Computer Architecture
Fall, 2017
Week 14
2017.12.11

1.

[group1]

1. 對於一個計算機結構學家來說觀念是很重要的，請幫初出茅廬的小明解釋以下觀念題-

(a) Larger blocks would both reduce miss rate and increased miss rate, Why?

(b) Three reasons for larger blocks would have larger miss penalty, one is more access time and transmit time, another is pollution, and the other is it can override benefit of reduced miss rate. Define the word “Pollution”.

Answer:

(a) 會減少是因為 spatial locality 的特性，把附近的 data 也都一起搬進來，而會增加是因為把 blocks 設定的很大的話，在 block size 固定的情形下，blocks 的數目會變小，會有更多的 memory block map 到這個位置，競爭增加，增加 miss rate

(b) block 有被寫的話，表示有被汙染(pollution)，需要把 cache 過程中更改的，寫回去 main memory，沒被汙染(pollution)，表示只有被 read，都沒被寫。

2.

[group4]

2. Assume that there is a 32 blocks cache and each block is 32 bytes. How many bits are used for tag, index, and offset in 32-bit memory address? To what block number does address 1280 map to? (direct-mapped, byte addressable)

Ans:

Block address = 1280/32 = 40

Block number = 40 % 32 = 8

Tag: 22 bits, Index: 5 bits, Offset: 5 bits

3.

[group3]

Q1：小明昨天晚上熬夜打 code，所以今天上課時一直昏昏沉沉很想睡覺，在半睡半醒之中，隱約聽到老師說 locality 有兩種，之後就睡著了。好心的同學，請你幫幫小明，請跟他解釋 locality 有兩種分別是哪兩種，解釋且各舉一個例子。

A1：

1.時間區域性 temporal locality：一筆資料若被存取，則可能很快地再次被存取

Ex : Loop

2.空間區域性 spatial locality：一筆資料若被存取，則其附近的資料也有即將被存取的傾向。

Ex : Array

4.

[group7]

1.請問 block size 的大小造成哪些優缺點?

ans:

如果 block size 大，會減少 miss rate,但太大會需要頻繁的 swap blocks,反而會減少效能

5.

[group9]

1. compare SRAM with DRAM

	power	price	speed	density	use for	address divided or not	dynamic or static
SRAM							
DRAM							

ans:

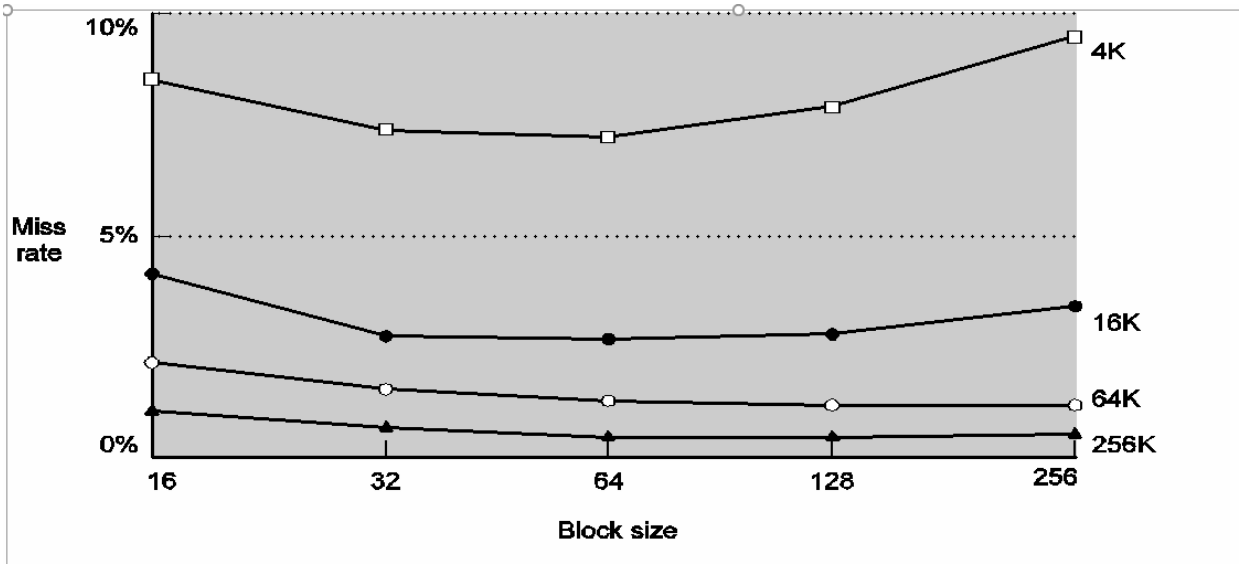
	power	price	speed	density	use for	address divided or not	dynamic or static
SRAM	high	expensive	fast	low	cache	not divided	static
DRAM	low	cheap	slow	high	main memory	address in 2 halves	dynamic

6.

[group11]

下圖是 miss rate 與 block size 的關係圖:

我們知道發現適當的增大 block size 可以降低 miss rate ，但 block size 只要增加到一定大小的時候，miss rate 在某些情況下反而上升了（如下圖 4k and 16k）。 請問這是為什麼呢？



Ans: 因為在同樣的 cache size 無限量增加 block size 的大小，反而會使 cache 內的 block 的數量太少，還有另一個提高 block size 會造成 miss penalty 變大，因為一旦 miss，你須要轉移更多的 memory。

7.
[group12]
1.
Briefly describe two write policy:
i) Write through
ii) Write back
Do we need dirty bit when using write through policy? Why?

Write through – 隨時與 cache 同步，cache 中的內容改變，也同時改變 memory 中的內容
Write back – 利用 dirty bit 為 flag，當 cache 中的內容要被替換掉時，檢查 dirty bit，如果內容有改才寫回 memory，若沒有則直接替換

不需要。因為 write through 隨時跟 memory 溝通，不需做記號。

8.
[group13]
Assume that there is a 64-block cache and each block is 16 bytes. How many bits are used for tag, index, and offset in 32-bit memory address? What block number does address 1120 map to?
Ans:
Tag:22
Index:6
Offset:4
Block:6

9.

[group8]

1. Which of the following statements are generally ture?

   (1) Cache take advantage of temporal locality.

   (2) On a read, the value returned depands on which blocks are in the cache.

   (3) Most of the cost of the memory hierarchy is at the highest level.

   (4) Most of the capacity of the memory hierarchy is at the lowest level.

<Sol> (1)(4)

10.

[group2]

Q2:

Cache: 64blocks, 16bytes/block

What does cache block number does address 1400 map?

Ans:

1400/16 = 87.5 =>floor(87.5) = 87, means 87th block

87 mod 64 = 23, so block number should be 23.