Last time in Lecture 3

- Microcoding became less attractive as gap between RAM and ROM speeds reduced
- Complex instruction sets difficult to pipeline, so difficult to increase performance as gate count grew
- Load-Store RISC ISAs designed for efficient pipelined implementations
  - Very similar to vertical microcode
  - Inspired by earlier Cray machines (more on these later)
- Iron Law explains architecture design space
  - Trade instructions/program, cycles/instruction, and time/cycle
An Ideal Pipeline

- All objects go through the same stages
- No sharing of resources between any two stages
- Propagation delay through all pipeline stages is equal
- The scheduling of an object entering the pipeline is not affected by the objects in other stages

These conditions generally hold for industrial assembly lines, but instructions depend on each other!

Pipelined RISC-V

- To pipeline RISC-V:
- First build RISC-V without pipelining with CPI=1
- Next, add pipeline registers to reduce cycle time while maintaining CPI=1
Lecture 3: Unpipelined Datapath for RISC-V

Lecture 3: Hardwired Control Table

<table>
<thead>
<tr>
<th>Opcode</th>
<th>ImmSel</th>
<th>Op2Sel</th>
<th>FuncSel</th>
<th>MemWr</th>
<th>RFWen</th>
<th>WBSel</th>
<th>WASel</th>
<th>PCSel</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALU</td>
<td>*</td>
<td>Reg</td>
<td>Func</td>
<td>no</td>
<td>yes</td>
<td>ALU</td>
<td>rd</td>
<td>pc+4</td>
</tr>
<tr>
<td>ALU i</td>
<td>IType2</td>
<td>Imm</td>
<td>Op</td>
<td>no</td>
<td>yes</td>
<td>ALU</td>
<td>rd</td>
<td>pc+4</td>
</tr>
<tr>
<td>LW</td>
<td>IType2</td>
<td>Imm</td>
<td>+</td>
<td>no</td>
<td>yes</td>
<td>Mem</td>
<td>rd</td>
<td>pc+4</td>
</tr>
<tr>
<td>SW</td>
<td>BType1</td>
<td>Imm</td>
<td>+</td>
<td>yes</td>
<td>no</td>
<td>*</td>
<td>*</td>
<td>pc+4</td>
</tr>
<tr>
<td>BEQ_true</td>
<td>BrType1</td>
<td>*</td>
<td>*</td>
<td>no</td>
<td>no</td>
<td>*</td>
<td>br</td>
<td></td>
</tr>
<tr>
<td>BEQ_false</td>
<td>BrType1</td>
<td>*</td>
<td>*</td>
<td>no</td>
<td>no</td>
<td>*</td>
<td>*</td>
<td>pc+4</td>
</tr>
<tr>
<td>J</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>no</td>
<td>no</td>
<td>*</td>
<td>*</td>
<td>jabs</td>
</tr>
<tr>
<td>JALR</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>no</td>
<td>yes</td>
<td>PC</td>
<td>X1</td>
<td>jabs</td>
</tr>
</tbody>
</table>

Op2Sel= Reg / Imm  WBSel = ALU / Mem / PC
WASel = rd / X1  PCSel = pc+4 / br / rind / jabs
Clock period can be reduced by dividing the execution of an instruction into multiple cycles

\[ t_c > \max \{ t_{IM}, t_{RF}, t_{ALU}, t_{DM}, t_{RW} \} \] (\( = t_{DM} \) probably)

However, CPI will increase unless instructions are pipelined
CPI Examples

Microcoded machine

<table>
<thead>
<tr>
<th>Time</th>
<th>Inst 1</th>
<th>Inst 2</th>
<th>Inst 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 cycles</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>5 cycles</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>10 cycles</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

3 instructions, 22 cycles, CPI=7.33

Unpipelined machine

<table>
<thead>
<tr>
<th>Inst 1</th>
<th>Inst 2</th>
<th>Inst 3</th>
</tr>
</thead>
</table>

3 instructions, 3 cycles, CPI=1

Pipelined machine

<table>
<thead>
<tr>
<th>Inst 1</th>
<th>Inst 2</th>
<th>Inst 3</th>
</tr>
</thead>
</table>

3 instructions, 3 cycles, CPI=1

5-stage pipeline CPI≠5!!!

Technology Assumptions

- A small amount of very fast memory (caches) backed up by a large, slower memory
- Fast ALU (at least for integers)
- Multiported Register files (slower!)

