

COMPUTER SCIENCE

Computer Organization
and Architecture

Cache Memory

Lecture_08



Vijay Agarwal sir



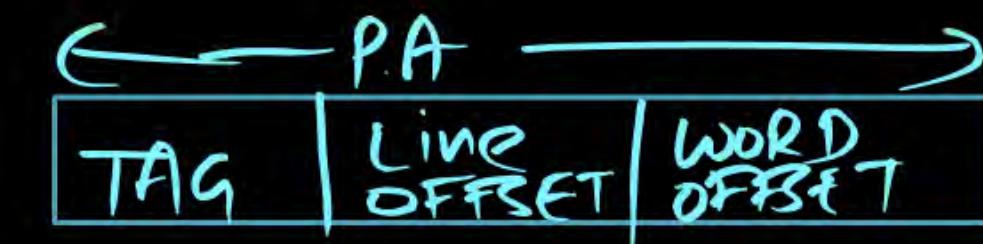


**TOPICS
TO BE
COVERED**

- o1 Mapping Techniques**
- o2 Replacement Algo & Updating Technique**

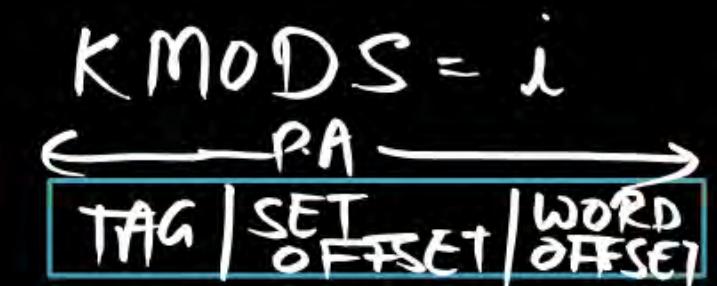
Mapping

① Direct Mapping

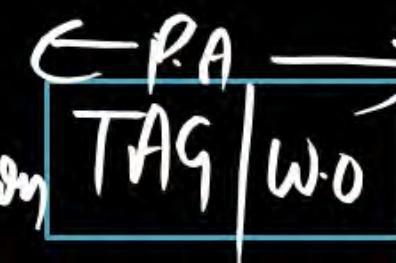


$$K \bmod N = i$$

② Set Associative Mapping



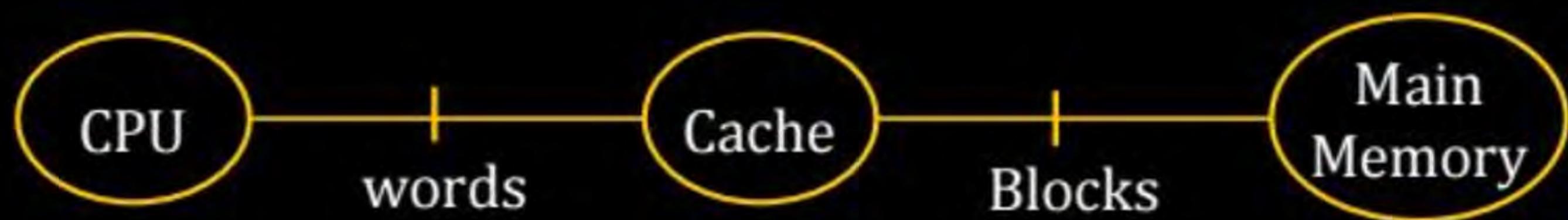
③ Fully Associative Mapping
No Mapping function



Replacement Algo

- ① FIFO
- ② LRU
- ③ Direct Mapping
- ④ Set Associative with LRU.

Memory Organization



Q.

Consider fully associative cache consists 8 Block & MM contain 128 Block & Request made by the CPU.

119, 84, 37, 0, 16, ~~0, 84~~, 120, 121, 93, ~~37, 0~~ 43, 39, 47, 48.

Calculate # of compulsory & capacity miss?



0	119 43
1	84 39
2	37 47
3	0 48
4	16
5	120
6	121
7	93

Hit: 0, 84, 37, 0

#Hits: 4

Total Miss = 12

Compulsory Miss
Capacity Miss

No Conflict Miss

If CM Contain 16 Block

No Capacity Miss

Replacement Algorithm

When Cache is Full, then replacement algorithm are required to replace the exist cache block with new block.

In the CM design 3 type of replacement algorithm is used.

- 1) Random Algorithm
- 2) FIFO Replacement
- 3) LRU Replacement

In the random algorithm, any cache block can be replaced based on the random selection.

Type of Miss

- ① Compulsory | Cold Start | First Reference Miss
- ② Capacity Miss.
- ③ Conflict | Collision Miss.

Types of Misses

In the CM design 3 types of misses are present.

- 1) Compulsory miss - (Cold start miss / first reference miss)

This miss will occur when the very first reference to the cache itself

a miss. (When Any MM Block, first time Access in Cache)

- 2) Capacity Miss - This miss will occur when cache is full.

- 3) Conflict Miss (Collision miss / reference miss)

This miss will occur when the too many blocks are placed into same

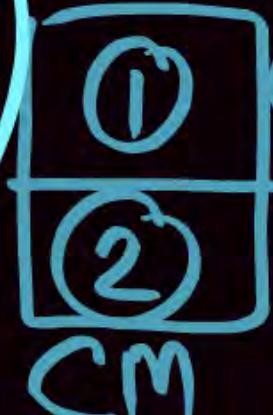
cache line or same cache SET.

How we can Reduce Miss

① Compulsory Miss :

Reking wadden
example
4 Person.

$$\text{BLOCKSIZE} = 32B$$



② Compulsory Miss

↳ we can Reduce the Compulsory Miss
by Increasing the Block Size.
[larger Block Size]

BS=64B
1 Compulsory Miss

64Byte
64Byte

② Capacity Miss \Rightarrow Capacity Miss occur when Cache is Full.

Capacity Miss can be Reduced by Increasing the Cache Size [Larger Cache]

③ Conflict Miss \Rightarrow Conflict / Collision Miss can be Reduced

By increasing Set Associativity.

2 Way Set Associative [2 mm Block Per Set]

4 Way Set Associative [4 mm Block Per Set]

8 Way Set Associative [8 mm Block Per Set]



If In GATE 2017 Question

If 4 way Set Associative.

