CS & IT ENGINEERING

COMPUTER ORGANIZATION AND ARCHITECTURE

Cache Organization

Lecture No.- 09

By- Vishvadeep Gothi sir

Recap of Previous Lecture

Topic Block Replacement

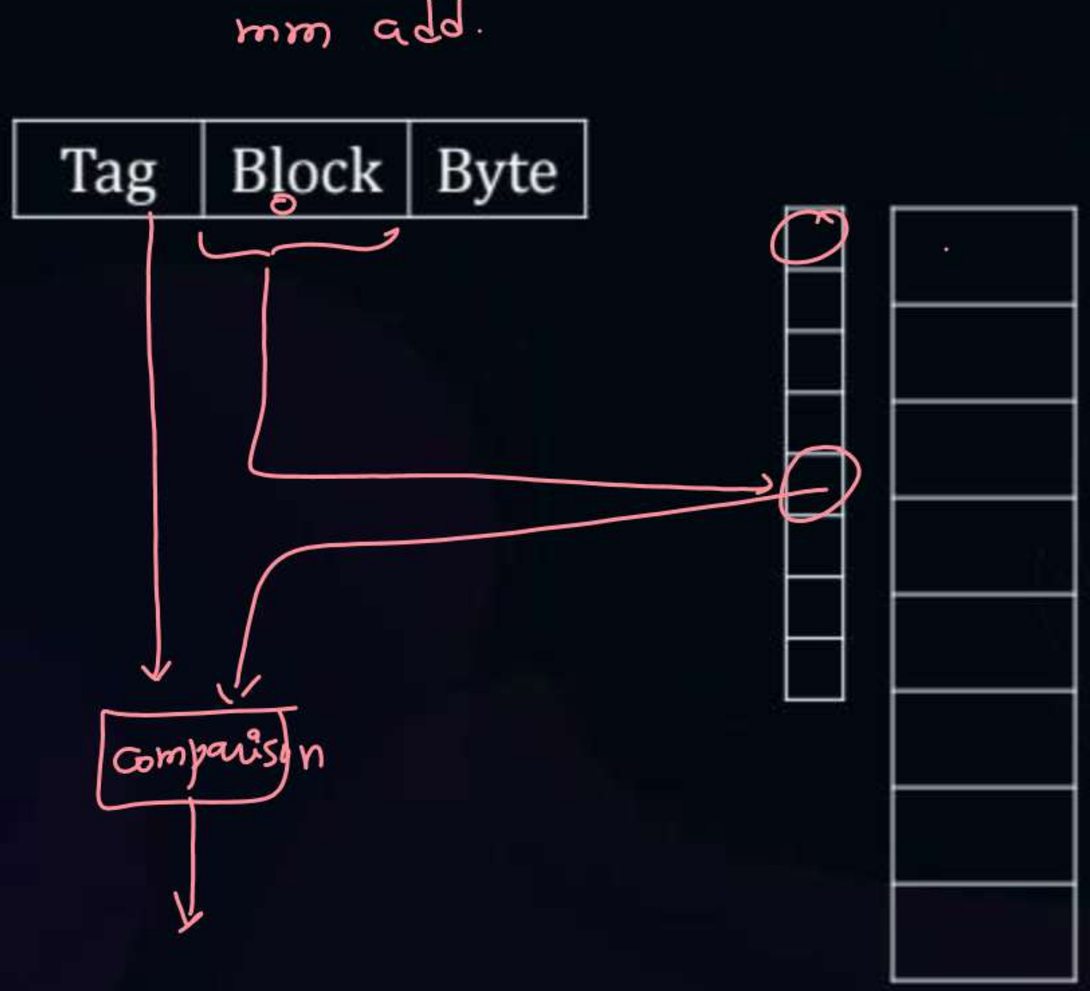
Topic Miss Penalty

Topic Types of Cache Miss

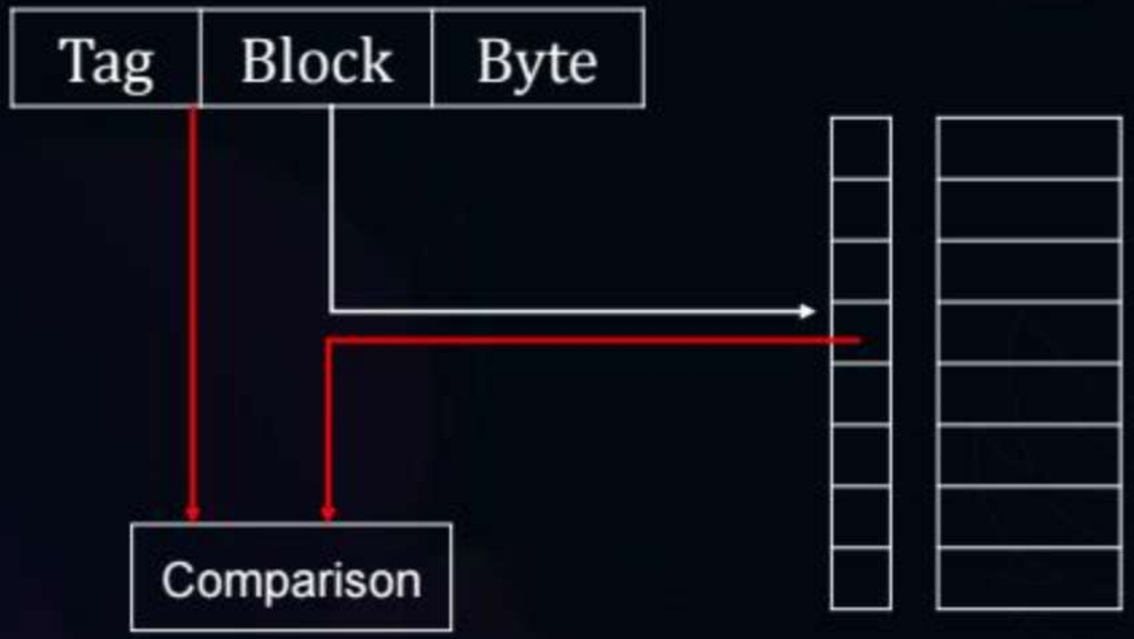
Topics to be Covered

