CS & IT ENGINEERING

COMPUTER ORGANIZATION AND ARCHITECTURE

Cache Organization

Lecture No.- 06

By- Vishvadeep Gothi sir

Recap of Previous Lecture

Topic Block Replacement

Topic Miss Penalty

Topic Types of Cache Miss

Topics to be Covered

Topic Cache Mapping Hardware

Topic Array Access with Cache

[NAT]

#Q. Consider a 2-way set associative cache with 256 blocks and uses LRU replacement, Initially the cache is empty. Conflict misses are those misses which occur due the contention of multiple blocks for the same cache set. Compulsory misses occur due to first time access to the block. The following sequence of accesses to memory blocks

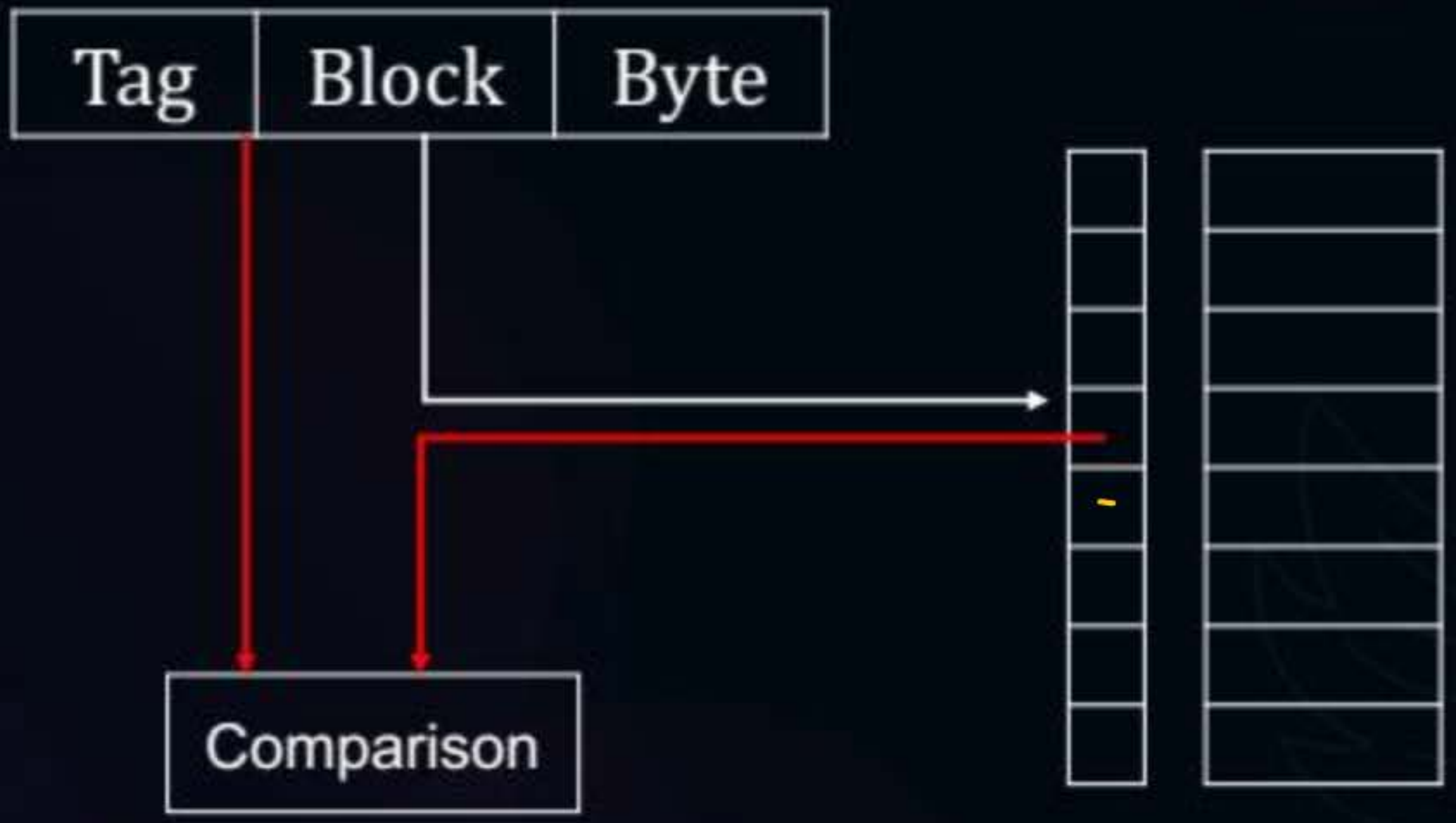
(0,128,256,128,0,128,256,128,1,129,257,129,1,129,257,129)

is repeated 10 times. The number of conflict misses experienced by the cache is ___76___?

