Recap: Microprogramming

- Microprogramming is a convenient method for implementing structured control state diagrams:
  - Random logic replaced by microPC sequencer and ROM
  - Each line of ROM called a microinstruction: contains sequencer control + values for control points
  - Limited state transitions: branch to zero, next sequential, branch to microinstruction address from dispatch ROM

- Design of a Microprogramming language
  1. Start with list of control signals
  2. Group signals together that make sense (vs. random): called “fields”
  3. Place fields in some logical order (e.g., ALU operation & ALU operands first and microinstruction sequencing last)
  4. To minimize the width, encode operations that will never be used at the same time
  5. Create a symbolic legend for the microinstruction format, showing name of field values and how they set the control signals

Recap: Multicycle datapath (book)
**Recap: Start with List of control signals**

<table>
<thead>
<tr>
<th>Signal name</th>
<th>Effect when deasserted</th>
<th>Effect when asserted</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALUSelA</td>
<td>1st ALU operand = PC</td>
<td>1st ALU operand = Reg[rs]</td>
</tr>
<tr>
<td>RegWrite</td>
<td>Reg is write</td>
<td>Reg is write</td>
</tr>
<tr>
<td>MemToReg</td>
<td>Reg. write data input = ALU</td>
<td>Reg. write data input = memory</td>
</tr>
<tr>
<td>RegDst</td>
<td>Reg. dest. no. = rt</td>
<td>Reg. dest. no. = rd</td>
</tr>
<tr>
<td>MemRead</td>
<td>None</td>
<td>Memory at address is read, MDR &lt;= Mem[addr]</td>
</tr>
<tr>
<td>MemWrite</td>
<td>None</td>
<td>Memory at address is written</td>
</tr>
<tr>
<td>IorD</td>
<td>Memory address = PC</td>
<td>Memory address = S</td>
</tr>
<tr>
<td>IRWrite</td>
<td>None</td>
<td>IR &lt;= Memory</td>
</tr>
<tr>
<td>PCWrite</td>
<td>None</td>
<td>PC &lt;= PCSource</td>
</tr>
<tr>
<td>PCSrc</td>
<td>PCSource = ALU</td>
<td>PCSource = ALUout</td>
</tr>
<tr>
<td>ExtOp</td>
<td>Zero</td>
<td>Sign Extended</td>
</tr>
</tbody>
</table>

**Recap: Group together related signals**

**Ideal**
- Memory
  - WrAdr: 32
  - Din: 32
  - RAdr: 32
- 32

**ALU**
- ALUOp: 32
  - ALU adds
  - ALU subtracts
  - ALU does function code
  - ALU does logical OR
- ALUSelB: 32
  - 2nd ALU input = Reg[rt]
- SRC1
  - PC: 1st ALU input <= PC
- SRC2
  - 2nd ALU input <= 4
- ExtOp: 32
  - 2nd ALU input <= sign ext. IR[15-0]
- Memreg
  - IR <= Mem
- MemRead
- MemWrite
- Memreg

