Last time in Lecture 4

- Pipelining increases clock frequency, while growing CPI more slowly, hence giving greater performance

<table>
<thead>
<tr>
<th>Time</th>
<th>Instructions Program</th>
<th>Cycles Instruction</th>
<th>Time Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Increases because of pipeline bubbles
Reduces because fewer logic gates on critical paths between flip-flops

- Pipelining of instructions is complicated by HAZARDS:
  - Structural hazards (two instructions want same hardware resource)
  - Data hazards (earlier instruction produces value needed by later instruction)
  - Control hazards (instruction changes control flow, e.g., branches or exceptions)

- Techniques to handle hazards:
  - Interlock (hold newer instruction until older instructions drain out of pipeline and write back results)
  - Bypass (transfer value from older instruction to newer instruction as soon as available somewhere in machine)
  - Speculate (guess effect of earlier instruction)
**Instruction to Instruction Dependence**

- What do we need to calculate next PC?
  - For Jumps
    » Opcode, offset and PC
  - For Jump Register
    » Opcode and Register value
  - For Conditional Branches
    » Opcode, PC, Register (for condition), and offset
  - For all other instructions
    » Opcode and PC
  • have to know it’s not one of above

---

**PC Calculation Bubbles**  
*(assuming no branch delay slots for now)*

```
<table>
<thead>
<tr>
<th>time</th>
<th>t0</th>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
<th>t6</th>
<th>t7</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I_1)</td>
<td>r1 ← (r0) + 10</td>
<td>IF_1</td>
<td>ID_1</td>
<td>EX_1</td>
<td>MA_1</td>
<td>WB_1</td>
<td>.....</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I_2)</td>
<td>r3 ← (r2) + 17</td>
<td>IF_2</td>
<td>ID_2</td>
<td>EX_2</td>
<td>MA_2</td>
<td>WB_2</td>
<td>.....</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I_3)</td>
<td></td>
<td>IF_3</td>
<td>ID_3</td>
<td>EX_3</td>
<td>MA_3</td>
<td>WB_3</td>
<td>.....</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I_4)</td>
<td></td>
<td>IF_4</td>
<td>ID_4</td>
<td>EX_4</td>
<td>MA_4</td>
<td>WB_4</td>
<td>.....</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

**Resource Usage**

```
<table>
<thead>
<tr>
<th>IF</th>
<th>t0</th>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
<th>t6</th>
<th>t7</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>nop</td>
<td>I_2</td>
<td>nop</td>
<td>I_3</td>
<td>nop</td>
<td>I_4</td>
<td>.....</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>t0</th>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
<th>t6</th>
<th>t7</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>nop</td>
<td>I_2</td>
<td>nop</td>
<td>I_3</td>
<td>nop</td>
<td>I_4</td>
<td>.....</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EX</th>
<th>t0</th>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
<th>t6</th>
<th>t7</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>nop</td>
<td>I_2</td>
<td>nop</td>
<td>I_3</td>
<td>nop</td>
<td>I_4</td>
<td>.....</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MA</th>
<th>t0</th>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
<th>t6</th>
<th>t7</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>nop</td>
<td>I_2</td>
<td>nop</td>
<td>I_3</td>
<td>nop</td>
<td>I_4</td>
<td>.....</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>WB</th>
<th>t0</th>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
<th>t6</th>
<th>t7</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>nop</td>
<td>I_2</td>
<td>nop</td>
<td>I_3</td>
<td>nop</td>
<td>I_4</td>
<td>.....</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


`nop` ⇒ *pipeline bubble*
Speculate next address is PC+4

PCSrc \((pc+4 / jabs / rind/ br)\)

\[ \text{stall} \]

- \[ \text{addr} \]
- \[ \text{inst} \]
- \[ \text{Inst Memory} \]
- \[ \text{IR} \]

I_1 096 ADD
I_2 100 J 304
I_3 104 ADD
I_4 304 ADD

A jump instruction kills (not stalls) the following instruction

How?

