Rules of the Examination

Closed book, closed notes. No calculators, cell phones, PDAs or other electronic devices. Just you, your writing instrument, and a pad of paper. Answer all questions. Please keep you answers concise and confined to the area provided. **Any answer not contained in the empty box following the question will not be graded!** Be sure to address the actual question asked.

Questions

1. **(8 points)** Describe the four types of relaxed consistency models.

2. **(8 points)** Describe briefly the operation of a snooping cache protocol with write invalidate. Describe specifically the various states that a cache block might go through and how various accesses by the CPU triggers changes to the state.

3. **(10 points)** Assume branches are predicted to occur with a frequency of 20% and data hazards are predicted to occur with a frequency of 17% of all instructions executed. Assume that you are evaluating the potential speedup of a 5 stage pipeline that must stall 3 cycles for each branch outcome and two cycles for each data hazard. Show quantitatively the expected speedup (from a non-pipelined machine) from pipelining in the presence of of these control and data hazards.

4. **(10 points)** Show the average memory access time for a multi-core platform with a snoopy cache. Normally the L1 cache has a hit time of 3 cycles and the L2 cache has a hit time of 8 cycles. However, cache writes to shared cache lines incur an additional two cycle delay. Assuming that 3% of L1 cache writes and 12% of L2 cache writes trigger a coherence delay derive an equation to show the average memory access time.

5. **(10 points)** Assume that you are deciding whether to include a new optimization into your computer system design. Furthermore assume that the optimization has a speedup of 3 on 7% of the computation, a slowdown of 1.3 on 12% of the computation but a slowdown of 1.7 on 15% of the computation. Will the system be faster with or without the optimization? You can assume that the percent impacted by each speedup/slowdown occur to completely unique portions of the computation.

6. **(10 points)** Show the impact of memory access performance on a NUMA platform. You should assume that non-local memory access times are 16 times longer than local accesses.

7. **(10 points)** Assume that you are evaluating a new optimization for your computer system design. If the optimization has a positive speedup of N on X% computation but a slowdown of M on Y% of the computation. Develop an equation that shows the overall speedup.

8. **(10 points)** The assume that a program can be parallelized as follows. The program will be broken down into 2 parts, namely A and B. Part A is 5% of the computation and part B is 95% of the computation. Part A is entirely serial and part B is fully parallel to any amount, however, there is a factor of $P + K\alpha$ synchronization cost for this part (where P and α are fixed costs and K is the amount of parallelism. Develop an equation showing the potential speedup from this parallization.

9. **(10 points)** Assume that your team is considering the addition of a vector processing unit to your next generation processor design that will increase the performance of vector operations by 70%. Assuming that, in general, X% of your programs statements are vectorizable. How much faster would programs be with the addition of this vector hardware?

10. **(10 points)** Following the above question, assume that your hardware vector register contains space for 128 64-bit words and that the average vector size of your is 512 words. Refine your answer to show the potential speedup with this additional limitation.

11. **(10 points)** Following from the above two question, assume that your vector hardware also supports lanes of 32, 16, and 8 bits with each operation. Assuming that 10% of your vector operations are on 32-bit operands, that 8% of your vector operations are on 16-bit operands, that 20% of your vector operations are on 8-bit operands, and that all the remainder are expected to be on 64-bit operands. Refine your above answer to show the potential speedup with this additional limitation.