CS5451 Exam #1, Fall 2023

This is a closed book closed notes exam..

Please answer all the following questions, and **always justify your answers**

Good Luck!

The questions are organized into two groups. Each question in the first group requires short (but concise) answers. The questions of the second group will require longer answers.

The exam has a total of 124 points. You need to get 100 points for a perfect score. Everything above that will be considered bonus points.

Group 1 (8pts each)

1. What is the cost of sending a message of size m between two processors that are l hops away from each other on a machine with store-and-forward routing and a machine with cut-through routing?
2. Define UMN, NUMA, and ccNUMA architectures.
3. Define serializability in the context of parallel processing.
4. Define false sharing and the type(s) of parallel architectures on which it occurs (make sure that you are very specific about the types of parallel architectures).
5. Describe the common approach for deriving a task decomposition from a data decomposition.
6. Define the task dependency and task interaction graphs.
7. What are the three primary sources of overheads in parallel programs?
8. What happens to the parallel efficiency when we fix the number of processors and increase the size of the problem being solved and why?

Group 2 (20pts each)

1. Consider the case in which one processor has p chucks of data each of size m, and it needs to send the ith chuck to the ith processor. Answer the following questions:
   - What is the name of this communication operation? What is the name of the reverse version of this operation?
   - How will this operation be performed on a hypercube-connected architecture with p processors (assuming that p is a power of two)?
   - Derive the complexity of this operation in terms of t_s and t_w.
   - Is this operation sensitive to the cross-bisection bandwidth of the architecture (i.e., can the same algorithm be implemented on a ring with cut-through routing and get the same complexity?)?
2. Answer the following questions related to the structure of collective communication operations:
   - If you focus on the set of point-to-point communications performed by the various collective communication operations on a hypercube-connected parallel system, you will notice that they fall into two patterns. Briefly describe those patterns and map the various collective operations that we discussed in each one of them.
   - Now consider the above two patterns on a multistage interconnection network. Will these sets of point-to-point communication operations be contention-free on a multistage network?
3. Consider the problem of multiplying two $n \times n$ matrices A, B to obtain C (i.e., $C = AB$) on a shared-memory system. Compute the amount of distinct data elements from A and B that a processor needs to access under the following scenarios.
   - Each processor is responsible for computing consecutive n/p columns of C.
   - Each processor is responsible for computing every pth column of C.
   - Each processor is responsible for computing a $(n/\sqrt{p}) \times (n/\sqrt{p})$ block of C.