Pipelining Jumps

To kill a fetched instruction -- Insert a mux before IR

IRSrc \(D = \text{Case opcode}_D\)

- J, JAL \(\Rightarrow\) nop
- \(\ldots\) \(\Rightarrow\) IM

Any interaction between stall and jump?
Jump Pipeline Diagrams

\[
\begin{array}{cccccccc}
& t0 & t1 & t2 & t3 & t4 & t5 & t6 & t7 \\
(I_1) & 096: ADD & IF_1 & ID_1 & EX_1 & MA_1 & WB_1 & & \\
(I_2) & 100: J 304 & IF_2 & ID_2 & EX_2 & MA_2 & WB_2 & & \\
(I_3) & 104: ADD & IF_3 & nop & nop & nop & nop & & \\
(I_4) & 304: ADD & IF_4 & ID_4 & EX_4 & MA_4 & WB_4 & & \\
\end{array}
\]

Resource Usage

- IF
- ID
- EX
- MA
- WB

 nop \rightarrow \text{pipeline bubble}

Pipelining Conditional Branches

- Condition is not known until the execute stage
- \textit{what action should be taken in the decode stage?}

I_1 096 ADD
I_2 100 BEQZ r1 +200
I_3 104 ADD
I_4 304 ADD

Branch condition is not known until the execute stage
Pipelining Conditional Branches

If the branch is taken

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I₁ 096 ADD</td>
<td>- kill the two following instructions</td>
</tr>
<tr>
<td>I₂ 100 BEQZ r₁ +200</td>
<td>- the instruction at the decode stage is not valid</td>
</tr>
<tr>
<td>I₃ 104 ADD</td>
<td>(\Rightarrow \text{stall signal is not valid})</td>
</tr>
<tr>
<td>I₄ 304 ADD</td>
<td>(\Rightarrow \text{stall signal is not valid})</td>
</tr>
</tbody>
</table>

2/5/2009
CS152-Spring’09

Pipelining Conditional Branches

If the branch is taken

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I₁ 096 ADD</td>
<td>- kill the two following instructions</td>
</tr>
<tr>
<td>I₂ 100 BEQZ r₁ +200</td>
<td>- the instruction at the decode stage is not valid</td>
</tr>
<tr>
<td>I₃ 104 ADD</td>
<td>(\Rightarrow \text{stall signal is not valid})</td>
</tr>
<tr>
<td>I₄ 304 ADD</td>
<td>(\Rightarrow \text{stall signal is not valid})</td>
</tr>
</tbody>
</table>

2/5/2009
CS152-Spring’09
New Stall Signal

\[
\text{stall} = ( (rs_D = ws_E) \cdot \text{we}_E + (rs_D = ws_M) \cdot \text{we}_M + (rs_D = ws_W) \cdot \text{we}_W) \cdot \text{re}_D
+ ((rt_D = ws_E) \cdot \text{we}_E + (rt_D = ws_M) \cdot \text{we}_M + (rt_D = ws_W) \cdot \text{we}_W) \cdot \text{re}_D
\]

Don’t stall if the branch is taken. Why?

Instruction at the decode stage is invalid

Control Equations for PC and IR Muxes

\[
\text{PCSrc} = \text{Case opcode}_E
\]

\[
\begin{align*}
\text{BEQZ}.z, \text{BNEZ}.!z & \Rightarrow \text{br} \\
\text{...} & \Rightarrow \\
\text{Case opcode}_D & \\
\text{J, JAL} & \Rightarrow \text{jabs} \\
\text{JR, JALR} & \Rightarrow \text{rind} \\
\text{...} & \Rightarrow \text{pc+4}
\end{align*}
\]

\[
\text{IRSrc}_D = \text{Case opcode}_E
\]

\[
\begin{align*}
\text{BEQZ}.z, \text{BNEZ}.!z & \Rightarrow \text{nop} \\
\text{...} & \Rightarrow \\
\text{Case opcode}_D & \\
\text{J, JAL, JR, JALR} & \Rightarrow \text{nop} \\
\text{...} & \Rightarrow \text{IM}
\end{align*}
\]

\[
\text{IRSrc}_E = \text{Case opcode}_E
\]

\[
\begin{align*}
\text{BEQZ}.z, \text{BNEZ}.!z & \Rightarrow \text{nop} \\
\text{...} & \Rightarrow \text{stall.nop} + \text{!stall.IRD}
\end{align*}
\]

Give priority to the older instruction, i.e., execute stage instruction over decode stage instruction.
Branch Pipeline Diagrams
(resolved in execute stage)

\[
\begin{align*}
\text{time} & \quad t_0 & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 & t_7 \ldots \\
(I_1) 096: \text{ADD} & & \text{IF}_1 & \text{ID}_1 & \text{EX}_1 & \text{MA}_1 & \text{WB}_1 \\
(I_2) 100: \text{BEQZ }+200 & & \text{IF}_2 & \text{ID}_2 & \text{EX}_2 & \text{MA}_2 & \text{WB}_2 \\
(I_3) 104: \text{ADD} & & \text{IF}_3 & \text{ID}_3 & \text{nop} & \text{nop} & \text{nop} \\
(I_4) 108: \quad & & \text{IF}_4 & \text{ID}_4 & \text{nop} & \text{nop} & \text{nop} \\
(I_5) 304: \text{ADD} & & \text{IF}_5 & \text{ID}_5 & \text{EX}_5 & \text{MA}_5 & \text{WB}_5
\end{align*}
\]

