CSE 141 – Computer Architecture
Spring 2005

Lectures 11
Exceptions and
Introduction to Pipelining

Pramod V. Argade
May 4, 2005

Announcements

- **Reading Assignment**
  - Sections 5.6, 5.9 The Processor Datapath and Control
  - Section 6.1, Enhancing Performance with Pipelining

- **Homework 2:**
  - None!

- **Quiz**
  **When:** Mon., May 16th, First 10 minutes of the class
  **Topic:** Pipeline Hazards, Chapter 6  **Need:** Paper, pen
Course Schedule

<table>
<thead>
<tr>
<th>Lecture #</th>
<th>Date</th>
<th>Day</th>
<th>Lecture Topic</th>
<th>Quiz Topic</th>
<th>Homework Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3/28</td>
<td>Monday</td>
<td>Introduction, Ch. 1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>3/30</td>
<td>Wednesday</td>
<td>Performance, Ch. 4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>4/4</td>
<td>Monday</td>
<td>ISA, Ch. 2</td>
<td>Performance</td>
<td>#1</td>
</tr>
<tr>
<td>4</td>
<td>4/6</td>
<td>Wednesday</td>
<td>Arithmetic, Ch. 3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>4/11</td>
<td>Monday</td>
<td>Arithmetic, Ch. 3</td>
<td>ISA</td>
<td>#2</td>
</tr>
<tr>
<td>6</td>
<td>4/13</td>
<td>Wednesday</td>
<td>Single cycle CPU, Ch. 5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>4/18</td>
<td>Monday</td>
<td>Single cycle CPU, Ch. 5</td>
<td>Arithmetic</td>
<td>#3</td>
</tr>
<tr>
<td>8</td>
<td>4/20</td>
<td>Wednesday</td>
<td>Multi-cycle CPU, Ch. 5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>4/25</td>
<td>Monday</td>
<td>Multi-cycle CPU, Ch. 5</td>
<td>Single Cycle CPU</td>
<td>#4</td>
</tr>
<tr>
<td>10</td>
<td>4/27</td>
<td>Wednesday</td>
<td>Review for the Midterm</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>5/2</td>
<td>Monday</td>
<td>Mid-term Exam Center 101</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>5/4</td>
<td>Wednesday</td>
<td>Exceptions, Ch. 5 and Pipelining, Ch. 6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>5/9</td>
<td>Monday</td>
<td>Pipelining, Ch. 6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>5/11</td>
<td>Wednesday</td>
<td>Data and control hazards, Ch. 6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>5/16</td>
<td>Monday</td>
<td>Data and control hazards, Ch. 6</td>
<td>Pipeline Hazards</td>
<td>#5</td>
</tr>
<tr>
<td>16</td>
<td>5/18</td>
<td>Wednesday</td>
<td>Memory &amp; cache design, Ch. 7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>5/23</td>
<td>Monday</td>
<td>Memory &amp; cache design, Ch. 7</td>
<td>Cache</td>
<td>#6</td>
</tr>
<tr>
<td>18</td>
<td>5/25</td>
<td>Wednesday</td>
<td>Virtual Memory &amp; cache design, Ch. 7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No Class</td>
<td>5/30</td>
<td>Monday</td>
<td>Memorial Day Holiday</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>6/1</td>
<td>Wednesday</td>
<td>Course Review</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>6/10</td>
<td>Friday</td>
<td>Final Exam</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Exceptions

- There are two sources of non-sequential control flow in a processor
  - Explicit branch and jump instructions
  - Exceptions
- **Branches** are synchronous and deterministic
- **Exceptions** are typically asynchronous and non-deterministic
- Guess which is more difficult to handle?

*(Control flow refers to the movement of the program counter through memory)*
Exceptions and Interrupts

- The terminology is not consistent, but we’ll refer to
  - Exceptions as any unexpected change in control flow
  - Interrupts as any externally-caused exception

So then, what is:
- Arithmetic overflow
- Divide by zero
- I/O device signals completion to CPU
- User program invokes the OS
- Memory parity error
- Illegal instruction
- Timer signal

For now...

- The machine we’ve been designing in class can generate two types of exceptions.
  - Arithmetic overflow
  - Illegal instruction
- On an exception, we need to
  - Save the PC (invisible to user code)
  - Record the nature of the exception/interrupt
  - Transfer control to OS
Handling exceptions

- PC saved in EPC (exception program counter), which the OS may read and store in kernel memory
- Two ways of signaling
  - A status cause register, and a single exception handler may be used to record the exception and transfer control, or
  - A vectored interrupt transfers control to a different location for each possible type of interrupt/exception

Supporting exceptions

- For our MIPS-subset architecture, we will add two registers:
  - EPC: a 32-bit register to hold the user’s PC
  - Cause: A register to record the cause of the exception
    - Undefined inst: Cause = 0
    - Overflow: Cause = 1
- We will also add three control signals:
  - EPCWrite (subtract 4 from PC)
  - CauseWrite
  - IntCause
- Need to force PC
  - Select the interrupt handler address into the PC.
Supporting exceptions in our FSM

Instruction Fetch, state 0
- MemRead
- ALUSelA = 0
- InstD = 0
- IRWrite
- ALUSelB = 01
- ALUOp = 00
- PCWrite
- PCSource = 00

