

When / Where

Monday, 8 May 2000, 10:00-12:00 CDT, Room 1112 CEBA.

Conditions

Closed Book, Closed Notes

Bring one 215×280 mm note sheet.

No communication devices.

Format

Two or three or maybe four problems.

One set of short answers.

Resources

Solved tests and homework: `href = "http://www.ee.lsu.edu/ee4720/prev.html"`

Study Recommendations

Solve Old Problems

Memorizing solutions is not the same as solving.

Following and understanding solutions is not the same as solving.

Use the solutions for brief hints and to check your own solutions.

Introductory Material

ISA v. Implementation.

Technological Factors: Transistor speed and quantity, memory speed and size.

Different factors influencing ISA and implementation.

Design principles: Amdahl's law, locality.

CPU Performance Equation

Benchmark types.

Instruction Set Design

Data Types: What to include, what to leave out.

Basic integer and floating point

Packed types: BCD, integer, saturating integer.

Size choices.

Memory and Register Organization: Why ≈ 32 registers?

Stack and accumulator architectures.

Memory/Memory, Register/Memory.

Addressing Modes: What they do, which ones to include.

Register, Immediate, Direct, Register Deferred (Register Indirect), Displacement, Indexed, Memory Indirect, Autoincrement, Autodecrement, Scaled.

Control Transfer Instructions: Types, when to use.

Branch, Jump, Jump & Link, Call, Return

Format of displacements in instruction.

Specification of condition: condition code registers, integer registers, loop counter.

Delayed and predicated instructions; prediction hints.

Instruction Coding.

Fixed-length, variable-length, and bundled instructions.

Splitting of opcode field (as in DLX type-R instructions).

ISA Classifications: RISC, CISC, VLIW, Stack, Accumulator

Synthetic Instructions

DLX

Classification: RISC

Goals: ISA should allow simple, high-speed implementation.

Instruction types.

Know how to read and write DLX programs.

Chapter-3 (Static Scheduled) DLX Implementations

Unpipelined Implementation

Pipelined Implementations

Basic (3-cycle branch delay).

One-cycle branch delay.

Bypassed.

Hazard Definitions

For a Given Pipelined Implementations

Show pipeline execution diagrams.

Show register contents at any cycle.

Determine control hardware.

Determine CPI.

Interrupts, Exceptions, Traps

Difference between interrupt, exception, trap.

Causes of exceptions, role of handler.

Privileged Mode.

Pipeline activity leading to execution of handler.

Vectored traps (using trap table).

Precise exceptions, achieving with floating-point operations.

Long Latency Operations

Types of operations. (Floating point and maybe load.)

Degree of pipelining: Initiation interval, latency.

Detecting functional unit structural hazards.

Detecting WB structural hazards: reservation register.

Detecting and handling RAW hazards: ID-stage v. pre-WB stall.

Handling WAW hazards.

Instruction-Level Parallelism

Definition, analysis.

Dependency definitions.

Basic block definition.

Loop Unrolling

Hand unrolling of loops.

Software pipeline using rotating registers in IA-64.

Dynamic Scheduling

Reservation stations.

Register Renaming

How renaming allows out-of-order execution.

Where registers get renamed.

Role of register map.

Register Names:

Reorder buffer entry number.

Reservation station number.

Physical register number.

Reorder Buffer

Normal Use

Issue: placement of instructions in reorder buffer.

Completion: updating reorder buffer entry

Commitment (retirement): removal from reorder buffer entry.

Recovery

Goal: Undo execution starting at some instruction.

Reasons: exception, branch misprediction.

Steps when register map backed up.

Steps when register map not backed up.

Load/Store Unit

Store/load ordering rules.

Load/store bypassing.

Multiple Issue

Superscalar

Duplication of Resources.

For n instructions / cycle: Fetch, decode, rename, writeback, commit.

For $< n$ instructions: load/store, floating-point units.

Added Complexity

Instruction fetch inefficiency.

Rules for fetching a new group.

VLIW

Difference with superscalar: instruction bundling.

Dependence information in bundles.

Branch Prediction

One-level branch prediction.

Two-level (gshare, gselect) branch-prediction.

Branch target prediction.

Branch folding.

Memory

General

Definitions: bus width (w), address space size (a), character size (c).

Connection of memory devices.

Caches

Set Associative Caches

Definitions: block (line), index, tag, alignment, associativity

Connection of tag and data memory.

Special Cases: Direct mapped, Fully associative.

Write Mechanisms

Write allocate or write around.

Write back or write through.

Victim (Block to replace) Selection

Least-Recently Used (LRU), Random (arbitrary).

Virtual Memory

Definitions: virtual address, physical address, page, virtual page number, physical page number.

Page Tables

Organization: one-level, multi-level.

Page table entry contents.

Virtual to Physical Address Mapping

Using TLB.

Using page tables.

Page Replacement Policy

Approximation of LRU using the clock algorithm.