

CS451/CS551/ECE441/ECE541 - SECOND EXAM
Fall 2019

Name (Please print): Joe Stanley

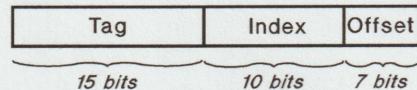
Take home exam. I don't want the "wikipedia" or internet answer, I want *your* answer. Be concise with your answers - excessive wording will result in a loss of points. Show your work for full credit. You may not consult anyone else (except the instructor) regarding answers to this test.

By signing your name below, you swear that you have not received help from anyone else (besides the instructor) in writing this test:

Name (Please Sign): Joe

There are 10 questions to 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 4-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?

$$2^7 = 128 \text{ bytes}$$

- How many sets are in the cache?

~~$$2^{10} = 1024 \text{ sets}$$~~

$$2^{10} = 1024 \text{ Sets}$$

- How many blocks are in the cache?

$$1024 \cdot 4 = 4096 \text{ blocks}$$

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

- What is the total cache size in bytes?

$$128 \cdot 4 \cdot 1024 = 524288 \text{ B}$$

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

- Increase Block Size. May adversely affect Conflict Misses
- Multilevel Cache Architecture

b) Capacity misses

- ~~Decrease Block Size~~
- Increase Cache Capacity
- Increase Associativity. May Adversely affect Compulsory Misses

c) Conflict misses

- Give Priority to Read Misses
- Avoid Address Translation During Indexing

3. (9 pts) We used several terms to describe memory in a modern processor. Describe each term below - be sure to distinguish each term from the others.

- Physical Memory

The real and physical location in memory where some data is stored. Often times, the CPU or OS doesn't even know this.

- Logical Memory

Is the logical "go between" that helps dereference virtual memory (seen by CPU/OS) and Physical Memory. Tells Memory System what physical address to look for.

- Virtual Memory

The memory seen and used by Programs and the OS, normally some named variable and associated reference.

4. (4 pts) Explain the purpose of the TLB in a virtual memory system.

The TLB is effectively an active and "living" look-up table that references the Physical Memory locations given the virtual memory address.

5. (4 pts) Some processors implement *data prefetch* instructions. Explain how these can speed up overall memory access.

These methods leverage ideas of prediction to "foresee" what data may be needed in the next few operation cycles. It is this predictive nature that allows the system to prepare for operations in the preceding few cycles.

6. (8 pts) Both write through and write back caches are in common use, which would suggest that there are pros and cons for each. List one advantage and one disadvantage for each type.

Write Through Methods ensure data security + integrity, but binds the system while that write occurs.

Write Back Methods allow for low latency, high-throughput data writes, but it introduces high-risk for data security and integrity.

7. (5 pts) Explain what a *victim buffer* does.

The victim buffer serves as a temporary location for data that has been removed from a layer of cache. It serves to improve efficiency by allowing quick recovery as needed.

8. (6 pts) With virtual memory, it is not necessary for the virtual address 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. Conversely, several models of the 64-bit Alpha processor have a 44-bit physical address. Explain how this is done.

The TLB offers some ability to provide "look-up" functionality to provide referencing capabilities to allow just the proper (and most effective) memory addressing.

9. (20 pts) You currently have a processor with a 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 1.5, with an instruction mix that includes 20% loads and 10% stores.

Original

$$\cancel{20\% \cdot 90\% \cdot 1 + 10\% \cdot 80} +$$

$$\text{Avg Mem Time} = \text{Hit Time} + \text{Miss Rate} \cdot \text{Miss Penalty}$$

$$\left[20\% (90\% \cdot 1 + 10\% \cdot 80) + 80\% \cdot 1 \right] \cdot 1.5 = 3.87$$

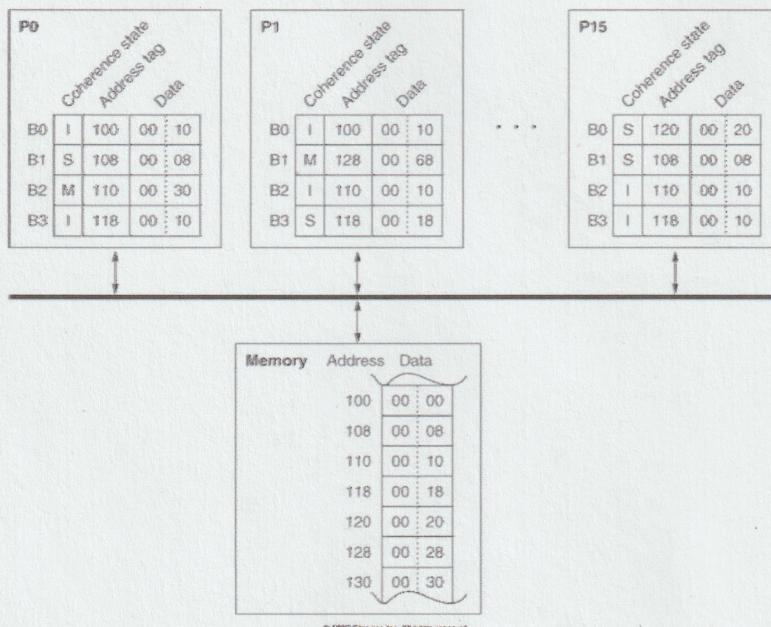
Modified

$$\left[20\% (90\% + 10\% (96\% \cdot 8 + 4\% \cdot 90)) + 80\% \right] \cdot 1.5 = 1.8084$$

$$\text{Speedup: } \frac{\text{w/ Enhancement}}{\text{w/o Enhancement}} = \frac{1.8084}{3.87} = 0.467 = \underline{\underline{46.7\%}}$$

46.7% Speedup

10. (20 pts) The figure below represents the current status of a bus-based multiprocessor that uses a snooping protocol similar to that described in the textbook. The local caches of processors P0, P1, etc. and main memory are shown. Each cache has room for 4 blocks (B0 thru B3), and each block holds 2 32-bit values. The current status of each block in the cache is shown in the figure (I - Invalid, S - for shared, and M - for Exclusive as described in the textbook). For each part below, assume that the initial cache state is as shown in the figure - that is, the operations are *not* cumulative. Describe what the state of the system will be as a result of the following CPU operations:



- a) CPU P0 reads address 120

I invalid

- b) CPU P0 writes the value 80 into address 120

M exclusive

- c) CPU P15 writes the value 80 into address 120

S shared

- d) CPU P0 reads address 110

M exclusive

- e) CPU P0 write the value 48 into address 108

M exclusive