Thus, the following timing assumption is reasonable

\[ t_{IM} = t_{RF} \approx t_{ALU} \approx t_{DM} \approx t_{RW} \]

A 5-stage pipeline will be focus of our detailed design
- some commercial designs have over 30 pipeline stages to do an integer add!
5-Stage Pipelined Execution

```
time  t0  t1  t2  t3  t4  t5  t6  t7  ....
```

- **Instruction 1:** 
  - **IF:** $I_1$ 
  - **ID:** $I_2$ 
  - **EX:** $I_3$ 
  - **MA:** $I_4$ 
  - **WB:** $I_5$

- **Instruction 2:** 
  - **IF:** $I_2$ 
  - **ID:** $I_3$ 
  - **EX:** $I_4$ 
  - **MA:** $I_5$ 
  - **WB:** $I_6$

- **Instruction 3:** 
  - **IF:** $I_3$ 
  - **ID:** $I_4$ 
  - **EX:** $I_5$ 
  - **MA:** $I_6$ 
  - **WB:** $I_7$

- **Instruction 4:** 
  - **IF:** $I_4$ 
  - **ID:** $I_5$ 
  - **EX:** $I_6$ 
  - **MA:** $I_7$ 
  - **WB:** $I_8$

- **Instruction 5:** 
  - **IF:** $I_5$ 
  - **ID:** $I_6$ 
  - **EX:** $I_7$ 
  - **MA:** $I_8$ 
  - **WB:** $I_9$

**Resources:**

- **IF:** $I_1$ 
- **ID:** $I_2$ 
- **EX:** $I_3$ 
- **MA:** $I_4$ 
- **WB:** $I_5$
Pipelined Execution:
ALU Instructions

Not quite correct!
We need an Instruction Reg (IR) for each stage
Instructions interact with each other in pipeline

- An instruction in the pipeline may need a resource being used by another instruction in the pipeline → structural hazard
- An instruction may depend on something produced by an earlier instruction
  - Dependence may be for a data value → data hazard
  - Dependence may be for the next instruction’s address → control hazard (branches, exceptions)

Resolving Structural Hazards

- Structural hazard occurs when two instructions need same hardware resource at same time
  - Can resolve in hardware by stalling newer instruction till older instruction finished with resource
- A structural hazard can always be avoided by adding more hardware to design
  - E.g., if two instructions both need a port to memory at same time, could avoid hazard by adding second port to memory
- Our 5-stage pipeline has no structural hazards by design
  - Thanks to RISC-V ISA, which was designed for pipelining
Resolving Data Hazards (1)

Strategy 1:

Wait for the result to be available by freezing earlier pipeline stages → interlocks

... 

x1 ← x0 + 10 
x4 ← x1 + 17 
...

x1 is stale. Oops!
Feedback to Resolve Hazards

- Later stages provide dependence information to earlier stages which can *stall (or kill) instructions*

- Controlling a pipeline in this manner works provided the instruction at stage $i+1$ can complete without any interference from instructions in stages 1 to $i$ (otherwise deadlocks may occur)

Interlocks to resolve Data Hazards

Stall Condition

$$x_1 \leftarrow x_0 + 10$$

$$x_4 \leftarrow x_1 + 17$$
Stalled Stages and Pipeline Bubbles

\[
\begin{align*}
(I_1) & \times 1 \leftarrow (x0) + 10 \text{\ IF} \\
(I_2) & \times 4 \leftarrow (x1) + 17 \text{\ IF} \\
(I_3) & \text{\ ID} \\
(I_4) & \text{\ ID} \\
(I_5) & \text{\ ID} \quad \ldots
\end{align*}
\]

\text{stalled stages}

Resource Usage

t\quad t0 \quad t1 \quad t2 \quad t3 \quad t4 \quad t5 \quad t6 \quad t7 \quad \ldots

\begin{array}{ccccccccc}
\text{IF} & I_1 & I_2 & \boxed{I_3} & I_3 & I_3 & I_3 & I_4 & I_5 \\
\text{ID} & I_1 & I_2 & I_2 & I_2 & I_2 & I_2 & I_3 & I_4 \\
\text{EX} & I_1 & - & - & - & I_2 & I_3 & I_4 & I_5 \\
\text{MA} & I_1 & - & - & - & I_2 & I_3 & I_4 & I_5 \\
\text{WB} & I_1 & - & - & - & I_2 & I_3 & I_4 & I_5 \\
\end{array}

\Rightarrow \text{pipeline bubble}

Interlock Control Logic

Compare the source registers of the instruction in the decode stage with the destination register of the uncommitted instructions.
Should we always stall if an rs field matches some rd?
not every instruction writes a register ⇒ we
not every instruction reads a register ⇒ re
Deriving the Stall Signal

\[
C_{\text{dest}} \quad \text{ws} = \text{Case opcode}
\]

