CS & IT ENGINEERING

COMPUTER ORGANIZATION AND ARCHITECTURE

Cache Organization

Lecture No.- 5

By- Vishvadeep Gothi sir

Recap of Previous Lecture

Topic Set Associative Mapping

Topic Fully Associative Mapping

Topics to be Covered

Topic Set Associative Mapping

Topic Fully Associative Mapping

Topic Block Replacement

Topic Cache Miss Penalty

[NAT]

no. of sets in cache = $\frac{64\,B}{8\,B * 2} = 4$

MM address = 12-bits

Cache size = 64 bytes

Block size = 8 bytes

2-way set associative Mapping

All values in Decimal			
MM Address	mm block ~~Tag~~ no.	Tag	Cache set Number where it maps
521	mm block no. $= \left\lfloor \frac{521}{8} \right\rfloor = 65$	$\left\lfloor \frac{65}{4} \right\rfloor = 16$	$65 \% 4 = 1$
298	$\left\lfloor \frac{298}{8} \right\rfloor = 37$	$\left\lfloor \frac{37}{4} \right\rfloor = 9$	$37 \% 4 = 1$
1000	$\left\lfloor \frac{1000}{8} \right\rfloor = 125$	$\left\lfloor \frac{125}{4} \right\rfloor = 31$	$125 \% 4 = 1$
917	$\left\lfloor \frac{917}{8} \right\rfloor = 114$	$\left\lfloor \frac{114}{4} \right\rfloor = 28$	$114 \% 4 = 2$

[MCQ]

#Q. Consider a 64 bytes direct mapped cache with a block size of 16 bytes. Main memory size is 256bytes. Currently in the cache, the blocks are having tags as follows:

Block	Tag
00	10
01	01
10	11
11	01

mm block no.

Tag	cm block no.

$(10\ 00)_2 = (8)_{10}$ ✓

$(0101)_2 \Rightarrow (5)_{10}$

$(1110)_2 \Rightarrow (14)_{10}$

$(0111)_2 = (7)_{10}$

Identify the correct statement with respect to the availability of the main memory data into cache? Note: Given addresses are in decimal

a) ✗ Main memory byte number 243 present in cache $\Rightarrow$ mm block no $= \left\lfloor \frac{243}{16} \right\rfloor = 15$

b) ✓ Main memory byte number 143 present in cache $\left\lfloor \frac{143}{16} \right\rfloor = 8$

c) ✗ Main memory byte number 43 present in cache $\left\lfloor \frac{43}{16} \right\rfloor = 2$

d) ✓ Main memory byte number 119 present in cache $\left\lfloor \frac{119}{16} \right\rfloor = 7$

[MCQ]

#Q. A computer system with a word length of 32 bits has a 16 MB byte- addressable main memory and a 64 KB, 4-way set associative cache memory with a block size of 256 bytes. Consider the following four physical addresses represented in hexadecimal notation.

A1 = 0x42C8A4 ⇒ 001000

A2 = 0x546888 ⇒ 101000

A3 = 0x6A289C ⇒ 101000

A4 = 0x5E4880 ⇒ 001000

Which one of the following is TRUE?

A. A1 and A4 are mapped to different cache sets.

B. ✓ A2 and A3 are mapped to the same cache set.

C. A3 and A4 are mapped to the same cache set.

D. A1 and A3 are mapped to the same cache set.

24-bits

Tag	set	byte
10	6	8

log 64K − log 4

16 − 2

14

[MCQ]

#Q. Cache size = 256KB 2^{18}

Block size = 32 Bytes

MM address = 30 bits

Calculate tag and tag directory size for direct, 4-way and fully associative cache?

Direct :-

30

Tag	block	byte
12	13	5

Tag directory size $= 2^{13} * 12$ bits

4-way :-

30

Tag	set	byte
14	11	5

$= 2^{13} * 14$ bits

fully :-

30

Tag	byte
25	5

$= 2^{13} * 25$ bits

Topic : Tag & Index in all Mappings

	Maximum	Minimum
Index	Direct	fully
Tag	fully	Direct

Topic : Byte vs Word Addressable Memory

#Q. Cache size = 64KB

Block size = 32 Bytes $= 2^5 B$.