Conflict Miss Reduced

$$\#SET = \frac{256}{4} = 64$$

$$K MOD S = i$$

$$K MOD 64 = i$$

\uparrow	-0	$\cdot 164 = 0 \rightarrow M$
	-128	$\cdot 164 = 0 \rightarrow m$
	-256	$\cdot 164 = 0 \rightarrow m$
\uparrow	128	$\cdot 164 = 0 \rightarrow HIT$
	0	$\cdot 164 \rightarrow H$
\uparrow	128	$\cdot 164 \rightarrow H$
\uparrow	256	$\cdot 164 \rightarrow H$
\uparrow	128	$\cdot 164 \rightarrow H$
<hr/>		
	1	$\cdot 164 = 0 \rightarrow M$
	129	$\cdot 164 = 0 \rightarrow M$
	257	$\cdot 64 = 0 \rightarrow m$
\uparrow	129	$\cdot 64 \quad \left. \begin{array}{l} \\ \end{array} \right\} HIT$
	1	$\cdot 64$
\uparrow	129	$\cdot 64 \quad \left. \begin{array}{l} \\ \end{array} \right\} HIT$
	257	$\cdot 64$
\uparrow	129	$\cdot 64$

	$MOD 12:$
0%	$=0 \rightarrow 0$
128%	$=0 \rightarrow 0$
256%	$=0 \rightarrow 0$
128%	$=0 \rightarrow 0$
0%	$=0 \rightarrow 1$
128%	$=0 \rightarrow 0$
256%	$=0 \rightarrow 0$
128%	$=0 \rightarrow 0$
	<hr/>
1%	$=$
129%	$=$
257%	$=$
129%	$=$
1%	$=$
129%	$=$
257%	$=$
129%	$=$

Compulsory Miss : A
Conflict miss : B

	O	—	—	4Wb Set
SET0	128	—	—	
	256	—	—	
	—	—	—	
SET1	129	—	—	BS
	257	—	—	
	—	—	—	
Cache Miss is Not Full.				
SET67				

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 to 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 ____.

$$\# \text{LINES(Block)} = 256$$

[GATE-2017(Set-1)-CS: 2M]

2 Way Set
 $\# \text{SET} = \frac{256}{2} = 128$

$$\begin{aligned} K \bmod S &= i \\ K \bmod 128 - i \end{aligned}$$

2 Way Set Association with LRU

\uparrow -0 $\cdot 128 = 0 \text{ M-} \textcircled{A}$

\uparrow -128 $\cdot 128 = 0 \text{ M-} \textcircled{A}$

\uparrow -256 $\cdot 128 = 0 - \text{M} \text{ } \textcircled{A}$

\uparrow 128 $\cdot 128 = 0 \rightarrow \text{HIT}$

\uparrow 0 $\cdot 128 = 0 - \text{M} \rightarrow \textcircled{B}$

\uparrow 128 $\cdot 128 = 0 \rightarrow \text{HIT}$

\uparrow 256 $\cdot 128 = 0 \rightarrow \text{M-} \textcircled{B}$

\uparrow 128 $\cdot 128 = 0 \rightarrow \text{HIT}$

\uparrow 1 $\cdot \cdot 128 = 1 - M - A$

\uparrow 129 $\cdot \cdot 128 = 1 - M - A$

\uparrow 257 $\cdot \cdot 128 = 1 - M - A$

\uparrow 129 $\cdot \cdot 128 = 1 \rightarrow HIT$

\uparrow 1 $\cdot \cdot 128 = 1 \rightarrow m \rightarrow B$

\uparrow 129 $\cdot \cdot 128 = 1 \rightarrow HIT$

\uparrow 257 $\cdot \cdot 128 = 1 \rightarrow m \rightarrow B$

\uparrow 129 $\cdot \cdot 128 = 1 \rightarrow HIT$

$0 \cdot 1 = 0 \rightarrow M \rightarrow B$ ✓

$128 \cdot 1 = 0 \rightarrow HIT$ ✓

$256 \cdot 1 = 0 \rightarrow M \rightarrow B$

$128 \cdot 1 = 0 \rightarrow HIT$ ✓

$0 \cdot 1 = 0 \rightarrow M \rightarrow B$

$128 \cdot 1 = 0 \rightarrow HIT$ ✓

$256 \cdot 1 = 0 \rightarrow M \rightarrow B$

$128 \cdot 1 = 0 \rightarrow HIT$ ✓

$1 \cdot 1 = 1 \rightarrow M \rightarrow B$

$129 \cdot 1 = 1 \rightarrow HIT$ ✓

$257 \cdot 1 = 1 \rightarrow M \rightarrow B$

$129 \cdot 1 = 1 \rightarrow HIT$ ✓

$1 \cdot 1 = 1 \rightarrow M \rightarrow B$

$129 \cdot 1 = 1 \rightarrow HIT$

$1 \cdot 1 = 1 \rightarrow M \rightarrow B$

$257 \cdot 1 = 1 \rightarrow M \rightarrow B$

$129 \cdot 1 = 1 \rightarrow HIT$ ✓

Compulsory Miss : A
Conflict miss : B

SET 0	∅ 256	∅ 256	∅ 256	128	1257
SET 1	X 257	X 257	X 257	129	1287
SET 2					
Set 3					
Set 4					
Set 5					
Set 6					
Set 7					
Set 8					
Set 9					
Set 10					
Set 11					
Set 12					
Set 13					
Set 14					
Set 15					
Set 16					
Set 17					
Set 18					
Set 19					
Set 20					
Set 21					
Set 22					
Set 23					
Set 24					
Set 25					
Set 26					
Set 27					
Set 28					
Set 29					
Set 30					
Set 31					

No Capacity Miss
Cache Is Not Full.

if there is only 4 CM Lines
2 way Set

$$\#SET = \frac{4}{2} = 2 \text{ set}$$

↓
Capacity Miss

First Access

Compulsory Miss = 6 Miss

Conflict Miss = 4 Miss

HIT = 6 HIT

2nd Access (2nd Iteration)

No Compulsory Miss

Conflict Miss = 8 Miss

HIT = 8 HIT

① First Access : Conflict Miss = 4

② 2nd to 10th Time

Total Conflict Miss = 8×9 time = 72 Miss

Total Conflict Miss = $4 + 8 \times 9$

$$= 4 + 72$$

$$= 76 \text{ Miss} \quad \text{Ans}$$

Types of Misses

In the CM design 3 types of misses are present.

