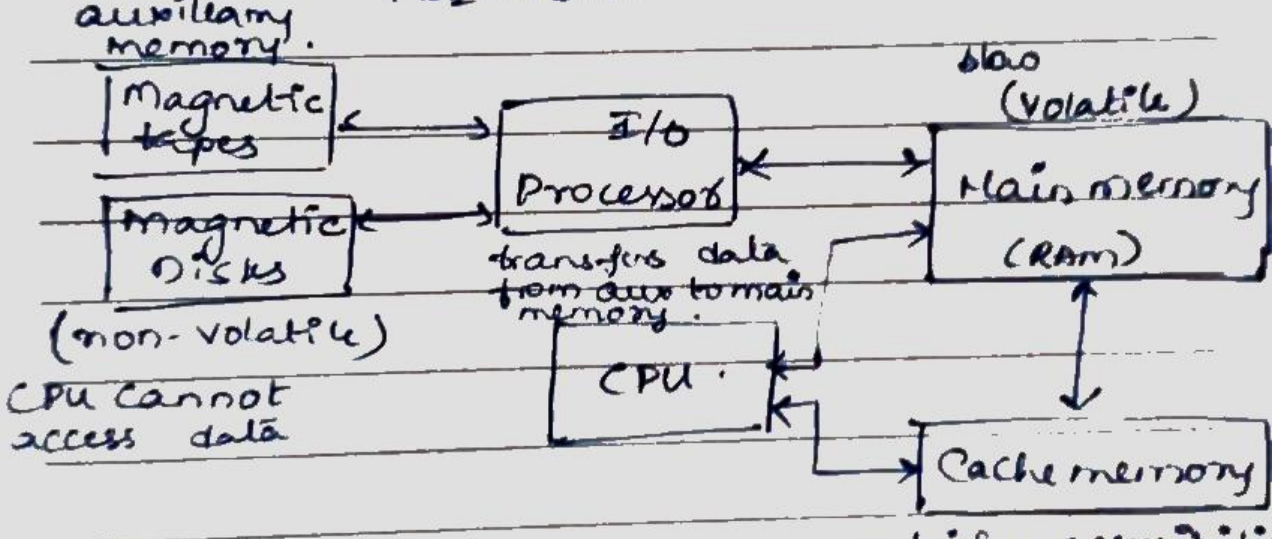
Memory Hierarchy -

Implement multiprogra
multiprocessi

* Memory is essential part in any system.
* Useful to store information in computer

Types of Memory :

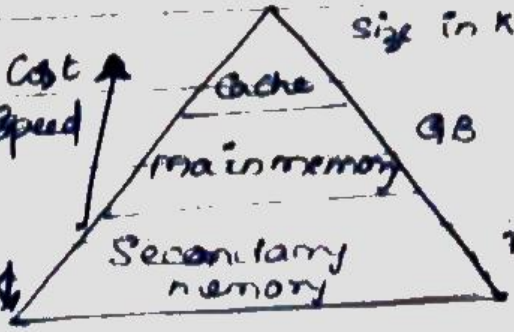
* Main memory - communicates with CPU
* Auxiliary " - (Secondary memory)
  Stores data permanent in computer.
  * Backup purpose.
* Cache memory
  ↓
  Store freq used instructions.



* higher accessibili
* faster Smaller

Properties :

1. Storage capacity
2. Cost per bit
3. Access Speed



as storage ↑ Cost ↓ Speed ↓

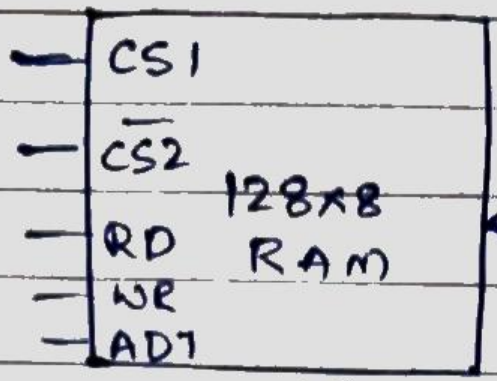
Main Memory

* Central storage unit in computer System
* Relatively large and fast memory
* Store prog and data during comp opern
* Major technology involved is based on Semiconductor Integrated Circuits

Integrated Ciruit RAM chips
- static (expensive)
  - f/f to store binary infm
  - extremely fast.
- Dynamic
  - Stores infm in form of electric charges applied to capacitors → provide with help of transistor
  - ⇒ Refresher
  - reduces power consumption.
  - Stores large data.

Primary Memory
- RAM (volatile)
- ROM (non-volatile)
  - ↓
  - Boot strap loader Pgm.

