Computer Architecture

Fall, 2023

Week 15

2023.12.18

組別：_____ 簽名：_____

[group4]

1.

Assume there is a machine with cache access time 1 clock cycle and main memory access time 50 clock cycles. The miss rate of this cache is 5%. What is the speedup if we add a secondary cache that has 10 clock cycles access time and miss rate 1%?

Ans:

Original: AMAT = 1 + 0.05 * 50 = 3.5 cycles

With secondary cache: AMAT = 1 + 0.05 * 10 + 0.01 * 50 = 1 + 0.5 + 0.5 = 2 cycles

Speedup = 3.5 / 2 = 1.75

[group6]

2.

(1). There are 3 types of memory design, what are they and how to distinguish them?

(2). Assume that we have 1 memory bus clock to send the address, 10 memory bus clocks for each DRAM access initiated, 1 clock for sending a word of data, and 8 words for a cache block (no overlap). Please calculate the miss penalty of (i) a four-word-wide bank of DRAMs, (ii) a two-bank, one-word-wide bus of DRAMs.

Ans:

(1).

One-word-wide memory organization： only access one word in memory and transfer one word to cache.

Wide memory organization： access n word in memory and transfer n words to cache at the same time.

Interleaved memory organization： Separate memory into several banks which can be accessed at the same time, only transfer one word to cache.

(2).

(i) 1+2x10+2=23

(ii) 1+4x10+8=49

[group5]
3.

I-cache miss rate	1%
D-cache miss rate	5%
Base CPI	2.5
Load and store instructions	40%
Miss penality	100 cycles

Consider the stats above, how much faster is an Ideal CPU?

Ans:
Miss cycle per instruction
I-cache -> 0.01 * 100 = 1
D-cache -> 0.4 * 0.05 * 100 = 2
Actual CPI = 2.5 + 1 + 2 = 5.5

Ideal CPU is 5.5 / 2.5 = 2.2 times faster

[group7]
4.
Which of the following statements are true.
(a) The main reason for Adding an L2 cache between the L1 cache and main memory is to reduce the average hit time.
(b) The main memory can be the “cache” for disk storage.
(c) We need more tag comparators in a n-way set associative cache than in a fully associative cache (assume the cache size are the same).
(d) For a data cache with 90% hit rate, 2 cycle hit latency and 100 cycle miss penalty, the average memory access time is 11 cycle.
Ans:
(b)

(a)The main reason is to reduce the miss rate.
(c)More tag comparators in fully associate cache.
(d)AMAT is 2+(1-0.9)*100=12 cycle

[group10]

5. 4096 blocks, 4-word block size, 32-bit addressing

Find the total number of tag bits for caches that are direct-mapped and two-way set associative.

Ans:

(a) Direct mapped

The number of entries in the cache = 4096 = 2^12

4 word = 2^4 bytes

Tag size = 32 - 12 - 4 = 16

Total number of tag bits = 16×2^12

(b) 2-way set associative

The number of entries in the cache = 2048 = 2^11

4 word = 2^4 bytes

Tag size = 32 - 11 - 4 = 17

Total number of tag bits = 2×17×2^11

[group8]

6. Assume that

- 2 memory bus clocks to send the address.
- 20 memory bus clocks for each DRAM access initiated.
- 2 memory bus clocks to send a word of data.
- A cache block contains 8 words.

What is the miss penalty of the three memory organizations below according to the top imformations :

1. Two-Word-Wide Bank of DRAMs
2. Eight-Word-Wide Bank of DRAMs:
3. Four-Bank, Two-Word-Wide Bus of DRAMs

ans :

1. Two-Word-Wide Bank of DRAMs :

Total Miss Penalty = Address sending time + DRAM access time + Data sending time = 2 + 80 + 8 = 90 clocks.

2. Eight-Word-Wide Bank of DRAMs:

Total Miss Penalty = Address sending time + DRAM access time + Data sending time = 2 + 20 +2 = 24 clocks.

3. Four-Bank, Two-Word-Wide Bus of DRAMs

Total Miss Penalty = Address sending time + DRAM access time + Data sending time = 2 + 20 +8= 30 clocks.

.

[group2]

7. Explain True and False

a. Virtual memory “block” is called a page.
b. There can be more than one process in the running state in the same moment
c. When calculating average memory access time, hit time is not important.
d. When CPU performance increased Miss penalty becomes more significant

Ans
a. True
b. False only one process in running state at the same time
c. False hit time is also important
d. true

[group12]

8. Which of the followings are true:

(A).To Improve cache performance,we decrease the miss ratio,miss penalty,and reduce the time to hit in the cache.
(B). LRU in hardware is costly due to the high hardware complexity.
(C). In fully associative cache, each memory block is placed in a specified cache location.
(D).In direct mapped cache,cache block is available after hit/miss.

Ans :
(A)(B) are true.
(C) In fully associative cache, each memory block can be placed in any cache location
(D) In direct mapped cache, cache block is available before hit/miss