COMPUTER SCIENCE

Computer Organization and Architecture

Memory Hierarchy

Cache Memory

Lecture_01

Vijay Agarwal sir

01 Memory Hierarchy

02 Cache Memory

1. Introduction of CoA.
2. Machine Instruction & Addressing Mode
3. Floating Point Representation
4. Micro operation & Control Unit
5. Pipelining & its Hazards.
6. Cache Memory
7. Disk & I/O org

Digital logic Subject

Flag

1. Carry Flag
2. Parity Flag
3. Auxilliary Carry Flag
4. Sign Flag
5. Zero Flag
6. Overflow Flag.

MUX

Decoder

2 Number → Multiplication
↳ Addition

CHIP select

∴

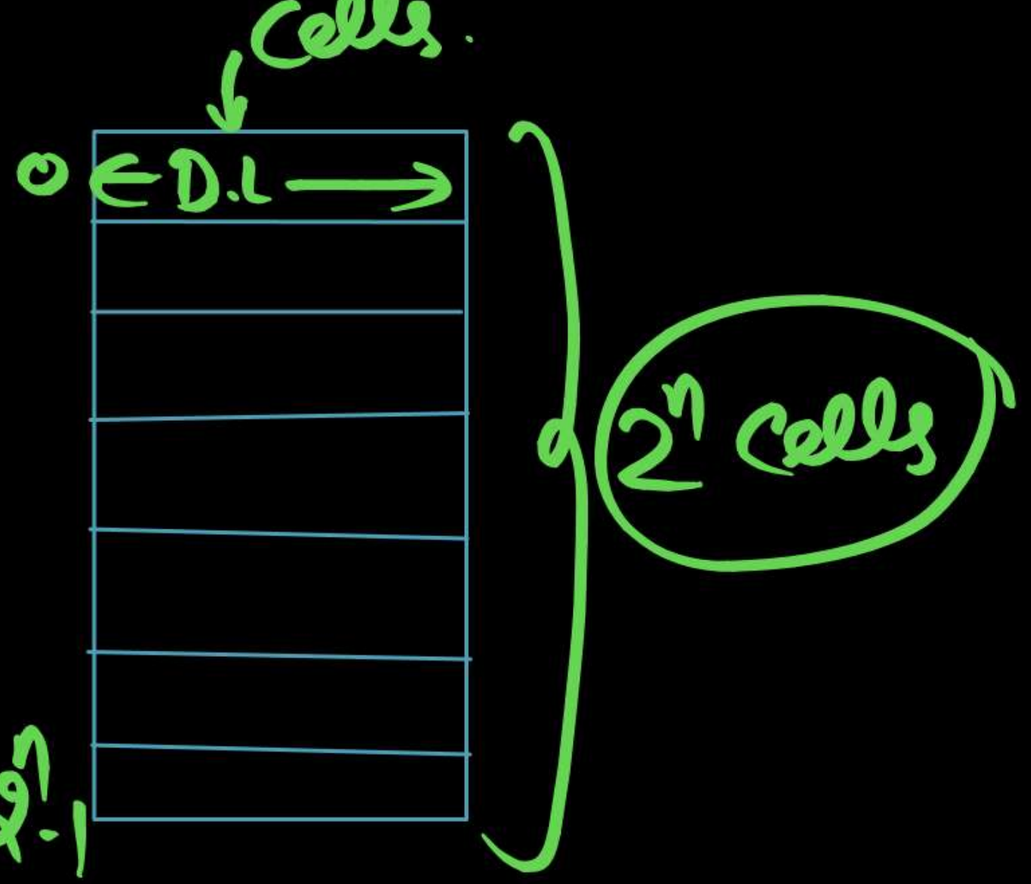
Memory

$2^n \times m$

n: Number of Address line [A.L]

m: Number of Data line [D.L]

- n bit A.L Can Represent 2^n Memory Cells. (0 to $2^n - 1$)

