Computer Science 703

Advance Computer Architecture

2006 Semester 1

Lecture Notes
13Apr06
Review

James Goodman

Test Information

- Time: Tuesday, May 2: 6-7.30
- Place: CS 279 (here)
- Open-book, notes
- Coverage: lecture material through Tuesday, 11Apr

Readings

- P. Sweazey and A.J. Smith, "A class of compatible cache consistency protocols and their support by the IEEE Futurebus," *Proc. Thirteenth International Symposium on Computer Architecture (ISCA-13)*, Tokyo, Japan, pp. 414-423, June 1986.
- M. Herlihy and J.E.B. Moss, "Transactional Memory: Architectural Support for Lock-Free Data Structures," *Proc. International Symposium on Computer Architecture (ISCA-93)*, ACM Press, 1993, pp. 289-300.
- R. Rajwar & J.R. Goodman, "Speculative Lock Elision: enabling highly concurrent multithreaded execution," *34th Annual International Symposium on Microarchitecture (MICRO-34)*, December 2001, pp. 294-305.
- Hennessy & Patterson, "Distributed shared-memory architectures," Section 6.5 from *Computer Architecture: A Quantitative Approach* (3rd Ed.), 2003.
- W.-H. Wang, J.-L Baer, & H.M. Levy, "Organization and performance of a two-level virtual-real cache hierarchy," *ISCA-16*, pp. 140-148, June 1989.
- Mark Hill, "Processors should support simple memory-consistency models," *IEEE Computer*, **31**(8), pp. 28-34, August 1998.

First-week topics

- Overview, Moore's Law
- Multiprocessing, Multithreading, & Multicores
- Multiprocessing Issues
  - Lost updates
  - Memory ordering
  - Cache coherence

Threads and Thread programming

From assignment, you should have experience with threads

- Creation
- Use of Mutex
- Condition variables
- Thread-safe, MT-safe functions

Interconnection networks & topologies

- Crossbar vs. buses
- Direct vs. indirect networks
- Trees, fat trees, mesh, torus, ring, hypercube
- Perfect shuffle, Omega
- Topological measures (diameter, degree, bisection bandwidth)
- Avoiding deadlock in routing, virtual circuits
- Store-and-Forward routing
- Wormhold & Cut-through routing

Simulation

- General Simulation techniques
  - time-based, event-based and process-based simulation
  - validation of results
- Architecture-specific simulation
  - instruction emulation
  - trace collection, reduction, processing
- Simulation tools

Scalable Memory Systems

Interaction of

- Virtually-addressed cache
- Multi-level cache
- Cache coherence
- Non-blocking cache

Scalable Memory Systems (2)

- Directory-based protocols vs. snooping
  - Snooping has serious limitations of scale
  - Directory-based is always slower, but scalable
  - Basic protocol is simpler (3 states), but requires more serial events
- Maintaining a sharing list in the directory
- Distributed writes are slow
- Dealing with races

Cache Coherence-MOESI Model

- Attributes
  - ownership
  - exclusiveness
  - validity
- Five states
  - Allowed state changes
  - Permitted combination of states
- Example of lock contention

2006 YEAR PRESENTATION The University of Auckland | New Zealand

Better programming Models

- Critical sections
- Atomic RMW operations
  - T&S, T&T&S, Atomic Swap
  - Compare & Swap
  - LL/SC
- Notion of transactional memory
  - Atomic insertion of a transaction (linearizability)
  - Hardware support (SLE)
  - Implementing Transactional Memory
    - Hardware
    - Software