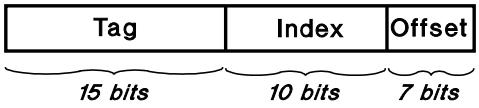
CS451/CS551/ECE441/ECE541 - SECOND EXAM
Fall 2015

Name:

Open book, open notes. Take home. There are 7 questions and 5 pages to this test.
NOTE: You may not consult any other person (other than the instructor) about the questions on this test!

1. (15 pts) The 32 bit address generated by a certain processor is divided up as shown to access its cache:



If this address is used to access a *2-way set associative* cache, answer the following questions. If you cannot determine the answer to a question from the information given, then so state. Assume the memory is byte addressable.

- What is the cache line (block) size in bytes?

- How many blocks are in the cache?

- How many sets are in the cache?

- Is the cache a virtually addressed or physically addressed cache?

- What is the total cache size in bytes?

2. (9 pts) When designing a cache memory, a designer can choose several different parameters for that cache - total cache size, cache line (block) size, and associativity. For a given set of these parameters, explain what you would need to do to reduce the following types of cache misses:

a) Compulsory misses

b) Capacity misses

c) Conflict misses

3. (7 pts) With virtual memory, it is not necessary for the *virtual a ddress space* to be the same as the *physical address space*. For example, the larger models of the 16-bit PDP11 processor had a 4 mB (22 bit) physical address space. Conv ersely, several models of the 64-bit Alpha processor have a 44-bit physical add ress. Explain how this is done.

4. (20 pts) You currently have a processor with an direct-mapped, L1 cache. Its hit rate is 90%, and the miss penalty to main memory is 80 cycles. You are considering adding an off-chip L2 cache. This cache will be sized so that its hit rate is 96%. The hit time in this cache is 8 cycles, and its miss penalty to main memory is 90 cycles. Determine the speedup you can expect to gain (if any) with this cache, if the base CPI of the processor is 1.6, with an instruction mix that includes 20% loads and 10% stores.

5. (12 pts) Explain briefly how the following techniques can improve the performance of a cache memory system.

- Merging write buffers

- Victim buffer

- Compiler-based prefetching

6. (12 pts) Strip-mining is a technique to efficiently use vector processors.

- (6 pts) Explain what strip-mining is.

- (6 pts) Explain why it is used.

7. (30 pts) Another way to evaluate the effectiveness of caches is to look at bus bandwidths. Consider a CPU with an IPC (Instructions per cycle) of 1.5 and a 200MHz clock. Measurements show that load instructions constitute 25% and stores 10% of all instructions. The CPU has separate instruction and data caches, each 8 KB and direct mapped. The hit rate in each cache is 90%. The block size in each cache is 32 bytes (or 8 words). Each instruction is 1 word (4 bytes) long. The D-cache is write-back, write allocate, and at any time 30% of the blocks are dirty. On a miss, the entire block must be read before the miss can be satisfied.

a) (10 pts) Determine the bandwidth required between the CPU and each cache, in MB/sec.

b) (10 pts) Determine the bandwidth required between the caches and main memory.

c) (10 pts) Now consider the addition of an L2 cache, which is write through, and has a hit rate of 95%. Now what is the bandwidth between the L2 cache and main memory?