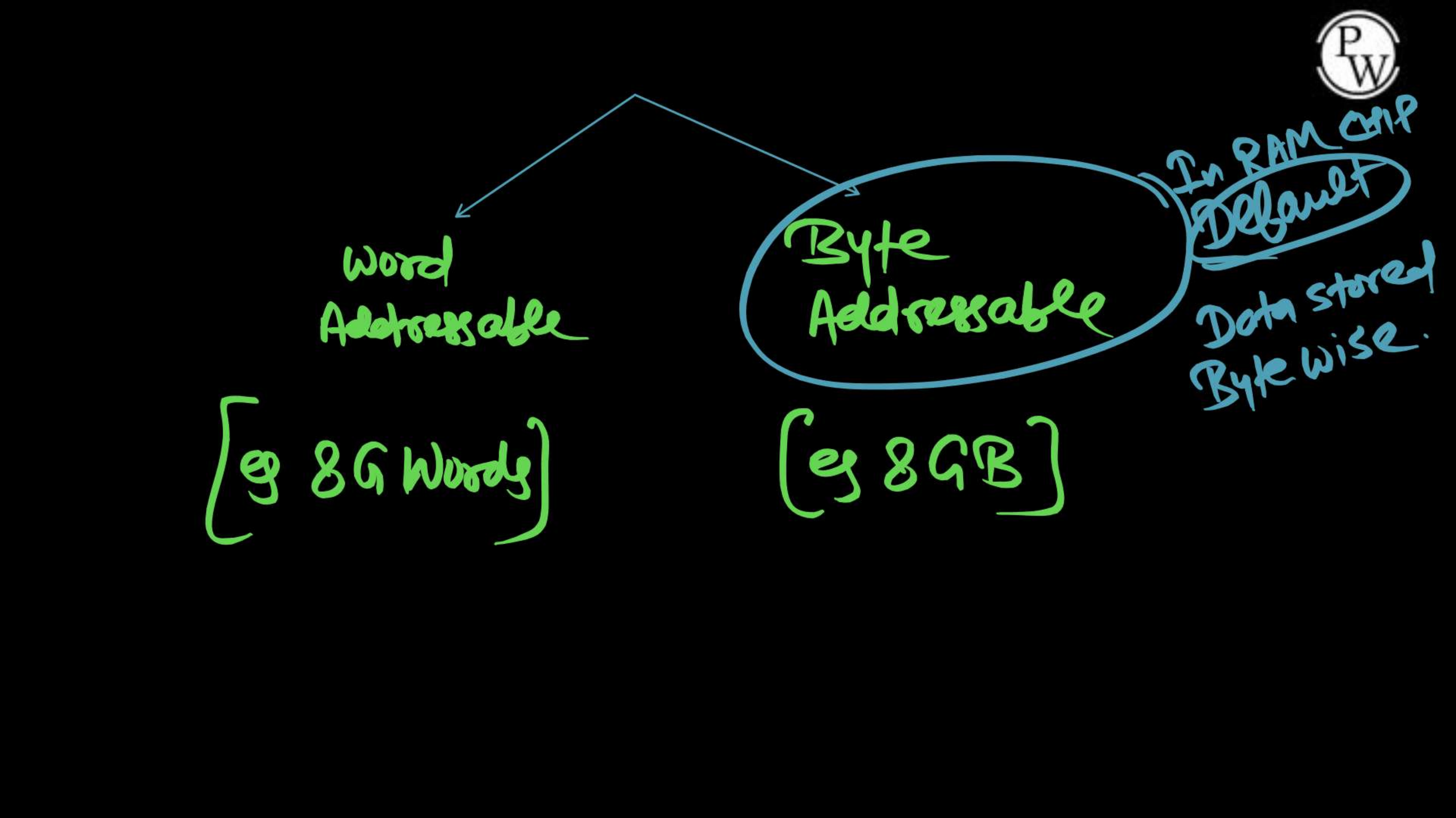


(Eg) Address line is 28 bit then what is the Memory Size ?