**Recap: Group into Fields, Order and Assign Names**

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Values for Field</th>
<th>Function of Field with Specific Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALU</td>
<td>Add</td>
<td>ALU adds</td>
</tr>
<tr>
<td></td>
<td>Subt.</td>
<td>ALU subtracts</td>
</tr>
<tr>
<td></td>
<td>Func</td>
<td>ALU does function code</td>
</tr>
<tr>
<td></td>
<td>Or</td>
<td>ALU does logical OR</td>
</tr>
<tr>
<td>SRC1</td>
<td>PC</td>
<td>1st ALU input &lt;= PC</td>
</tr>
<tr>
<td></td>
<td>rs</td>
<td>1st ALU input &lt;= Reg[rs]</td>
</tr>
<tr>
<td>SRC2</td>
<td>4</td>
<td>2nd ALU input &lt;= 4</td>
</tr>
<tr>
<td></td>
<td>Extend</td>
<td>2nd ALU input &lt;= sign ext. IR[15-0]</td>
</tr>
<tr>
<td></td>
<td>Extend0</td>
<td>2nd ALU input &lt;= zero ext. IR[15-0]</td>
</tr>
<tr>
<td></td>
<td>ExtShift</td>
<td>2nd ALU input &lt;= sign ext. sl IR[15-0]</td>
</tr>
<tr>
<td></td>
<td>rt</td>
<td>2nd ALU input &lt;= Reg[rt]</td>
</tr>
<tr>
<td>dest(ination)</td>
<td>rd ALU</td>
<td>Reg[rd] &lt;= ALUout</td>
</tr>
<tr>
<td></td>
<td>rt ALU</td>
<td>Reg[rt] &lt;= ALUout</td>
</tr>
<tr>
<td></td>
<td>rt Mem</td>
<td>Reg[rt] &lt;= Mem</td>
</tr>
<tr>
<td>Mem(ory)</td>
<td>Read PC</td>
<td>Read memory using PC</td>
</tr>
<tr>
<td></td>
<td>Read ALU</td>
<td>Read memory using ALUout for addr</td>
</tr>
<tr>
<td></td>
<td>Write ALU</td>
<td>Write memory using ALUout for addr</td>
</tr>
<tr>
<td>Memreg</td>
<td>IR</td>
<td>IR &lt;= Mem</td>
</tr>
<tr>
<td>PCWrite</td>
<td>PCWr</td>
<td>PC &lt;= PCSrc</td>
</tr>
<tr>
<td></td>
<td>PCSrc</td>
<td>IF Zero then PC &lt;= PCSrc &lt;= ALUout</td>
</tr>
<tr>
<td></td>
<td>PCWriteCond</td>
<td>IF Zero then PC &lt;= PCSrc &lt;= ALUout</td>
</tr>
<tr>
<td>Seq(uencing)</td>
<td>Seq</td>
<td>Go to next sequential instruction</td>
</tr>
<tr>
<td></td>
<td>Fetch</td>
<td>Go to the first microinstruction</td>
</tr>
<tr>
<td></td>
<td>Dispatch</td>
<td>Dispatch using ROM</td>
</tr>
</tbody>
</table>

**Recap: Quick check: what do these fieldnames mean?**

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Values for Field</th>
<th>Function of Field with Specific Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALUSelB</td>
<td>00</td>
<td>ALU adds</td>
</tr>
<tr>
<td></td>
<td>01</td>
<td>ALU subtracts</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>ALU does function code</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>ALU does logical OR</td>
</tr>
<tr>
<td>ExtOp</td>
<td>000</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>001</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>010</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>011</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>111</td>
<td>X</td>
</tr>
</tbody>
</table>

**Destination:**

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>RegWrite</th>
<th>MemToReg</th>
<th>RegDest</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>---</td>
<td>0</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>01</td>
<td>rd ALU</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>rt ALU</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>rt MEM</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**SRC2:**

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>ALUSelB</th>
<th>ExtOp</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>---</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>001</td>
<td>4</td>
<td>0</td>
<td>X</td>
</tr>
<tr>
<td>010</td>
<td>rt</td>
<td>01</td>
<td>X</td>
</tr>
<tr>
<td>011</td>
<td>ExtShift</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>Extend</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>111</td>
<td>Extend0</td>
<td>11</td>
<td>0</td>
</tr>
</tbody>
</table>
Recap: Finite State Machine (FSM) Spec

**IR <= MEM[PC]**

*instruction fetch*

**PC <= PC + 4**

**R-type**

**ALUout <= A fun B**

**ORi**

**ALUout <= A or ZX**

**R[rt] <= ALUout**

**ORi**

**ALUout <= A + SX**

**R[rt] <= M**

**M <= MEM[ALUout]**

**LW**

**ALUout <= A + SX**

**MEM[ALUout] <= B**

**SW**

**ALUout <= A + SX**

**MEM[ALUout] <= B**

Recap: Microprogram it yourself!

<table>
<thead>
<tr>
<th>Addr</th>
<th>ALU SRC1 SRC2 Dest.</th>
<th>Memory</th>
<th>Mem. Reg.</th>
<th>PC Write Sequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>Add     PC  4        Read PC IR ALU Seq</td>
<td>Dispatch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0001</td>
<td>Add     PC Extshft   Read PC IR ALU Seq</td>
<td>Fetch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0010</td>
<td>Subt. rs rt ALUOutCond. Fetch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0100</td>
<td>Func rs rt rd ALUSeq</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0101</td>
<td>Read ALU Seq Fetch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0110</td>
<td>Or rs Extend0 Seq Fetch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0111</td>
<td>rt ALU Seq Fetch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>Add rs Extend Seq Fetch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>Read ALU Seq Fetch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1010</td>
<td>rt MEM Seq Fetch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1011</td>
<td>Add rs Extend Seq Fetch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1100</td>
<td>Write ALU Seq Fetch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1101</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Recap: Specific Sequencer from last lecture

