

CS151B/EE116C - Homework #8, Due 5/30/2018 (no late homeworks)

Reading Assignment:

- Appendix B: B-58 – B-65
- Chapter 5: pp. 374-418
- Preparation for next week's material: Chapter 5: pp. 427-444

Problems:

- (1) Compute the effective access time, in seconds, of a memory system consisting of a cache and main memory with a processor running at 2GHz. The cache hit time is one cycle. The miss penalty is 55 cycles. The hit rate is 94%.
- (2) Compute the **total** number of bits of storage required to implement a cache that is similar to the cache shown in Figure 5.18 on page 408, except that the associativity of the cache is 8 instead of 4. All the other parameters (cache size, block size) are the same as in Figure 5.18. Note that the number you are being asked to compute is different from the "size of the cache," which refers only to the number of bits of data stored in the cache.
- (3) A CPU generates 48 bit addresses. The memory system is word-addressable. The cache contains 2048 block frames. The tag size is 33 bits. Based on this information, what is the block size? If a range of block sizes is possible, specify this range.
- (4) 5.2.2 in the book. Note that the problem refers to a reference trace shown higher on the page (problem 5.2).

In addition to what is asked in the book, show the final contents of the cache. This means that, for each block frame that contains a valid block, show the block number, the word addresses of the words in the block, and the tag of that block.

- (5)-(7) 5.5.1, 5.5.2, 5.5.3

Note: The minimum addressable unit of the processor is a byte and the addresses shown in the address stream are byte addresses.

- (8) The minimum addressable unit of a memory system is a 32-bit word. The memory system includes a four-way set-associative cache. The block size is eight words.

The **data** part of the cache is implemented using a single conventional $64K \times 32$ RAM chip. We will call this chip **C1**. The connections to **C1** are as shown on page 10.24 in the notes. You cannot modify **C1** in any way.

For each address generated by the CPU, the cache controller, using the directory part of the cache, maps that address to an address in **C1**.

- A) The CPU generates the address 0x321. This access is a **hit** in the cache. Specify the address in **C1** where the word is found. If more than one answer is possible for your assumed implementation of the cache controller and directory, you must specify all possible answers.

Addresses must be specified in hex

- B) Repeat part A for address 0x9876543.

- (9) Consider the multicycle MIPS implementation described in Figures 5.28 and 5.37 in Chapter 5, **3rd Ed.** This processor executes a program with the following instruction mix:

loads	27%	branches	20%
stores	13%	jumps	2%
R-format	38%		

Unlike figure 5.28, the memory subsystem of the system you are considering consists of a **write-back, write-around** unified instruction/address cache and main memory. The access time of the cache is 1 cycle. The access time to main memory is 6 cycles. The block size is 8 bytes (2 words). The main memory word size is 32-bits. On average, 30% of the blocks are dirty. There is no write buffer.

- A) Compute the CPI of the system assuming the cache hit rate is 100%.
- B) For what cache hit rate will performance decrease (i.e., execution time increase) by a factor of 2. State every assumption and be sure that it is clear how you arrive at your answer.

Practice problems: You do not need to hand in a solution to the problems below.

- (10) Consider the following series of references:
2, 3, 11, 16, 21, 13, 64, 48, 19, 11, 3, 22, 4, 27, 6, 11
Assuming a direct-mapped cache with 16 one-word blocks that is initially empty, label each reference in the list as a hit or a miss and show the final contents of the cache.
- (11) Using the series of references given in problem 10, show the hits and misses and final cache contents for a two-way set-associative cache with one-word blocks and a total size of 16 words. Assume LRU replacement.
- (12) Using the series of references given in problem 10, show the hits and misses and final cache contents for a fully associative cache with one-word blocks and a total size of 16 words. Assume LRU replacement.
- (13) You have been given 18 $32K \times 16$ -bit SRAMs to build an instruction cache for a processor with a 32-bit address. What is the largest size (i.e., the largest size of the data storage area in bytes) direct-mapped instruction cache that you can build with two-word (64-bit) blocks? Show the breakdown of the address into its cache access components (for an example, see Figure 5.10 on page 389) and describe how the various SRAM chips will be used. (Hint: You may not need all of them.)
- (14) This problem is similar to the previous problem, except that this time you decide to build a direct-mapped cache with four-word blocks, similar to Figure 5.12, page 396 (but note the different block sizes). Once again show the breakdown of the address and describe how the chips are used.
- (15) A processor issues 24-bit addresses. Memory is byte-addressable. The memory system consists of main memory and a 128K bytes 8-way set associative write-back cache. The block size is 16 bytes. The access time to the cache is 30 nanoseconds. The access time to the memory is 120 nanoseconds. The data bus between the cache and main memory is 128 bits wide. The hit rate on the cache is 95%. 25% of the cache blocks are dirty.
 - I. What is the size, in *bits*, of the tags in the mapping part of the cache? Explain your answer.
 - II. What is the average access time of the memory system (as seen by the processor)? Explain your answer.
- (16) A system includes a unified instruction/data cache that handles **all** accesses to memory by the CPU. With a perfect cache (no misses **and** no need to update main memory), the CPI of the system is 2.6. 25% of the instructions are loads. 15% of the instructions are stores.

The CPU generates word addresses. The bus width between the CPU and cache is one word. The main memory word size and the bus width between the cache and main memory are two CPU words. The main memory access time is five cycles. The cache block size is eight words. The system does **not** include a write buffer.

Your task is to compare the performance of the system under two different cache update policies: I) write-through write-around, and II) write-through write allocate. With policy I, the hit rate is 85%. With policy II, the hit rate is 92%.

 - A) Compute the CPI for policy I.
 - B) Compute the CPI for policy II.

- C) If the execution time with a perfect cache is 200 seconds, what will it be under policy I?
- D) If the execution time with a perfect cache is 200 seconds, what will it be under policy II?

Be sure to explain every step and clearly state every assumption.