- **Topic** Hardware Implementation of Mappings
- **Topic** Hit Latency
- **Topic** Mux & Comparators in Cache
- **Topic** Array Access with Cache

Topic : Checking Hit/Miss in Direct Mapping



Topic : Checking Hit/Miss in Direct Mapping



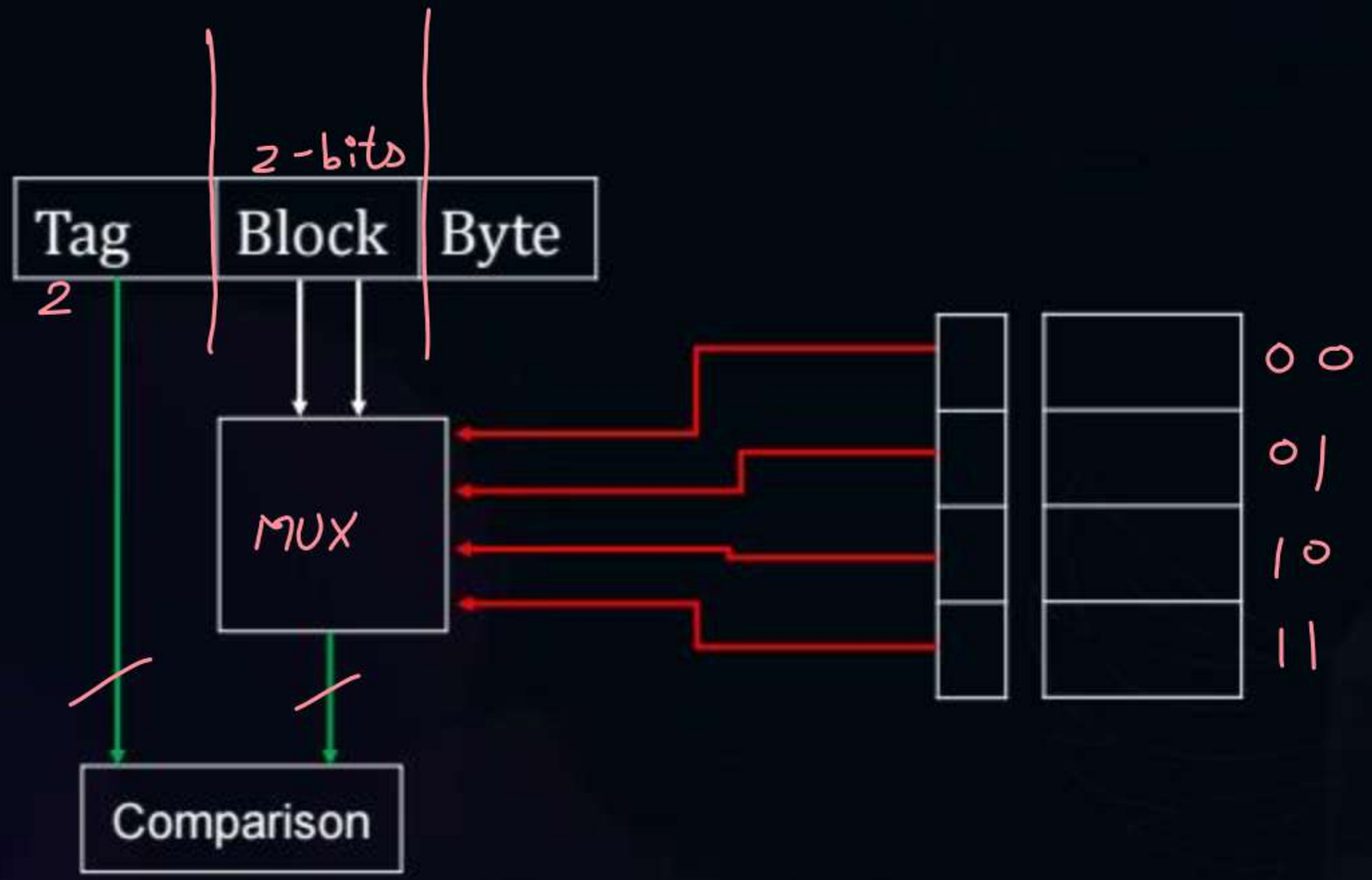
Topic : Hardware Implementation in Direct Mapping

