Seat Number _____ Name _____

CS 3853 Midterm Exam 2 — Fall 2013

This is a closed book exam. You may use a calculator, as long as it is not capable of accessing a network. No cell phones.
This is a 75-minute exam. The questions are divided into 2 parts. The first part is worth 80 points. Everyone should be able to finish Part 1. Part 2 is worth 20 points. If you cannot do the questions are Part 1 quickly, you may not have time to finish Part 2.
Put your answers in the space provided. You must show your work to get credit. If you cannot fit your answers in the space provided, use the back of the last page or ask for an additional sheet of paper. You will be given a page of diagrams to refer to for this exam. You must turn in the diagram sheet with your exam.

Part 1: 80 points

1) (4 points) Suppose that a cache has a hit time of 1 cycle, a hit ratio of 93% and a miss penalty of 9 cycles. What is the average access time in cycles? **Put a box around your final answer.**

2) (5 points) Suppose we have a standard MIPS 5-stage pipeline which has a CPI of 1 when all memory accesses are cache hits. The instruction cache has a hit rate of 97% and the data cache has a hit rate of 94%. Both caches have a miss penalty of 17 cycles. Loads and stores are 28% of all instructions. What is the CPI when cache misses are taken into account? **Put a box around your final answer.**

3) (5 points) What is the most important reason for using separate caches for instructions and data? Explain.

4) (4 points) A 32K cache has a block size of 128 bytes. Assume addresses are 32 bits. Draw a diagram of an address showing the names and sizes of the fields if the cache is:

a) direct mapped

b) 2-way set associative

c) 4-way set associative

d) fully associative

5) (2 points) For problem 4, find the block address and offset in hexadecimal for the address `0x1b7456`.
Note: The answer is the same for all four parts of problem 4.

6) (8 points) For the byte at address given in problem 5, find the tag and set number in hexadecimal. Show your calculations and label your answers.

a)

b)

c)

d)

7) (11 points) This problem is about caches.
   a) What is meant by write-back and write-through?
   b) What is meant by write allocate and no-write allocate?
   c) What does a dirty bit represent, when is it used, and why is it useful?

8) (8 points) A virtual memory system has a virtual address of 34 bits and a physical address of 41 bits. The page size is 8K. Draw diagrams of the virtual and physical addresses showing the names and sizes of the fields. What is the maximum amount of physical memory for this system in gigabytes? What is the maximum size of a process in gigabytes?

9) (6 points) A memory system uses three caches. The caches have hit rates and times given. Main memory has an access time of 50 ns. For a memory access, first the L1 cache is checked. If there is a miss, the L2 cache is checked. If there is a miss in the L2 cache, the L3 cache is checked. Finally, a miss in the L3 caches causes an access to main memory. What is he effective memory access time? **Do all calculations to 5 decimal places, show your work, and put a box around your final answer.**

Cache	hit rate	hit time
L1	.91	1 ns.
L2	.93	6 ns.
L3	.99	17 ns.

10) (8 points) This problem is about virtual memory.

a) What is a page table? How is a page table used to translate a virtual addresses to a physical addresses?

b) What is a TLB? What is it used for? Why is it necessary? What is typically stored in a TLB entry?

11) (6 points) A virtual memory system has a 64-bit virtual address, a 43-bit physical address, a page size of 32K, and a TLB with 32 entries. Each TLB entry has protection bits for read/write, user/superuser, and no execute. How many total bits are required for the TLB if the TLB is:

a) direct mapped b) 4-way set associative c) fully associative

Show your work, label you answers and put boxes around them.

12) (13 points) The diagram shows the memory organization of a processor with a TLB and 2 levels of cache. The number of bits in each field is labeled with a lower case letter, a through r. For each blank field in the table below, fill in the value, or write ND, for "not determined by the information given." Assume the following:

Virtual addresses are 64 bits
Physical addresses are 43 bits
The page size is 16K
The TLB has 32 entries and is direct mapped
The L1 cache is 64K bytes and is 4-way set associative with a block size of 32 bytes.
The L2 cache is 4MB and 8-way set associative
the L1 and L2 caches use the same block size

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r

Part 2: 20 points

13) (10 points) We want to determine whether a particular 2-way set associative cache is better than a direct mapped cache. Assume that there is a CPI of 1 when there are no cache misses, and that the miss penalty is 42 ns. Assume that the direct mapped cache system has an cycle time of .52 ns and a miss rate of 8%. The 2-way set associative cache system has an cycle time of .59 ns and a miss rate of 7%. There are 1.27 memory accesses per instruction. Calculate each of the following for each type of cache. **Show your work** and fill in your answers in the table. Then write a statement about which cache is better and why.

	direct	2-way
a) miss penalty (in cycles, rounded up to an integer)		
b) CPI (4 decimal places)		
c) time/instruction (in ns, 4 decimal places)		

14) (10 points) This problem asks you to give explanations relating to a memory system. Each part is worth only 2 points and full credit will only be given for good explanations. A short answer for a part is unlikely to receive more than one point of credit. You may need to use the back of this sheet for the answers to this question. You will want to complete the rest of the exam before starting on these questions.

a) Explain what is meant by access time and bandwidth. How are they related? Explain how bandwidth might be calculated from access time under some circumstances and why this calculation is often not correct.

b) Virtual memory always uses write-back. Explain why.

c) A modern microprocessor will typically have several megabytes of cache on chip. Explain why the L1 cache is usually a very small fraction of the size of the total cache.

d) Multilevel page tables may require several memory accesses for a single address translation. Explain why multilevel page tables are needed and why this added translation time is not a significant problem.

e) Explain what is meant by virtually indexed and physically indexed (in relation to cache design) and why there is an advantage to using a virtually indexed design.

Seat Number _____ Name _____