- 1) Compulsory miss - (Cold start miss / first reference miss)

This miss will occur when the very first reference to the cache itself a miss.

- 2) Capacity Miss - This miss will occur when cache is full.

- 3) Conflict Miss (Collision miss / reference miss)

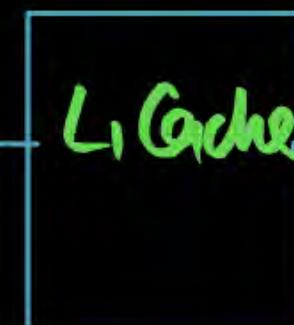
This miss will occur when the too many blocks are placed into same cache line or same cache SET.

Multilevel Cache

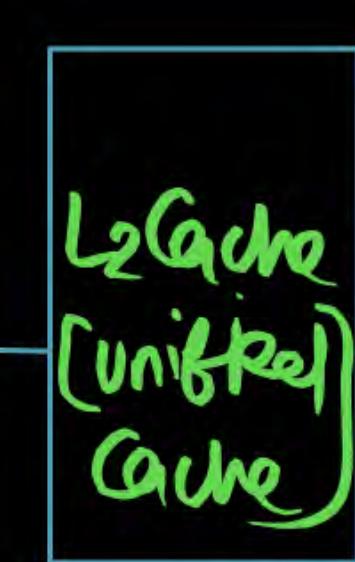
① L₁ Cache

② L₂ Cache

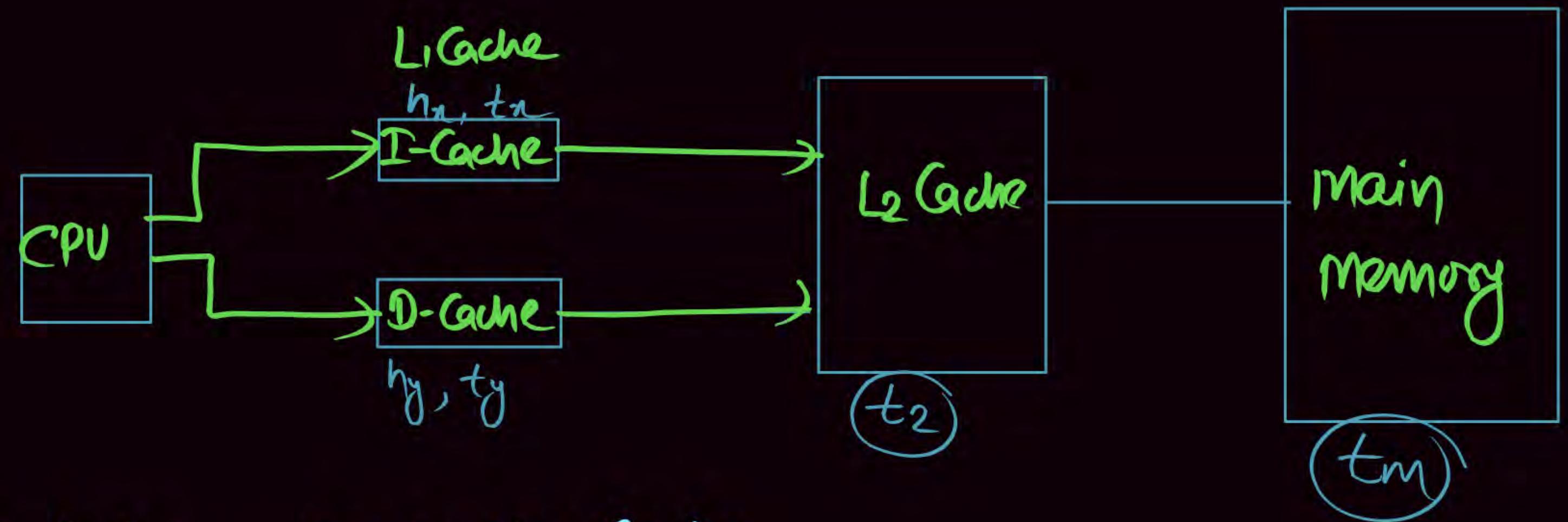
Generally:



L₁ Cache → I-Cache
L₁ Cache → D-Cache



Size of: L₂ > L₁
Speed: L₁ > L₂



I-Cache : Instruction Cache

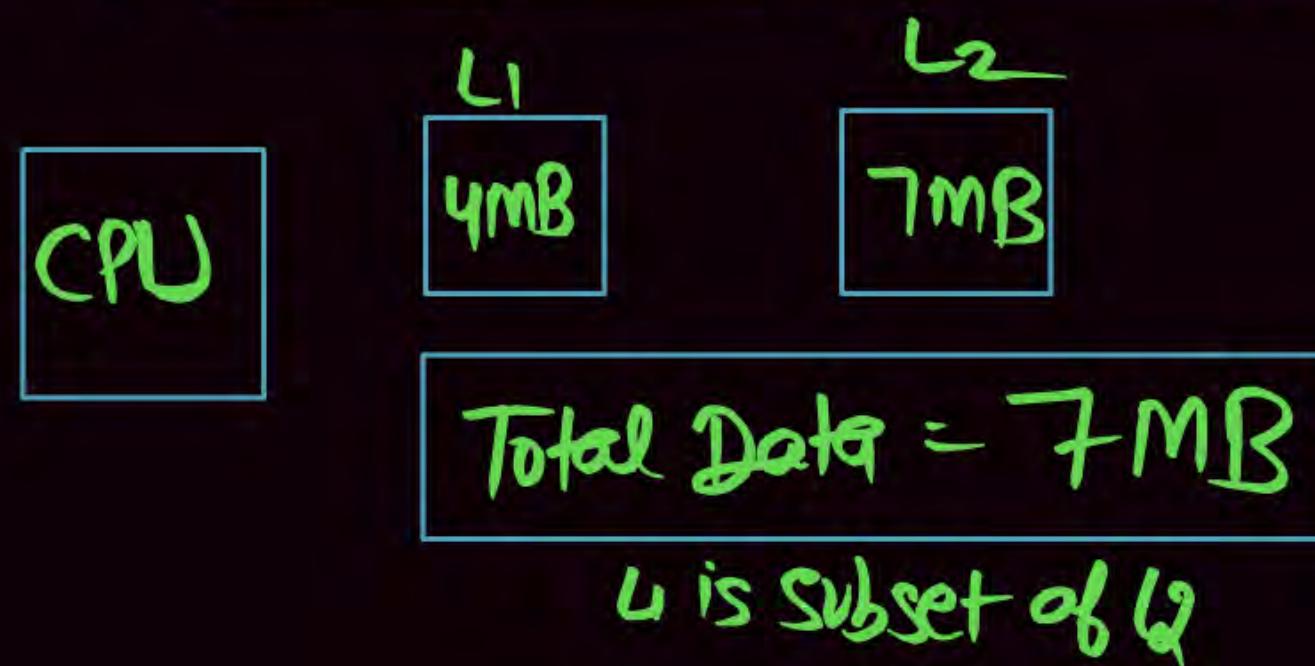
D-Cache : Data Cache

Working Principle of Multi Level Cache:

① Principle of Inclusion

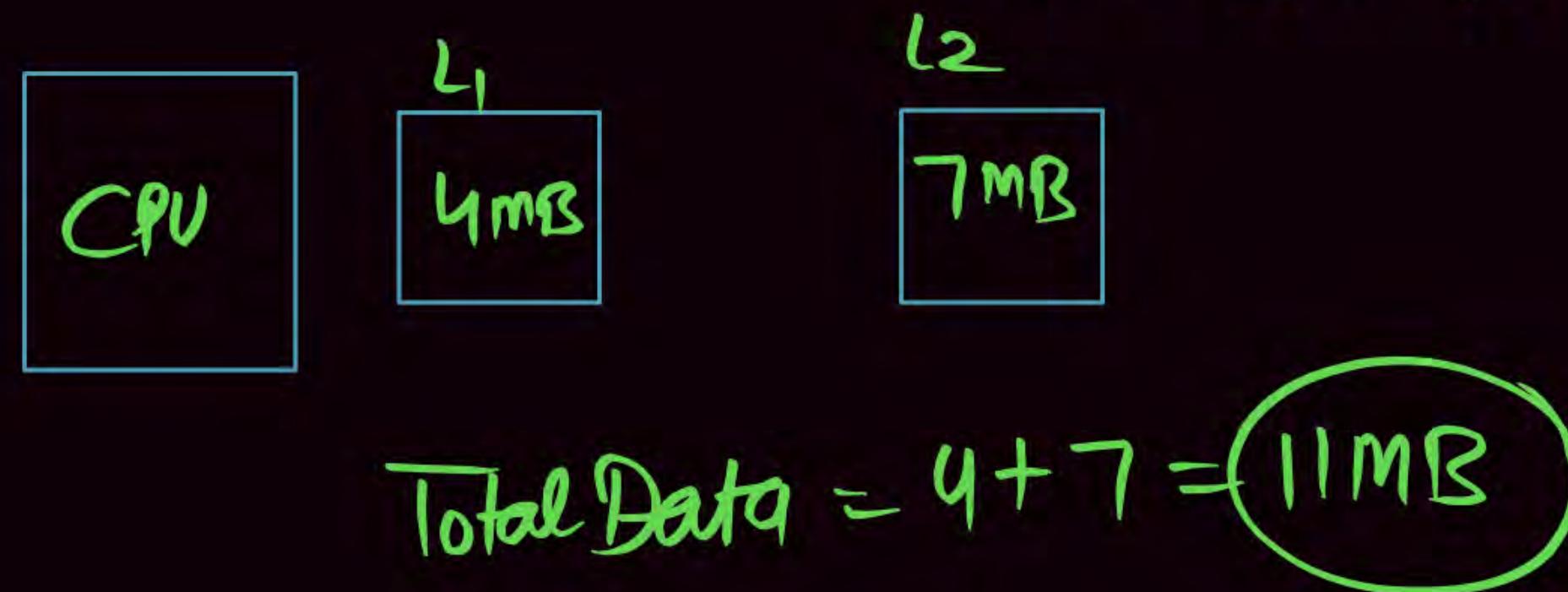
② Principle of Exclusion.

① Principle of Inclusion: Data Present in L₁ Cache Must be Present in L₂ Cache.



OR
L₁ Cache [Content/Data] is a subset of L₂ Cache.
∴ L₁ Contains Copying of L₂ Data.

② Principle of Exclusion : Data Present in L1 Cache
is differ from L2 Cache.



Q) Why we used Multi level Cache ?

Ans

L₁ Cache Access = Long

MM Access time = 300 nsec

L₁ Cache Access = long

L₂ Cache Access = long

MM Access time = 300 nsec

Tag will Reduce

CPU generates
100 Reference



40 miss

then Here
40 times Reference
to Main Memory



[40 Times Main Memory]



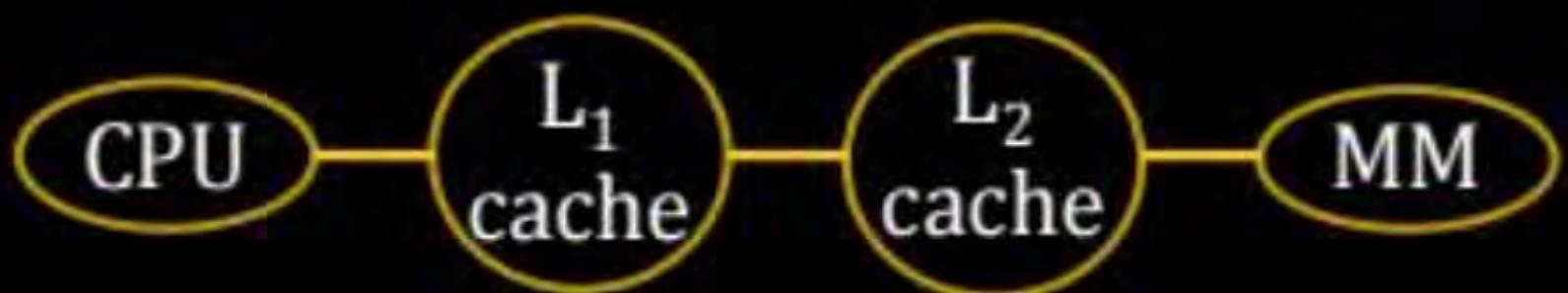
Here Memory is Access
Only 20 times.