Blocks in cache = 4

Block size = 16 bytes

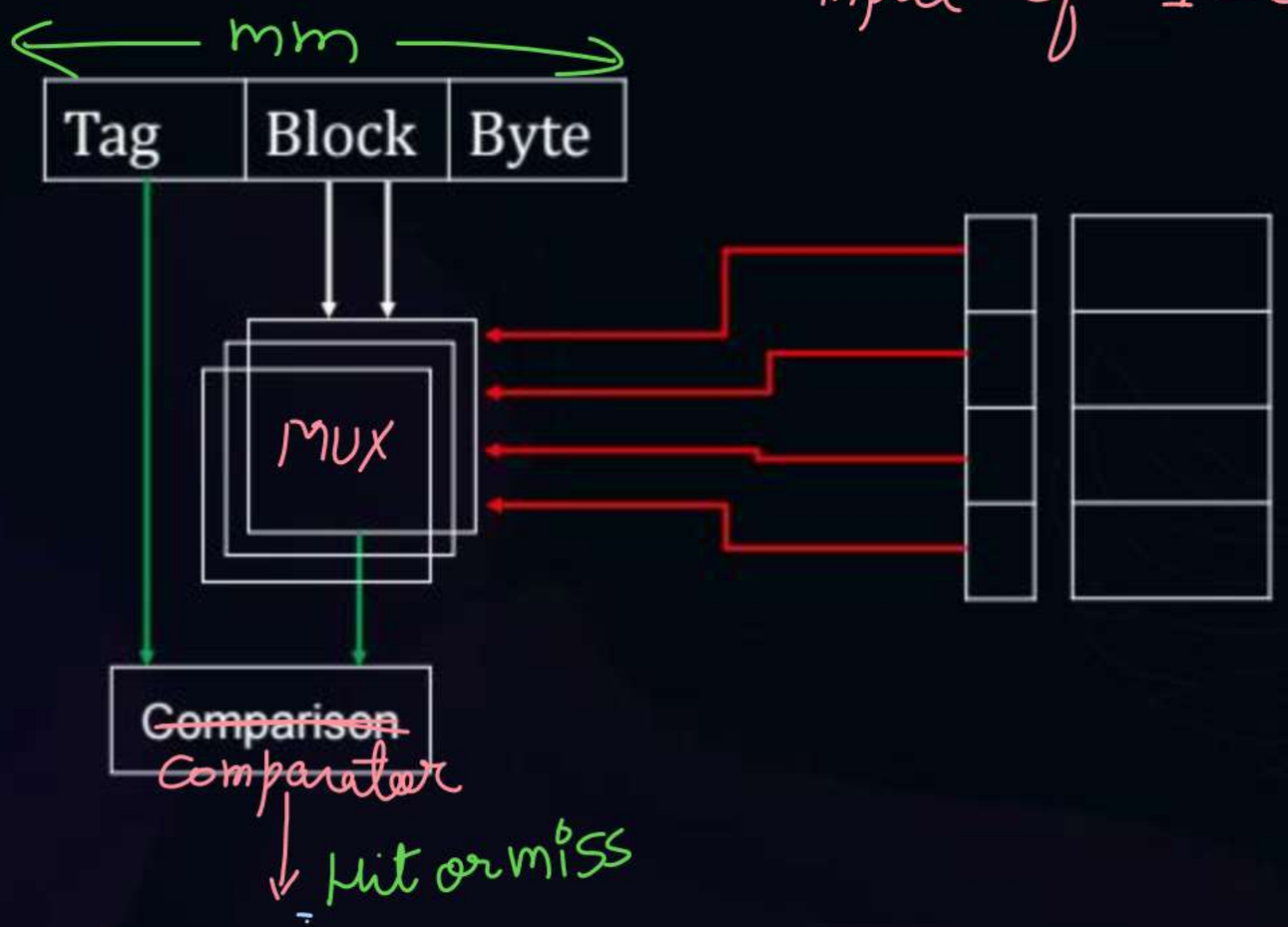
Main memory address = 8-bits

Topic : Hardware Implementation in Direct Mapping



Topic : Hardware Implementation in Direct Mapping

one MUX can accept each input of 1-bit only.