Start

Instruction Decode/ Register Fetch, state 1
- ALUSelA = 0
- ALUSelB = 11
- ALUOp = 00
- Opcode = anything else

- Opcode = LW or SW
- Opcode = R-type
- Opcode = B-type
- Opcode = JMP

Memory Inst FSM
R-type Inst FSM
Branch Inst FSM
Jump Inst FSM

to state 10
Supporting exceptions in our FSM

from state 1
R-type instructions

ALUSrcA = 1
ALUSrcB = 00
ALUOp = 10

RegDst = 1
RegWrite
MemtoReg = 0

overflow
To state 11

To state 0

State to Support Exceptions

IntCause=1
CauseWrite
ALUSrcA = 0
ALUSrcB = 01
ALUOp = 01
EPCWrite
PCWrite
PCSource=11

IntCause=0
CauseWrite
ALUSrcA = 0
ALUSrcB = 01
ALUOp = 01
EPCWrite
PCWrite
PCSource=11

to state 0 (fetch)
Overview of Pipelining
Pipelining: Its Natural!

- Laundry Example
- Ann, Brian, Cathy, Dave each have one load of clothes to wash, dry, and fold
- Washer takes 30 minutes
- Dryer takes 40 minutes
- “Folder” takes 20 minutes

Sequential Laundry

- Sequential laundry takes 6 hours for 4 loads
- If they learned pipelining, how long would laundry take?
Pipelined Laundry: Start work ASAP

- Pipelined laundry takes 3.5 hours for 4 loads
- Sequential laundry takes 6 hours for 4 loads

Pipelining Overview

- What is pipelining?
  - Multiple instructions are overlapped in execution
- Notes:
  - Time for completion of a single instruction is not shorter
  - Multiple tasks operate simultaneously
  - Pipelining does not change latency
  - Pipelining increases the throughput
  - Pipelining rate is limited by the slowest stage
  - Potential speedup = number of pipeline stages
  - Time to “fill” pipeline and time to “drain” it reduces speedup
Pipelining

- Requires separable jobs per stage
- Requires separate resources
- Achieves parallelism with replication
- Pipeline efficiency (keeping the pipeline full) critical to performance
- Time between instructions $\text{pipelined} = \frac{(\text{Time between instructions non-pipelined})}{(\# \text{ Pipe Stages})}$
- Fundamentally invisible to the programmer

Pipelining: Instructions Covered

- Only following instructions will be implemented
  - Memory: LW, SW
  - Arithmetic: ADD, SUB, AND, OR, SLT
  - Branch: BEQ

<table>
<thead>
<tr>
<th>Instruction Class</th>
<th>Instruction Fetch</th>
<th>Register Read</th>
<th>ALU Operation</th>
<th>Data Access</th>
<th>Register Write</th>
<th>Total Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Load Word (lw)</td>
<td>200 ps</td>
<td>100 ps</td>
<td>200 ps</td>
<td>200 ps</td>
<td>100 ps</td>
<td>800 ps</td>
</tr>
<tr>
<td>Store Word (sw)</td>
<td>200 ps</td>
<td>100 ps</td>
<td>200 ps</td>
<td>200 ps</td>
<td></td>
<td>700 ps</td>
</tr>
<tr>
<td>R-Format</td>
<td>200 ps</td>
<td>100 ps</td>
<td>200 ps</td>
<td></td>
<td>100 ps</td>
<td>600 ps</td>
</tr>
<tr>
<td>Branch (beq)</td>
<td>200 ps</td>
<td>100 ps</td>
<td>200 ps</td>
<td></td>
<td></td>
<td>500 ps</td>
</tr>
</tbody>
</table>
Non-Pipelined vs. Pipelined Execution

Register file writes in the first half and reads in the second half of the cycle

Non-Pipelined: Time for 3 instructions 3 * 800 ps

Pipelined: Time for 3 instructions 3 * 200 ps

Pipelining offers 4x improvement in this example

Pipelining for MIPS

- All MIPS instructions are the same length.
  - Makes it easier to fetch them in the first pipeline stage
  - Decode stage has to always process 32 bits
- MIPS has only a few instruction formats
  - Various fields are in the same location in different formats
    - e.g. Opcode, Rs, Rt
  - Register file can be read in parallel with decoding
- Memory operands for MIPS appear only in load/store
  - ALU can be used to calculate memory address
  - Memory can be accessed in the following stage
Pipeline Hazards

- What are hazards?
  - Next instruction cannot execute in the following cycle
- Types of hazards
  - Structural hazards
    - Hardware does not support combination of instructions
  - Data hazards
    - One instruction must wait for another to complete
  - Control hazards
    - Need to make a decision based on the result of one instruction while others are executing

Announcements

- **Reading Assignment**
  - Sections 5.6, 5.9 The Processor Datapath and Control
  - Section 6.1, Enhancing Performance with Pipelining
- **Homework 2:**
  - None!
- **Quiz**
  - **When:** Mon., May 16th, First 10 minutes of the class
  - **Topic:** Pipeline Hazards, Chapter 6
  - **Need:** Paper, pen