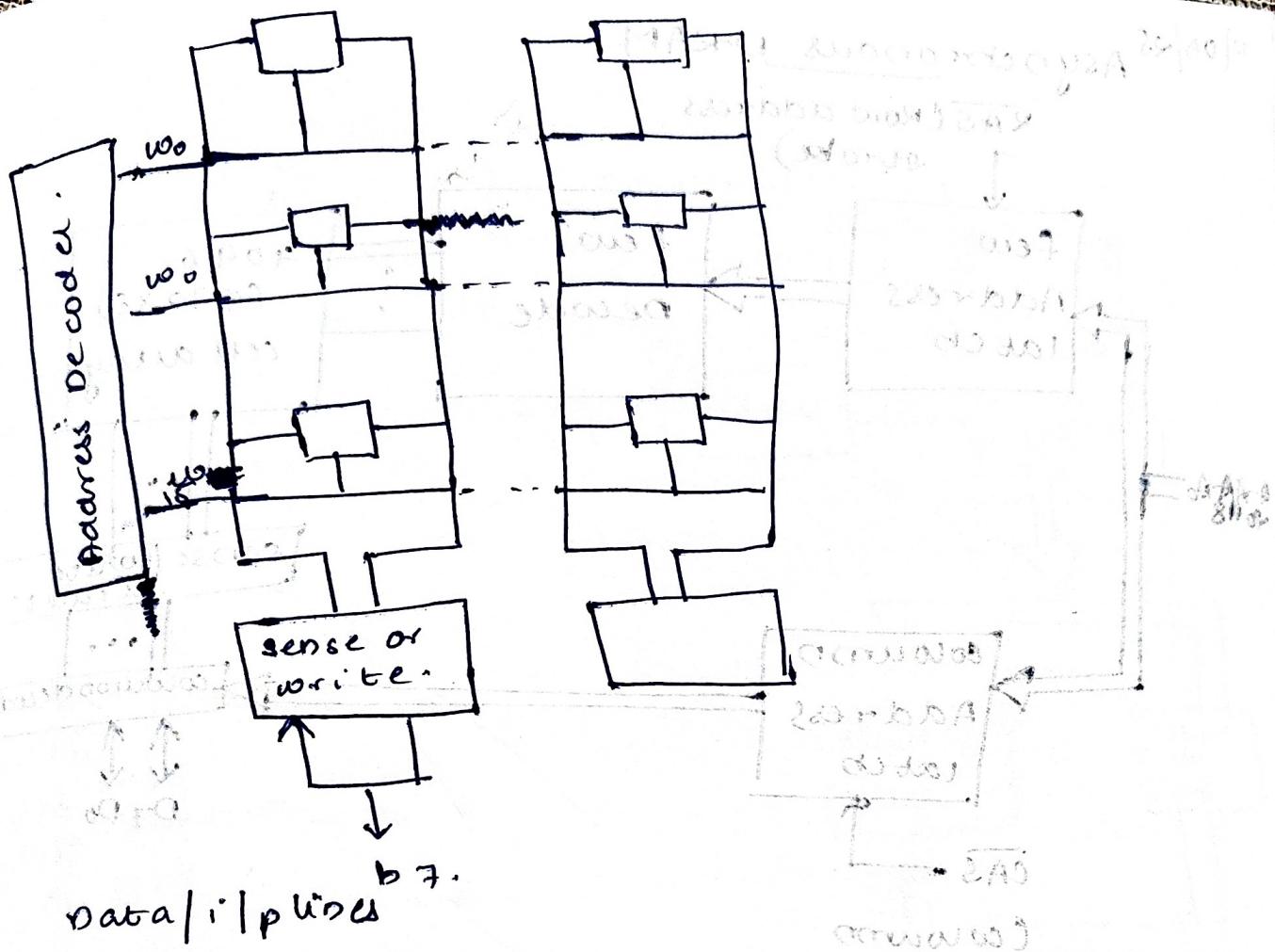
Intel processor  
⇒ Endian

addressable. The maximum size of memory is determined by the addressing bits.