no. of MUX needed for tag selection = no. of bits in tag

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

No. of comparators = 1

size of comparator = tag bits comparator

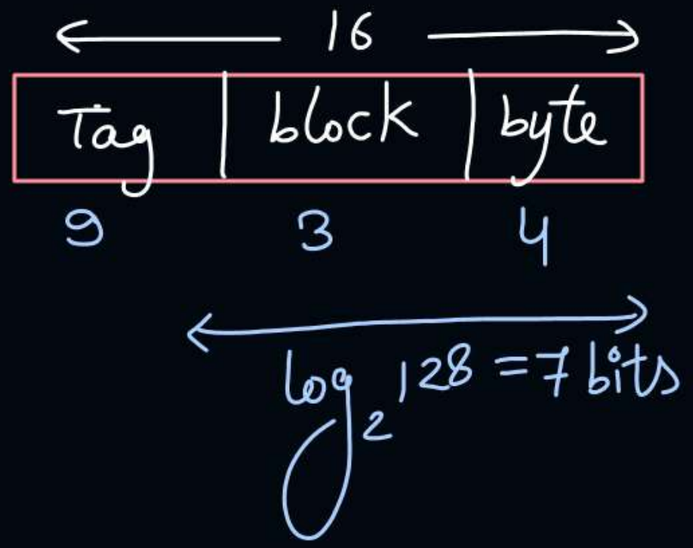
ex:-

Direct mapped cache

cm size = 128 B

Block = 16 B

MM = 16 bits



no. of blocks in cache = $\frac{128\,B}{16\,B} = \frac{2^7}{2^4} = 2^3 = 8$

H/w :-

1. MUX size = 8 : 1
2. no. of MUX = 9
3. No. of comparators = 1
