Computer Architecture

Fall, 2019

Week 16

2019.12.23

[group6] **(對抗賽)**

1. Fill in the blanks below with same memory access sequence but different format of cache.
   Memory access sequence: 1, 6, 2, 7, 3
   Block amount of cache: 4

Ans:

Direct mapped

Block address	Cache index	Cache content after access			
		0	1	2	3
1	1		Mem[1]		
6	2		Mem[1]	Mem[6]	
2	2		Mem[1]	Mem[2]	
7	3		Mem[1]	Mem[2]	Mem[7]
3	3		Mem[1]	Mem[2]	Mem[3]

time ↓

2-way set associative (LRU)

Block address	Cache index	Cache content after access			
		Set 0		Set 1	
1	1			Mem[1]	
6	0	Mem[6]		Mem[1]	
2	0	Mem[6]	Mem[2]	Mem[1]	
7	1	Mem[6]	Mem[2]	Mem[1]	Mem[7]
3	1	Mem[6]	Mem[2]	Mem[3]	Mem[7]

time ↓

Fully associative (LRU)

Block address	Cache content after access			
1	Mem[1]			
6	Mem[1]	Mem[6]		
2	Mem[1]	Mem[6]	Mem[2]	
7	Mem[1]	Mem[6]	Mem[2]	Mem[7]
3	Mem[3]	Mem[6]	Mem[2]	Mem[7]

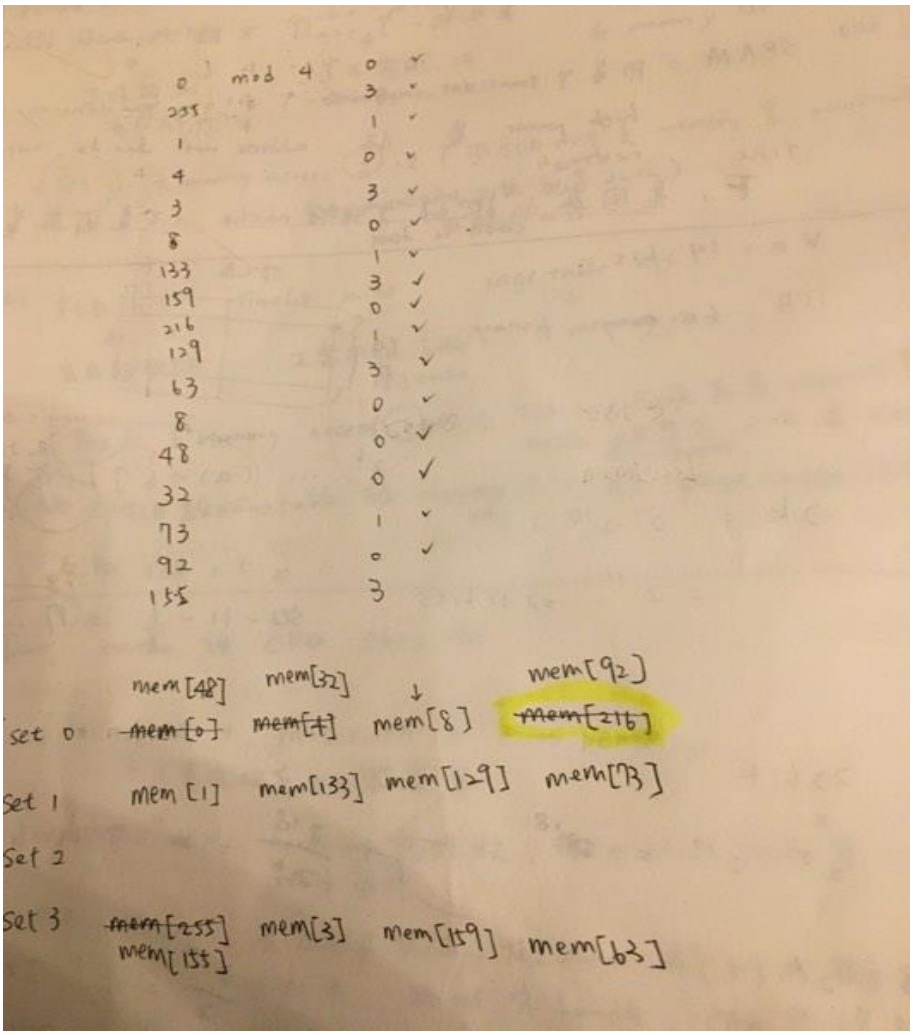
time ↓

[group7] **(對抗賽)**

2. 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?
   (a) 3
   (b) 8
   (c) 129
   (d) 216

Ans: (d)

4-way set associative so 16 block will be divided in 4 sets of 4 blocks each.
We apply(Address mod 4) function to decide set.



[group9] **(對抗賽)**

3. Revise these improper statement

   a. When we use a multilevel cache to improve the performance, the most important thing for the level-2 cache is to minimize the hit time.

   b. There are no disadvantage of using a set associativity cache to improve performance.

   c. In an interleaved memory organization, since there are several memory banks, once the data is accessed(ready), the data can be transfer concurrently.

   d. When a new process is created and is waiting for the CPU scheduler dispatch, this process is in waiting state.