(i) 28 bit A.L $= 2^{28}$ Cells

Byte Addressable $= 2^{28}$ Byte

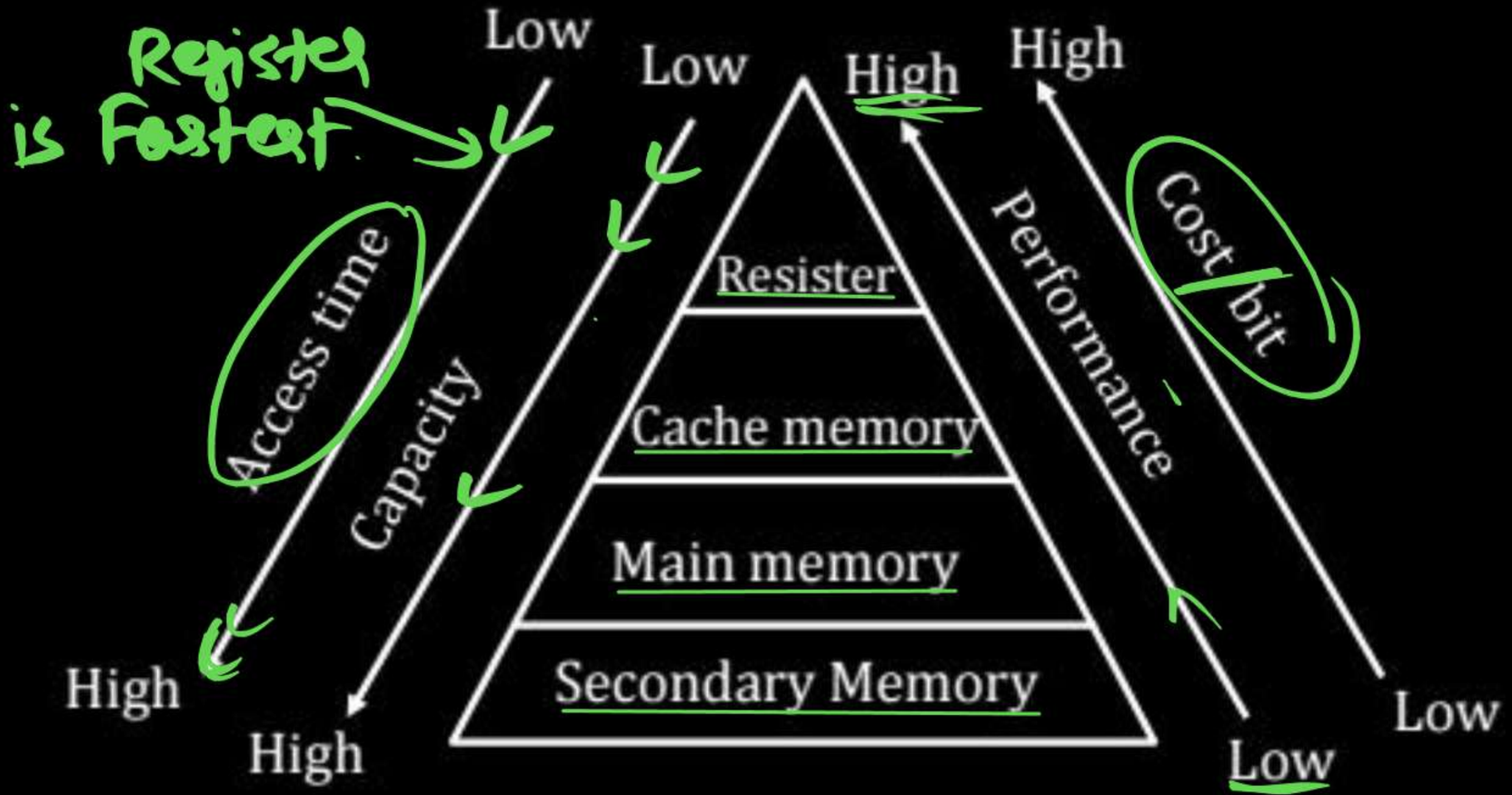
$\Rightarrow$ 256 MByte


Word Addressable
[eg 8G Words]
Byte Addressable
[eg 8GB]
In RAM comp
Default
Data stored Byte wise.