4. size of " = 9-bits comparator

Topic : Hit Latency in Direct Mapping

Time taken by cache to check hit or miss

Hit latency = 1 MUX delay for tag selection + 1 comparator delay

Topic : Checking Hit/Miss in Set Associative Mapping

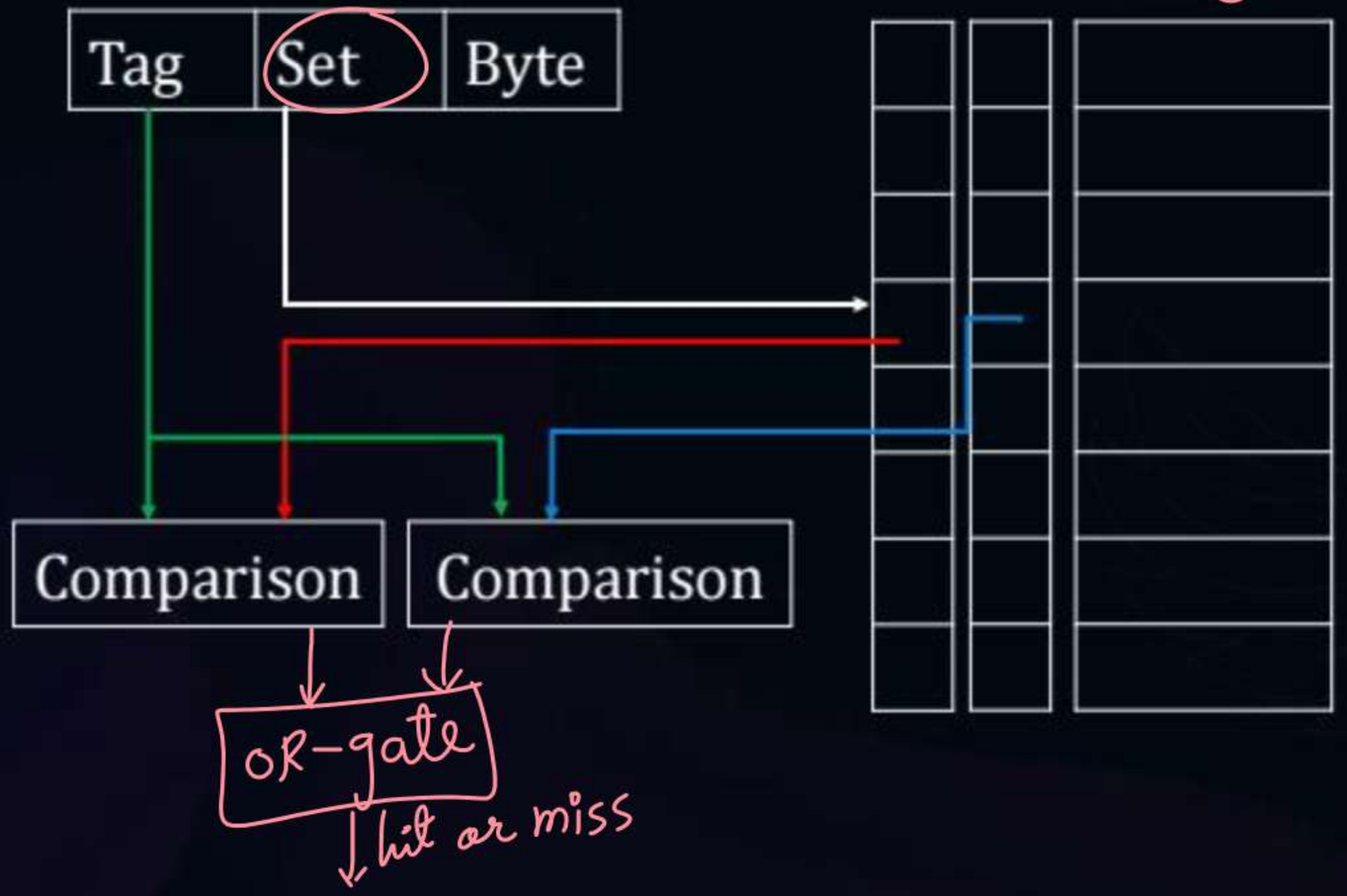
ex:- 2-way

Tag	Set	Byte

Topic : Checking Hit/Miss in Set Associative Mapping

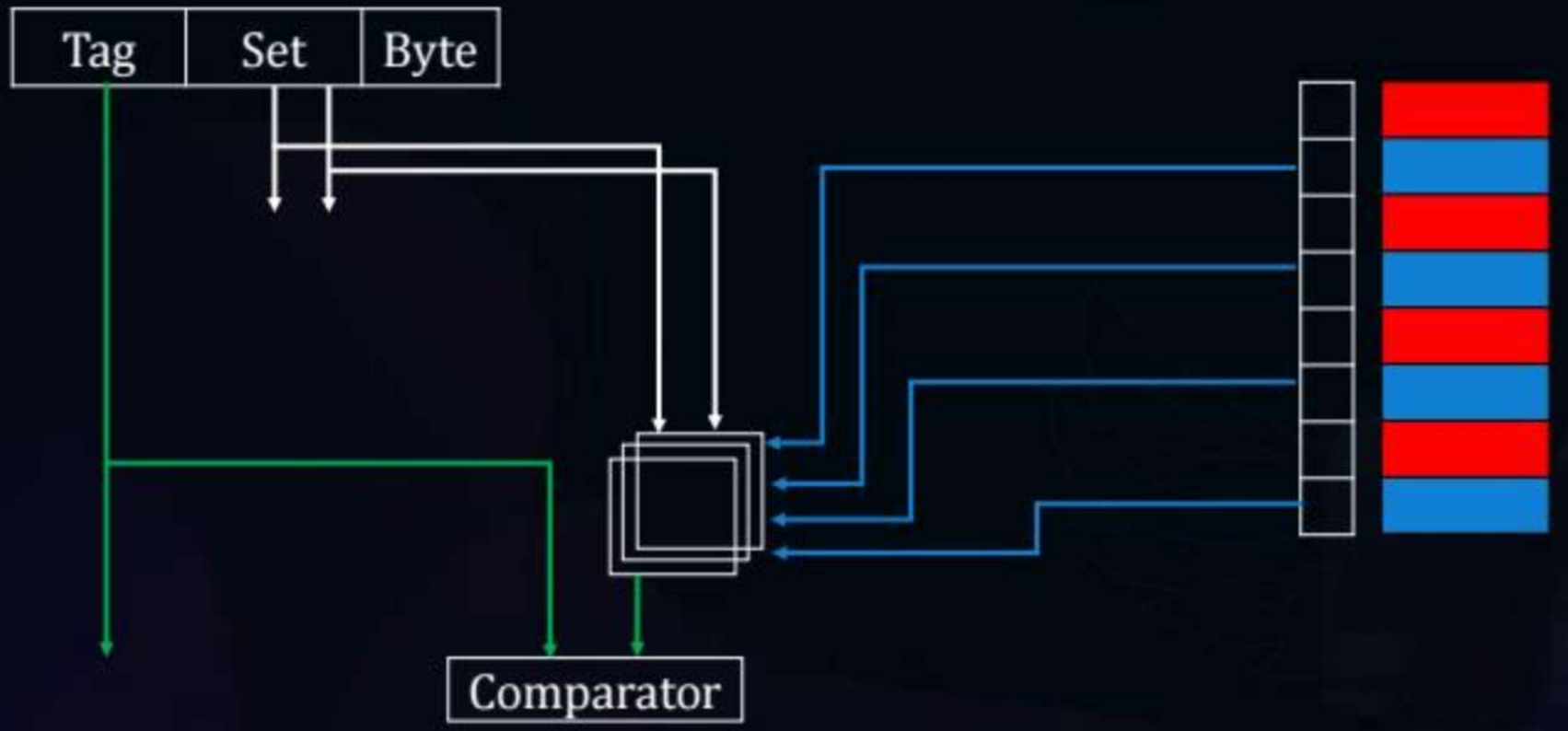
mm.

2-way set ass.

