CS & IT ENGINEERING

COMPUTER ORGANIZATION AND ARCHITECTURE

Cache Organization

Lecture No.- 08

By- Vishvadeep Gothi sir

Recap of Previous Lecture

Topic Direct Mapping

Topic Set Associative Mapping

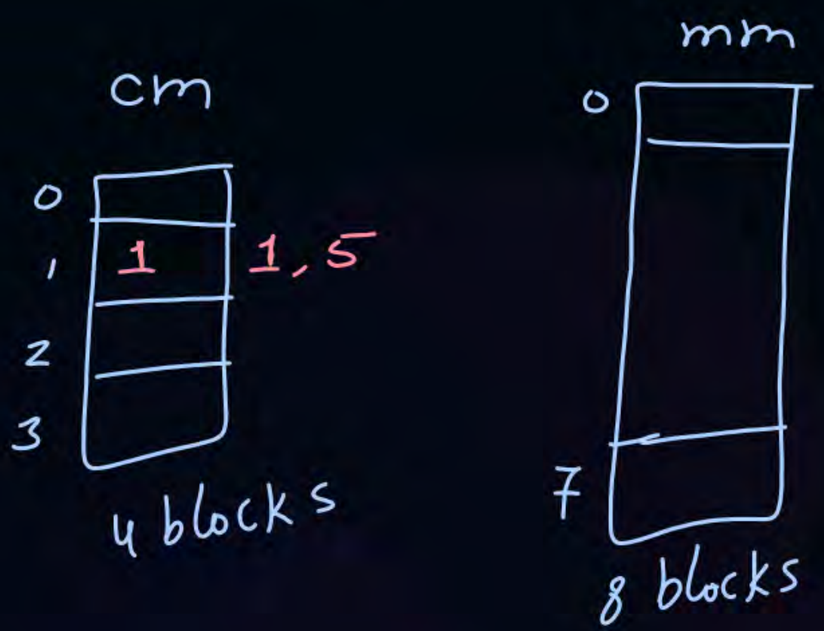
Topic Fully Associative Mapping

Topics to be Covered

- Topic: Block Replacement
- Topic: Miss Penalty
- Topic: Types of Cache Miss

Topic : Block Replacement

Direct mapping :-



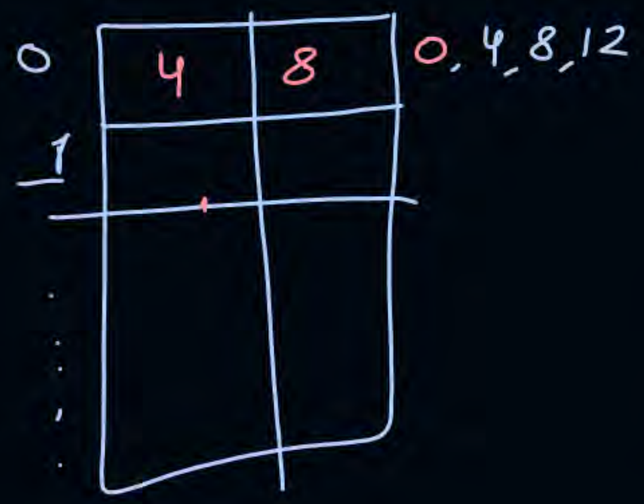
CPU requests MM block no.

1, 5

↓

replace block 1 from CM

2-way set ass. :-



4, 8 are already present in cache and CPU requests for block 12

⇓

miss; so bring block 12 from mm to cache.

⇓

To replace one of block 4, 8; replacement policy is used.

Direct mapping ⇒ no any replacement policy needed

set ass. or fully ass. ⇒ replacement policy is needed

Replacement policies:-

1. FIFO (First In first out)
2. optimal
3. ☆ LRU (Least Recently Used) ⇒ Replace the block which has not been used for 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: (note :- cache initially empty)

8, 12, 0, 12, 8

miss, miss, miss, hit, miss

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

Cache

0	8 ~~0~~ 8	12
1		

8 % 2 ⇒ 0

12 % 2 ⇒ 0

0 % 2 ⇒ 0

A 2

B 3

C ✓ 4

D 5

[MCQ]

Cm block no. = mm block no. % no. of blocks in Cm

#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

3 % 8 = 3

5 % 8 = 5

2 % 8 = 2

8 % 8 = 0

0 % 8 = 0

[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

0 % 4 = 0

255 % 4 = 3

[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?