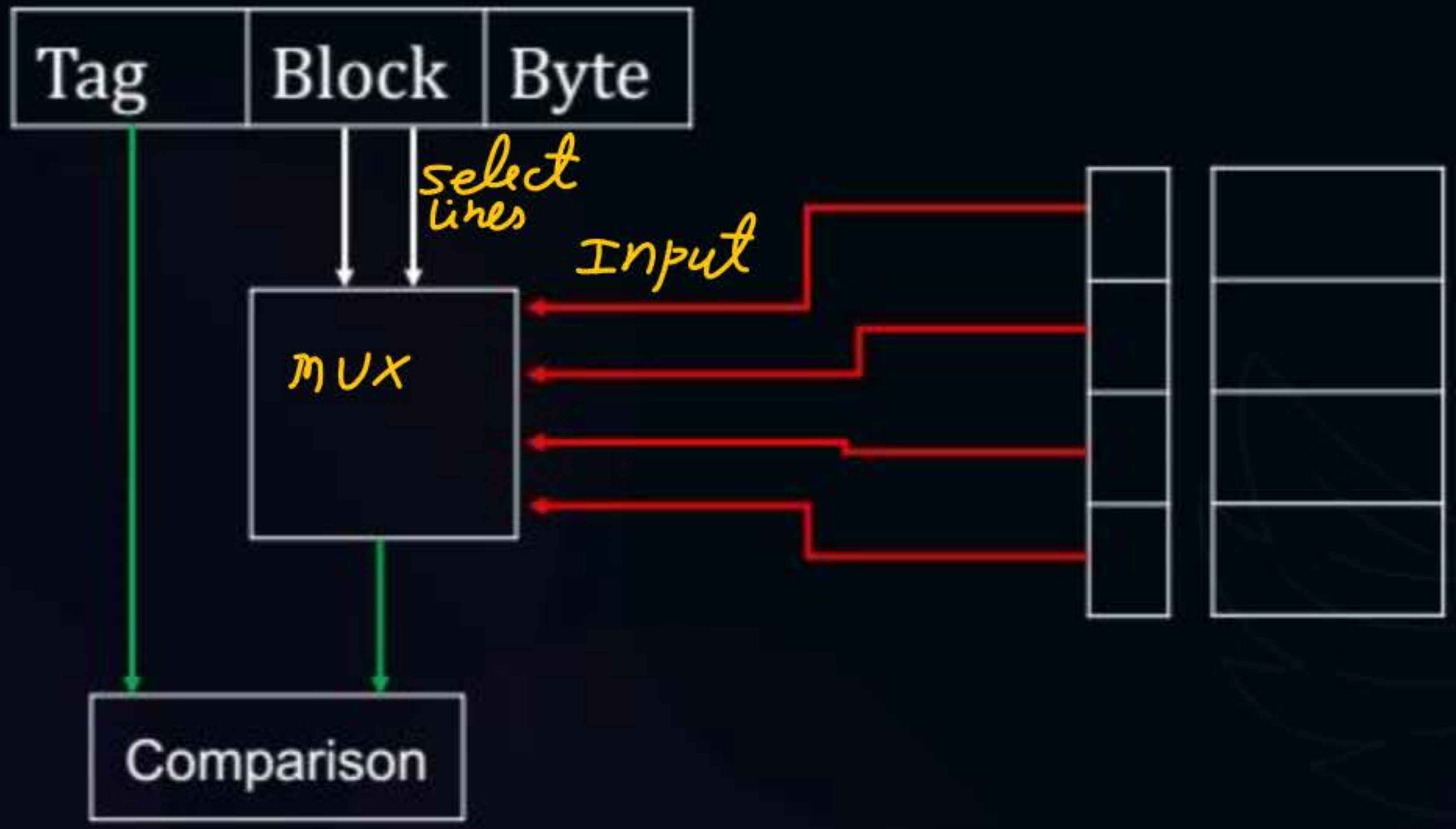
$$\frac{1^{st}}{4} \qquad \frac{2^{nd} - 9^{th}}{8 * 9}$$

Topic : Checking Hit/Miss in Direct Mapping

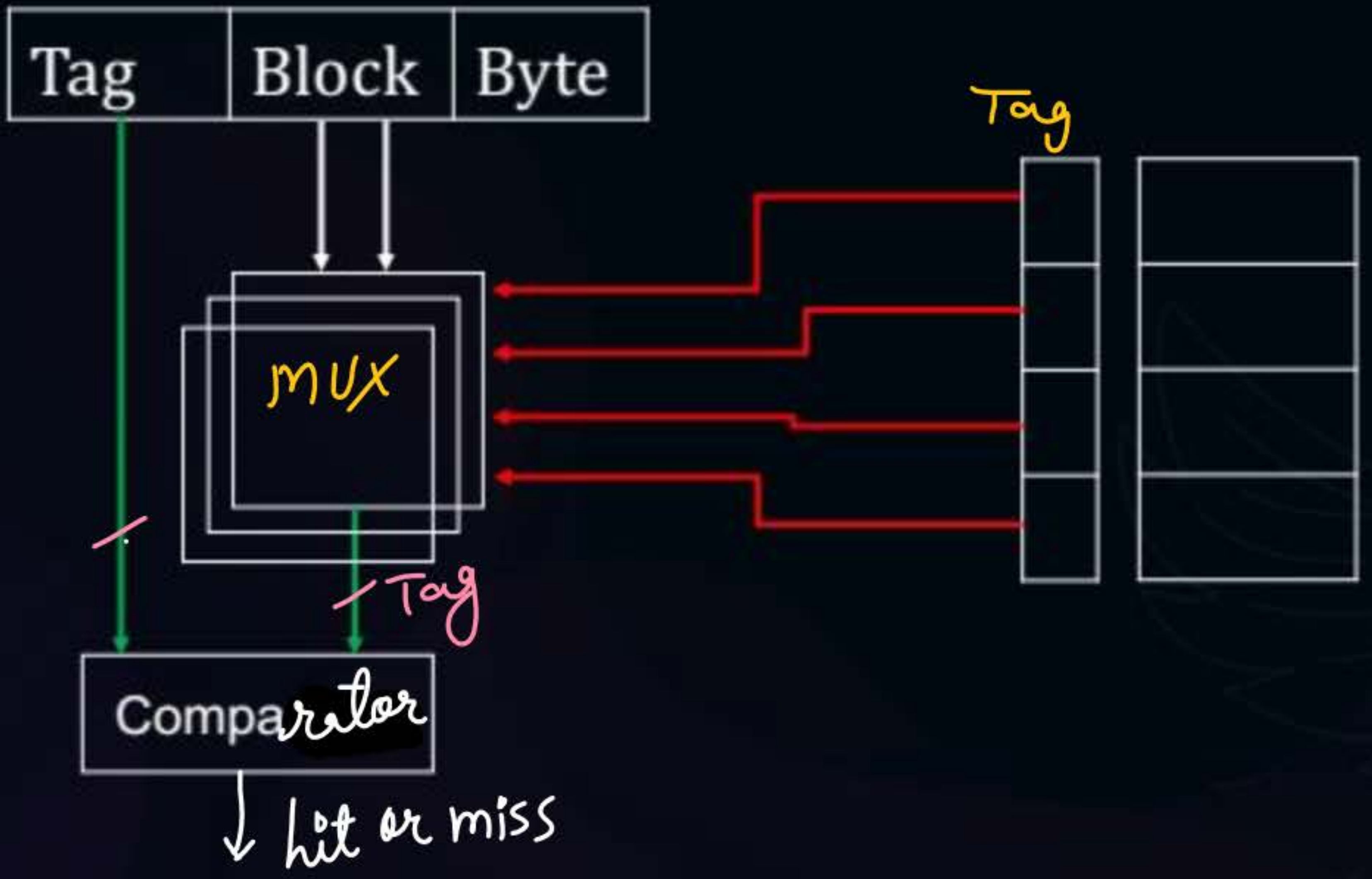
mm add.