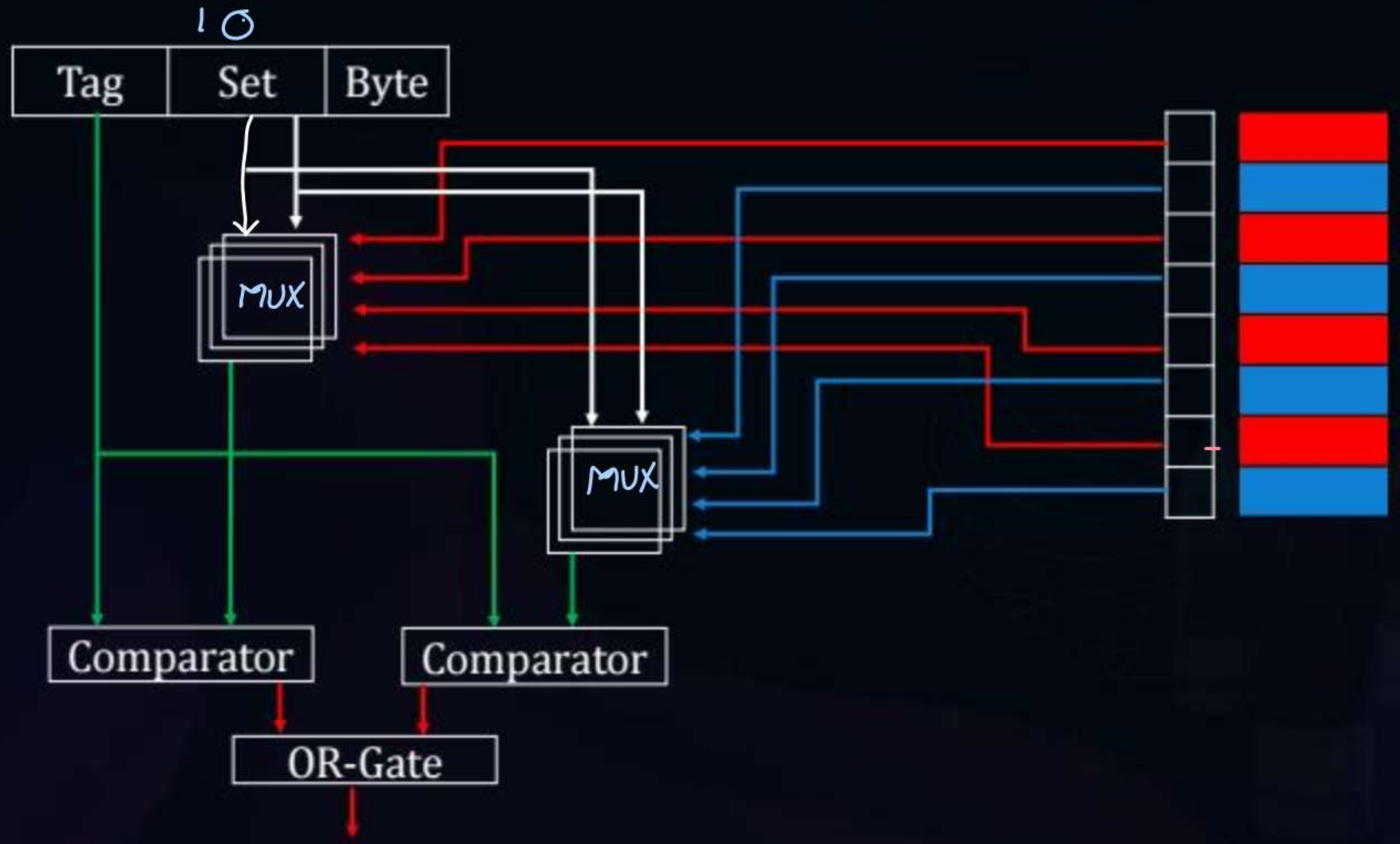


Topic : Hardware Implementation in Set Associative Mapping

Topic : Hardware Implementation in Set Associative Mapping



Topic : Hardware Implementation in Set Associative Mapping



for k-way set ass. cache:-

No. of MUX needed for tag selection = k * no. of tag bits

size of —"— —"— = No. of sets in cache : 1

No. of comparators = k

size of comparators = tag bits comparator

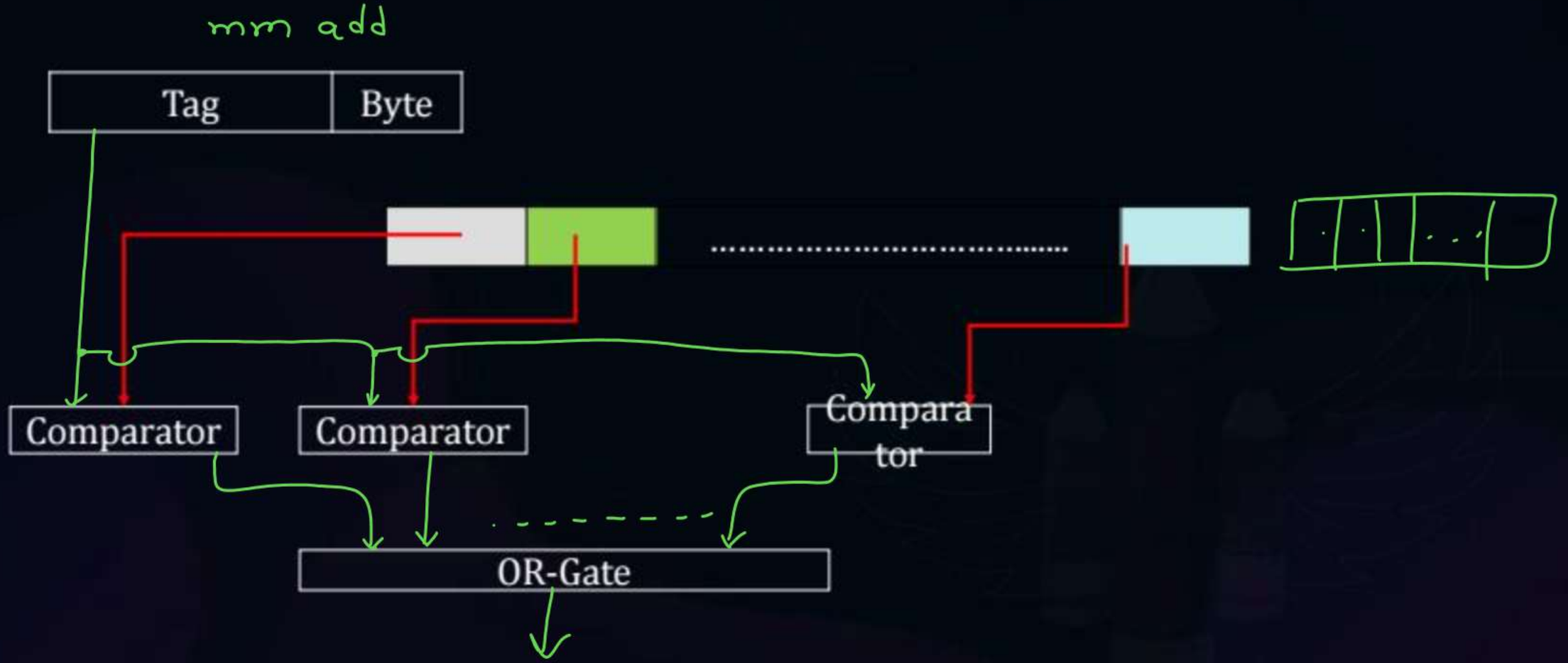
No. of OR gate = 1

size of OR gate = k input OR gate

Topic : Hit Latency in Set Associative Mapping

Hit latency = 1 MUX delay for tag selection + Comparator delay + OR-gate delay

Topic : Checking Hit/Miss in Fully Associative Mapping



Topic : Hit Latency in Fully Associative Mapping

No. of comparators = No. of blocks in cm

Size of Comparators = tag bit comparator

No. of OR gates = 1

No. of inputs in OR gate = no. of blocks in cache

Hit latency = Comparator delay + OR gate delay

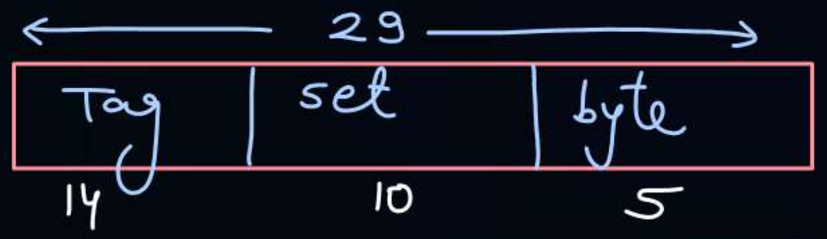
[NAT]

#Q. Cache Size = 128KB

Block size = 32 bytes

Main memory address = 29-bits

4-way set associative cache



$\log 128k - \log 4$

$= 17 - 2$