If Only Level L Cache.
then if there is a Miss in
Level L then Refer to
Main Memory (main memory
takes very higher)
Access time

if Level L , Level 2 Cache is used.
then if there is a Miss in Level L
then go to Level 2 (we got Some
hit in L2)
if there is a Miss in Level L2
then Only that Case go to Main
memory
(so in this MM Reference will Reduced
(Bcz in Miss in L1 then Found in L2.)

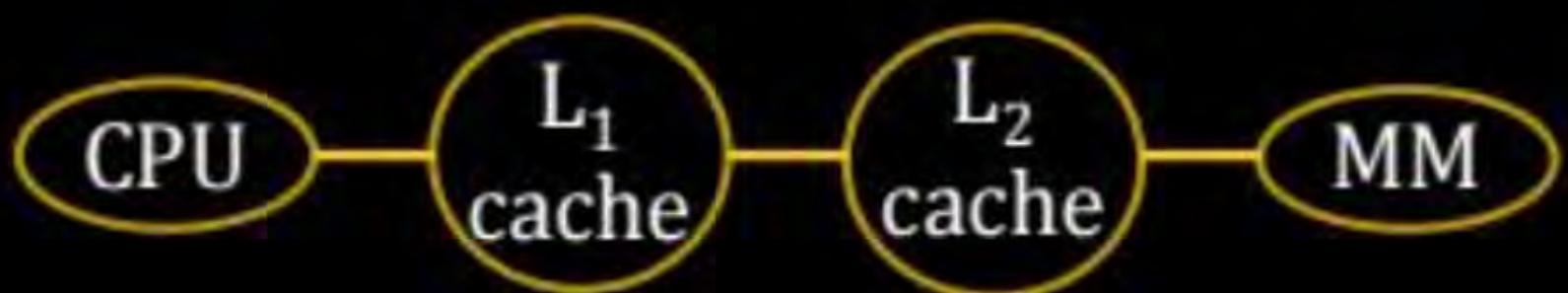
Multi level cache

- ❑ To reduce the miss penalty multi-level caches are used in the system design.
- ❑ The number of cycles required to transfer the data from higher levels to L_1 due to miss operation is called as miss penalty



Multi level cache

- ❑ To reduce the miss penalty multi-level caches are used in the system design.
- ❑ The number of cycles required to transfer the data from higher levels to L_1 due to miss operation is called as miss penalty



Local Miss Rate [LMR]

Global Miss Rate [GMR]

$$\text{Local Miss Rate} = \frac{\text{#misses in the cache}}{\text{(LMR)} \quad \text{\# accesses to that cache}}$$

$$\text{Global Miss Rate} = \frac{\text{#misses in the cache}}{\text{(GMR)} \quad \text{Total #CPU generated reference}}$$

CPU generates
100 Reference



40 Miss



20 Miss



$$LMR = \frac{\# \text{miss in cache}}{\text{Total # Access that Cache}}$$

$$GMR = \frac{\# \text{miss in cache}}{\text{Total # CPU Request}}$$

$$LMR_{L1} = \frac{40}{100} = 0.4$$

$$GMR_{L1} = \frac{40}{100} = 0.4$$

$$LMR_{L2} = \frac{20}{40} = 0.5$$

$$GMR_{L2} = \frac{20}{100} = 0.2$$

Access Time in Multi Level Cache

Hierarchical Access : L1 level

$$T_{avg} = h t_1 + (1-h)(t_m + t_1)$$

$$\cancel{ht_1} + \cancel{t_m} + t_1 - \cancel{ht_m} - \cancel{ht_1}$$

$$t_L + t_m(1-h)$$

$$T_{avg} = t_L + (1-h) t_m$$

h_1 : Hit Rate of L1

h_2 : Hit Rate of L2

$(1-h_1)$: Miss Rate of L1

$(1-h_2)$: Miss Rate of L2

Case I: If L Level Cache

Hierarchical Access

$$T_{avg} = h t_1 + (1-h) [t_m + t_L]$$

(OR)

$$T_{avg} = t_1 + (1-h) (t_m)$$

Case II: If 2 Level Cache

$$T_{avg} = h_1 t_1 + (1-h_1) h_2 (t_2 + t_1) + (1-h_1) (1-h_2) [t_m + t_2 + t_1]$$

(OR)

$$T_{avg} = \underline{t_1} + (1-h_1) \left[t_2 + (1-h_2) \underline{t_m} \right]$$

↑
same
only expansion.

Case I: If L Level Cache

Hierarchical Access

$$T_{avg} = h t_1 + (1-h) [t_m + t_L]$$

OR

$$T_{avg} = t_1 + (1-h) (t_m)$$

Case II: If 2 Level Cache

$$T_{avg} = h_1 t_1 + (1-h_1) h_2 (t_2 + t_1) + (1-h_1) (1-h_2) [t_m + t_2 + t_1]$$

OR

$$T_{avg} = \frac{t_1}{\text{Hit Time in } L_1} + (1-h_1) \left[\frac{t_2}{\text{Hit time in Level 2}} + (1-h_2) \frac{t_m}{\text{Miss Rate in Level 2}} \right]$$

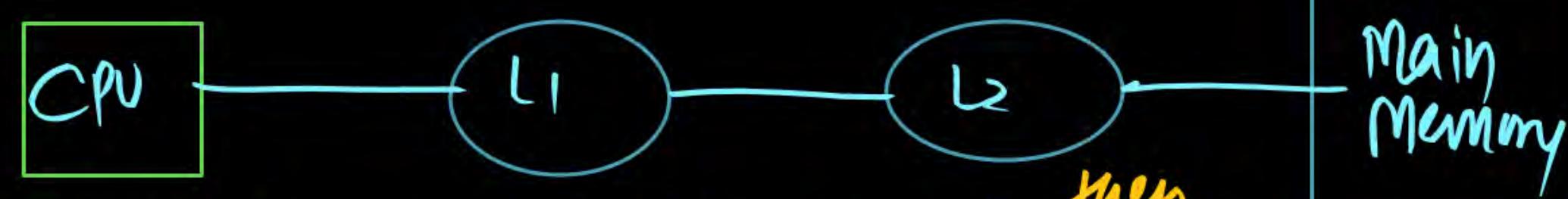
↓
SAM
Only expansion.