Memory Hierarchy or classifications of memory







Horizonas à 32 bits para escrita e leitura

Adressamento é de 960, só posso ordenar que

horizontais horizontais ou 330 da diagonal

horizontal e 330 vertical 256x8bit · EAS é base basta

que aí pra significado só pra ler e só pra

escrever só pra ler e só pra ler e só pra

escrever só pra ler e só pra ler e só pra

escrever só pra ler e só pra ler e só pra

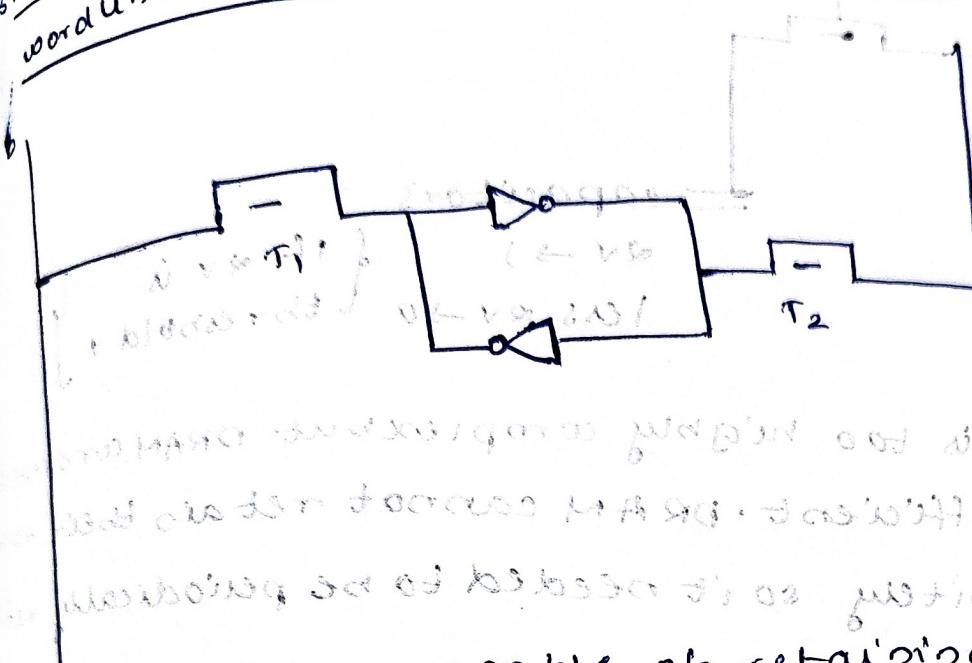
(EAS) significa 256x8bit



13/05

SRAM  
wordline

SRAM



the circuits are capable of retaining their state  
as long as power is applied (volatile)

Read  
when address is applied to wordline  
the wordline is activated if the transistors are on  
then mode, the sense of write circuit at bottom

monitor the state of  $b$  &  $\bar{b}$

Write  
writings are done by placing appropriate  
value on bit line  $b$ . It is fast, but  
it is expensive