MM size = 2GB

Direct mapping

Word size = 4 bytes

no. of blocks in cache $= \frac{\cancel{64}^{2} KB}{\cancel{32 B}} = 2K = 2^{11}$

Byte addressable memory:-

mm size $= 2GB = 2^{31} B$.

$\Downarrow$

mm add = 31 bits

31

Tag	block	byte
15	11	5

word addressable:-

$\longleftarrow$ 29 $\longrightarrow$

Tag	block	word
15	11	3

mm size $= 2GB = \frac{2GB}{4B} = 2^{29}$

$\Downarrow$

add. = 29 bits

block size $= \frac{32 B}{4 B}$.

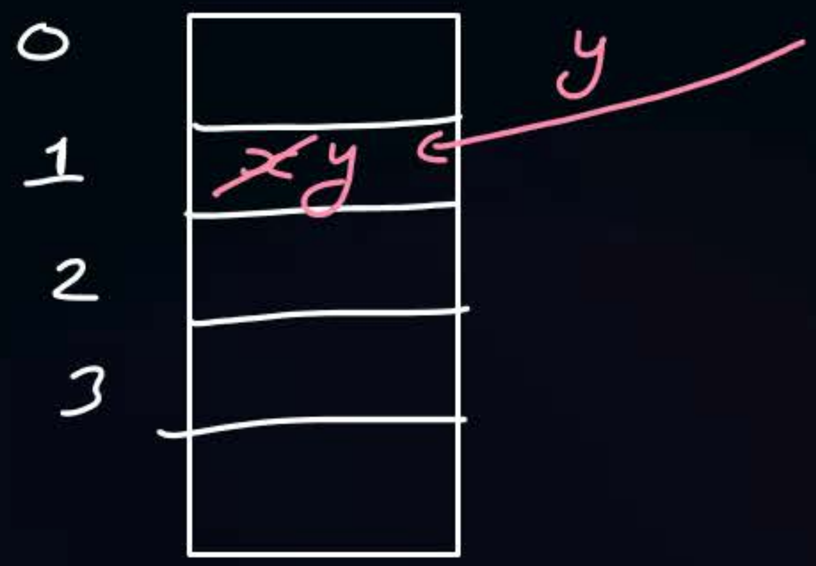
$= 8$ words $= 2^3$

$$\text{cm size} = 64KB = \frac{64KB}{4B} = 16K = 2^{14}$$

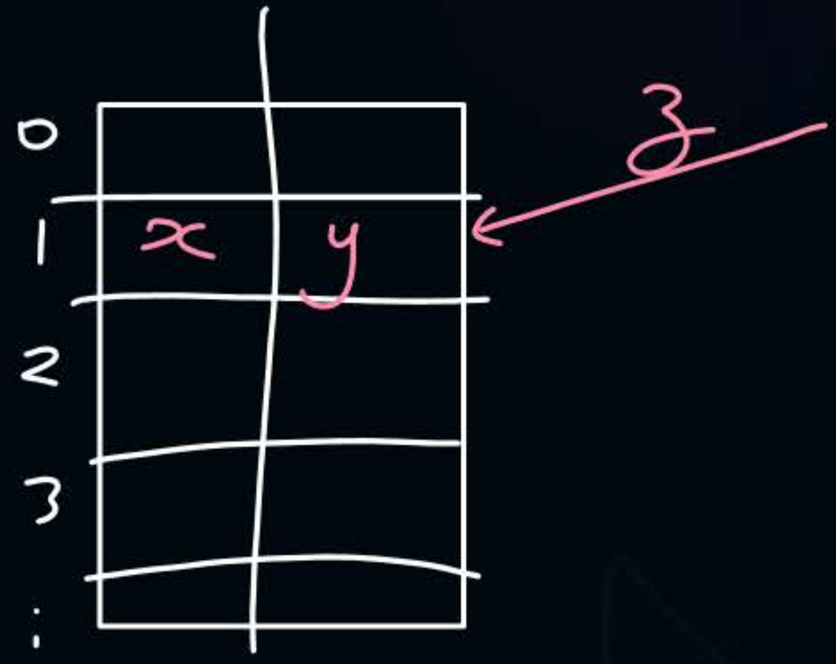
$$\text{no. of blocks in cache} = \frac{2^{14}}{2^{3}} = 2^{11}$$

Topic : Block Replacement

Direct



2-way



x, y ⇒ which to be replaced

⇓

replacement policy needed

Replacement policy :-

In cache LRU (Least Recently Used) is used.