$= 15$ bits

1. Tag size? 14
2. Tag Directory size? $2^{12} * 14$ bits
3. Comparator required? 4, 14-bit comparators needed
4. MUX required? $= 14 * 4 = 56$

56 MUX of size 1024 : 1

no. of blocks in cache

$= \frac{128KB}{32B} = \frac{2^{17}}{2^5}$

$= 2^{12}$

Assume:-

k-bits comparator has delay $\Rightarrow \frac{k}{2}$ ns

for 16-bits comparator delay = ? $= \frac{16}{2} = 8$ ns

9-bits —11— ——— = ? $= \frac{9}{2} = 4.5$ ns

[MCQ] GATE- PYQ

as 2-input OR gate

#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:

no. of blocks in cache $= \frac{32KB}{32B} = 1k = 2^{10}$

no. of sets in cache $= \frac{2^{10}}{2} = 2^9$

A ✓ 2.4 ns

B 2.3 ns

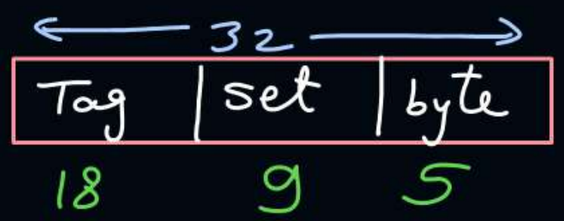
C 1.8 ns

D 1.7 ns

cm size = 32kB

2-way

block = 32B



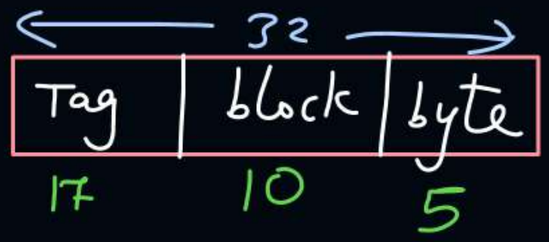
$$h_1 = 0 + \frac{18}{10} + 0.6$$

$$= 2.4\,ns$$

cm = 32kB = 2^{15} B

direct mapped

block = 32B



[MCQ]

#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

$h_2 = 0 + \frac{17}{10} = 1.7$ ns

Topic : Questions on Cache and Array

- Cache Size = 16 bytes
- Block size = 4 bytes
- Array in main memory int A[10], each element is 2 bytes

