

Parallel Systems Theory Exam

January 19th 2010

1. One of the major trends is the use of graphics cards (GPUs) for general purpose high-performance computing. Explain the success of GPUs by comparing them with CPUs and optionally other processors. Also explain the drawbacks of the philosophy behind GPUs: what could be the reason of the failure of GPUs to get widely used (and replace for instance multi- and manycore CPUs)?

2. (a) Compare the possible overheads of blocking and non-blocking send and receive operations. What are the possible overheads? What are the differences of both modes?
(b) Analyse the overheads in the following data-parallel application: the master partitions the data in N packages and sends the packages one by one to the N slaves which after receiving process the received package. Assume that the data can be partitioned in any number of packages and that each package can be processed independently. The amount of data is considerable so that the overheads have a big impact on the performance. What are the overheads coming from the communication? Only considers one-port communication. *Hint*: draw the execution profile.
(c) How could the algorithm be optimized in order to reduce the communication overhead?

3. (a) Consider a parallel system in which the 2^d processing elements are linked by a hypercube communication network (e.g. $d=3$ resulting in 8 nodes). Consider a parallel application such as a Discrete Optimization Problem (DOP) which needs a termination detection algorithm. Dijkstra's algorithm with a white/black token is chosen. Explain how in your implementation the nodes would communicate: for each node explain with which other node(s) it would communicate and how each node calculates the ranks of these nodes. Make sure you minimize the communication overhead! So you do NOT have to explain the algorithm or give a detailed algorithm to calculate the ranks!
(b) In the algorithm, after the token has been changed to black it goes through the ring to P_0 and the token has to go through the whole ring again. Is it a valid (correct) optimization to send the black token *directly* to P_0 instead of sending it throughout the whole ring? Is it an effective optimization (will it reduce average completion time) ? Why (not)? Which would be the optimal communication paths of this optimization in the hypercube network?