Sequencer-based control unit from last lecture
- Called “microPC” or “µPC” vs. state register

<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>fetch</td>
<td>µaddress = 0</td>
</tr>
<tr>
<td>01</td>
<td>dispatch</td>
<td>µaddress = dispatch ROM</td>
</tr>
<tr>
<td>10</td>
<td>seq</td>
<td>µaddress = µaddress + 1</td>
</tr>
</tbody>
</table>

Initial Representation

- **Finite State Diagram**
- **Microprogram**

Sequencing Control

- **Explicit Next State Function**
- **Microprogram counter + Dispatch ROMs**

Logic Representation

- **Logic Equations**
- **Truth Tables**

Implementation Technique

- **PLA**
- **“hardwired control”**
- **“microprogrammed control”**
**Exceptions**

- **Exception** = unprogrammed control transfer
  - system takes action to handle the exception
    - must record the address of the offending instruction
    - record any other information necessary to return afterwards
  - returns control to user
  - must save & restore user state
- **Allows construction of a “user virtual machine”**

**Two Types of Exceptions: Interrupts and Traps**

- **Interrupts**
  - caused by external events:
    - Network, Keyboard, Disk I/O, Timer
  - asynchronous to program execution
    - Most interrupts can be disabled for brief periods of time
    - Some (like “Power Failing”) are non-maskable (NMI)
  - may be handled between instructions
  - simply suspend and resume user program
- **Traps**
  - caused by internal events
    - exceptional conditions (overflow)
    - errors (parity)
    - faults (non-resident page)
  - synchronous to program execution
    - condition must be remedied by the handler
    - instruction may be retried or simulated and program continued
    - or program may be aborted

**Precise Exceptions**

- **Precise** ⇒ state of the machine is preserved as if program executed up to the offending instruction
  - All previous instructions completed
  - Offending instruction and all following instructions act as if they have not even started
  - Same system code will work on different implementations
  - Position clearly established by IBM
  - Difficult in the presence of pipelining, out-of-order execution, ...
  - MIPS takes this position
- **Imprecise** ⇒ system software has to figure out what is where and put it all back together
- **Performance goals often lead designers to forsake precise interrupts**
  - system software developers, user, markets etc. usually wish they had not done this
- **Modern techniques for out-of-order execution and branch prediction help implement precise interrupts**

**Big Picture: user / system modes**

- **Two modes of execution (user/system)**:
  - operating system runs in privileged mode and has access to all of the resources of the computer
  - presents “virtual resources” to each user that are more convenient than the physical resources
    - files vs. disk sectors
    - virtual memory vs physical memory
  - protects each user program from others
  - protects system from malicious users.
  - OS is assumed to “know best”, and is trusted code, so enter system mode on exception
- **Exceptions allow the system to taken action in response to events that occur while user program is executing**:
  - Might provide supplemental behavior (dealing with denormal floating-point numbers for instance).
  - “Unimplemented instruction” used to emulate instructions that were not included in hardware (i.e. MicroVax)
**Administrivia**

- Midterm I on Wednesday 3/10
  - 5:30 - 8:30 in 306 Soda Hall
  - Bring a Calculator!
  - One 8 1/2 by 11 page (both sides) of notes
  - Make up exam: Tuesday 5:30 – 8:30 in 606 Soda Hall
- Meet at LaVal’s afterwards for Pizza
  - North-side Lavals!
  - I’ll Buy!
- Materials through Chapter 5, some of Chapter 6, Appendix A, B & C
  - Should understand single-cycle processor
  - Multicycle processor, including exceptions
- Get moving on Lab 3!
  - This is a hard lab, since there are so many new things

**Addressing the Exception Handler**

- **Traditional Approach: Interrupt Vector**
  - PC <- MEM[IV_base + cause || 00]
  - 370, 68000, Vax, 80x86, . . .
- **RISC Handler Table**
  - PC <- IT_base + cause || 0000
  - Saves state and jumps
  - Sparc, PA, M88K, . . .
- **MIPS Approach: fixed entry**
  - PC <- EXC_addr
  - Actually very small table
    - RESET entry
    - TLB
    - other

