

CS 6761: Parallel Computer Architecture II
Midterm: May 5, 1998 9:30 to 11:00 (Max points = 80)

OPEN BOOK AND NOTES

CONCISE BULLETS GET MORE CREDIT THAN WORDY SENTENCES
LEGIBILITY IS A REQUIREMENT NOT AN OPTION...THIS IS NOT A
JOKE...PLEASE TAKE THIS SERIOUSLY!

1. (1 point, 1 min) The newly appointed president of the European union's bank is of the following nationality
 - (a) Dutch
 - (b) French
 - (c) British
 - (d) American
 - (e) Japanese
 - (f) None of the above

2. (29 points, 30 min) A processor architecture has an atomic *fetch&inc* instruction. In a split transaction bus, show how you can implement this instruction atomically across all processors of a cache-coherent SMP. Assume that the commands available on the bus are *read-req*, *write-req*, *inv-req*, and *data-resp*. The *data-resp* is in response to the *read-req*, while the other two requests don't need any response back to indicate completion. Assume that the caches in the multiprocessor implement a simple invalidation based coherence protocol (MSI protocol in Fig 5.13 of text).

Note: Convince yourself that your design works by considering a series of requests and stepping through your design.

3. (10 points, 10 min) Do you need intermediate states for implementing cache coherence protocols if the bus that connects the processors and the memories is *atomic*? If yes, explain with an example why you need them. If not, why not?

4. (20 points, 20 min) Assume that for a computation involving a $n \times n \times n$ matrix, the memory requirement on a uniprocessor is 8MBytes for a $100 \times 100 \times 100$ problem size, with each element requiring 8 bytes of storage. Assume that the uniprocessor execution time grows as $O(n^3)$, where n is the number of elements in the matrix.
 - (a) If we have a 64-node *properly* scaled machine (refer to Chapter 4 for machine scaling), what is the maximum problem size that can be accommodated for this computation?
 - (b) Assume that the execution time of the scaled problem in the previous question on the scaled machine is 4 hours. If we increase the memory of the uniprocessor such that it can accommodate the scaled problem, and assuming that the computation is perfectly parallelizable, how much time will the same scaled problem take on a uniprocessor?

5. (20 points, 20 min) Consider the following execution:

P1	P2	P3	P4
Read Y:1	Write Y:1	Read X:1	Read Z:1
Write X:1		Read Y:0	Read X:0
Write Z:1			

The notation “Read X:1”, denotes that the read of memory location X returned a value of 1. The notation “Write X:1” denotes that the write stored a value of 1 into memory location X. Assume that X, Y, and Z are 0 prior to the start of the execution. Is the above execution SC, PC, TSO, or PSO? For each of these possible memory consistency models, explain why this execution DOES or DOES NOT conform to that particular model.