CMSC411 Fall 2009 Final Exam

Name: _____

Instructions

- You have 120 minutes to take this exam.
- You do not need to provide a number if you can show the appropriate fraction. E.g., 1/13 is acceptable in place of .0769.
- This is a closed book exam. No notes or other aids are allowed.
- If you have a question, please raise your hand and wait for the instructor.
- Answer essay questions concisely using 2-3 sentences. Longer answers are not necessary and a penalty may be applied.
- In order to be eligible for partial credit, show all of your work and clearly indicate your answers.
- Write neatly. Credit cannot be given for illegible answers.

	Problem	Score
1	Architectures	/28
2	Branch Prediction	/18
3	Instruction Scheduing	/26
4	Cache and VM	/16
5	Storage, Reliability, Performance, Amdahl's Law	/28
6	Multiprocessor Cache	/30
	Total	/146

1. (28 pts) Architectures
    a. Give an example of an area of computer architecture where bandwidth has improved faster than latency. How has this gap affected performance?

    b. Describe how speculation can improve performance where dynamic scheduling cannot.

    c. Give an example of a software technique for improving ILP, and briefly describe why it is effective.

    d. Describe how mirroring may be used in RAID systems to improve availability.

    e. Describe how snooping protocols differ from directory-based protocols.

    f. Describe why coherency is necessary for multiprocessor caches.

    g. Describe the difference between connecting computers using shared media vs. switched media.

2. (18 pts) Branch prediction

For a loop containing two branches B1 & B2 (branch actions provided) for each loop iterations, show on each loop iterations the state of the branch predictors (and branch history tables, if needed), the predictions. Assume that all predictors are initialized to not taken, and that the correlation bits are initially set to not taken. When multiple predictors may be used, circle (or underline) the predictors used to make a prediction

a. (12 pts) (2,2) predictor w/ local history + branch address (standard 2-bit counter)

Loop Iteration	Branch B1			Branch B2		
	predictor	prediction	action	predictor	prediction	action
1			T			T
2			NT			T
3			T			T
4			NT			T
5			T			T
Exit						

b. (6 pts) What is the size in bits of a 128-entry (2,2) predictor w/ local history? Show your work.

3. (26 pts) Instruction scheduling

Consider the following sequence of instructions. The table on the right represents the latency incurred by each instruction. For instance, the execution stage of a +0 latency instruction completes in one cycle.

I1: LD F1, 0(Rx)
I2: MULT.D F2, F1, F1
I3: LD F3, 8(Rx)
I4: ADD.D F4, F1, F1
I5: ADD.D F1, F3, F3

Instruction	Latency
Memory LD	+2
ADD.D	+0
MULT.D	+3
Branches	+0

a. (6 pts) List all RAW, WAR, and WAW hazards found in the code.

b. (8 pts) Use the classic MIPS five-stage integer pipeline, show the timing of this instruction sequence for two loop iterations. Assume all memory accesses take 1 clock cycle, conditional branch is handled by predicted to be taken, and a register may be read and written in the same clock cycle. Assume normal forwarding and bypassing hardware

	1	2	3	4	5	6	7	8	9	10	11
I1: LD F1, 0(Rx)	IF	ID	EX	MEM	WB						
I2: MULT.D F2, F1, F1											
I3: LD F3, 8(Rx)											
I4: ADD.D F4, F1, F1											
I5: ADD.D F1, F3, F3											

I1: LD F1, 0(Rx)
I2: MULT.D F2, F1, F1
I3: LD F3, 8(Rx)
I4: ADD.D F4, F1, F1
I5: ADD.D F1, F3, F3

Instruction	Latency
Memory LD	+2
ADD.D	+0
MULT.D	+3
Branches	+0

c. (12 pts) Now schedule the code for a single-issue Tomasulo-style pipeline processor. **Assume only one each of the load buffer, integer unit, floating point adder, and floating point multiplier**. Assume an unlimited number of reservation stations for every functional unit & load buffer.

Describe when each instruction is issued, executed, and writes its result to the CDB. If its execution or completion stalls, **list the length of stall and provide a reason for the stall(s)**.