Cache Memory

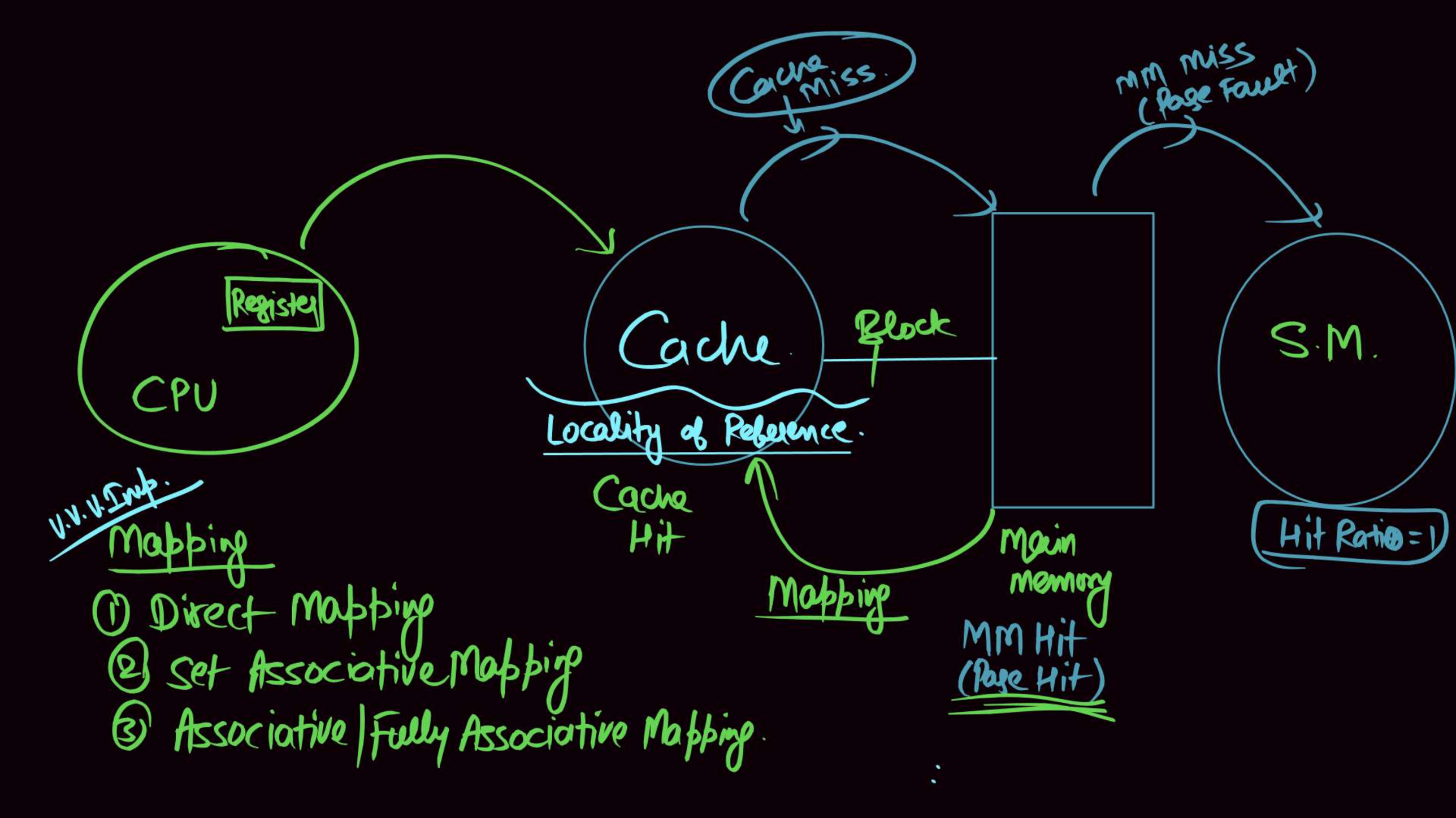
Memory Hierarchy

- Hierarchy design organize the system supported memory into 4 levels to minimize the Accessing times. They are:



Register Capacity is Low & its Performance is Fast.