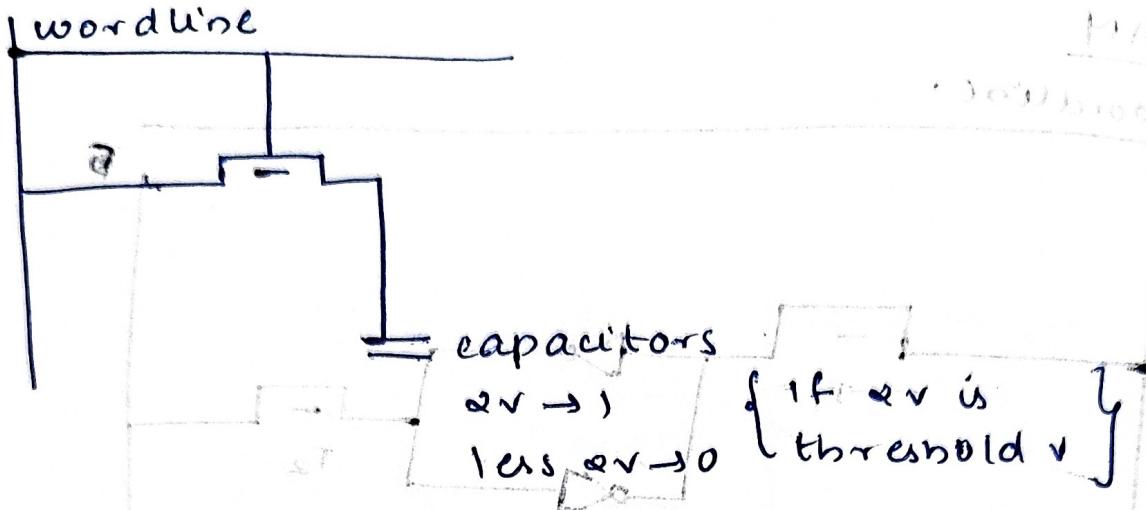
- expensive

- complex

- volatile

- used to make cache memory

## DRAM



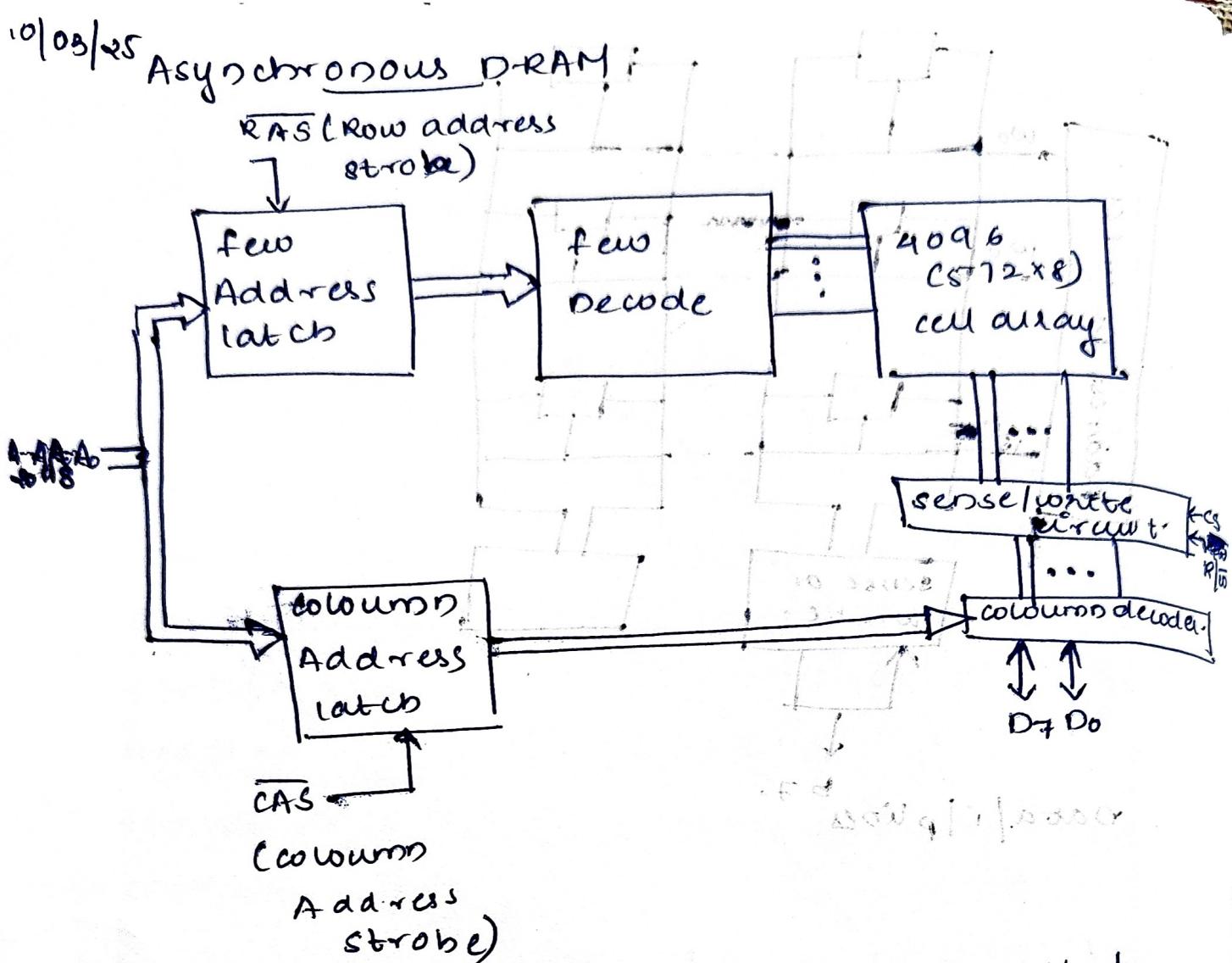
SRAM is too highly complex but DRAM are cheap, area efficient. DRAM cannot retain their state indefinitely so it needed to be periodically refresh.

