JOINT INSTITUTE
交大密西根学院

Ve370 Introduction to Computer Organization

Homework 7

1. Exercise 5.3.5

There are many different design parameters that are important to a cache's overall performance. The table below lists parameters for different direct-mapped cache designs.

	Cache Data Size	Cache Block Size	Cache Access Time
a.	32 KB	2 words	1 cycle
b.	32 KB	4 words	2 cycle

5.3.5 [20] <5.2, 5.3> Generate a series of read requests that have a lower miss rate on a 2 KB 2-way set associative cache than the cache listed in the table. Identify one possible solution that would make the cache listed in the table have an equal or lower miss rate than the 2 KB cache. Discuss the advantages and disadvantages of such a solution.

HINT: keep requesting two blocks with the same index in cache.

2. Exercise 5.4.1
3. Exercise 5.4.2
4. Exercise 5.4.3
5. Exercise 5.4.4
6. Exercise 5.4.5
7. Exercise 5.4.6

Exercise 5.4

For a direct-mapped cache design with a 32-bit address, the following bits of the address are used to access the cache.

	Tag	Index	Offset
a.	31–10	9–5	4–0
b.	31–12	11–6	5–0

5.4.1 [5] <5.2> What is the cache line size (in words)?

5.4.2 [5] <5.2> How many entries does the cache have?

5.4.3 [5] <5.2> What is the ratio between total bits required for such a cache implementation over the data storage bits?

Starting from power on, the following byte-addressed cache references are recorded.

Address	0	4	16	132	232	160	1024	30	140	3100	180	2180

5.4.4 [10] <5.2> How many blocks are replaced?

5.4.5 [10] <5.2> What is the hit ratio?

5.4.6 [20] <5.2> List the final state of the cache, with each valid entry represented as a record of <index, tag, data>.

Solution 5.4

5.4.1

a.	8
b.	16

5.4.2

a.	32
b.	64

5.4.3

a.	1 + (22/8/32) = 1.086
b.	1 + (20/8/64) = 1.039

5.4.4 3

Address	0	4	16	132	232	160	1024	30	140	3100	180	2180
Line ID	0	0	1	8	14	10	0	1	9	1	11	8
Hit/miss	M	H	M	M	M	M	M	H	H	M	M	M
Replace	N	N	N	N	N	N	Y	N	N	Y	N	Y

5.4.5 0.25

5.4.6 <Index, tag, data>

$<000001_2, 0001_2, \text{mem}[1024]>$
$<000001_2, 0011_2, \text{mem}[16]>$
$<001011_2, 0000_2, \text{mem}[176]>$
$<001000_2, 0010_2, \text{mem}[2176]>$
$<001110_2, 0000_2, \text{mem}[224]>$
$<001010_2, 0000_2, \text{mem}[160]>$

8. Exercise 5.6.1
9. Exercise 5.6.2

Exercise 5.6

Media applications that play audio or video files are part of a class of workloads called "streaming" workloads; i.e., they bring in large amounts of data but do not reuse much of it. Consider a video streaming workload that accesses a 512 KB working set sequentially with the following address stream:

0, 2, 4, 6, 8, 10, 12, 14, 16, ...

5.6.1 [5] <5.5, 5.3> Assume a 64 KB direct-mapped cache with a 32-byte line. What is the miss rate for the address stream above? How is this miss rate sensitive to the size of the cache or the working set? How would you categorize the misses this workload is experiencing, based on the 3C model?

5.6.2 [5] <5.5, 5.1> Re-compute the miss rate when the cache line size is 16 bytes, 64 bytes, and 128 bytes. What kind of locality is this workload exploiting?

0	2	4	6	8	10	...	30	32
M	H	H	H	H	H	H	H	M

5.6.1:

$\frac{1}{16}$ miss rate. It does not have any relationship with the size of the cache or the working set. They are cold misses, which cannot be avoided.

0	2	4	6	8	10	12	14	16
M	H	H	H	H	H	H	H	M

5.6.2:

16 bytes: $\frac{1}{8}$ 64 bytes: $\frac{1}{32}$ 128 bytes: $\frac{1}{64}$

- Spatial locality