Ans:

a. Minimize the miss rate

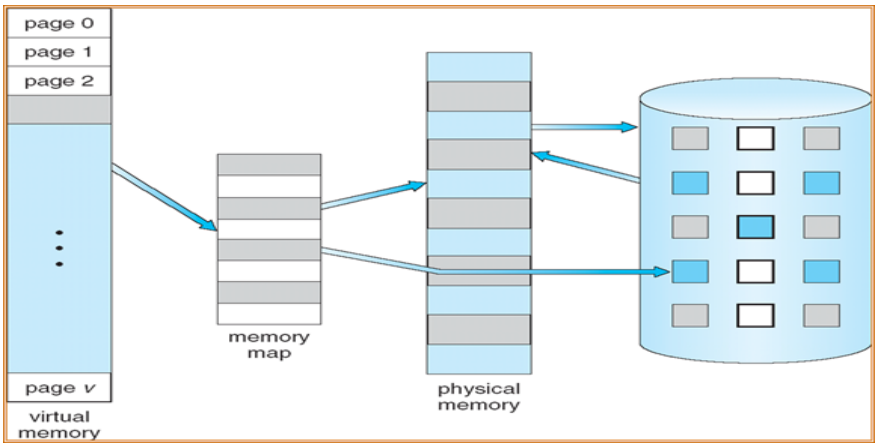
b. Data is ready after hit/miss detection and there are extra MUX delay.

c. data still transfer sequentially.

d. ready state.

[group11] **(對抗賽)**

4. Explain the process of the memory mapping in the following picture.



consider three scenarios

1. The memory we want is already in the physical memory
2. The memory we want is NOT in the physical memory and there still has space in physical memory
3. If the memory we want is NOT in the physical memory and there has NO space in physical memory

Ans:

1. if the memory we want is already in the physical memory, we can simply access in physical memory.
2. if the memory we want is NOT in the physical memory and there still has space in physical memory => accessing memory from disk and moving it to physical memory.
3. if the memory we want is NOT in the physical memory and there has NO space in physical memory => moving one memory from physical memory, then accessing memory from disk and move to physical

memory.

[group3] **(對抗賽)**

5. 考慮 4 block caches, 6 bits address 有三種 associativity 方式, 分別是 Direct mapped, 2-way set associative, Fully associative 請根據下面的 input 情況, 比較三種方式在此情況下的優劣順序。(Hint: 請用 LRU replacement policy and the LSB indexing scheme 來計算 miss/hit)

Input
010010
110010
010010
000001
000100
110010

Ans:

Direct mapped	2-way set associative	Fully associative
Miss	Miss	Miss
Miss	Miss	Miss
Miss	Hit	Hit
Miss	Miss	Miss
Miss	Miss	Miss
Miss	Miss	Hit

Miss rate:

Direct mapped = 6 / 6 = 1

2-way set associative = 5 / 6

Fully associative = 4 / 6 = 2 / 3

此種情況的優劣為 Fully associative > 2-way set associative > Direct mapped

[group2]

6. 請選出錯的選項。

(A)在 VM Translation 中，”Miss” 被稱為 Page fault

(B)在 n-way associativity 中，當 n 越大時，降低 miss rate 的效果越低

(C)Average Memory Access Time = hit rate*hit time + miss rate*miss penalty

(D)N-Way Set-Associative Cache’s data comes after hit/miss decision

Ans:

C. Average Memory Access Time = 1*hit time + miss rate*miss penalty

[group12] **(對抗賽)**

7. Consider an Intel P4 microprocessor with a 16 Kbyte unified L1 cache. The miss rate for this cache is 3% and the hit time is 2 clock cycles. The processor also has an 8 Mbyte, on-chip L2 cache. 95% of the time, data requests to the L2 cache are found and the hit time is 15 clock cycles. If data is not found in the L2 cache, a request is made to a 4 Gbyte main memory. The time to service a memory request is 100,000 clock cycles (page fault) and the hit time is 200 clock cycles (memory latency). On average, it takes 3.5 clock cycles to process a memory request. How often is data found in main memory?

Ans:

AMAT = Hit time + (Miss Rate × Miss Penalty), so

AMAT= Hit time of L1 + (Miss Rate of L1 × Miss Penalty of L1)

Miss Penalty of L1 = Hit time of L2 + (Miss Rate of L2 × Miss Penalty of L2)

= 15 + (1-0.95) × Miss Penalty of L2

Miss Penalty of L2 = Hit time of Main + (Miss Rate of Main × Miss Penalty of Main)

= 200 + X(suppose) × 100000

Therefore,

3.5 = AMAT = 2 + 0.03 (15 + 0.05 × Miss Penalty of L2)

= 2 + 0.03 (15 + 0.05 × (200 + 100000X))

= 2 + 0.03 (15 + 10 + 5000X)

= 2 + 0.75 + 150X

150 X = 0.75

X = 3/600 = 1/200 = 0.5%

Thus, (100-0.5)% of the time, we find the data we are looking for in main memory.

[group5]

8. The following statements, please choose which are false, and why it is false.
   A. The basic unit of virtual memory is page.
   B. The number of virtual pages is always equal to the number of physical pages.
   C. Virtual memory is implemented and managed by the operating system.
   D. When the program is interrupted, its state goes from running to waiting.
   E. Virtual memory maps cache to memory.

Ans:

B, D, E

B: The number of virtual pages can be larger than that of physical pages.

D: It should be go from running to ready.

E: Maps memory to disk.