Resource Usage:
- IF
- ID
- EX
- MA
- WB

\[\text{nOP} \Rightarrow \text{pipeline bubble}\]

Reducing Branch Penalty
(resolved in decode stage)

- One pipeline bubble can be removed if an extra comparator is used in the Decode stage

Pipeline diagram now same as for jumps
Branch Delay Slots
(expose control hazard to software)
• Change the ISA semantics so that the instruction that follows a jump or branch is always executed
  – gives compiler the flexibility to put in a useful instruction where normally a pipeline bubble would have resulted.

| $I_1$ | 096 | ADD          |
| $I_2$ | 100 | BEQZ r1 +200 |
| $I_3$ | 104 | ADD          |
| $I_4$ | 304 | ADD          |

- Other techniques include more advanced branch prediction, which can dramatically reduce the branch penalty... to come later

Branch Pipeline Diagrams
(branch delay slot)

<table>
<thead>
<tr>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>t0 t1 t2 t3 t4 t5 t6 t7 . . .</td>
</tr>
</tbody>
</table>

(I) 096: ADD
IF $I_1$
ID $I_1$
EX $I_1$
MA $I_1$
WB $I_1$

(I) 100: BEQZ +200
IF $I_2$
ID $I_2$
EX $I_2$
MA $I_2$
WB $I_2$

(I) 104: ADD
IF $I_3$
ID $I_3$
EX $I_3$
MA $I_3$
WB $I_3$

(I) 304: ADD
IF $I_4$
ID $I_4$
EX $I_4$
MA $I_4$
WB $I_4$

Resource
Usage

<table>
<thead>
<tr>
<th>IF</th>
<th>ID</th>
<th>EX</th>
<th>MA</th>
<th>WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_1$</td>
<td>$I_2$</td>
<td>$I_3$</td>
<td>$I_4$</td>
<td>$I_1$</td>
</tr>
</tbody>
</table>

2/5/2009 CS152-Spring'09 16
Why an Instruction may not be dispatched every cycle (CPI>1)

• Full bypassing may be too expensive to implement
  – typically all frequently used paths are provided
  – some infrequently used bypass paths may increase cycle time and counteract the benefit of reducing CPI

• Loads have two-cycle latency
  – Instruction after load cannot use load result
  – MIPS-I ISA defined load delay slots, a software-visible pipeline hazard (compiler schedules independent instruction or inserts NOP to avoid hazard). Removed in MIPS-II (pipeline interlocks added in hardware)
    » MIPS: “Microprocessor without Interlocked Pipeline Stages”

• Conditional branches may cause bubbles
  – kill following instruction(s) if no delay slots

_Machines with software-visible delay slots may execute significant number of NOP instructions inserted by the compiler. NOPs not counted in useful CPI (alternatively, increase instructions/program)_

CS152 Administrivia

• PS 1 due Tuesday Feb 10 in class
• Section covering PS 1 on _Wednesday Feb 11_
  – Room/time TBD

• First Quiz on Thursday Feb 12
  – In class, closed-book, no computers or calculators
  – Covers lectures 1-5 (today is last lecture in quiz 1)

• Lecture 7, Tuesday Feb 17 in _320 Soda_
• Lecture 8, Thursday Feb 19 back in 306 Soda

• See website for full schedule
**Interrupts:**
altering the normal flow of control

An *external or internal event* that needs to be processed by another (system) program. The event is usually unexpected or rare from program’s point of view.

**Causes of Interrupts**

**Interrupt:** an *event* that requests the attention of the processor

- **Asynchronous:** an *external event*
  - input/output device service-request
  - timer expiration
  - power disruptions, hardware failure
- **Synchronous:** an *internal event (a.k.a. exceptions)*
  - undefined opcode, privileged instruction
  - arithmetic overflow, FPU exception
  - misaligned memory access
  - *virtual memory exceptions:* page faults, TLB misses, protection violations
  - *traps:* system calls, e.g., jumps into kernel
History of Exception Handling

• First system with exceptions was Univac-I, 1951
  – Arithmetic overflow would either
    » 1. trigger the execution a two-instruction fix-up routine at address 0, or
    » 2. at the programmer's option, cause the computer to stop
  – Later Univac 1103, 1955, modified to add external interrupts
    » Used to gather real-time wind tunnel data

• First system with I/O interrupts was DYSEAC, 1954
  – Had two program counters, and I/O signal caused switch between two PCs
  – Also, first system with DMA (direct memory access by I/O device)