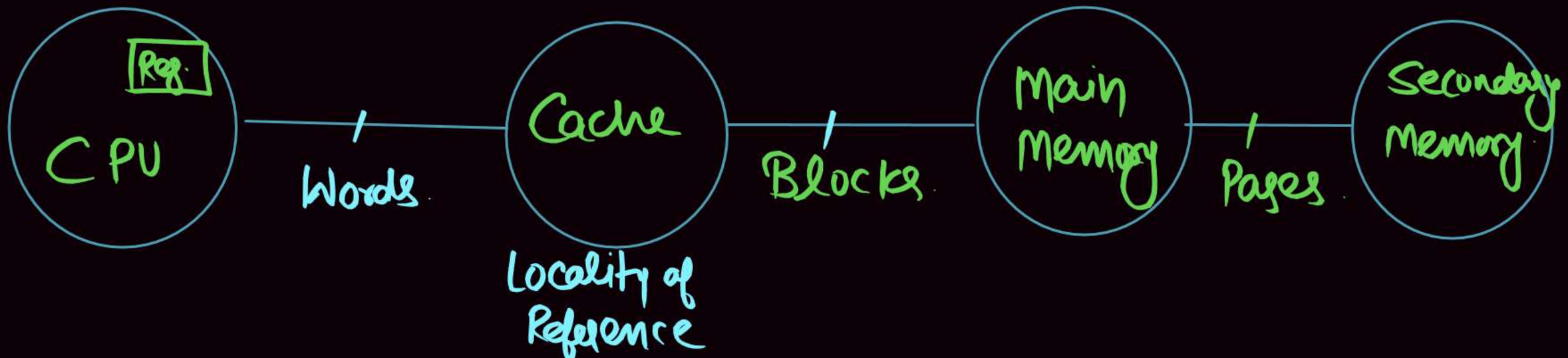
$$\text{Performance} \propto \frac{1}{\text{Time}}$$