### Read op

transistor is turned on, sensor checks voltage of capacitor. If voltage is  $<$  threshold value, it represents logical 0; if the voltage is  $>$  threshold it represents logical 1.

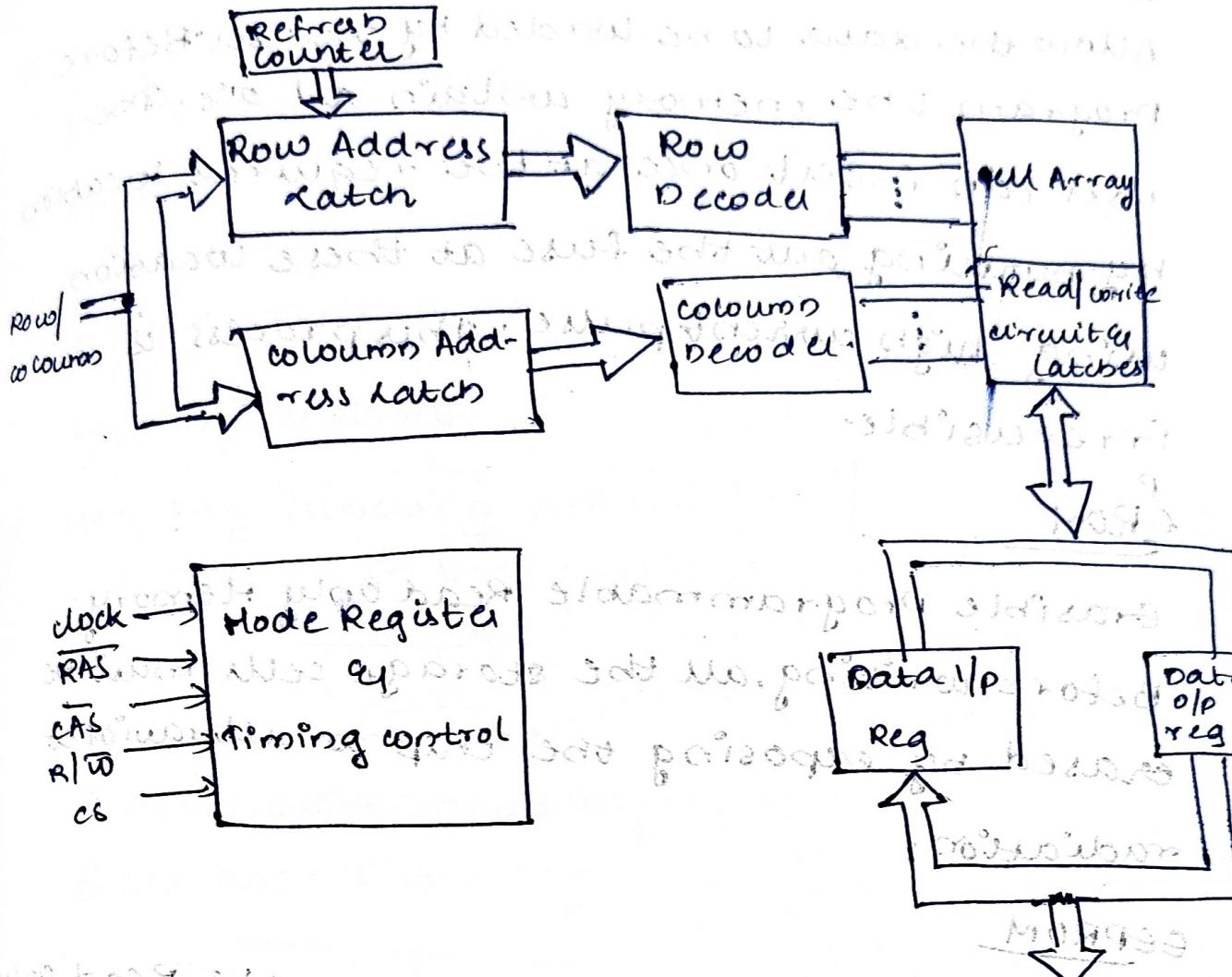
### Write

transistor is turned on, voltage is applied or removed on the bitline. DRAMs are also less expensive & simpler. It is used to build main memory.