⇓

replace the block which has not been refered since longest period of time.

[MCQ]

#Q. Consider a small two-way set-associative cache memory, consisting of four blocks. For choosing the block to be replaced, use the least recently used (LRU) scheme. The number of cache misses for the following sequence of block addresses is:

8, 12, 0, 12, 8

no. of sets $= \frac{4}{2} = 2$

$8 \% 2 = 0$

$12 \% 2 = 0$

$0 \% 2 = 0$

0	~~8~~ ~~0~~ 8	12
1		

A) 2

B) 3

C) 4

D) 5

[MCQ]

cm block no. = (mm block no.) % no. of blocks in cache

#Q. Consider a Direct Mapped Cache with 8 cache blocks (numbered 0-7). If the memory block requests are in the following order

3, 5, 2, 8, 0, 63, 9, 16, 20, 17, 25, ~~18~~, 30, 24, 2, 63, 5, 82, 17, 24.

Which of the following memory blocks will not be in the cache at the end of the sequence?

A. 3

B. 18

C. 20

D. 30

0	~~8~~ ~~0~~ ~~16~~ 24
1	~~9~~ ~~17~~ ~~25~~ 17
2	~~2~~ ~~18~~ ~~2~~ 82
3	3
4	20
5	5
6	30
7	63

no. of hits = 3

no. of miss = 17

hit ratio = $\frac{3}{20}$

miss ratio = $\frac{17}{20}$

[MCQ]

no. of sets $= \frac{16}{4} = 4$

#Q. Consider a 4-way set associative cache (initially empty) with total 16 cache blocks. The main memory consists of 256 block and the request for memory blocks is in the following order:

0, 255, 1, 4, 3, 8, 133, 159, 216, 129, 63, 8, 48, 32, 73, 92, 155

Which one of the following memory block will not be in cache if LRU replacement policy is used?

A 3

B 8

C 129

D 216

0	~~0~~ 48	~~4~~ 32	8	~~216~~ 92
1	1	133	129	73
2				
3	~~255~~ 155	3	159	63

[MCQ]

#Q. Consider a fully associative cache with 8 cache blocks (numbered 0–7) and the following sequence of memory block requests:

4,3,25,8,19,6,25,8,16,35,45,22,8,3,16,25,7

If LRU replacement policy is used, which cache block will have memory block 7?

0	1	2	3	4	5	6	7
~~4~~	~~3~~	25	8	~~19~~	~~6~~	16	35
45	22			3	7		

A. 4

B. 5

C. 6

D. 7

[MCQ]

#Q. Consider a Direct Mapped Cache with size of 256 bytes and 32 bytes blocks. The main memory size is 2Kbytes. If the memory address requests (in decimal) are in the following order

mm block no.

112, 97, 161, 127, 32, 15, 190, 363, 97 $\Rightarrow$ 3 ✓, 3 ✗, 5 ✓, 3 ✗, 1 ✓, 0 ✓, 5 ✗, 11 ✓, 3 ✓

The number of cache misses, at the end of the sequence is __6__?

no. of blocks in cache $= \frac{256B}{32B}$

$= 8$

0	0
1	1
2	
3	~~3~~ 11
4	
5	5
6	
7	

Ans $= \underline{6}$

Topic : Cache Miss Penalty

Time required to bring a missed block from main memory to cache

block size smaller ⇒ smaller miss penalty

—〃— larger ⇒ larger —〃—

Topic : Cache Miss Penalty

Assume:

Cycles required to send address to memory : 2 cycles

Cycles required to access 1 main memory cell : 10 cycles

Cycles required to transfer 1 cell data to cache : 2 cycles

Cache Block Size	Main memory cell size	Miss Penalty
4 bytes	1 byte	$2+(4*10)+(4*2) = 50$ cycles $= 25$ ns
4 bytes	2 bytes	$2+(2*10)+(2*2) = 26$ cycles $= 13$ ns
4 bytes	4 bytes	$2+(1*10)+(1*2) = 14$ cycles $= 7$ ns