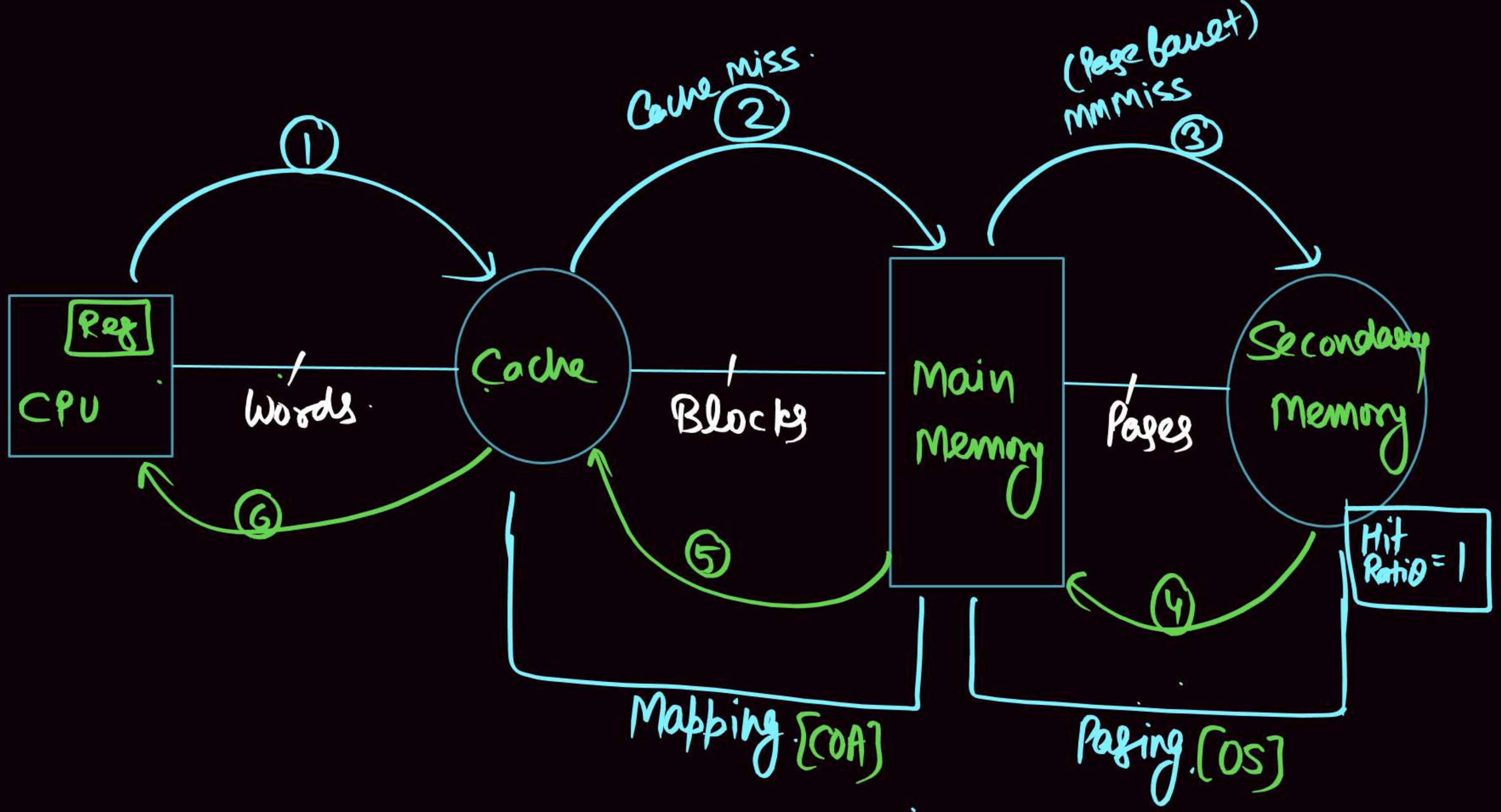

Cache Miss
MM Miss (Page Fault)
Register
CPU
Cache
Block
S.M.
Locality of Reference.
Cache Hit
V.V.V.Imp.
Mapping
Mapping
Main memory
MM Hit (Page Hit)
Hit Ratio = 1
① Direct Mapping
② Set Associative Mapping
③ Associative / Fully Associative Mapping.

$$\text{Hit Ratio} = \frac{\text{Number of Hit}}{\text{Total Number of Access.}}$$

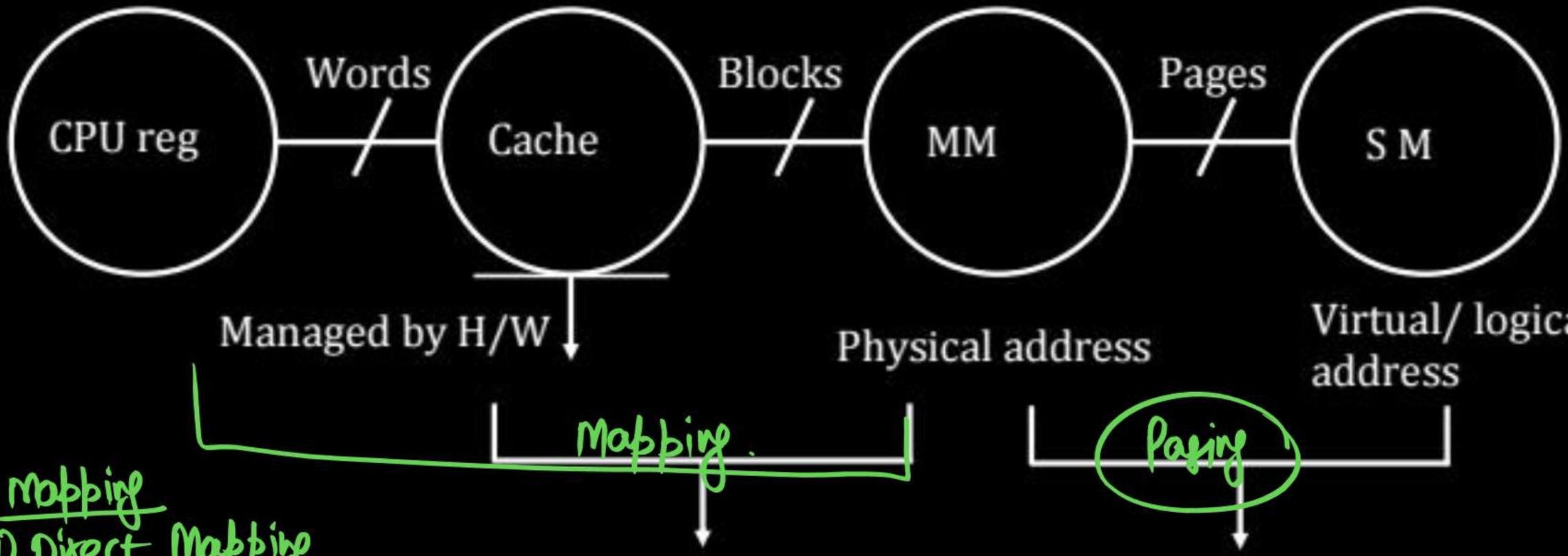
(e.g.) If Cache Hit Ratio 80%.