Average access time of the memory is always calculated in terms of Hit time, miss rate and miss penalty is as follows:

$$T_{avg} = \text{Hit time } L_1 + (\text{Miss Rate } L_1 * \text{Miss penalty } L_1) \quad \text{eq1}$$

$$\text{Miss penalty } L_1 = \text{Hit time } L_2 + (\text{Miss rate } L_2 * \text{Miss penalty } L_2) \quad \text{eq2}$$

$$\text{Miss penalty } L_2 = \text{MM Access Time} \quad \text{eq3}$$



When we put eq2 & eq3 into eq1 then we getting same formula.

- ① Hit Time
- ② Miss Rate
- ③ Miss Penalty

$$T_{avg} = \text{Hit time in } L_1 + \frac{\text{Miss Rate of } L}{\text{Hit time in } L_2} + \frac{\text{Miss Rate of } L_2 \times \text{MM Access time}}{\text{Hit time in } L_2}$$

$$T_{avg} = T_1 + (1-h_1) \left[T_2 + (1-h_2) \times t_m \right]$$

Average access time of the memory is always calculated in terms of Hit time, miss rate and miss penalty is as follows:

$$T_{avg} = \text{Hit time } L_1 + (\text{Miss Rate } L_1 * \text{Miss penalty } L_1)$$

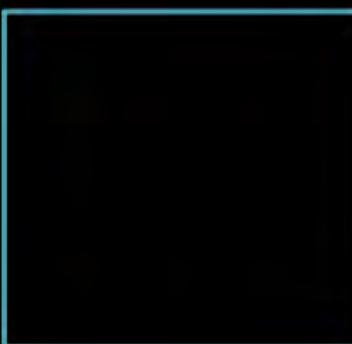
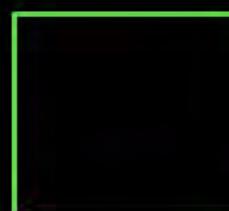
eq1

$$\text{Miss penalty } L_1 = \text{Hit time } L_2 + (\text{Miss rate } L_2 * \text{Miss penalty } L_2)$$

eq2

$$\text{Miss penalty } L_2 = \text{MM Access Time}$$

eq3



Types of Misses

In the CM design 3 types of misses are present.

- 1) Compulsory miss - (Cold start miss / first reference miss)

This miss will occur when the very first reference to the cache itself a miss.

- 2) Capacity Miss - This miss will occur when cache is full.

- 3) Conflict Miss (Collision miss / reference miss)

This miss will occur when the too many blocks are placed into same cache line or same cache SET.

Cache Update Technique:

- ① Write Through
- ② Write Back



PYQ's & Home Work Questions

Home Work Questions

- Q.1** Consider a Direct Mapping if the size of Cache memory is 512KB & Main Memory 512 KB & Cache line size (Block) is 64KB the calculate the number of bit required for
- (i) P.A
 - (ii) TAG
 - (iii) B.O
 - (iv) W.O
 - (v) #LINES
 - (vi) TAG Memory Size.

Q.2

Consider a Direct Mapping, Cache size = 64 byte, Line Size = 8

Byte. MM = 256 Byte then #bits for P.A, TAG, L.O, W.O Tag
memory size?

Q.3 Consider a Direct Mapping, Cache size = 128 KB, Line Size = 64

Byte. Main Memory is 1MB then what is the line number of physical address $(ABCDE)_{16}$?

Q.4

Consider a 2-way set associative if the size of cache memory is 512KB & Main Memory 512MB & Cache line size is 64KB then calculate the Number of bit Required for

Q.5

Consider a 2-way set associative Cache Size = 256 KB, Line size = $\frac{P}{W}$,
32 Byte, MM = 1MB, then what is the set number of Physical
address $(ABCDE)_{16}$?



GATE PYQ's

An 8-way set associative cache of size 64 KB (1 KB = 1024 bytes) is used in a system with 32-bit address. The address is sub-divided into TAG, INDEX, and BLOCK OFFSET.

The number of bits in the TAG is _____.

[GATE-2023-CS: 2M]

Consider a computer system with a byte-addressable primary memory of size 2^{32} bytes. Assume the computer system has a direct-mapped cache of size 32 KB ($1 \text{ KB} = 2^{10}$ bytes), and each cache block is of size 64 bytes.

The size of the tag field is _____ bits.

[GATE-2021(Set-1)-CS: 1M]

Consider a machine with a byte addressable main memory of 2^{32} bytes divided into blocks of size 32 bytes. Assume that a direct mapped cache having 512 cache lines is used with this machine. The size of the tag field in bits is ____.

[GATE-2017(Set-1)-CS: 2M]

Consider a machine with a byte addressable main memory of 2^{20} bytes, block size of 16 bytes and a direct mapped cache having 2^{12} cache lines. Let the addresses of two consecutive bytes in main memory be $(E201F)_{16}$ and $(E2020)_{16}$. What are the tag and cache line address (in hex) for main memory address $(E201F)_{16}$?

[GATE-2015(Set-3)-CS: 1M]

A E, 201

B F, 201

C E, E20

D 2, 01F

Q.5

[Common Data for this and next question]

A computer has a 256 Kbyte, 4-way set associative. Write back data cache with block size of 32 Bytes. The processor sends 32 bit address to the cache controller. Each cache tag directory entry contains, in addition to address tag, 2 valid bits, 1 modified bit and 2 replacement bit.

The number of bits in the tag field of an address is

- (a) 11
- (b) 14
- (c) 16
- (d) 27

[GATE - 2012: 2 Marks]

Q.6

[Common Data from previous question]

A computer has a 256 Kbyte, 4-way set associative. Write back data cache with block size of 32 Bytes. The processor sends 32 bit address to the cache controller. Each cache tag directory entry contains, in addition to address tag, 2 valid bits, 1 modified bit and 2 replacement bit.