Topic : Hardware Implementation in Direct Mapping



Topic : Hardware Implementation in Direct Mapping



Topic : Hardware Requirements

No. of MUX needed to extract tag = no. of bits in tag

size of MUX —— " ———— = No. of blocks in cache : 1

No. of comparators needed = 1

size of —— " ———— = (No. of bits in tag) – bits comparator

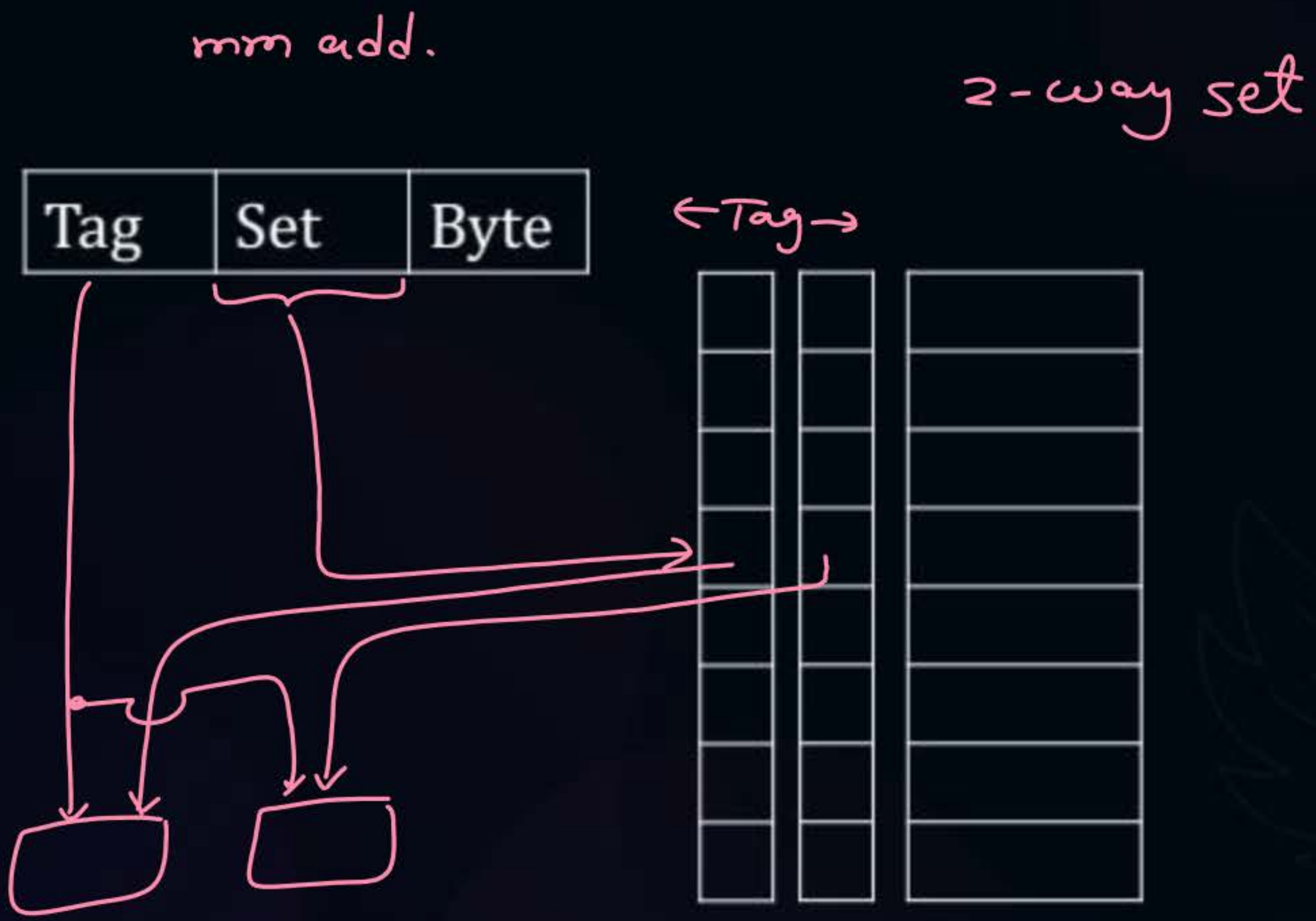
Topic : Hit Latency in Direct Mapping

↓

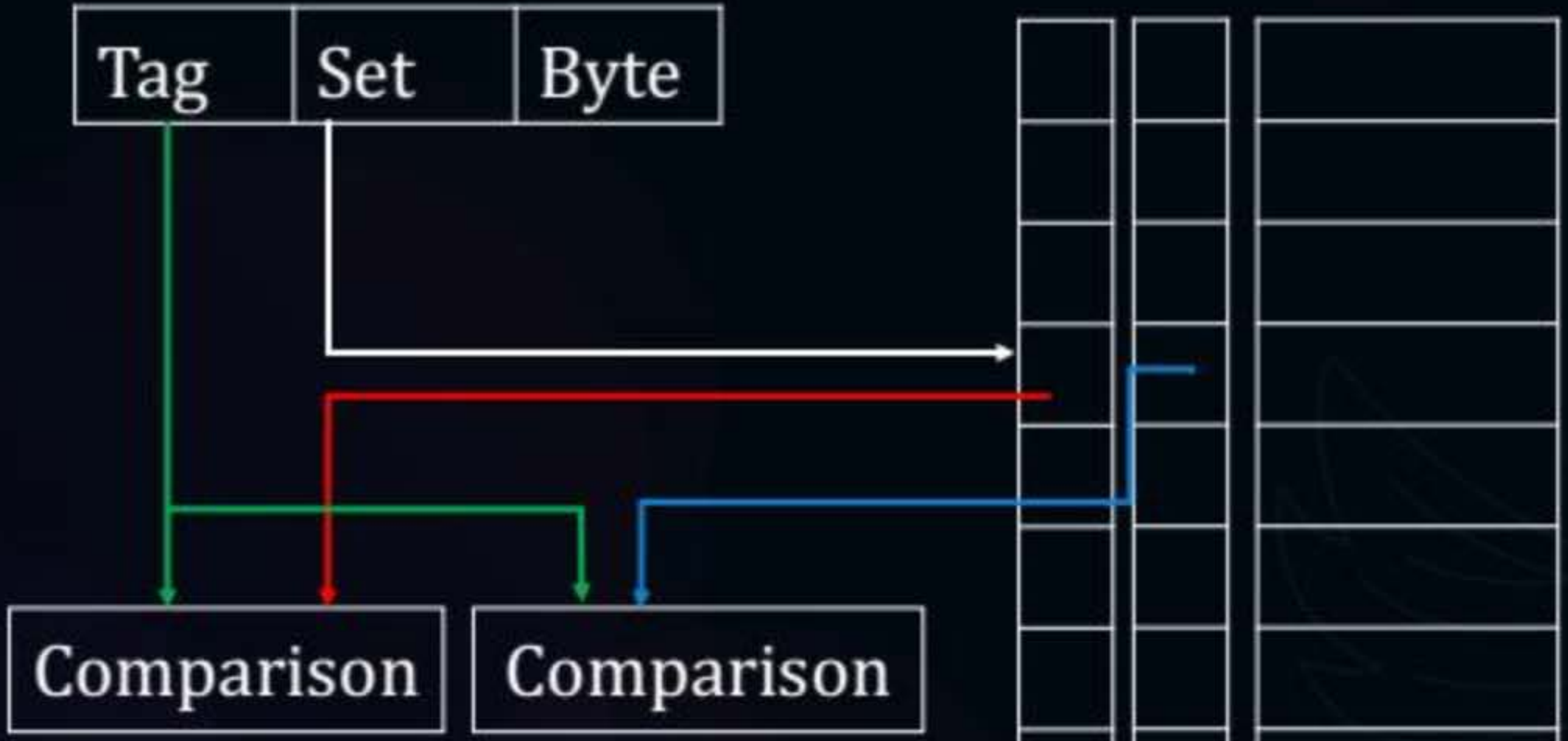
Time to check hit or miss

= MUX delay + comparator delay

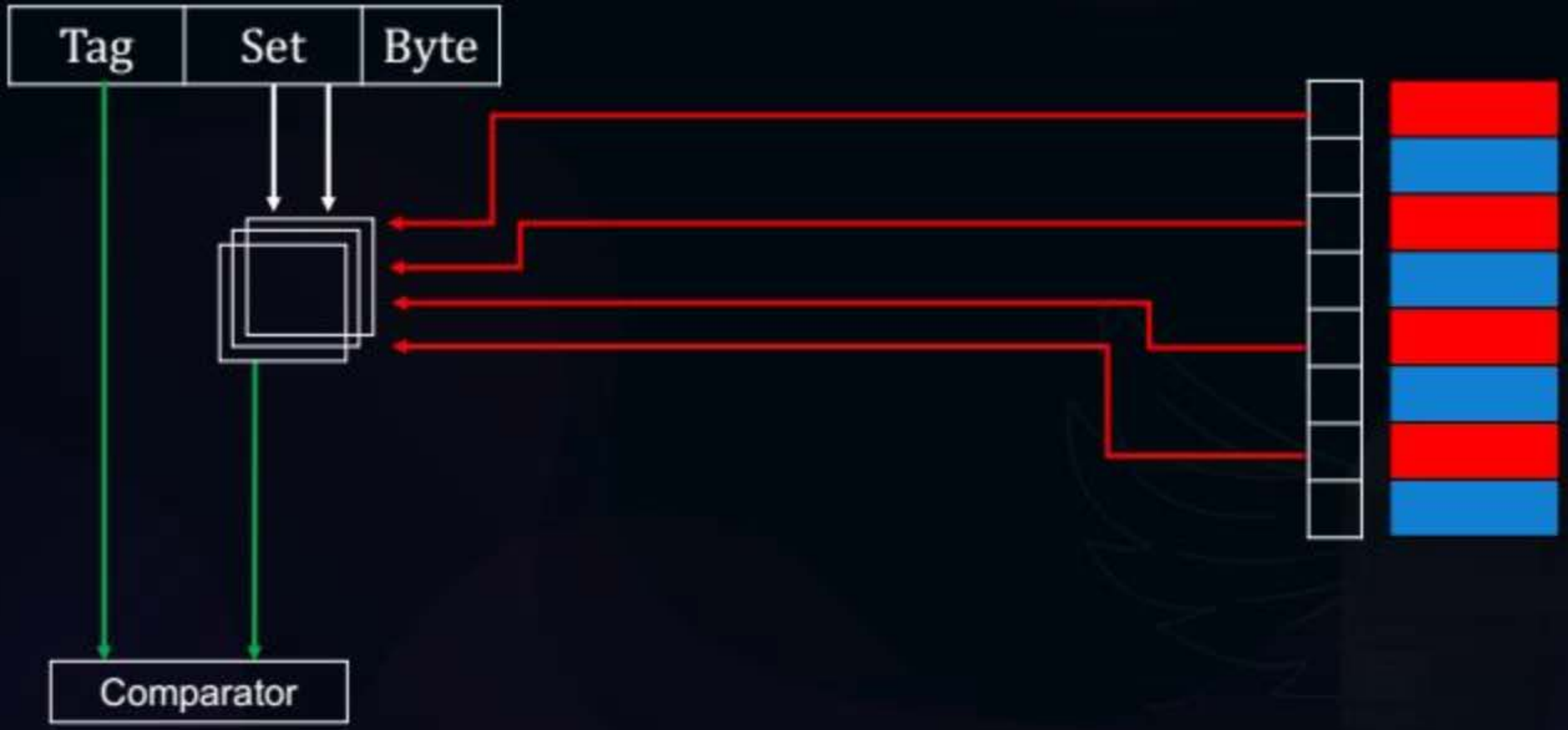
Topic : Checking Hit/Miss in Set Associative Mapping