If CPU runs on 2 GHz Clock rate $\Rightarrow$ 1 cycle time $= \frac{1}{2\,GHz} = 0.5$ ns

[NAT]

Ans = 160

$8 * 4 = 32\,B$

#Q. A certain processor deploys a single-level cache. The cache block size is 8 words, and the word size is 4 bytes. The memory system uses a 60 MHz clock. To service a cache-miss, the memory controller first takes 1 cycle to accept the starting address of the block, it then takes 3 cycles to fetch all the eight words of the block, and finally transmits the words of the requested block at the rate of 1 word per cycle. The maximum bandwidth for the memory system when the program running on the processor issues a series of read operations is ___160___ $\times 10^6$ bytes/sec?

miss penalty $= 1 + 3 + (8 * 1) = 12$ cycles $= 12 * \frac{1}{60\,MHz} = \frac{1}{5}\,\mu sec$

$= 0.2\,\mu sec$

in $0.2\ \mu sec$, data = 32 Bytes

in 1 sec, data $= \dfrac{32\ B}{0.2 * 10^{-6}\ sec}$

$= 160 * 10^{6}\ B/sec$

Topic : Types of Cache Misses

1. Cold or Compulsory Miss → first access
2. Capacity Miss
3. Conflict Miss

} Repeat access

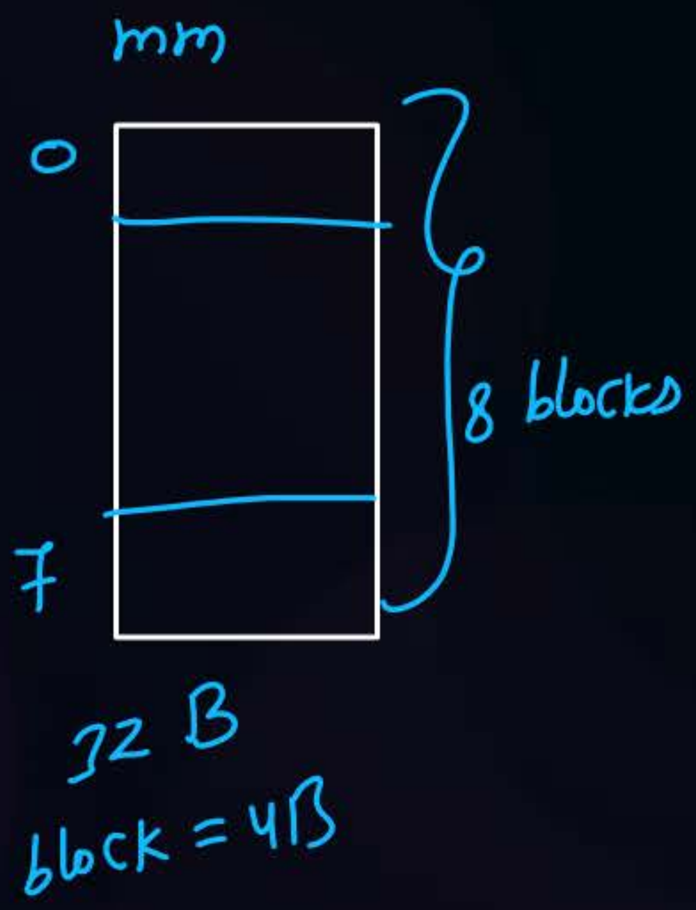
	max	min
conflict miss	Direct mapping	fully ass. cache

