Final Exam Topics and Question Types

The final exam will cover parts of Chapter 7 again, parts of my lecture notes covering network topologies, and GPUs and GPU programming. Selected topics from Chapter 7 that were not on the second exam will be included. You are expected to know:

- The difference between strong and weak scaling;
- The reduction algorithm for summing a sequence of numbers on a shared memory multiprocessor;
- The reduction algorithm for summing a sequence of numbers on a message-passing multiprocessor;
- Flynn's taxonomy: what it is and the differences between SIMD and MIMD architectures in particular;
- The components of network performance;
- The difference between static and dynamic network topologies and examples of each;
- The difference between blocking and non-blocking networks and examples of each;
- The definitions of network diameter, bisection width, bisection bandwidth, and total bandwidth;
- The definitions of crossbar networks, buses, and multistage networks;
- The definitions and properties of meshes, toruses, rings, buses, fully-connected networks, butterfly and omega networks, including their diameters, bisection widths, maximum edges per node and maximum edge lengths, as a function of the network parameters (such as its dimension or number of nodes.)
- The purpose of a GPU and the ways in which a GPU differs from a CPU;
- The difference between GPU programming and general purpose GPU programming;
- The basic architectural design of the NVidia GPU family (organization in terms of SMs, cores, shared memory versus device memory, connections, etc;
- What the CUDA language is;
- The differences between kernel, device, and host functions in CUDA;
- How to read a CUDA program and identify the different parts, such as block and grid dimensions, thread IDs, block IDs, and what is executed synchronously, what is executed on the host, and what, on the device;
- The basic terminology of GPU programming including grids, thread blocks, threads, and warps.

Some questions may be short code fragments to be analysed. Some questions may contain representations of network topologies, about which various questions will be asked. Some may ask for a short algorithm or piece of an algorithm to be completed. Some questions will be short answer, some multiple choice, and some true-false. The format will be similar to those of the first and second exams.