Computer Architecture

Fall, 2018

Week 13

2018.12.3

[group10]

1. Choose the statement that is wrong and explain why.
   (a) Spatial locality stating that if a data location is referenced, data locations with nearby addresses will tend to be referenced soon.
   (b) Temporal locality stating that if a data location is referenced then it will tend to be referenced again soon.
   (c) Memory hierarchy is a structure that uses multiple levels of memories; as the distance from the processor increases, the size of the memories and the access time both increase.
   (d) Miss penalty, the time required to fetch a block into a level of the memory hierarchy from the lower level, including the time to access the block, transmit it from one level to the other, insert it in the level that experienced the miss, and then pass the block to the requestor.
   (e) Larger blocks exploit spatial locality to lower miss rates. Increasing the block size always decreases the miss rate.

Ans:

(e) is wrong. Reason:不一定，如果 block size 最後占了 cache size 的一大部分,則 miss rate 會上升因為能在 cache 裡面被利用的 blocks 數會減少，在讀 words 前某些 block 會被擠壓出去，造成 miss rate 的效益降低

[group9] **(對抗賽)**

2. 根據下表格 (A 8-blocks, 1 word/block, direct mapped cache.)

Index	Valid	Tag	Data
000	0		
001	0		
010	1	10	Mem[10010]
011	0		
100	0		
101	1	01	Mem[01101]
110	0		
111	1	10	Mem[10111]

完成表格

Word addr	Binary addr	Hit/Miss	Cache block
16			
23			
19			
31			
4			
23			

Ans:

Word addr	Binary addr	Hit/Miss	Cache block
16	**10 000**	**Miss**	**000**
23	**10 111**	**Hit**	**111**
19	**10 011**	**Miss**	**011**
31	**11 111**	**Miss**	**111**
4	**00 100**	**Miss**	**100**
23	**10 111**	**Miss**	**111**

[group4] **(對抗賽)**

3. 以下關於 memory hierarchy 的敘述何者錯誤？
   (a) Block is the basic unit of information transfer
   (b) Loop is an example of spatial locality
   (c) Hierarchy between cache and memory is managed by hardware
   (d) DRAM is slow, cheap, and dense while SRAM is fast, expensive, and not very dense
   (e) Miss penalty is the time to deliver the block to the processor

Ans:

(B) 應改為 temporal locality

(E) Miss penalty is the time to deliver the block to the processor + time to replace a block in the upper level

[group6] **(對抗賽)**

4. 假設有一個 cache，存取時間為 a，而存取下一階記憶體的時間為 b (b = miss penalty)，cache 目前的內容如下：

Index	Val	Tag	Data
000	Y	10	Mem[10000]
001	N		
010	Y	11	Mem[11010]
011	Y	00	Mem[00011]
100	N		
101	N		
110	Y	10	Mem[10110]
111	N		

然後接受了以下記憶體位址的存取要求，請列出所耗的時間。

10 001

11 010

01 110

00 101

10 110

10 001

Ans:

6a + 4b

P.S. 這題答案與我不同(我不太懂題目意思)

[group1]

5. Use SRAM and DRAM to fill up the blank.

	High	Low
Density		
Power		
Cost		
Speed		

Ans:

	High	Low
Density	**DRAM**	**SRAM**
Power	**SRAM**	**DRAM**
Cost	**SRAM**	**DRAM**
Speed	**SRAM**	**DRAM**

[group11] **(對抗賽)**

6. How many total bits are required for a direct-mapped cache with 16 KB of data and 4word blocks, assuming a 32-bit address? (Hints, 4 word per block; 4 bytes per word)

Ans:

We know that 16 KB is 4K (2^12) words. With a block size of 4 words (2^2), there are 1024 (2^10) blocks. Each block has 4 × 32 or 128 bits of data plus a tag, which is 32 – 10 – 2 – 2 bits, plus a valid bit. Thus, the total cache size is

2^10 × (4 × 32 + (32 - 10 - 2 - 2) + 1) = 2^10 × 147 = 147 Kbits

or 18.4 KB for a 16 KB cache. For this cache, the total number of bits in the cache is about 1.15 times

as many as needed just for the storage of the data.

[group2] **(對抗賽)**

7. (a) Name the five memory components in the memory hierarchy from fastest to slowest.
   (b) Name the four memory data transferring blocks from biggest sized to smallest sized.
   (c) On what is the effectiveness of the cache memory based on? Explain the principle of it and name the two types of it.

Ans:

(a) Register -> Cache -> Memory -> Disk -> Tape

(b) Files -> Pages -> Blocks -> Operands

(c)

***Principle of Locality*:**

- **Program access a relatively small portion of the address space at any instant of time**
- **90/10 rule: 10% of code executed 90% of time**

Two types of locality:

- **Temporal locality: if an item is referenced, it will tend to be referenced again soon, e.g., loop**
- **Spatial locality: if an item is referenced, items whose addresses are close by tend to be referenced soon., e.g., instruction access, array data structure**

[group3] **(對抗賽)**

8. A computer system uses 16-bit memory addresses. It has a 2K-byte cache organized in a direct-mapped manner with 64 bytes per cache block. Assume that the size of each memory word is 1 byte.
   (a) Calculate the number of bits in each of the Tag, Block, and offset of the memory address.
   (b) When a program is executed, the processor reads data sequentially from the following word addresses: 128, 144, 2176, 2180
   All the above addresses are shown in decimal values. Assume that the cache is initially empty. For each of the above addresses, indicate whether the cache access will result in a hit or a miss.

Ans:

(a) Block size = 64 bytes = 2^6 bytes = 2^6 words (since 1 word = 1 byte)

Therefore, Number of bits in the Word field = 6

Cache size = 2K-byte = 2^{11} bytes Number of cache blocks = Cache size / Block size = $2^{11}/2^6 = 2^5$

Therefore, Number of bits in the Block field = 5

Total number of address bits = 16

Therefore, Number of bits in the Tag field = 16 - 6 - 5 = 5

For a given 16-bit address, the 5 most significant bits, represent the Tag, the next 5 bits represent the Block, and the 6 least significant bits represent the Word.

(b) The cache is initially empty. Therefore, all the cache blocks are invalid.

Access # 1:

Address = $(128)_{10}$ = $(0000000010000000)_2$

For this address, Tag = 00000, Block = 00010, Word = 000000

Since the cache is empty before this access, this will be a cache miss

After this access, Tag field for cache block 00010 is set to 00000

Access # 2:

Address = $(144)_{10}$ = $(0000000010010000)_2$

For this address, Tag = 00000, Block = 00010, Word = 010000

Since tag field for cache block 00010 is 00000 before this access, this will be a cache hit (because address tag = block tag)

Access # 3:

Address = $(2176)_{10}$ = $(0000100010000000)_2$

For this address, Tag = 00001, Block = 00010, Word = 000000

Since tag field for cache block 00010 is 00000 before this access, this will be a cache miss (address tag ≠ block tag)

After this access, Tag field for cache block 00010 is set to 00001

Access # 4:

Address = $(2180)_{10}$ = $(0000100010000100)_2$

For this address, Tag = 00001, Block = 00010, Word = 000100

Since tag field for cache block 00010 is 00001 before this access, this will be a cache hit (address tag = block tag)