A 4

B 5

C 6

D 7

Topic : Cache Miss Penalty

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

Topic : Cache Miss Penalty

Assume:

needed only once

Cycles required to send address to memory : 1 cycle

Cycles required to access 1 main memory cell : 10 cycles

Cycles required to transfer 1 cell data to cache : 1 cycle

Cache Block Size	Main memory cell size	Miss Penalty
4 bytes	1 byte	$1+(4*10)+(4*1)=45$ cycles
4 bytes	2 bytes	$1+(2*10)+(2*1)=23$ cycles
4 bytes	4 bytes	$1+(1*10)+(1*1)=12$ cycles

#a if in prev. questⁿ cycle time is 2 ns, then miss penalty in ns will be ?

Solⁿ

$45 * 2 = 90$ ns

$23 * 2 = 46$ ns

$12 * 2 = 24$ ns

#a In prev. questⁿ, CPU clock rate = 2 GHz, then miss penalty in ns will be ?

Solⁿ Cycle time $= \frac{1}{2\text{GHz}} = 0.5$ ns

Miss penalty :- $45 * 0.5 = 22.5$ ns

$23 * 0.5 = 11.5$ ns

$12 * 0.5 = 6$ ns

[NAT]

#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?

Block size = 8 words
word size = 4 bytes

8 * 4 = 32B

miss penalty = 1 + 3 + 8 = 12 cycles = $\frac{12 * 1}{60\,MHz}$

$= \frac{1}{5}$ usec

$= 0.2$ usec

in 0.2 μsec time, data transfer = 32 Bytes

in 1 sec —— 11 —— $= \frac{32\ B}{0.2 * 10^{-6}\ sec}$

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

Ans = 160

Topic : Types of Cache Misses

1. Cold or Compulsory Miss
2. Capacity Miss
3. Conflict Miss

Topic : Types of Cache Misses

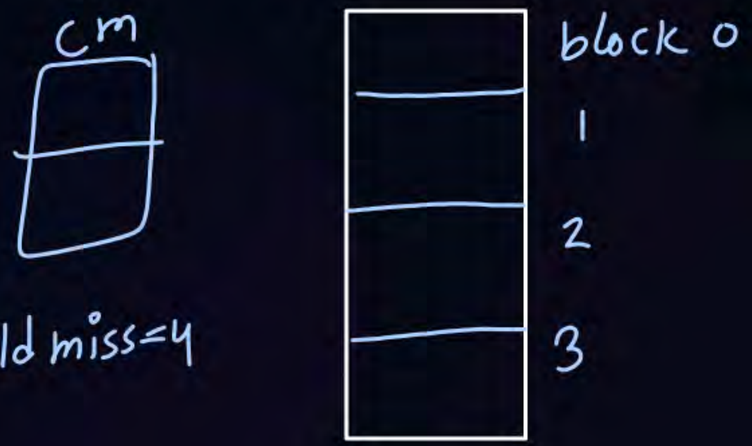
1. Cold or Compulsory Miss

First time access of a block will always cause a miss

To reduce Cold misses: Increase block size

ex:-

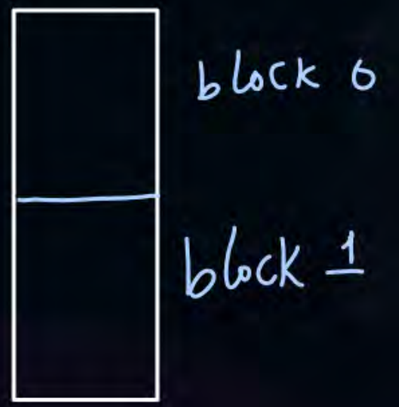
block size = 4B
mm size = 16 Bytes



max no. of cold miss = 4

ex:-

mm size = 16 bytes
block size = 8B



max no. of cold miss = 2

Topic : Types of Cache Misses

2. Capacity Miss

If cache is full and hence miss occurs.(which is not cold miss)

To reduce Capacity misses:

→ Increase cache size

Topic : Types of Cache Misses

3. Conflict Miss

If cache is not full and hence miss occurs due to the current set is full (and also it's not the cold miss)

To reduce Conflict misses:

↳ Increase associativity

Fully associative cache:- It has only single set

↳ Hence in fully ass. cache no. of conflict misses are zero.

Topic : Example

- No. of blocks in cache = 4
- 2-way set associative cache
- LRU replacement policy
- CPU requests for main memory blocks

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

cm set no. = (mm block no.) % no. of sets in cm

0, 4, 0, 8, 0, 4, 1, 3, 1, 5, 1, 3

Cold, Cold, Cold, Conflict, Cold, Cold, Capacity

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

2 mins Summary

Topic Block Replacement

Topic Miss Penalty

Topic Types of Cache Miss

Happy Learning

THANK - YOU