CS & IT ENGINEERING

COMPUTER ORGANIZATION AND ARCHITECTURE

Cache Organization

Lecture No.- 02

By- Vishvadeep Gothi sir

Recap of Previous Lecture

Topic — Associative Memory

Topic — Locality of Reference

Topic — Cache Memory

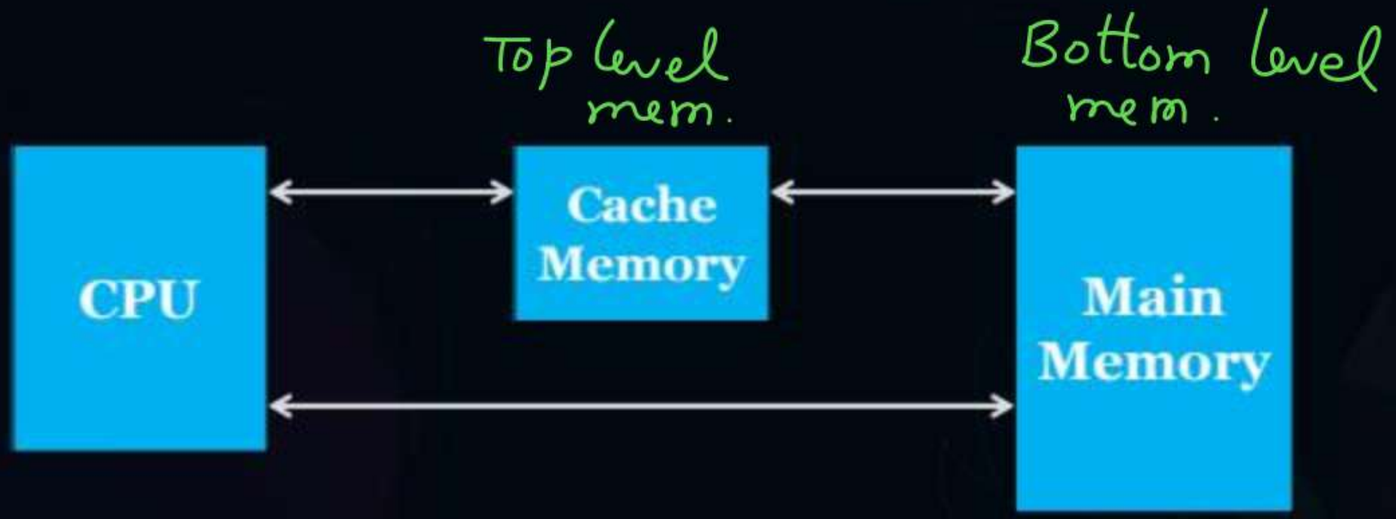
Topics to be Covered

Topic Cache Memory

Topic Average Memory Access Time

Topic Cache Write

Topic : Cache Memory



Topic : Working of Cache Memory

1. Cache Hit
2. Cache Miss
3. Hit Ratio

Topic : Average Memory Access Time

Mem. Reference

- Hit
- Miss

Avg. mem. access time = H * Time needed to access mem. when hit + (1 − H) * Time needed to access memory when miss ..... (1)

ex:-

⇒ 100 times mem. reference → 80 times hit
→ 20 times miss

⇒ for every hit, mem. access time = 10 ns

⇒ for —11— miss, ——11—— = 100 ns

Total time, for all hit = 80 * 10 = 800 nsec

——11—— ——11—— miss = 20 * 100 = 2000 nsec

Total = 2800 nsec

avg for 1 mem. access = $\frac{2800 \text{ nsec}}{100}$ = 28 nsec

$$\text{hit ratio} = \frac{80}{100} = 0.8$$

$$\text{miss ratio} = \frac{20}{100} = 0.2$$

$$\text{Avg. mem. access time} = 0.8 * 10 + 0.2 * 100$$

$$= 8 + 20$$

$$= 28 \text{ nsec}$$

$$\frac{80 * 10 + 20 * 100}{100} = \frac{80 * 10}{100} + \frac{20 * 100}{100}$$

$$= 0.8 * 10 + 0.2 * 100$$

Topic : Types of Cache Accesses

Simultaneous Access: (Parallel access)

Request for cache and main-memory are generated simultaneously

Avg. mem. access time = $H * t_{cm} + (1-H)\, t_{mm}$ ..... (2)

(Tavg.)

t_{cm} = cache mem. access time

t_{mm} = main —11—

Topic : Types of Cache Accesses

Hierarchical Access: (Serial access)

Only cache is accessed first

$$T_{avg} = H * t_{cm} + (1-H)(t_{cm} + t_{mm})$$ ..... (3)

$$H * t_{cm} + (1-H)\,(t_{cm} + t_{mm})$$

$$= \cancel{H * t_{cm}} + t_{cm} - \cancel{H t_{cm}} + (1-H)\,t_{mm}$$

$$T_{avg} = t_{cm} + (1-H)\,t_{mm}$$

Topic : When to Use Which Formula

if t_{cm}, t_{mm} given

No → Use general formula

(1)

Yes → if words 'Hierarchy' or 'Level' given

No → simultaneous access

(2)

Yes → Hierarchical access

(3)

simultaneous :- Cache searching time is zero (negligible)

cache look-up time — 11 ——

[NAT]

Ans = 95

#Q. Assume that for a certain processor, a read request takes 200 nanoseconds on a cache miss and 25 nanoseconds on a cache hit. Suppose while running a program, it was observed that 60% of the processor's read requests result in a cache hit. The average read access time in nanoseconds is_____?

