Parallel Computing

Sample Questions for CIE 1 **December 2012**

1. Suppose you have a computation that uses two vector inputs to compute a vector output, where each vector is stored in consecutive memory locations. Each input and output location is unique, but data is loaded/stored from cache in 4-word transfers. Suppose you have P processors and N data elements, and execution time is a function of time L for a load from memory and time C for the computation. Compare parallel execution time for shared memory architecture with a bus (Nehalem) versus a full crossbar (Niagara) from Lecture 3, assuming a write back cache that is larger than the data footprint.
2. For the given network topologies determine the bisection width, diameter, degree, and total number of switching nodes.
3. To send a message of size m from one processor to another processor directly linked in a network, which of the following statement is correct?
   a) Cut-through routing is faster if m is large;
   b) Store-and-forward routing is faster if m is large;
   c) Cut-through and store and forward routing have similar performance.
   d) Cut-through routing may be faster if the link is bidirectional.
4. What are the major advantages and disadvantages of the crossbar switching networks and the bus-based networks?
5. What is a NUMA system? What are its features?
6. Write a (rough) parallel algorithm for the following problems along with a description about the different design choices possible and their relative benefits, and their algorithmic complexity. (You have to use the PCAM design methodology).
   a. All pairs shortest path (Floyd's algorithm)
   b. Matrix Matrix multiplication
   c. Matrix Vector multiplication
   d. Quick sort
   e. N-body simulation
   f. Monte Carlo methods
   g. Stencil computations (Finite difference methods)

   (The quality of your algorithm design depends on the extent of your understanding of the Foster design methodology). You may benefit reading the article series: http://www.drdobbs.com/parallel/designing-parallel-algorithms-part-1/223100878
   Parallel patterns:
   https://wiki.engr.illinois.edu/display/ppp/Home
   http://parlab.eecs.berkeley.edu/wiki/patterns/patterns
7. Why the following algorithms are called parallel primitives?
   a. Reduction or broadcast (tree based efficient parallel algorithms)
   b. Parallel prefix sum (scan)
8. Very briefly explain the following types of parallelism along with at least one application example.
   a. Embarrassingly parallel
   b. Data Parallelism

c. Geometric decomposition
d. Pipelined parallelism
e. Speculative parallelism
f. Dataflow network
g. Task graph based
h. Recursive splitting
i. Wavefront parallelism
j. Discrete Event based parallelism

9. What are different types of synchronization primitives used in shared memory programming? What are their relative advantages and drawbacks?
    a. Atomic operation
    b. Mutex
    c. Semaphore
    d. Transaction
    e. Barrier
    f. Conditional
    g. Speculative update (aka. Compare-and-swap, CAS)
    h. Busy waiting

    (Some of these ideas are covered in chapter 4 of Pacheco's book).
10. Given a communication pattern, how to select a network topology that enables optimal mapping. When you say that a mapping is optimal? What are the associated metrics?