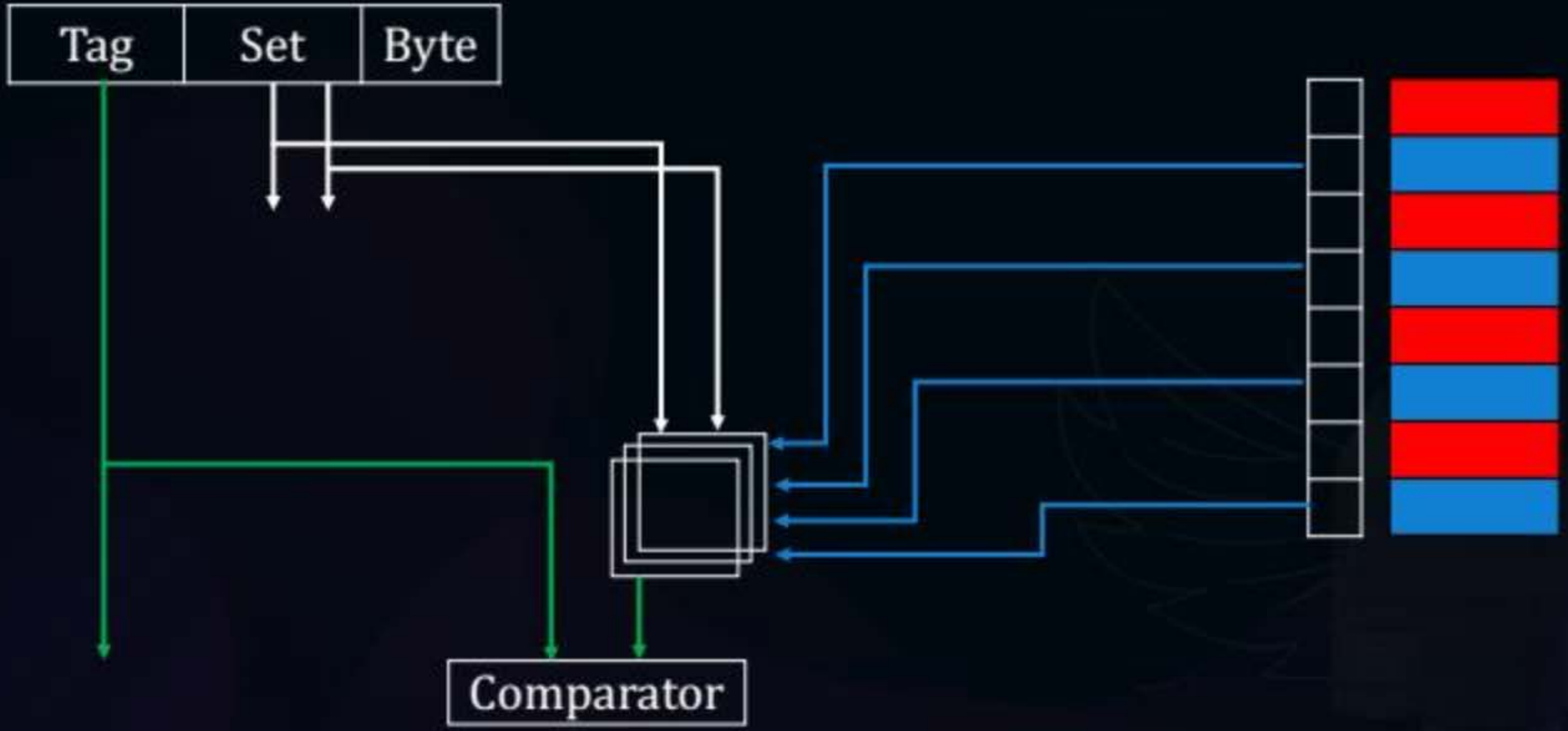
Topic : Checking Hit/Miss in Set Associative Mapping



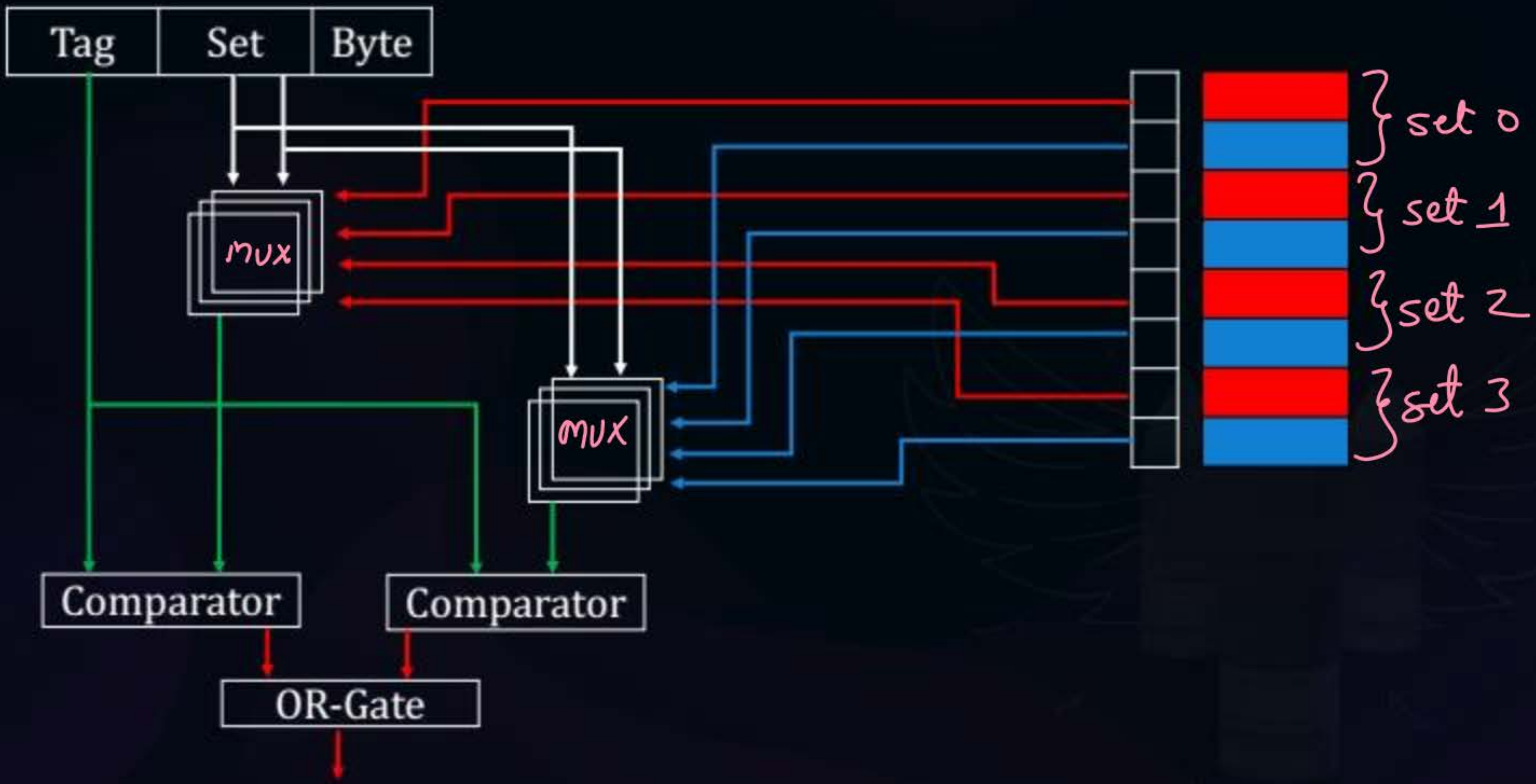
Topic : Hardware Implementation in Set Associative Mapping