General formula:-

$$T_{avg} = 0.6 * 25ns + 0.4 * 200$$

$$= 15 + 80$$

$$= 95 \text{ nsec}$$

[NAT]

#Q. In a two-level hierarchy, if the top level has an access time of 8 ns and the bottom level has an access time of 60 ns, what is the hit rate on the top level required to give an average access time of 10ns?

$t_{cm} = 8\ ns$

$t_{mm} = 60\ ns$

$t_{avg} = 10\ ns$

$H = ?$

Hierarchical access :-

$$10 = 8 + (1-H)\ 60$$

$$H = 0.96 \text{ or } 96\%$$

[NAT]

#Q. In previous question if hit rate of the top-level memory is 100%, then the access time of bottom level memory will be 60 ns?

bottom level memory → main mem.

$$T_{avg} = 1 * 8 + 0 * (8 + 60) = 8 = t_{cm}$$

$$T_{avg\ min} = T_{cm}$$

Ques) Top mem. access time = 12 ns

Bottom — 11 ——— = 200 ns

T_{avg} = 10 nsec

H = ?

Hierarchical access

→ invalid Quest^n because

$t_{avg} < t_{cm}$

[NAT]

Ans = 0.23

#Q. A computer system contains a cache. Uncached memory access takes 7 times longer than access to cache. If cache has a hit ratio 0.9. The ratio of cached memory access time to uncached memory access time is?

cache mem. ⇒ The cache

cached mem. ⇒ mem. system with cache ⇒ cache & main memory

Uncached mem. ⇒ mem. system without cache ⇒ only main mem.

$t_{cm} = x$

$t_{mm} = 7x$

$$= \frac{\text{cached mem. access time}}{\text{uncached mem. access time}}$$

$$= \frac{t_{avg}}{t_{mm}}$$

$$= \frac{0.9 * x + 0.1 * 7x}{7x}$$

$$= \frac{1.6}{7} \cong \underline{\underline{0.23}} \quad \text{Ans.}$$

Ques) $t_{cm} = 15\ ns$

$t_{mm} = 200\ ns$

$H = 90\%$

Hierarchical access

---

1 Mem. access time without cache = 200 nsec

—— // —— with cache = $15 + 0.1 * 200 = 35$ nsec

Performance gain (speed up) by using cache = $\frac{200\ ns}{35\ ns} \cong 5.7$

Topic : T_{avg} When Locality of Reference is Used

replace t_{mm} by (block transfer time from mm to cm :

↳ t_{bt}

Sim. access :-

$$t_{avg} = H * t_{cm} + (1-H)\, t_{bt}$$

Hier. access :-

$$t_{avg} = H * t_{cm} + (1-H)(t_{cm} + t_{bt})$$

or

$$= t_{cm} + (1-H)\, t_{bt}$$

[NAT]

#Q. In a two-level hierarchy, the top level has an access time of 10 ns and hit rate of 90%. If the block transfer from main memory to cache takes 500ns in case of miss then average memory access time is ____?

$$t_{avg} = 0.9 * 10 + 0.1 * (10 + 500) = 9 + 51 = 60 \text{ ns}$$

or

$$= 10 + 0.1(500)$$

$$= 60 \text{ ns}$$

#Q. From mm to cache 1 byte transfer time = 50 ns

block size = 16 bytes

---

t_{bt} = ?

t_{bt} = 16 * 50 ns = 800 ns

#2. From mm to cache 1 byte transfer time = 50ns

block size = 8 words = 8 * 4 bytes = 32 bytes

1 word = 4 bytes

---

t_{bt} = 32 * 50

= 1600 ns

#Q. from mm to cache, 1 word transfer time = 100 ns

block size = 128 bytes = $\frac{128B}{4B}$ = 32 words

1 word = 4 bytes

---

t_{bt} = 32 * 100 = 3200 ns

#Q.) block size = 16 bytes

first byte transfer from mm to cache = 10 nsec

remaining each byte —11— = 2 nsec

---

t_{bt} = 10 nsec + (15 * 2) ns

= 40 ns

[NAT]

GATE-PYQ

#Q. A ~~direct mapped~~ cache memory of 1 MB has a block size of 256 bytes. The cache has an access time of 3 ns and a hit rate of 94%. During a cache miss, it takes 20ns to bring the first word of a block from the main memory, while each subsequent word takes 5 ns. The word size is 64 bits. The average memory access time in ns (round off to 1 decimal place) is _____ ?

8 bytes

block size = 256 B = $\frac{256\,B}{8\,B}$ = 32 words

t_{cm} = 3ns

H = 94%

t_{bt} = 20ns + (31 * 5)

= 175 nsec

$$t_{avg\ sim} = 0.94 * 3 + 0.06 * 175\ ns$$
$$= 13.3$$

$$t_{avg\ hier.} = 3 + 0.06 * 175$$
$$= 13.5$$

2 mins Summary

Topic Cache Memory

Topic Average Memory Access Time

Topic Cache Write

Happy Learning

THANK - YOU