RAM Chip



RAM chip stores 128 words and each word stores ↔ 8 bits.

$128 = 7\ 2^7$

7 bits are required to store address

2 chip select

4 combination → we get

depending on no of RAM CS will be there or vice verca

eg; if we take 4 ram chips

$4 = 2^{2\rightarrow} = 2 \Rightarrow$ Chip selects

$8 = 2^3 = 3 \Rightarrow$ Chip selects

CS1	$\overline{CS2}$	RD	WR	memory function	databus
0	0	x	x	Inhibit	High Impedance
0	1	x	x	"	"
1	0	x	x	"	"
1	0	0	1	write	i/p data to RAM
1	0	1	0	Read	o/p data from RAM
1	1	x	x	Inhibit	high impedance

Associative Memory

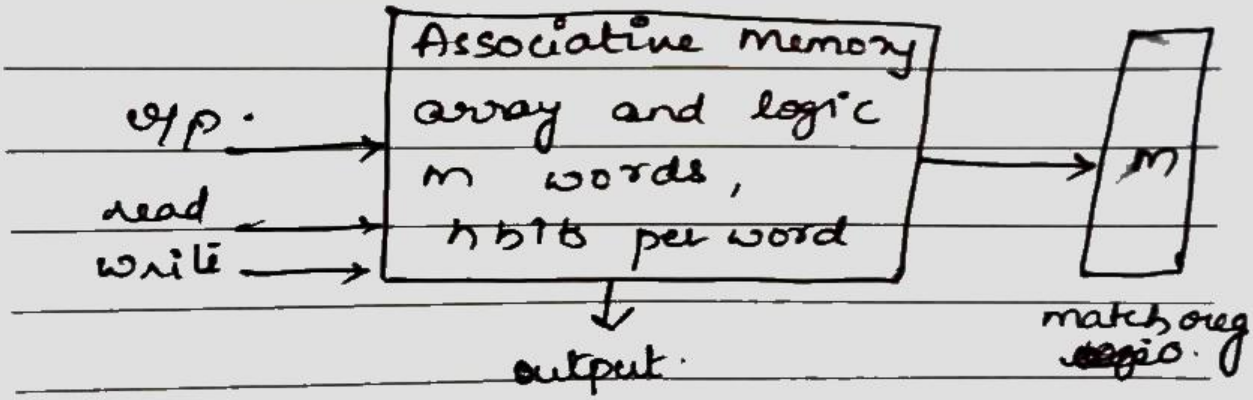
Also called as CAM (Content addressable memory)

CPU executes instr based on address - but " " " " " Content

Y AM? → because time required to search an item or execute an item can be substantially reduced

(A) Augment Register → Contains the content

(K) Key Register → Comparision purpose



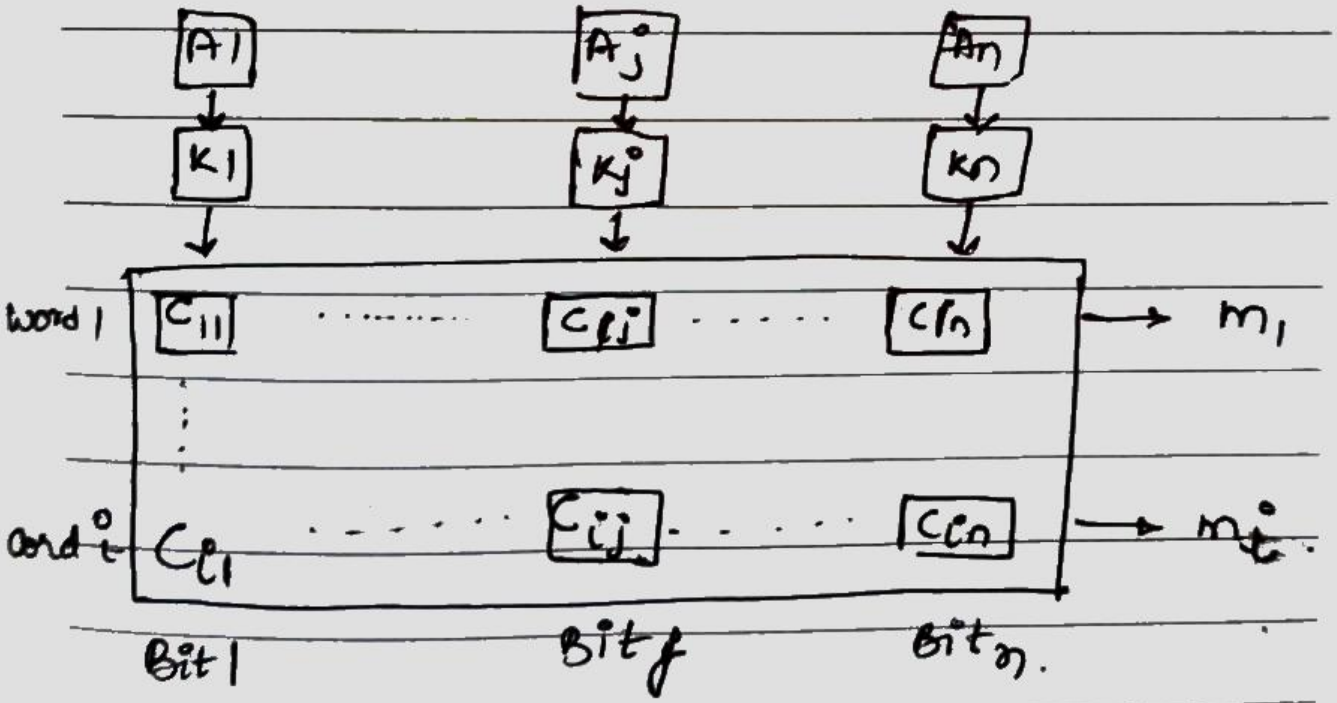
Note:

1. The content in augment reg is compared with data in associative memory only if the key register contains 1's.
   → If key regists contains 0's comparision is not done.

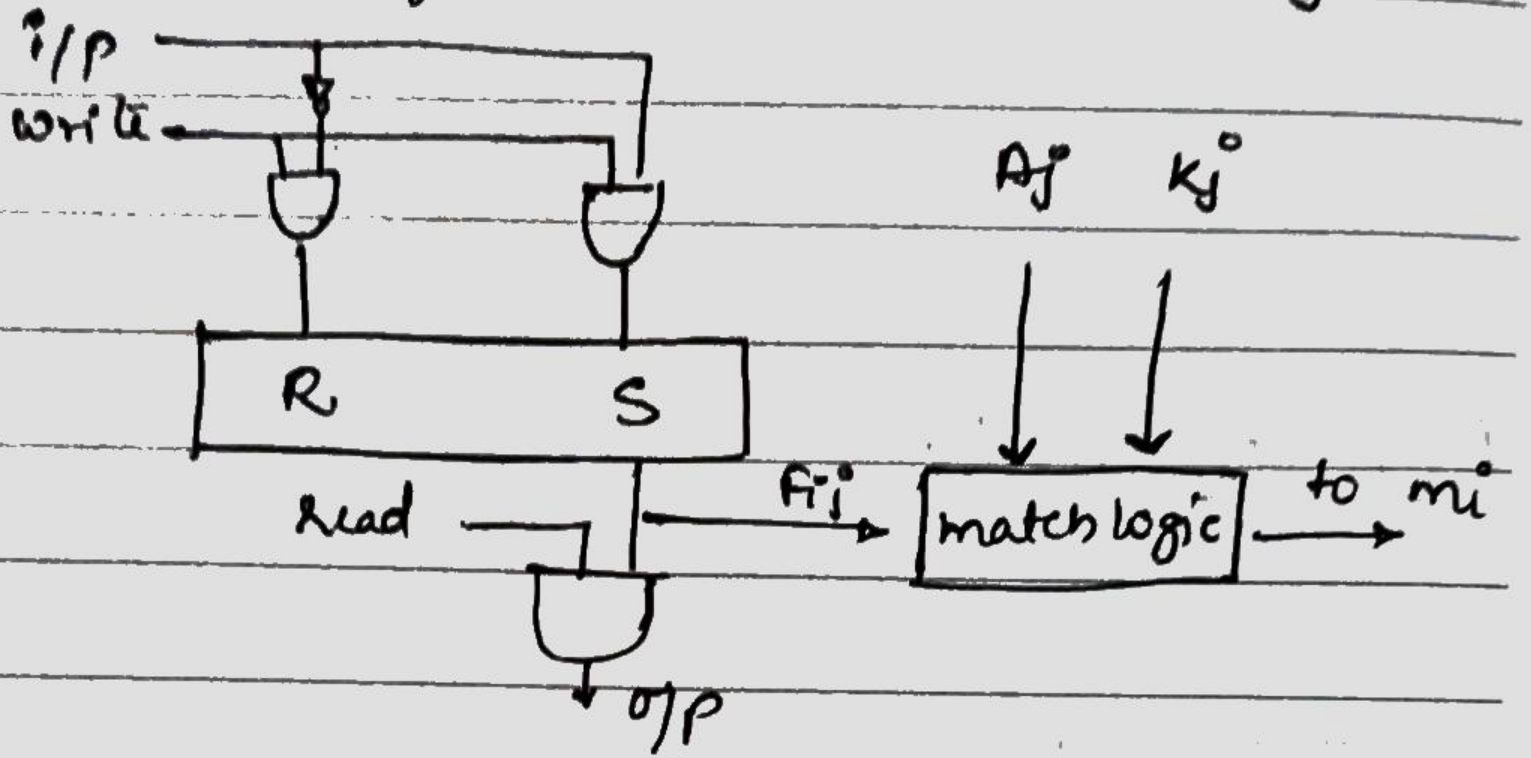
2. Size of match logic depends on the no. of words in associative memory.
for each word - match I/O has to be maintain