[GATE - 2012: 2 Marks]

The size of the cache tag directory is

- (a) 160 Kbits
- (b) 136 Kbits
- (c) 40 Kbits
- (d) 32 Kbits

MCQ**Q.7**

An 8KB direct-mapped write back cache is organized as multiple blocks, each of size 32 bytes. The processor generates 32-bit addresses. The cache controller maintains the tag information for each cache block comprising of the following.

1Valid bit

1Modified bit

As many bits as the minimum needed to identify the memory block mapped in the cache.

What is the total size of memory needed at the cache controller to store meta-data (tags) for the cache? [GATE-2011-CS: 2M]

A 4864 bits

B 6144 bits

C 6656 bits

D 5376 bits

Q.8

Q.9

Common Data for next two questions:

Consider a computer with a 4-ways set-associative mapped cache of the following characteristics: a total of 1 MB of main memory, a word size of 1 byte, a block size of 128 words and a cache size of 8 KB.

While accessing the memory location 0C795H by the CPU, the contents of the TAG field of the corresponding cache line is

[GATE-2008-CS: 2M]

- A 000011000
- C 00011000

- B 110001111
- D 110010101

The number of bits in the TAG, SET and WORD fields, respectively are:

[GATE-2008-CS: 2M]

A 7, 6, 7

B 8, 5, 7

C 8, 6, 6

D 9, 4, 7

Q.10

Consider a 4-way set associative cache consisting of 128 lines with a line size of 64 words. The CPU generates a 20-bit address of a word in main memory. The number of bits in the TAG, SET and WORD fields are respectively.

- (a) 9, 6, 5
- (b) 7, 7, 6
- (c) 7, 5, 8
- (d) 9, 5, 6

[GATE - 2007]

P
W

A certain processor uses a fully associative cache of size 16 kB. The cache block size is 16 bytes. Assume that the main memory is byte addressable and uses a 32-bit address. How many bits are required for the Tag and the Index fields respectively in the addresses generated by the processor?

[GATE-2019-CS: 1M]

A 24-bits and 0-bits

B 28-bits and 4-bits

C 24-bits and 4-bits

D 28-bits and 0-bits

The width of the physical address on a machine is 40 bits. The width of the tag field In a 512 KB 8-way set associative cache is _____ bits.

[GATE-2016(Set-2)-CS: 2M]

A 4-way set-associative cache memory unit with a capacity of 16 KB is built using a block size of 8 words. The word length is 32 bits. The size of the physical address space is 4 GB. The number of bits for the TAG field is ____.

[GATE-2014(Set-2)-CS: 1M]

A cache memory unit with capacity of N words and block size of B Words is to be designed. If it is designed as a direct mapped cache, the length of the TAG field is 10 bits. If the cache unit is now designed as a 16-way set-associative cache, the length of the TAG field is _____ bits.

[GATE-2017(Set-1)-CS: 2M]

Q.15

The main memory of a computer has 2^m blocks while the cache has 2^c blocks. If the cache uses the set associative mapping scheme with 2 blocks per set, then block k of the main memory maps to the set.

[GATE - 1999]

- (a) $(k \bmod m)$ of the cache
- (b) $(k \bmod c)$ of the cache
- (c) $(k \bmod 2^c)$ of the cache
- (d) $(k \bmod 2^m)$ of the cache

The size of the physical address space of a processor is 2^P bytes. The word length is 2^W bytes. The capacity of cache memory is 2^N bytes. The size of each cache block is 2^M words. For a K-way set-associative cache memory, the length (in number of bits) of the tag field is

[GATE-2018-CS: 2M]

A $P - N - \log_2 K$

B $P - N + \log_2 K$

C $P - N - M - W - \log_2 K$

D $P - N - M - W + \log_2 K$

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.

$A_1 = 0 \times 42C8A4, A_2 = 0 \times 546888, A_3 = 0 \times 6A289C, A_4 = 0 \times 5E4880$

Which one of the following is TRUE?

[GATE-2020-CS: 2M]

- A A₁ and A₃ are mapped to the same cache set.
- 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 A₄ are mapped to different cache sets.

Consider a set-associative cache of size 2 kb ($1\text{ KB} = 2^{10}$ bytes) with cache block size of 64 bytes. Assume that the cache is byte - addressable and a 32-bit address is used for accessing the cache. If the width of the tag field is 22 bits, the associativity of the cache is ____.

[GATE-2021(set-2)-CS: 1M]

Consider a 4-way set associative cache (initially empty) with total 16 cache blocks. The main memory consists of 256 blocks 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?

[GATE-2009-CS: 2M]

A 3

B 8

C 129

D 216

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 to 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 _____.

[GATE-2017(Set-1)-CS: 2M]