that means 80% Reference found in the Cache.


Reg.
CPU
Words
Cache
Locality of Reference
Blocks
Main Memory
Pages
Secondary Memory


Cache miss
(Page fault)
miss
1
2
3
Reg
CPU
Words
Cache
Blocks
Main Memory
Pages
Secondary Memory
6
5
4
Hit Ratio = 1
Mapping [COA]
Paging [OS]

Memory



Mapping

① Direct Mapping
② Set Associative Mapping
③ Fully Associative Mapping

Memory

- CPU generated Request initially Refer to the Cache.
- If the Reference [Respective Data] found in the Cache that is Called Cache Hit. [operation is called Hit] then Respective Data is given from Cache to CPU in the form of Words.
- If Reference Not found in the Cache Called Cache Miss, then Reference is Forwarded to Main Memory.

Memory

- If the Reference found in the Main Memory then its Called Main Memory Hit (OR) Page Hit, then Respective Data is transfered from Main Memory to Cache in the form of Blocks & Cache to CPU in the form of Words.

- If the Reference is Not found in the Main Memory then its Called MM Miss (OR) Page fault then Reference forward to Secondary Memory.
- Secondary Memory is the Last Level of Memory in which Hit Ratio always 1.

Memory

So Respective Data is transferred from Secondary Memory to Main Memory in the form of Pages, main memory to Cache Memory in the form of Block & Cache to CPU in the form of Words.

The Process of transfer the Data from Main Memory to Cache Memory is called Mapping.

Type of Memory Org

1. Simultaneous Access Memory Org.

2. Hierarchical Access Memory Org.

Type of Memory Org

1. Simultaneous Access Memory Org.

Type of Memory Org

2. Hierarchical Access Memory Org.

Q.

Calculate the average Access time with the cache access time 1ns, and main memory access time 100ns, Hit ratio 90%?
Using Hierarchical Access?

Q.

In a 2 level memory, level 1 memory is 5 times faster than level 2. and its access time is 10ns < Average Access Time. Let level 1 Access time is 20ns, What is the hit ratio? Using simultaneous Access org?

Q.

Consider a system with 2 levels. Level 1 Access time is 20ns Level 2 Access time = 150ns $T_{avg} = 30$ using simultaneous Access.

(i) What is the Hit Ratio?

(ii) If the Hit Ratio is mode to 100% then what is the Access time of L_1 & L_2 Memory?

Q. If the above Question if T_{avg} is increased by 10% then what is % of change in Hit Ratio?

Q.

Assume that for a certain processor, a read request takes 50 nanoseconds on a cache miss and 5 nanoseconds on a cache hit. Suppose while running a program, it was observed that 80% of the processor's read requests result in a cache hit. The average read access time in nanoseconds is _____. [GATE – 2015]

THANK YOU!