Eg; A 101 111100
K 111 000000

word 1 100 111100 no match
word 2 101 000001 match.



One cell of associative memory.



SR flf is used

S - set ⟶ accepts through and gate.

R - Reset

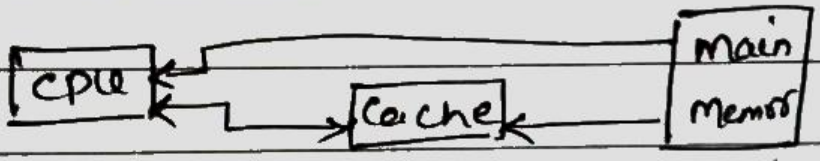
Cache Memory

- extremely fast
- high accessibility
- expensive
- Small in Size.

Property;- locality of refernce.
(freq executed instructions in the user prog)

∴ Avg memory access time is reduced.

placed in between CPU and main memory



performance can be analyzed by hit ratio

$$\text{hit ratio} = \frac{\text{no. of hits}}{\text{hits + miss}}$$

Mapping → Process through which the words in main memory are transferred to cache memory

3 types of mapping

1) Associative
2) Direct
3) set Associative

1. associative mapping:

assume CPU size is 32K x 12
(no of words ↑, each word contains ↓ no of bits)

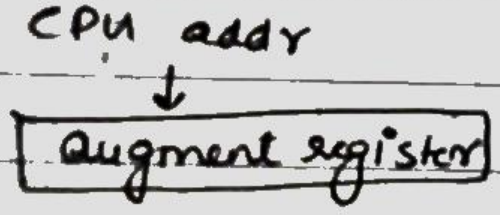
$2^5 \times 2^{10}$

$= 2^{15}$

15 → address generated by CPU.

Hence, Each address generated is stored in augment register.

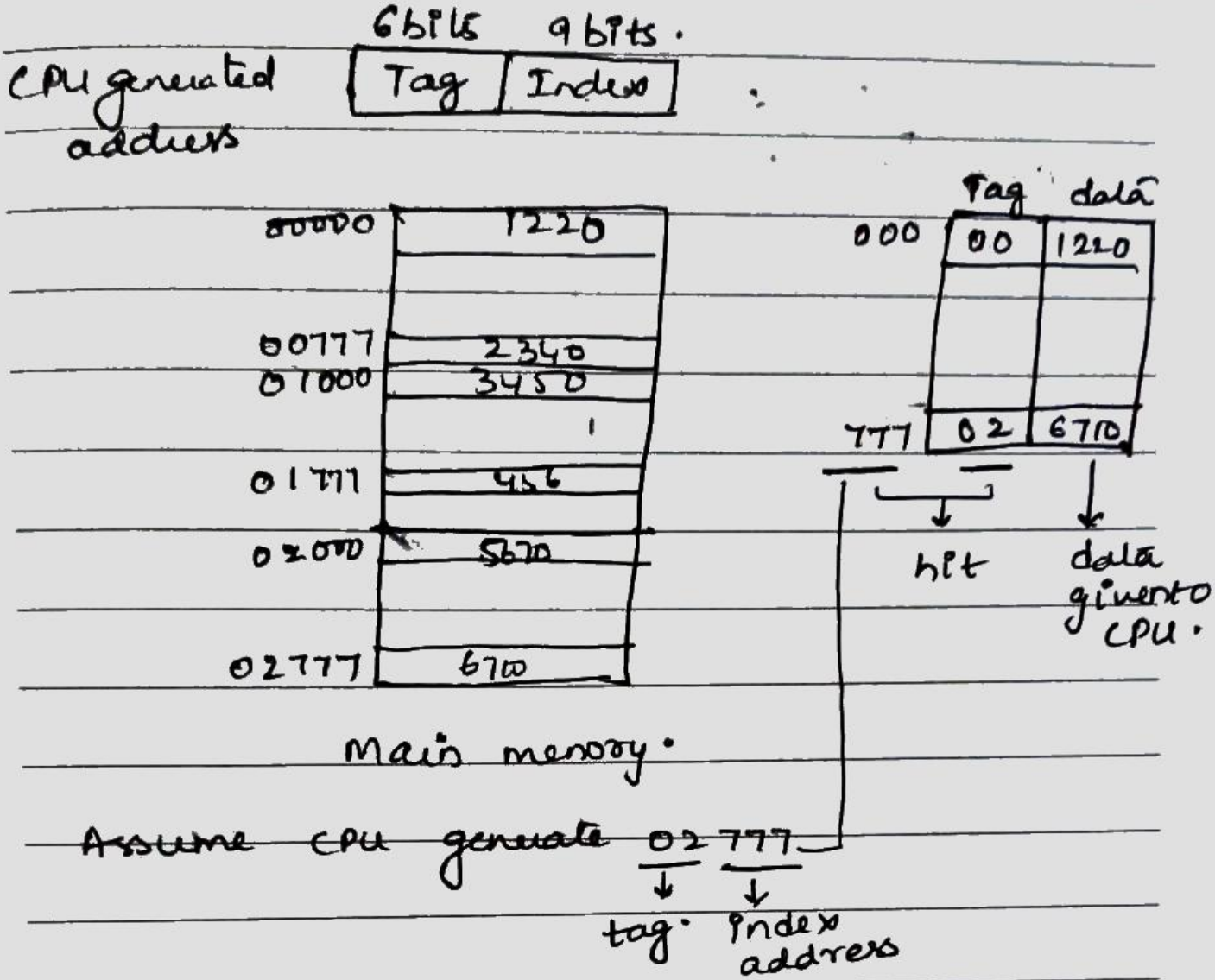
Cache memory size 512 x 12.



