Computer Architecture

Fall, 2024
Week 13
2024.11.25

組別：_____ **簽名：**_____

[group 1]

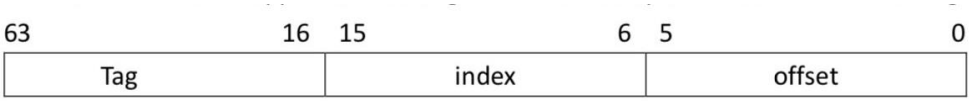
1. Why might increasing block size increase conflict misses?

Ans:

Larger block sizes reduce the number of blocks that can fit into the cache for a given cache size. This increases the likelihood that multiple addresses will map to the same cache set (or block, in a direct-mapped cache), causing evictions of useful data and resulting in conflict misses.

[group 2]

2. Consider a direct-mapped cache design with a 64-bit byte-address of the following format



What is the cache size in kibibytes (i.e. ,KiB) for the data storage?

Note that: 2**10 bytes=1KiB

Ans.

(2^10) * (2^6) = 2^6 KiB

[group 3]

3. How is the memory hierarchy managed in computer systems?

Ans

The memory hierarchy in computer systems is managed as follows:

1. Registers <-> Memory
    - Managed by the complier, and possibly by the programmer.
2. Cache <-> Memory
    - Managed by the hardware.
3. Memory <-> Disks
    - Managed by the hardware and operating system through virtual memory.
    - Managed by the programmer for file storage and retrieval.

[group 4]

4. Choose the correct answers and also explain if it is false?

(A) Large blocks maybe reduce miss rate due to temporal locality.

(B) DRAM need to be refreshed regularly.

(C) A write buffer operates in a LIFO manner.

(D) DRAM has slower access times compared to SRAM.

Ans: B D

(A)Large blocks maybe reduce miss rate due to spatial locality.

(C)A write buffer operates in a FIFO manner.

[group 5]

5. 請詳細說明並比較 **SRAM（靜態隨機存取記憶體）** 與 **DRAM（動態隨機存取記憶體）** 的差異：

1) **存取時間**：兩者在存取時間上的表現如何？
2) **密度與功耗**：SRAM 與 DRAM 在記憶體密度和功耗上有什麼不同？
3) **價格與應用**：哪一種記憶體較昂貴？它們各自常用於哪些應用場景？
4) **記憶體內容保持**：SRAM 的資料如何保持？DRAM 為什麼需要定期刷新？
5) **地址處理方式**：SRAM 與 DRAM 在地址處理上的差異是什麼？特別是 DRAM 如何使用 RAS/CAS 來存取資料？

Ans:

1) **存取時間：**

- **SRAM**：存取時間較短，速度快。因為 SRAM 不需要刷新資料，並且結構上不需要經過額外的資料存取流程，適合用於對速度要求較高的場景。
- **DRAM**：存取時間較長，速度較慢。由於 DRAM 的資料儲存在電容中，且需要經過刷新，因此速度不如 SRAM。

2) **密度與功耗：**

- **SRAM**：低密度、高功耗。每個記憶單元由多個電晶體組成，這使得它的密度較低。

- **DRAM**：高密度、功耗大部分時間較低。

3) **價格與應用**：

- **SRAM**：價格較昂貴。由於其較為複雜的結構和較低的密度，生產成本較高。通常用於需要快速存取的場合，例如 CPU 的**快取記憶體**。
- **DRAM**：價格較便宜。由於其高密度和低成本，最常見的應用是**主記憶體**。

4) **記憶體內容保持**：

- **SRAM**：資料是**靜態**的，只要有電源供應，資料就會一直保持下去，不需要刷新，因此適合用於需要長期保留資料且速度要求高的環境。
- **DRAM**：資料是**動態**的，必須不斷**刷新**（定期充電），因為電容中的電荷會隨時間流失。如果不刷新，資料會丟失。

5) **地址處理方式**：

- **SRAM**：地址是完整的，不需要分成兩半處理。
- **DRAM**：地址被分成兩半，透過 RAS/CAS（Row Access Strobe/Column Access Strobe，列/行選擇訊號）進行存取。這意味著 DRAM 的記憶體是一個 2D 矩陣，這種結構雖然增加了存取時間，但使得 DRAM 能以較低的成本提供大容量的存儲。

[group 6]

6. True or false. If false, write down the correct answer.

   A. Hit time is always greater than the miss penalty.

   B. Miss penalty includes the time to replace a block in the upper level and the time to deliver the block to the processor.

   C. Miss rate is equal to the hit rate.

   D. Hit time only refers to the time to determine a hit or miss.

Ans:

Correct Answer: B

Explanation:

a) A is incorrect: The hit time is significantly smaller than the miss penalty, as mentioned in the slide ("Hit Time << Miss Penalty").
b) B is correct: Miss penalty indeed includes two components: (1) the time to replace a block in the upper level and (2) the time to deliver the block to the processor.
c) C is incorrect: The miss rate is the complement of the hit rate, calculated as Miss Rate=1−Hit Rate.
d) D is incorrect: Hit time includes both the RAM access time and the time to determine whether it's a hit or miss.

[group 7]

7. Regarding block size considerations in a cache memory system, which of the following statements are correct?

A. Larger blocks generally reduce the miss rate due to increased spatial locality.

B. Increasing block size in a fixed-sized cache will always decrease the miss rate.

C. Larger blocks in a fixed-sized cache mean fewer blocks, leading to more competition and potentially increasing the miss rate.

D. Larger blocks reduce the miss penalty because more data can be accessed at once.

E. Larger blocks may increase cache pollution, where unnecessary data replaces useful data.

Ans: A, C, E

B: In a fixed-sized cache, larger blocks mean fewer blocks, which may lead to more competition and, consequently, an increased miss rate.

D: Larger blocks increase transmission time and access time, thereby increasing the miss penalty.

[group 14]

8. How do the principles of temporal and spatial locality influence the design choices for cache memory in terms of block size and associativity?

Ans:

For temporal locality: Temporal locality implies that if a memory location is accessed, it is likely to be accessed again soon. This influences cache design by encouraging the reuse of data in smaller, faster levels of the memory hierarchy.

For spatial locality: Spatial locality suggests that when one memory location is accessed, nearby locations are likely to be accessed soon. This motivates larger block sizes in caches, as fetching a block of data brings in adjacent data, leveraging spatial locality. However, increasing block size may also increase miss penalties due to fewer blocks fitting in the cache, leading to higher conflict misses.

[group 13]

9. Which of the following statements are correct?
(多選題)
A. In the memory hierarchy, the lower level is typically larger, slower, and uses less expensive technology than the upper level.
B. Temporal locality refers to the tendency of programs to access data that is physically close in memory addresses.
C. A block is the smallest unit of data that can be transferred between adjacent levels in the memory hierarchy.
D. Spatial locality is observed when accessing items that are close in memory addresses, such as array elements.
Ans：A、C、D

[group 11]

10. In the Memory Hierarchy, what does the upper level refer to, and what are its typical characteristics compared to lower levels?

Ans:

The upper level in the Memory Hierarchy refers to the memory levels closer to the processor. Its typical characteristics are:

- Higher Speed: It is faster in accessing data compared to lower levels.
- Smaller Capacity: It has a smaller storage size.
- Higher Cost: It is more expensive per unit of storage.

[group 9]

11. 請依照與 processor 的距離有遠至近排列

a.Disk b.Tape c.Register d.Memory e.Cache

越靠近 processor 是越快還是越慢 ？

Ans: c-e-d-a-b,快