Topic : Hardware Implementation in Set Associative Mapping



Topic : Hardware Implementation in Set Associative Mapping



Topic : Hardware Requirements

for k-way set ass. cache

No. of MUX for tag selection = k * (no. of bits in tag)

Size of —— 11 ———— = No. of sets in cache : 1

No. of comparators needed = k

size of —— 11 ———— = (no. of bits in tag) bits comparator

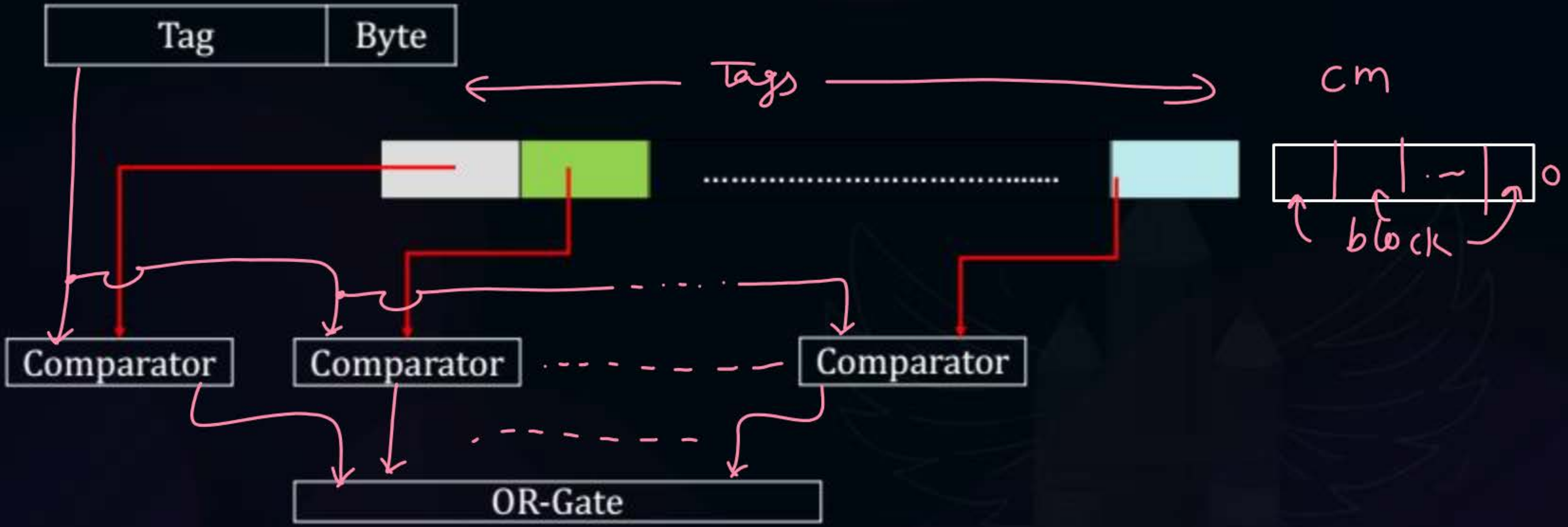
no. of OR gate = 1

no. of inputs = k

Topic : Hit Latency in Set Associative Mapping

= mux delay for tag selection + comparator delay + OR gate delay

Topic : Checking Hit/Miss in Fully Associative Mapping



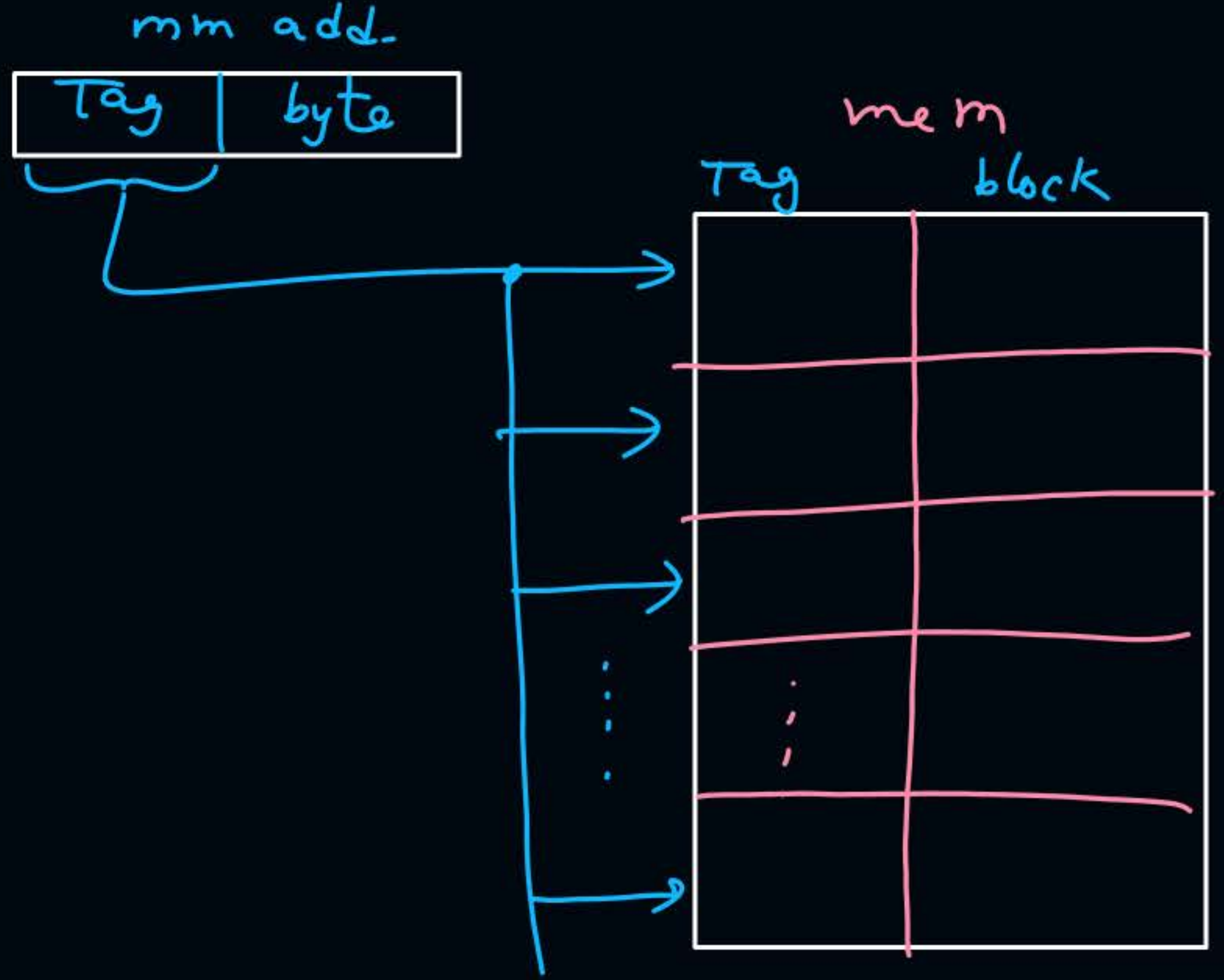
Topic : Hardware Requirements

No. of comparators = no. of blocks in cache

size —" ——— = (no. of bits in tag) – bits comparator

No. of OR gates = 1

no. of inputs in OR gate = no. of blocks in cache fully ass. cache is implemented using associative memory :-



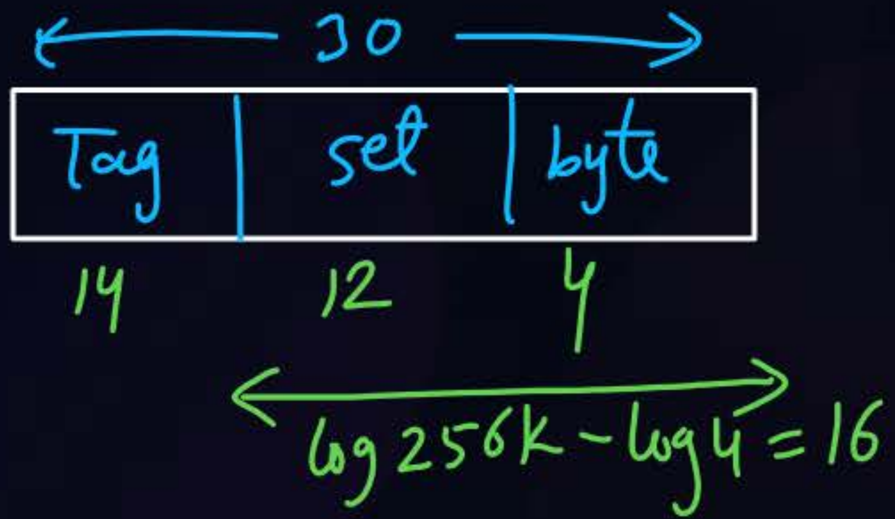
Topic : Hit Latency in Fully Associative Mapping

= comparator delay + OR gate

[NAT]

#Q. Consider a 4-way set associative cache with size 256Kbytes and block size of 16 bytes. The main memory address size is 30 bits. Find the following:

1. Tag bits in main memory address 14 bits
2. Number of MUX needed to extract the tag $= 4 * 14 = 56$
3. MUX size to extract the tag $2^{12} : 1$
4. Number of comparators needed 4
5. Comparator size 14 bits Comparator



[MCQ]

H.ω.

#Q. Consider two cache organizations. First one is 32 KB 2-way set associative with 32-bytes block size, the second is of same size but direct mapped. The size of an address is 32 bits in both cases. A 2-to-1 multiplexer has latency of 0.6 ns while a k-bit comparator has latency of $\frac{k}{10}$ns. The hit latency of the set associative organization is h_1 while that of direct mapped is h_2.

The value of h_1 is:

A 2.4 ns

B 2.3 ns

C 1.8 ns

D 1.7 ns

[MCQ]

H.W.

#Q. Consider two cache organizations. First one is 32 KB 2-way set associative with 32-bytes block size, the second is of same size but direct mapped. The size of an address is 32 bits in both cases. A 2-to-1 multiplexer has latency of 0.6 ns while a k-bit comparator has latency of $\frac{k}{10}$ns. The hit latency of the set associative organization is h_1 while that of direct mapped is h_2.

The value of h_2 is:

A 2.4 ns

B 2.3 ns

C 1.8 ns

D 1.7 ns

Topic : Questions on Cache and Array

- Cache Size = 32 bytes
- Block size = 4 bytes

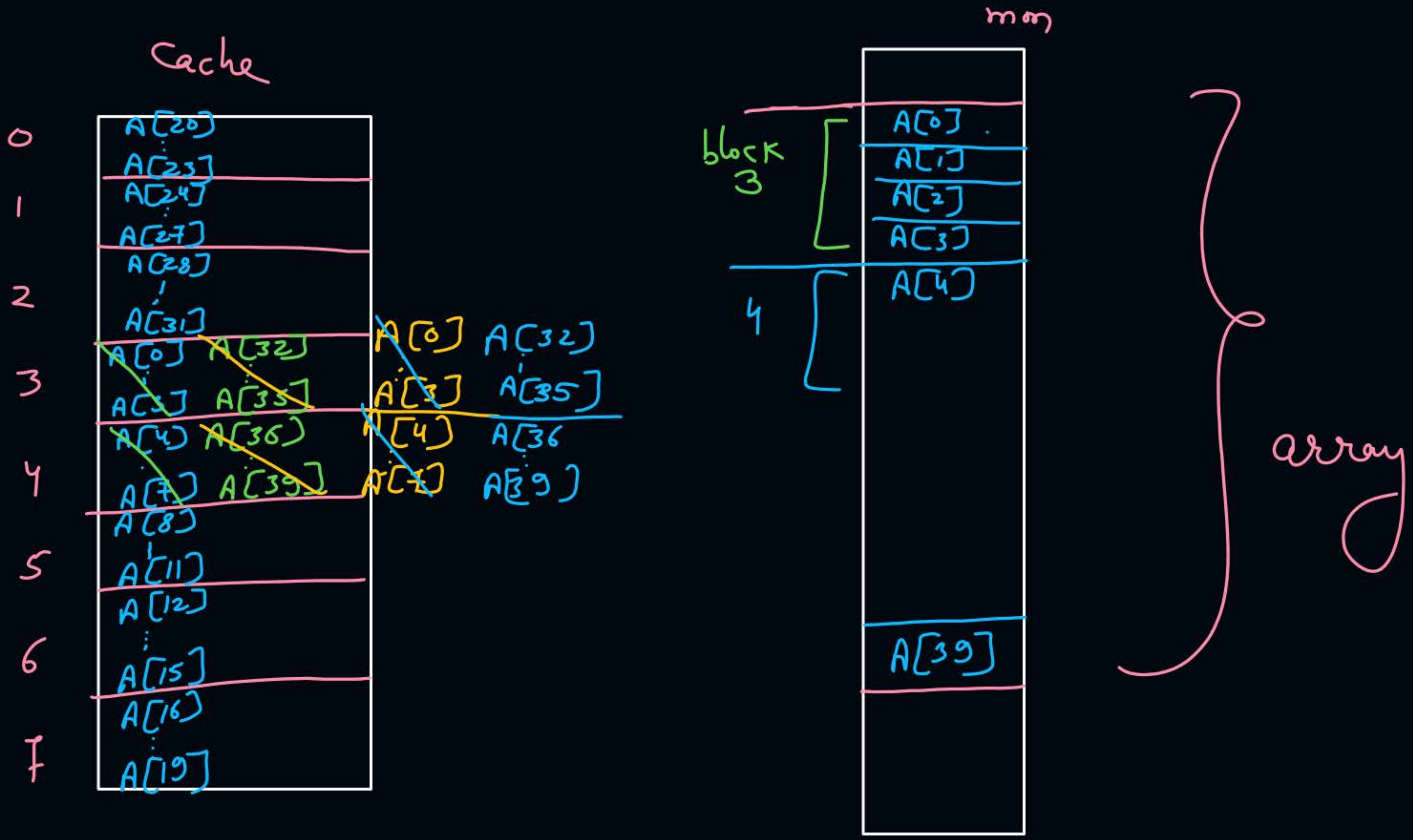
$\Big\}$ no. of blocks in cache $= \frac{32B}{4B} = 8$

- Array in main memory [redacted] A[40], each element is 1 bytes
- Direct mapping

array size $= 40 * 1B = 40B$

no. of blocks needed to store array in mm $= \frac{40B}{4B} = 10$

no. of Array elements per block $= \frac{4B}{1B} = 4$


Cache
0
A[20]
A[23]
1
A[24]
A[27]
2
A[28]
A[31]
3
A[0]
A[3]
A[32]
A[35]
4
A[4]
A[7]
A[36]
A[39]
A[0]
A[3]
A[32]
A[35]
A[4]
A[7]
A[36
A[39]
5
A[8]
A[11]
6
A[12]
A[15]
7
A[16]
A[19]
mem
block
3
4
A[0]
A[1]
A[2]
A[3]
A[4]
A[39]
array

CPU access	hit/miss
A[0]	miss
A[1]	hit
A[2]	hit
A[3]	hit
A[4]	miss
A[5] A[6] A[7]	hit
⋮	
A[32]	miss
A[33] -- A[35]	hit
A[36]	miss
A[37] .- A[39]	hit

2nd time access	
A[0]	miss
A[4]	miss
⋮	
A[32]	miss
A[36]	miss

no. of extra blocks in array = 10 - 8 = 2

1st time access of array	2nd time access of array	3rd time access
no. of miss = no. of blocks to store array	= 2 * 2	
= 10	= 4	= 4
no. of hits = array elements - 10	hits = 40 - 4	= 36
= 40 - 10	= 36	
= 30		

Ques) no. of blocks to store array = 21

no. of blocks in cache = 16

no. of extra blocks in array = 21 − 16

= 5

	first access of array	2nd access	3rd access	- - -
no. of miss	21	5*2=10	5*2=10	

Topic : Questions on Cache and Array

- Direct mapping
- Cache Size = 64 bytes
- Block size = 4 bytes

} no. of blocks in cache $= \frac{64B}{4B} = 16$

- Array in main memory int A[40], each element is 2 bytes

no. of hits = $\underline{84}$?

no of miss = $\underline{36}$?

If array is accessed 3 times

hit ratio $= \frac{84}{(36+84)} = 0.7 = 70\%$

miss ratio $= 30\%$

array size $= 40 * 2 = 80B$

no. of blocks in array $= \frac{80B}{4B} = 20$

extra blocks in array $= 20 - 16 = 4$

no. of miss $= 20 + (4*2) + (4*2) = 36$

no. of hits $= (40-20) + (40-8) + (40-8) = 84$

[NAT]

#Q. Consider a machine with a byte addressable main memory of 2^{16} bytes. Assume that a direct mapped data cache consisting of 32 lines of 64 bytes each is used in the system. A 50 × 50 two-dimensional array of bytes is stored in the main memory starting from memory location 1100H. Assume that the data cache is initially empty. The complete array is accessed twice. Assume that the contents of the data cache do not change in between the two accesses.

How many data cache misses will occur in total?

array size $= 50 * 50 = 2500$ elements $= 2500 * 1B = 2500B$

no. of blocks to store array $= \left\lceil \frac{2500B}{64B} \right\rceil = 40$

no. of extra blocks in array = 40 - 32 = 8

no. of miss = 40 + (8 * 2)

= 56

[MCQ]

#Q. Which of the following lines of the data cache will be replaced by new blocks in accessing the array for the second time?

[✓ A] line 4 to line 11

B line 4 to line 12

C line 0 to line 7

D line 0 to line 8

← 16 →

Tag	cm block no.	byte
5	5	6

$(1100)_{16}$

00010	00100	000000

cm block no. $= (00100)_2 = (4)_{10}$

from 4 to 11 (8 blocks replaced)

[NAT]

H.W.

#Q. Consider a direct mapped write back data cache of size 2KB with block size of 128 bytes. The cache is considered to be empty initially. The byte addressable main memory has size 1Mbytes. Further consider that there is an array A[30][25] with each element occupies 4 bytes. The base address of array is $(1A300)_{16}$. The array is accessed 4 times. And between the accesses, there is no any data cache changes happen. Hit ratio of cache for the array access is ____ %?

2 mins Summary

Topic Cache Mapping Hardware

Topic Array Access with Cache

Happy Learning

THANK - YOU