addr	data
01000	3450
00777	3710
23451	7734

address of augment reg is compared with associative mapping of cache and instructions are executed.

I. Direct mapping



Problem Several address can be generated by CPU with same index addr but different tags.

Hence, hit ratio is drastically ↓

To overcome this set associative mapping

(ii) set associative mapping

In single index, multiple words are stored.

	tag	data	tag	data
000	00	1220	01	1234
777	02	6710	04	4523

Auxillary Memory (Secondary memory)

↓ low cost
High storage capacity
low access speed

Eg; magnetic tapes, magnetic Disks (HDD), optical disks (CD).

* depends on magnetics, electronics, optical & electro mechanical systems

Access time = Seek time + Transfer time

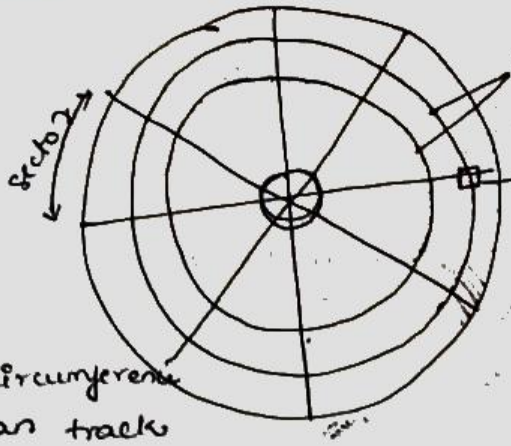
↓ position read/write head to location where data is residing

↓ time to transfer data to the destination

transfer rate = no. of characters/words that device can transfer per second

Magnetic disks

- made of circular plates (made with metal/plastic) and coated with magnetized material
- many disks are stacked on one spindle.
- read/write heads available on each surface.
- all disks rotate together at high speed
- Hence, bits are stored on magnetized surface.
- Concentric circles are called as tracks and these are divided into sectors



* track near circumference is longer than track near centre of disk.
* If bits are recorded with equal density, some tracks will have more bits than others.
* Hence, variable recording density => preferred

Read/Write Process:

1) Min quantity of info which can be transferred is a sector
2) Single read/write head used for each surface.
   ∴ When track address is given, read/write head is moved to that particular location (specified track).

   OR

   Separate read/write heads can be also maintained for each track on each surface. ∴ address bits can select particular track head.

Magnetic Tape

→ Strip of plastic coated with magnetic recording medium.
→ Bits are recorded as magnetic spots on tape along several tracks.
   1 character – 7/9 bits are used.
   multiple read/write heads are used to read data as a seqn of characters.
→ Information recorded in blocks are called as records (fixed / variable)
   • Each record has id (bit pattern).
   Hence addressed through the bit pattern.