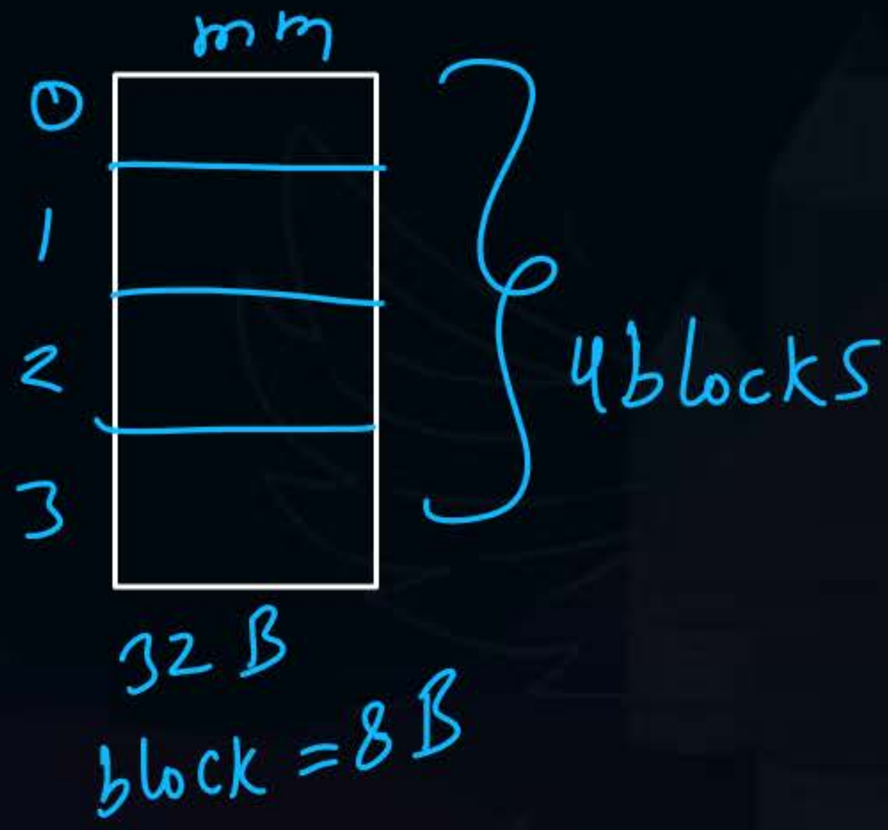
Topic : Types of Cache Misses

1. Cold or Compulsory Miss

First time access of a block will always cause a miss

To reduce Cold misses: Increase block size





Topic : Types of Cache Misses

2. Capacity Miss

if it is not cold miss.

If cache is full and hence miss occurs.

To reduce Capacity misses: Increase cache size

Topic : Types of Cache Misses

There is no any conflict miss fully ass. cache

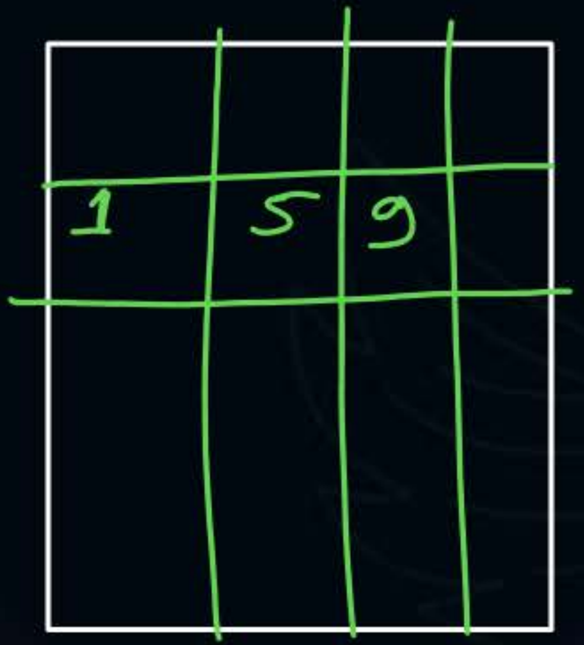
⇓

3. Conflict Miss

if it is not cold and capacity

If cache set is full and hence miss occurs [redacted]

To reduce Conflict misses: ⇒ Increase associativity



if set is full

⇓

means cache is full

⇓

capacity miss

Topic : Example

- 2-way set associative cache has 4 blocks
- LRU replacement policy
- CPU requests for main memory blocks

cold	cold	hit	cold	hit	conflict			hit		hit	
0,	4,	0,	8,	0,	4,	1,	3,	1,	5,	1,	3
						cold	cold		cold		capacity

0	0	~~4~~ ~~8~~ 4
1	1	~~3~~ ~~5~~ 3

no. of cold miss = 6

—"— capacity " = 1

—"— conflict miss = 1

[NAT]

H.W.

#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 ________?

2 mins Summary

Topic Set Associative Mapping

Topic Fully Associative Mapping

Topic Block Replacement

Topic Cache Miss Penalty

Happy Learning

THANK - YOU