CDA5155 Exam 1 Study Topics

Covers chapter 1, appendix B, chapter 2, appendix A, assignments 1 & 2, and the papers:

1. "Improving Direct-Mapped Cache Performance by the Addition of a Small Fully-Associative Cache and Prefetch Buffers"
2. "Speculative Tag Access for Reduced Energy Dissipation in Set-Associative L1 Data Caches"
3. "Reducing Set-Associative L1 Data Cache Energy by Early Load Data Dependence Detection (ELD3)"
4. "Practical Way Halting by Speculatively Accessing Halt Tags"

1. Chapter 1
   1.1. Classes of Computers
   1.2. Trends
      1.2.1. Memory Access Gap
      1.2.2. Power Wall
   1.3. Terms
      1.3.1. Bandwidth
      1.3.2. Latency
   1.4. Measures
      1.4.1. Comparing Performance
      1.4.2. Clock Speed
      1.4.3. Methods for Evaluation
   1.5. Principles
      1.5.1. Locality of Reference
      1.5.2. Amdahl’s Law
      1.5.3. CPU Time

2. Appendix B
   2.1. Memory Hierarchy Principles
   2.2. Terms
   2.3. Equations
   2.4. Caches
      2.4.1. Organizations
         2.4.1.1. direct mapped
         2.4.1.2. fully associative
         2.4.1.3. set associative
      2.4.2. Replacement Policies
         2.4.2.1. random
         2.4.2.2. LRU
         2.4.2.3. FIFO
      2.4.3. Write Policies
         2.4.3.1. write through/no write allocate
         2.4.3.2. write back/write allocate
      2.4.4. Unified Caches
      2.4.5. Reducing Cache Miss Rate
2.4.5.1. larger block size
2.4.5.2. larger cache size
2.4.5.3. higher associativity

2.4.6. Cache Miss Categories
2.4.6.1. compulsory
2.4.6.2. capacity
2.4.6.3. conflict

2.4.7. Reducing Cache Miss Penalty
2.4.7.1. multi-level caches
2.4.7.2. giving priority to read misses over write misses

2.4.8. Reducing Cache Hit Time
2.4.8.1. virtually addressed caches
2.4.8.2. virtually indexed and physically tagged caches

2.5. Overlays

2.6. Virtual Memory
2.6.1. virtual memory terms
2.6.2. page tables
2.6.3. translation lookaside buffers

2.7. Multiprogramming

3. Chapter 2

3.1. Advanced Cache Optimizations
3.1.1. small and simple first-level caches
3.1.2. way prediction
3.1.3. pipelined cache access
3.1.4. nonblocking data caches
3.1.5. multibanked cache
3.1.6. critical word first and early restart
3.1.7. merging write buffer
3.1.8. compiler optimizations
3.1.8.1. code positioning
3.1.8.2. array merging
3.1.8.3. loop interchange
3.1.8.4. blocking
3.1.9. hardware prefetching
3.1.10. compiler controlled prefetching

3.2. Memory Technologies
3.2.1. SRAM
3.2.2. DRAM
3.2.3. flash memory
3.2.4. enhancing dependability
3.2.5. protection via virtual memory

3.3. Cross Cutting Issues
3.3.1. I/O and the memory hierarchy
3.3.2. stale data
3.3.2.1. problems
3.3.2.2. solutions
4. Appendix A

4.1. Classes of Instruction Sets
   4.1.1. Stack (zero address)
   4.1.2. Accumulator (one address)
   4.1.3. General-Purpose (two and three address)

4.2. Addressing Issues
   4.2.1. Size of a Word
   4.2.2. Byte Order (little endian or big endian)
   4.2.3. Alignment Requirements
   4.2.4. Extension of Bytes and Halfwords

4.3. Special Addressing Modes and Operations

4.4. Transfers of Control

4.5. Calling Conventions

4.6. Instruction Encoding Tradeoffs and Properties

4.7. Computer Architecture Periods