The timing of memory device is controlled asynchronously i.e., not on a normal clock pulse. It needs an external control signal called RAS & CAS. Address lines are divided into two parts of multiplexed (i.e. A<sub>9</sub> to A<sub>20</sub>). The upper half of address is loaded into row address latch using RAS. The lower half of address is loaded into column address latch using column address strobe (CAS).

## Synchronous D-RAM



Operations is directly synchronised with processor clock signal. The output of the sense amplifiers are connected to a latches.

## ROM (Read Only Memory)

Non volatile involves reading of data.

Types:

- PROM
- GROM
- EEPROM

## PROM

Allow the data to be loaded by the user. Before programs, the memory contain all 0's. The user can insert one at the required location by burning out the fuse at these locations using high current pulse. This process is irreversible.

## EPROM

Erasable Programmable Read Only Memory.

Before writing, all the storage cells must be erased by exposing the chip to ultraviolet radiation.

## EEPROM

Electrically Erasable Programmable Read Only Memory is programmed & erased electrically. It can be erased & reprogrammed many times. Both erasing & programming takes 4 to 10 ms.

12/03/25

## Cache Memory

A cache is a smaller, faster memory located closer to a processor core which stores copies of the data from frequently used main memory locations.

Cache memory works based on the principle of locality of reference.

There are two types:- Temporal & spatial.

(i) Temporal : - Recently executed instructions are likely to be relatively executed very soon.

(ii) spatial: - Instruction in close proximity to  
already executed instruction are likely  
to be executed soon.

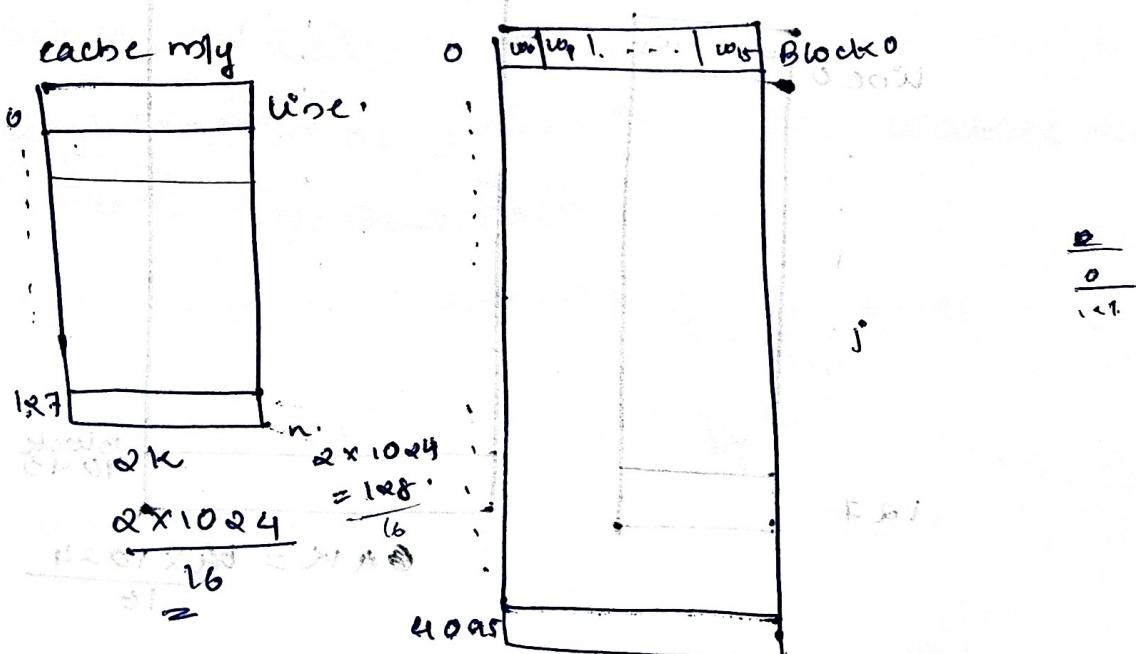
## Mapping Functions

Mapping function determine how memory blocks are placed in the cache.