[Courtesy Mark Smotherman]
**Asynchronous Interrupts:**

*invoking the interrupt handler*

- An I/O device requests attention by asserting one of the *prioritized interrupt request lines*

- When the processor decides to process the interrupt
  - It stops the current program at instruction $I_i$, completing all the instructions up to $I_{i-1}$ (*precise interrupt*)
  - It saves the PC of instruction $I_i$ in a special register (EPC)
  - It disables interrupts and transfers control to a designated interrupt handler running in the kernel mode

---

**Interrupt Handler**

- Saves EPC before enabling interrupts to allow nested interrupts $\Rightarrow$
  - need an instruction to move EPC into GPRs
  - need a way to mask further interrupts at least until EPC can be saved

- Needs to read a *status register* that indicates the cause of the interrupt

- Uses a special indirect jump instruction RFE (*return-from-exception*) which
  - enables interrupts
  - restores the processor to the user mode
  - restores hardware status and control state
Synchronous Interrupts

- A synchronous interrupt (exception) is caused by a particular instruction.

- In general, the instruction cannot be completed and needs to be restarted after the exception has been handled.
  - requires undoing the effect of one or more partially executed instructions.

- In the case of a system call trap, the instruction is considered to have been completed.
  - a special jump instruction involving a change to privileged kernel mode.

Exception Handling 5-Stage Pipeline

- How to handle multiple simultaneous exceptions in different pipeline stages?
- How and where to handle external asynchronous interrupts?
Exception Handling 5-Stage Pipeline

- Hold exception flags in pipeline until commit point (M stage)

- Exceptions in earlier pipe stages override later exceptions *for a given instruction*

- Inject external interrupts at commit point (override others)

- If exception at commit: update Cause and EPC registers, kill all stages, inject handler PC into fetch stage
Speculating on Exceptions

• Prediction mechanism
  – Exceptions are rare, so simply predicting no exceptions is very accurate!

• Check prediction mechanism
  – Exceptions detected at end of instruction execution pipeline, special hardware for various exception types

• Recovery mechanism
  – Only write architectural state at commit point, so can throw away partially executed instructions after exception
  – Launch exception handler after flushing pipeline

• Bypassing allows use of uncommitted instruction results by following instructions

Exception Pipeline Diagram

```
<table>
<thead>
<tr>
<th>time</th>
<th>t0</th>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
<th>t6</th>
<th>t7</th>
<th>. . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I_1) 096: ADD</td>
<td>IF_1</td>
<td>ID_1</td>
<td>EX_1</td>
<td>MA_1</td>
<td>nop</td>
<td>overflow!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I_2) 100: XOR</td>
<td>IF_2</td>
<td>ID_2</td>
<td>EX_2</td>
<td>nop</td>
<td>nop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I_3) 104: SUB</td>
<td>IF_3</td>
<td>ID_3</td>
<td>nop</td>
<td>nop</td>
<td>nop</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I_4) 108: ADD</td>
<td>IF_4</td>
<td>ID_5</td>
<td>EX_5</td>
<td>MA_5</td>
<td>WB_5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(I_5) Exc. Handler code</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Resource Usage

```
<table>
<thead>
<tr>
<th>Resource</th>
<th>t0</th>
<th>t1</th>
<th>t2</th>
<th>t3</th>
<th>t4</th>
<th>t5</th>
<th>t6</th>
<th>t7</th>
<th>. . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF</td>
<td>I_1</td>
<td>I_2</td>
<td>I_3</td>
<td>I_4</td>
<td>I_5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>I_1</td>
<td>I_2</td>
<td>I_3</td>
<td>nop</td>
<td>I_5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EX</td>
<td>I_1</td>
<td>I_2</td>
<td>I_3</td>
<td>nop</td>
<td>nop</td>
<td>I_5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>I_1</td>
<td>I_2</td>
<td>I_3</td>
<td>nop</td>
<td>nop</td>
<td>nop</td>
<td>I_5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WB</td>
<td>I_1</td>
<td>I_2</td>
<td>I_3</td>
<td>nop</td>
<td>nop</td>
<td>nop</td>
<td>nop</td>
<td>I_5</td>
<td></td>
</tr>
</tbody>
</table>
```

2/5/2009 CS152-Spring’09
Acknowledgements

- These slides contain material developed and copyright by:
  - Arvind (MIT)
  - Krste Asanovic (MIT/UCB)
  - Joel Emer (Intel/MIT)
  - James Hoe (CMU)
  - John Kubiatowicz (UCB)
  - David Patterson (UCB)

- MIT material derived from course 6.823
- UCB material derived from course CS252