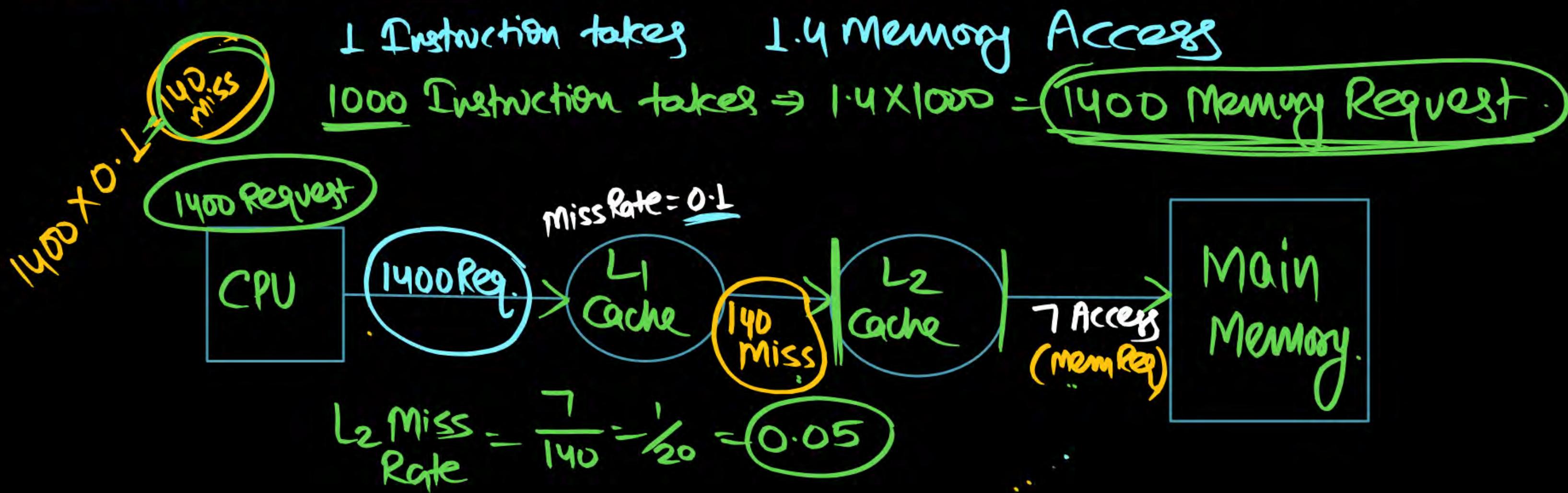
If the associativity of a processor cache is doubled while keeping the capacity and block size unchanged, which one of the following is guaranteed to be NOT affected?

[GATE-2014(Set-2)-CS: 2M]

- A Width of tag comparator
- B Width of set index decoder
- C Width of way selection multiplexer
- D Width of processor to main memory data bus

Consider a two-level cache hierarchy with L_1 and L_2 caches. An application incurs 1.4 memory accesses per instruction on average. For this application, the miss rate of L_1 cache is 0.1; the L_2 cache experiences on average, 7 misses per 1000 instructions. The miss rate of L_2 expressed correct to two decimal places is _____.

[GATE-2017(Set-1)-CS: 1M]



In a two-level cache system, the access times of L_1 and L_2 caches are 1 and 8 clock cycles, respectively. The miss penalty from the L_2 cache to main memory is 18 clock cycles. The miss rate of L_1 cache is twice that of L_2 . The average memory access time (AMAT) of this cache system is 2 cycles. The miss rates of L_1 and L_2 respectively are:

[GATE-2017(Set-2)-CS: 2M]

- A 0.111 and 0.056
- B 0.056 and 0.111
- C 0.0892 and 0.1784
- D 0.1784 and 0.0892

Q.24



Common Data for next two questions:

Consider a machine a 2-way set associative data cache of size 64Kbytes and block size 16 bytes. The cache is managed using 32 bit virtual addresses and the page size is 4 Kbytes. A program to be run on this machine begins as follows:

Double ARR [1024] [1024]

Int i, j;

```
/* Initialize array ARR to 0.0 */  
for (i = 0; i < 1024; i++)  
    for (j = 0; j < 1024; j++)  
        ARR [i] [j] = 0.0;
```

The size of double 8 bytes. Array ARR is in memory starting at the beginning of virtual page 0xFF000 and stored in row major order. The cache is initially empty and no pre-fetching is done. The only data memory references made by the program are those to array ARR.

MCQ

The total size of the tags in the cache directory is

[GATE-2008-CS: 2M]

- A 32 kbits
- B 34 kbits
- C 64 kbits
- D 68 kbits

Q.25

[Common Data for this and next question]

Consider two cache organization. The first one is 32 KB 2-way set associative with 32-byte block size. The second one is of the 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 a latency of $k/10$ ns. The hit latency of the set associative organization is h_1 while that of the direct mapped one 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

[GATE - 2006: 2 Marks]

Q.26

[Common Data from previous question]

Consider two cache organization. The first one is 32 KB 2-way set associative with 32-byte block size. The second one is of the 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 a latency of $k/10$ ns. The hit latency of the set associative organization is h_1 while that of the direct mapped one 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

[GATE - 2006: 2 Marks]

Q.27

A computer system has a level - 1 instruction cache (1-cache), a level-1 data cache(D-cache) and a level-2 cache(L2-cache) with the following specifications.

P
W

	Capacity	Mapping method	Block size
1-cache	4K words	Direct mapping	4 words
D-cache	4K words	2-way set associative mapping	4 words
L2-cache	64K words	4-way set associative mapping	16 words

Capacity mapping method block size 1-cache 4K words direct mapping 4 words D-cache 4 k words 2 way set associative mapping 4 words L2-cache 64K words 4-way set associative mapping 16 words. The length of the physical address of a word in the main memory is 30 bits. The capacity of the tag memory in the 1-cache, D-cache & L2-cache is. Respectively,

(a) $1K \times 18\text{-bit}$, $1K \times 19\text{-bit}$, $4K \times 16\text{-bit}$ (b) $1K \times 16\text{-bit}$, $1K \times 19\text{-bit}$, $4K \times 18\text{-bit}$
(c) $1K \times 16\text{-bit}$, $512 \times 18\text{-bit}$, $1K \times 16\text{-bit}$ (d) $1K \times 16\text{-bit}$, $512 \times 18\text{-bit}$, $1K \times 18\text{-bit}$

[GATE - 2006: 2 Marks]

Q.28 Consider a small two-way set-associative cache memory, consisting of 4 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

- (a) 2
- (b) 3
- (c) 4
- (d) 5

[GATE - 2004]

Q.29

Consider a system with 2 KB direct mapped data cache with a block size of 64 bytes. The system has a physical address space of 64 KB and a word length of 16 bits. During the execution of a program, four data words P, Q, R and S are accessed in that order 10 times (i.e., PQRSPQRS....) Hence, there are 40 accesses to data cache altogether. Assume that the data cache is initially empty and no other data words are accessed by the program. The addresses of the first bytes of P, Q, R and S are 0xA248, 0xC28A, 0xCA8A and 0xA262, respectively. For the execution of the above program, which of the following statements is/are TRUE with respect to the data cache? [2022: MSQ 2M]

- A Every access to S is a hit.
- B Once P is brought to the cache it is never evicted.
- C At the end of the execution only R and S reside in the cache.
- D Every access to R evicts Q from the cache.



Any Doubt ?

**THANK
YOU!**