~~Cache my~~

- ① Direct mapping
  - ② Associative mapping
  - ③ Set associative "

## Direct Mapping



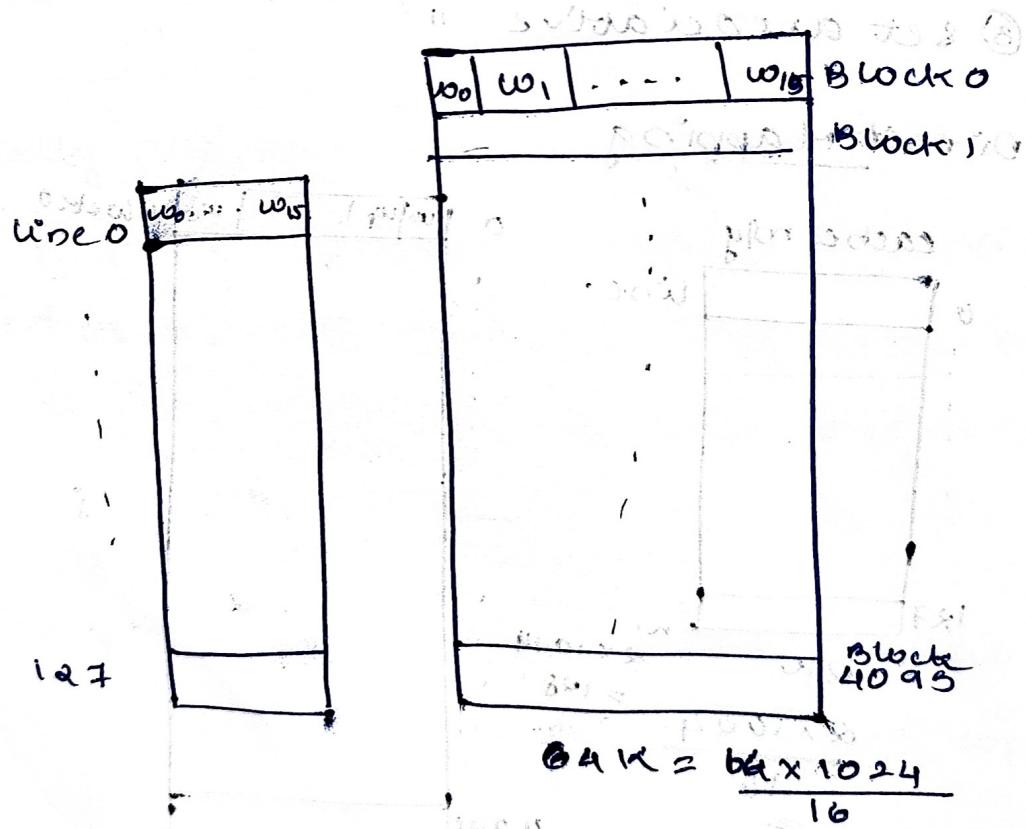
5%  $\approx$  0.05

$$64 \text{ kg} \\ 64 \times 1024 = 64096$$

## 8/15 Direct Mapping

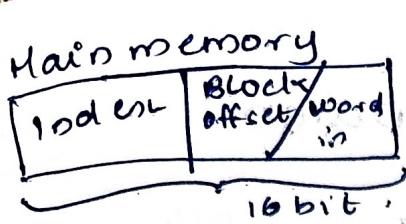
It is the simple & commonly used memory mapping technique where each block of main memory is mapped to exactly one location in cache.

- Index: It represents the block number. Index field bits tells us the location of block where a word can be.
- Block offset: It represents word in a memory block. These bit determine the location of the word in a memory block.



$$\text{Block Size} = \text{Line Size}$$

$$= 4096$$



thus tends to consolidate

Tag	line no.	block offset word size
-----	----------	---------------------------

cache my address bit.

problem. Second, there is a significant lack of data on the

3000 sq ft. 1000 sq ft. 3000 sq ft.

line no:- It represent cache line no.

**frag** :- Remaining part of the address which

$i \equiv j \pmod n$  and  $\exists q \in \mathbb{Z}$  so  $j = i + qn$

Associative mapping of entities

No restrictions. Block of main memory can reside any position in cache.

Set Associative mapping

Combination of direct & Associative mapping.

K-way set Association

Total no. of ~~set~~ = No. of lines

estimate requires much of your time.

By using direct mapping, i.e.  $k$  bits helps to identify the location in cache.