- JAL \implies X1
- else \implies rd

\[
C_{\text{re}}
\]

- re1 = Case opcode
  - ALU, ALUi
  - \( \text{LW, SW, Bcond, JALR} \implies \text{on} \)
  - J, JAL \implies \text{off}

- re2 = Case opcode
  - ALU, SW, Bcond \implies \text{on}
  - ... \implies \text{off}

\[
\begin{align*}
\text{C}_{\text{stall}} \quad \text{stall} &= ((\text{rs}_1D = \text{ws}_E).\text{we}_E + \\
& (\text{rs}_1D = \text{ws}_M).\text{we}_M + \\
& (\text{rs}_2D = \text{ws}_W).\text{we}_W).\text{re}_1D + \\
& ((\text{rs}_2D = \text{ws}_E).\text{we}_E + \\
& (\text{rs}_2D = \text{ws}_M).\text{we}_M + \\
& (\text{rs}_2D = \text{ws}_W).\text{we}_W).\text{re}_2D
\end{align*}
\]

Hazards due to Loads & Stores

Stall Condition

Is there any possible data hazard in this instruction sequence?

\[ \text{M}[x1+7] \leftarrow x2 \\
x4 \leftarrow \text{M}[x3+5] \]

What if \( x1+7 = x3+5 \)?
Load & Store Hazards

...  
M[x1+7] ← x2  
x4 ← M[x3+5]  

x1+7 = x3+5 ⇒ data hazard  

However, the hazard is avoided because our memory system completes writes in one cycle!

Load/Store hazards are sometimes resolved in the pipeline and sometimes in the memory system itself.

More on this later in the course.

CS152 Administrivia

- Quiz 1 on Feb 19 will cover PS1, Lab1, lectures 1-5, and associated readings.
Resolving Data Hazards (2)

Strategy 2:

Route data as soon as possible after it is calculated to the earlier pipeline stage \( \Rightarrow \text{bypass} \)

Bypassing

Each \textit{stall or kill} introduces a bubble in the pipeline

\[\Rightarrow CPI > 1\]

A new datapath, i.e., \textit{a bypass}, can get the data from the output of the ALU to its input
Adding a Bypass

When does this bypass help?

(l₁) \( x_1 \leftarrow x_0 + 10 \)  
(l₂) \( x_4 \leftarrow x_1 + 17 \)  

\( x_1 \leftarrow M[x_0 + 10] \)  
\( x_4 \leftarrow x_1 + 17 \)  

The Bypass Signal

Deriving it from the Stall Signal

\[
\text{stall} = ((\text{rs}_1 = \text{ws}_1).\text{we}_1 + (\text{rs}_1 = \text{ws}_M).\text{we}_M + (\text{rs}_1 = \text{ws}_W).\text{we}_W).\text{re}_1 + ((\text{rs}_2 = \text{ws}_E).\text{we}_E + (\text{rs}_2 = \text{ws}_M).\text{we}_M + (\text{rs}_2 = \text{ws}_W).\text{we}_W).\text{re}_2
\]

\[\text{ws} = \begin{cases} \text{Case opcode} & \text{JAL} \Rightarrow X1 \\ \text{else} & \Rightarrow \text{rd} \end{cases}\]

\[\text{we} = \begin{cases} \text{Case opcode} & \text{ALU, ALUI, LW, JALR} \Rightarrow (\text{ws} \neq 0) \\ \text{JAL} & \Rightarrow \text{on} \\ \text{...} & \Rightarrow \text{off} \end{cases}\]

\[\text{ASrc} = (\text{rs}_1 = \text{ws}_E).\text{we}_E.\text{re}_1\]

Is this correct?

No because only ALU and ALUI instructions can benefit from this bypass.

Split \( \text{we}_E \) into two components: we-bypass, we-stall
Bypass and Stall Signals

Split $we_E$ into two components: we-bypass, we-stall

$$we\text{-bypass}_E = \text{Case } \text{opcode}_E$$
$$\text{ALU, ALUi} \Rightarrow (ws \neq 0)$$
$$\ldots \Rightarrow \text{off}$$

$$we\text{-stall}_E = \text{Case } \text{opcode}_E$$
$$\text{LW, JALR} \Rightarrow (ws \neq 0)$$
$$\text{JAL} \Rightarrow \text{on}$$
$$\ldots \Rightarrow \text{off}$$

$$\text{ASrc} = (rs_1 = ws_E).we\text{-bypass}_E \cdot \text{re}_1_D$$

$$\text{stall} = ((rs_1 = ws_E).we\text{-stall}_E +$$
$$\quad (rs_1 = ws_M).we_M + (rs_1 = ws_W).we_W).\text{re}_1_D$$
$$\quad + ((rs_2 = ws_E).we_E + (rs_2 = ws_M).we_M + (rs_2 = ws_W).we_W).\text{re}_2_D$$

