COMP 4690 – Computer Systems and Architecture

Calendar Description: Investigation of today's modern computer architecture and system design concepts, including requirements, specifications, and implementation. Instruction sets, instruction-level parallelism, speculative execution, multi-threaded architectures, memory hierarchy, multiprocessors, storage design and implementation, and interconnection networks.
Prerequisite: COMP 3370

Outline

1) Fundamentals and Techniques for Computer Design (1 ½ weeks).
   Review of Design Issues (instruction sets, addressing, caches, system performance, etc.), Implementation Technologies, Tools and Trends (custom vs. programmable logic, SOC, CMP, hardware-software co-design, multi-core processors and PIM, etc.), Performance Prediction (trace-based hardware simulation, etc.)
2) Advanced Instruction-Level Parallelism (2 ½ weeks).
   Review of pipelining and multiple functional units, Out of Order Execution and Register Renaming, Branch Prediction, Speculation and Value Prediction
3) Introduction to Parallel Architectures (1 ½ weeks)
   Types of Parallel Machines (e.g. shared vs. distributed memory, NUMA, vector and array processing) and their applications and tradeoffs, Connecting Multiple Processors, Tradeoffs: Scalability, Programmability and Cost, etc.
4) Multiprocessor Architectures (2 weeks).
   Simple shared memory machines and snoopy cache coherence, Programming Model and Issues, Directory Based coherence and coherence protocols (MESI, MOESI, etc.), CC-NUMA, COMA
5) Cluster Architectures and Interconnection Networks (1 ½ weeks).
   Clusters – their uses and limitations, Programming Model and Issues, The importance of intercommunication, Overview of Interconnection Fabrics (Gig Ether vs. Infiniband vs. Myrinet vs. Quadrics)
6) Memory Systems Design (1 ½ weeks).
   Meeting bandwidth and latency requirements, climbing the "Memory Wall" – interleaving, pipelining, latency tolerance, etc., Memory-Processor Interconnects, Scalability Issues
7) High Performance I/O and Storage Systems (1 ½ weeks).
   What is Parallel I/O and what are its implications, Meeting bandwidth and latency requirements, Scalability Issues (especially with clusters)
8) New and Evolving Architectures (1 week)
   Multithreaded machines and thread-level parallelism, cellular architectures, dynamically reconfigurable hardware

Text: John Hennessy and David Patterson, *Computer Architecture: A Quantitative Approach,* Third Edition, Morgan Kaufmann, 2003