2016/Fall CS510 Computer Architecture PhD Qualifying Exam

Student ID _____ Name _____

1. [15pts] Pipelining (State the reason for the answer shortly)

   A. [3] What is the CPI (cycle-per-instruction) of a single-cycle processor, which can execute an instruction every cycle?
   B. [3] What is the speedup of the ideally pipelined processor compared to the single-cycle design? Explain the reason for your answer using the execution time decomposition (Execution time = instruction count * CPI * clock cycle time). Suppose the number of pipeline stages is N.
   C. [3] Explain two reasons that the clock cycle time of a pipelined processor can be longer than the ideally pipelined processor.
   D. [6] Explain any possible reasons that the CPI of a pipelined processor can be longer than the ideally pipelined processor.

2. [10pts] Suppose the width of physical address of a system is 30 bits (the maximum physical address space = 2^{30}B). Calculate the size of storage (in bits) for the tags for the following cache configurations. Each tag needs extra 2 bits for valid and modified bits. Ignore extra bits for LRU replacement for set-associative caches. You must show how you calculate the final answer (possibly with figures)

   A. 32KB direct-mapped cache, block size = 32B
   B. 64KB 4-way associative cache, block size = 32B
   C. 1KB fully associative cache, block size = 128B

3. [10pts] Cache Replacement Policy.
   A. Explain the LRU replacement policy for associative caches.
   B. Show an example sequence of memory accesses which always miss a 4-way associative cache. Discuss what replacement policy can minimize cache misses for such a cache.

4. [15pts] This problem assumes the following system configurations

   - 32-bit architecture, which uses 4GB address space for each process
   - Page size: 4KB page, the size of each page table entry : 4B

   A. [5pts] What is the page table size for each process for one-level page table (flat page table)?

   B. [5pts] What are the minimum and maximum page table sizes for a process, if the system uses two-level page tables? (For the minimum case, a process uses the memory which fits in one page.)

   C. [5pts] To hide the access latency for TLBs, TLBs may be accessed in parallel with accesses to L1 caches. To support such parallel accesses to TLBs and L1 caches, the organization of L1 caches may be restricted to meet certain conditions. If the maximum associativity is limited to 8 ways, what is the largest cache capacity for such L1 caches? (a physical address must be mapped to only one set in the cache)

5. [10pts] Suppose there are two different branch predictor designs, BF1 and BF2. BF1 can predict a certain pattern of branch sequences very well, and BF2 can predict a different type of patterns well. Explain how you will add both predictors to predict both type of branch patterns. Assume that you have 2048 bits of extra storage available for that, in addition to the two predictors.

6. [15pts] **Cache Coherence**: The following table shows a simple MSI protocol for snoop-based coherence

State	CPU		Snoop transition (by Bus command)	
	Command	Next state (Action)	Read	Next state and Action
Invalid	Read	Shared (BusRead)	BusRead	Invalid
	Write	Modified (BusRfo)	BusRfo	Invalid
			BusUp	Invalid
Shared	Read	Shared	BusRead	Shared
	Write	Modified (BusUp)	BusRfo	Invalid
			BusUp	Invalid
Modified	Read	Modified	BusRead	Shared (WriteBack)
	Write	Modified	BusRfo	Invalid (WriteBack)
			BusUp	Error

   A. [5pts] Explain why "Modified" state cannot receive "BusUp" command. ("Error" state in the table)

   B. [5pts] In which state transition, the data must be transferred directly from a cache to another cache, instead of transferring data from the memory. List all such state transitions.

   C. [5pts] Extend the protocol to add Exclusive state. In Exclusive state, the cache block is clean (not modified), but no other caches have a copy of the cache block. 1) Explain why the addition of exclusive state is beneficial by describing a scenario. 2) Explain extra changes in snoop responses to implement the extra state.

7. [10pts] Multi-threading, SIMD, VLIW

   A. [5pts] Discuss the differences between HW multi-threading and multi-cores.

   B. [5pts] Discuss the differences in SIMD and VLIW. Explain how the two approaches can improve system performance.

8. [15pts] Multi-processors

   A. [7pts] Discuss the pros and cons of updated-based coherence protocols compared to invalidation-based protocols.

   B. [8pts] Discuss the pros and cons of shared L2 caches, compare to private L2 caches. Assume that the total cache capacity on a chip does not change.