Fully Bypassed Datapath

Is there still a need for the stall signal?

$$\text{stall} = (rs_1 = ws_E). (\text{opcode}_E = \text{LW}_E). (ws_E \neq 0). \text{re}_1_D$$
$$\quad + (rs_2 = ws_E). (\text{opcode}_E = \text{LW}_E). (ws_E \neq 0). \text{re}_2_D$$
**Pipeline CPI Examples**

Measure from when first instruction finishes to when last instruction in sequence finishes.

- **Example 1:**
  - Inst 1
  - Inst 2
  - Inst 3
  - 3 instructions finish in 3 cycles
  - CPI = 3/3 = 1

- **Example 2:**
  - Inst 1
  - Inst 2
  - Bubble
  - Inst 3
  - 3 instructions finish in 4 cycles
  - CPI = 4/3 = 1.33

- **Example 3:**
  - Inst 1
  - Bubble 1
  - Inst 2
  - Bubble 2
  - Inst 3
  - 3 instructions finish in 5 cycles
  - CPI = 5/3 = 1.67

**Resolving Data Hazards (3)**

*Strategy 3: Speculate on the dependence!*

*Two cases:*

- *Guessed correctly ➔* do nothing
- *Guessed incorrectly ➔* kill and restart

.... We’ll later see examples of this approach in more complex processors.
Speculation that load value=zero

\[
\text{Guess_zero} = (\text{rs}_{1D} = \text{ws}_E). (\text{opcode}_E = \text{LW}_E). (\text{ws}_E = 0). \text{re}_{1D}
\]

Also need to add circuitry to remember that this was a guess and flush pipeline if load not zero!

Not worth doing in practice – why?

Control Hazards

What do we need to calculate next PC?

- For Jumps
  - Opcode, PC and offset
- For Jump Register
  - Opcode, Register value, and PC
- For Conditional Branches
  - Opcode, Register (for condition), PC and offset
- For all other instructions
  - Opcode and PC (and have to know it’s not one of above)
PC Calculation Bubbles

\[
\begin{align*}
(I_1)_x 1 &\leftarrow x0 + 10 \\
(I_2)_x 3 &\leftarrow x2 + 17 \\
(I_3) &\leftarrow \\
(I_4) &\leftarrow
\end{align*}
\]

\[
\begin{array}{cccccccc}
time & t0 & t1 & t2 & t3 & t4 & t5 & t6 & t7 \\
\text{IF} & I_1 & - & I_2 & - & I_3 & - & I_4 & - \\
\text{ID} & I_1 & - & I_2 & - & I_3 & - & I_4 & - \\
\text{EX} & I_1 & - & I_2 & - & I_3 & - & I_4 & - \\
\text{MA} & I_1 & - & I_2 & - & I_3 & - & I_4 & - \\
\text{WB} & I_1 & - & I_2 & - & I_3 & - & I_4 & - \\
\end{array}
\]

Resource Usage:

→ pipeline bubble

Speculate next address is PC+4

A jump instruction kills (not stalls) the following instruction

How?
Pipelining Jumps

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

Any interaction between stall and jump?

IRSrcD = Case opcodeD
J, JAL ⇒ bubble
... ⇒ IM

Resource Usage

- ⇒ pipeline bubble

Jump Pipeline Diagrams

<table>
<thead>
<tr>
<th></th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t0</td>
</tr>
<tr>
<td>(I₁) 096: ADD</td>
<td>IF₁</td>
</tr>
<tr>
<td>(I₂) 100: J 304</td>
<td>IF₂</td>
</tr>
<tr>
<td>(I₃) 104: ADD</td>
<td>IF₃</td>
</tr>
<tr>
<td>(I₄) 304: ADD</td>
<td>IF₄</td>
</tr>
</tbody>
</table>

1/31/2013 CS152, Spring 2013
Pipelining Conditional Branches

Branch condition is not known until the execute stage

what action should be taken in the decode stage?

If the branch is taken
- kill the two following instructions
- the instruction at the decode stage is not valid ⇒ stall signal is not valid
Pipelining Conditional Branches

If the branch is taken:
- kill the two following instructions
- the instruction at the decode stage is not valid ⇒ stall signal is not valid

Branch Pipeline Diagrams
(resolved in execute stage)

Resource Usage

- ⇒ pipeline bubble
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