Instruction	Cycle #			Stalls
	Issue	Exec	Write CDB	
I1: LD F1, 0(Rx)	1	2	5	
I2: MULT.D F2, F1, F1				
I3: LD F3, 8(Rx)				
I4: ADD.D F4, F1, F1				
I5: ADD.D F1, F3, F3				

4. (16 points) Cache and virtual memory organization

Suppose we have a byte addressable memory of size 4GB (2^{32} bytes).

a. (8 pts) We are given a cache of size 1MB (2^{20} bytes, not including tag bits) and a cache block size of 256 (2^{8}) bytes. Compute for a 4-way associative cache the length in number of bits for the tag, index and offset fields of the 32-bit memory address (show your calculations):

b. (8 pts) Suppose pages are 8KB (2^{13} bytes) each, and the machine has 512MB (2^{29} bytes) of physical memory. Computer the number of page table entries needed if all the pages are being used. Compute the size of the page table if each page table entry also required 4 additional bits (valid, protection, dirty, use).

5. (28 pts) Storage, reliability, performance, and Amdahl's Law

One approach for improving the performance of disk storage is *disk striping*, where data is interleaved between k disks. Disk striping has a number of advantages and disadvantages that are affected by the value of k.

a. (4 pts) Disk striping is most effective when accessing contiguous blocks of data on disk. When is disk striping least effective? Explain.

b. (6 pts) For accessing contiguous blocks of data on disk, disk striping can reduce transfer costs by a factor k for k disks. Assuming transfer costs make up 40% of data access time, how much is overall performance improved when data is striped across 4 disks (when all data accesses are contiguous)?

c. (6 pts) How much can overall performance be improved when data is striped across an infinite number of disks (when all data accesses are contiguous)?

d. (6 pts) Unfortunately not all application data accesses are contiguous. As a result not all applications improve the full amount calculated above when disk striping is applied. Given the application performance in the table below using disk striping, what is the geometric mean of the improvement for the applications on the list?

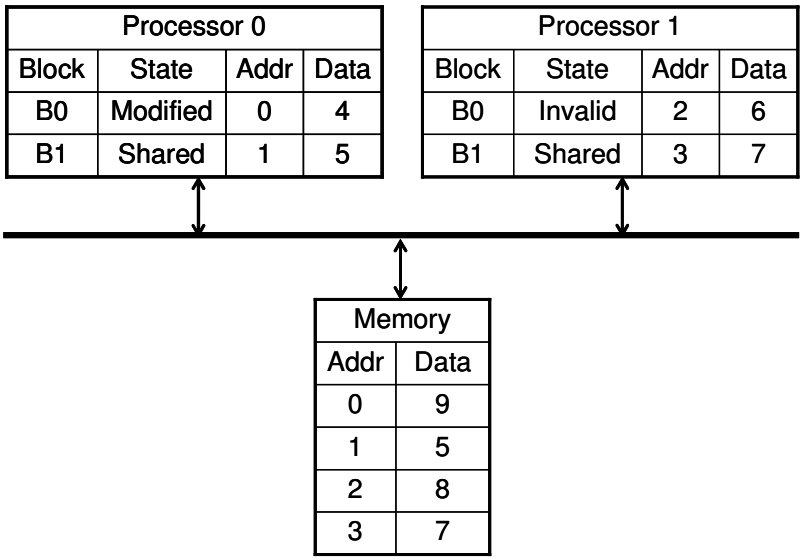
	Application 1	Application 2
Original performance	90	100
Performance w/ disk striping	80	90

e. (6 pts) A disadvantage of disk striping is that without additional measures, the entire system fails as soon as one disk fails. If the mean time to failure (MTTF) of a single disk is 100, and the MTTF of the rest of the system is 1000, what is the MTTF of the overall system when 8 disks are used for disk striping?

6. (30 pts) Multiprocessor cache coherency

Consider a simple bus-based symmetric multiprocessor, where each processor has a single private cache with coherence maintained with a snooping, write-back protocol. Each cache is direct mapped, with 2 blocks each holding 1 word (e.g., addrs 0 & 2 are mapped to block B0, addrs 1 & 3 are mapped to block B1).

For each question, assume the initial cache state is shown below. Show the resulting state of the caches and memory after each action. Show only the blocks that change. For instance, after [P0: read 3], the changes are [P0.B1: (Shared, 3, 7)], indicating Processor 0's block B1 now has the state = Shared, addr = 3, and data = 7. Also indicate the value returned by each read operation.



Processor 0

Block	State	Addr	Data
B0	Modified	0	4
B1	Shared	1	5

Processor 1

Block	State	Addr	Data
B0	Invalid	2	6
B1	Shared	3	7

Memory

Addr	Data
0	9
1	5
2	8
3	7

a. Processor 1 reads 2

b. Processor 1 writes 2 <- 10

c. Processor 0 writes 1 <- 11

d. Processor 1 writes 0 <- 12

e. Processor 0 reads 2