Topic : Questions on Cache and Array

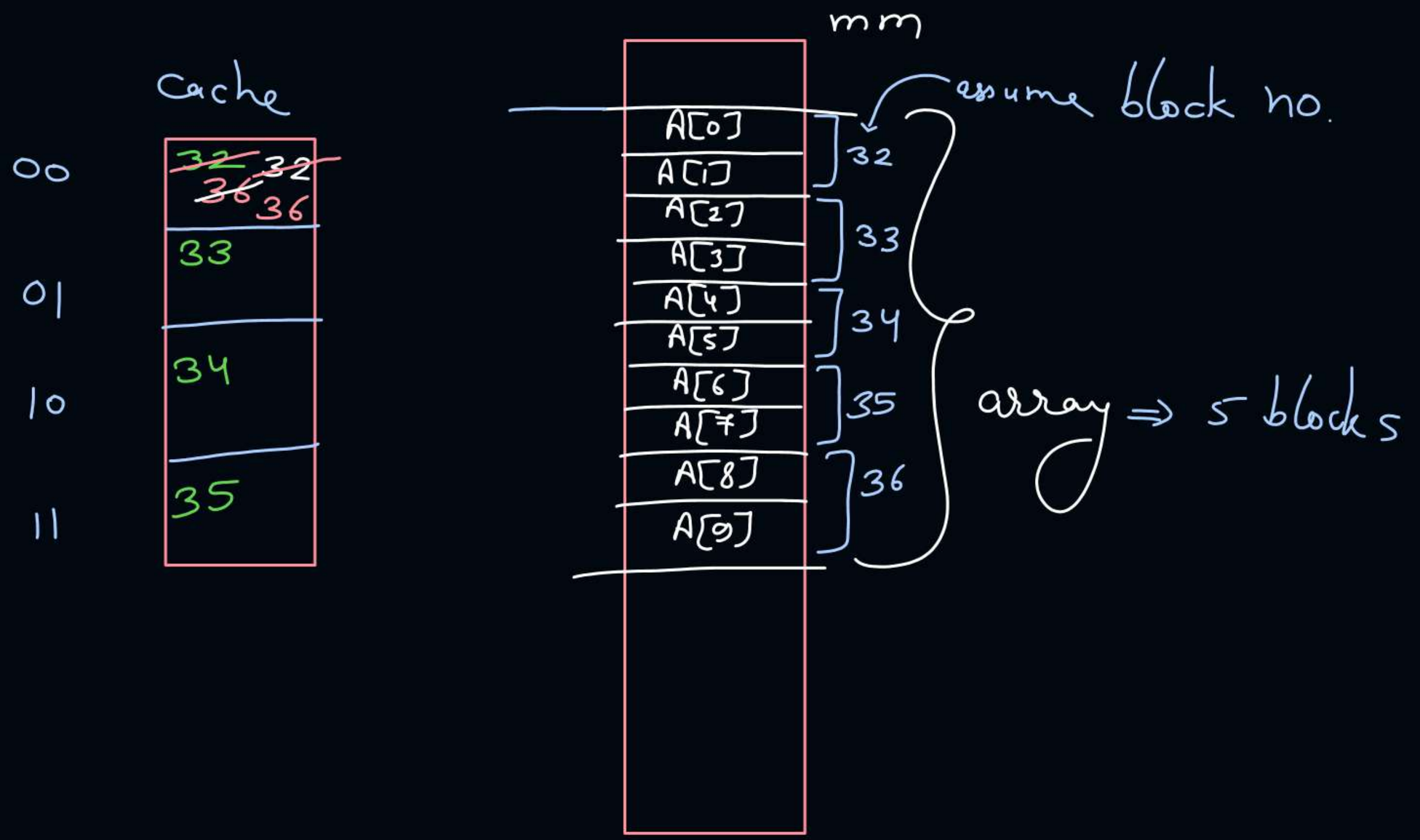
- Cache Size = 16 bytes ⇐ direct mapped
- Block size = 4 bytes
- Array in main memory int A[10], each element is 2 bytes

array size = 10 * 2 = 20B ⟶ $\frac{20B}{4B}$ ⇒ 5 blocks needed to store array

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

---

no. of elements per block = $\frac{4B}{2B}$ = 2


Cache
00
01
10
11
32
32
36
36
33
34
35
mm
assume block no.
A[0]
A[1]
A[2]
A[3]
A[4]
A[5]
A[6]
A[7]
A[8]
A[9]
32
33
34
35
36
array ⇒ 5 blocks

```
for (x = 0; x < 3; x++)
{
  for (i = 0; i < 10; i++)
  {
    printf("%d", A[i]);
  }
}
```

	first access	second access
A[0] →	Miss	Miss
A[1] →	Hit	Hit
A[2] →	Miss	Hit
A[3] →	Hit	Hit
A[4] →	miss	Hit
A[5] →	Hit	Hit
A[6] →	miss	Hit
A[7] →	Hit	Hit
A[8] →	Miss	Miss
A[9] →	Hit	Hit

	Hit	Miss
1st time access	5	5
2nd time —11—	8	2
3rd —11—	8	2
Total	21	9

$$\text{Hit ratio} = \frac{21}{30} = 0.7$$

$$\text{Miss ratio} = \frac{9}{30} = 0.3$$

2 mins Summary

Topic Hardware Implementation of Mappings

Topic Hit Latency

Topic Mux & Comparators in Cache

Topic Array Access with Cache

Happy Learning

THANK - YOU