**Saving State**

- Push it onto the stack
  - Vax, 68k, 80x86
- **Shadow Registers**
  - M88k
  - Save state in a shadow of the internal pipeline registers
- **Save it in special registers**
  - MIPS EPC, BadVaddr, Status, Cause

**Additions to MIPS ISA to support Exceptions?**

- **Exception state is kept in “coprocessor 0”**.
  - Use mfc0 read contents of these registers
  - Every register is 32 bits, but may be only partially defined
- **BadVAddr (register 8)**
  - Register contained memory address at which memory reference occurred
- **Status (register 12)**
  - Interrupt mask and enable bits
- **Cause (register 13)**
  - The cause of the exception
  - Bits 5 to 2 of this register encodes the exception type (e.g. undefined instruction=10 and arithmetic overflow=12)
- **EPC (register 14)**
  - Address of the affected instruction (register 14 of coprocessor 0).
- **Control signals to write BadVAddr, Status, Cause, and EPC**
  - Be able to write exception address into PC (8000 0080; low)
  - May have to undo PC = PC + 4, since want EPC to point to offending instruction (not its successor): PC = PC - 4
Details of Status register

<table>
<thead>
<tr>
<th>15</th>
<th>8</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mask</td>
<td>k</td>
<td>e</td>
<td>k</td>
<td>e</td>
<td>k</td>
<td>e</td>
<td></td>
</tr>
</tbody>
</table>

° Mask = 1 bit for each of 5 hardware and 3 software interrupt levels
  • 1 => enables interrupts
  • 0 => disables interrupts
° k = kernel/user
  • 0 => was in the kernel when interrupt occurred
  • 1 => was running user mode
° e = interrupt enable
  • 0 => interrupts were disabled
  • 1 => interrupts were enabled
° When interrupt occurs, 6 LSB shifted left 2 bits, setting 2 LSB to 0
  • run in kernel mode with interrupts disabled

Details of Cause register

<table>
<thead>
<tr>
<th>Status</th>
<th>Pending</th>
<th>Code</th>
</tr>
</thead>
</table>

° Pending interrupt 5 hardware levels: bit set if interrupt occurs but not yet serviced
  • handles cases when more than one interrupt occurs at same time, or while records interrupt requests when interrupts disabled
° Exception Code encodes reasons for interrupt
  • 0 (INT) => external interrupt
  • 4 (ADDRL) => address error exception (load or instr fetch)
  • 5 (ADRS) => address error exception (store)
  • 6 (IBUS) => bus error on instruction fetch
  • 7 (DBUS) => bus error on data fetch
  • 8 (Syscall) => Syscall exception
  • 9 (BKPT) => Breakpoint exception
  • 10 (RI) => Reserved Instruction exception
  • 12 (OVF) => Arithmetic overflow exception

Part of the handler in trap_handler.s

```assembly
.entry:      # Exceptions/interrupts come here
    set noat
    move $k1 $at    # Save $at
    set at
    sw $v0 s1      # Not re-entrant and we can’t trust $sp
    sw $a0 s2
    mf0 $k0 $13    # Cause      # Grab the cause register
    li $v0 4        # syscall 4 (print_str)
    la $a0 __m1_syscall
    li $v0 1        # syscall 1 (print_int)
    srl $a0 $k0 2   # shift Cause reg
    syscall

    ret:          # Exceptions/interrupts come here
    lw $v0 s1
    lw $a0 s2
    mf0 $k0 $14    # EPC      # Get the return address (EPC)
    .set noat
    move $at $k1   # Restore $at
    .set at
    rfe            # Return from exception handler
    addiu $a0 $k0 4 # Return to next instruction
    jr $k0
```

Example: How Control Handles Traps in our FSD

° Undefined Instruction—detected when no next state is defined from state 1 for the op value.
  • We handle this exception by defining the next state value for all op values other than lw, sw, 0 (R-type), jmp, beq, and ori as new state 12.
  • Shown symbolically using “other” to indicate that the op field does not match any of the opcodes that label arcs out of state 1.
° Arithmetic overflow—detected on ALU ops such as signed add
  • Used to save PC and enter exception handler
° External Interrupt – flagged by asserted interrupt line
  • Again, must save PC and enter exception handler
° Note: Challenge in designing control of a real machine is to handle different interactions between instructions and other exception-causing events such that control logic remains small and fast.
  • Complex interactions makes the control unit the most challenging aspect of hardware design
How add traps and interrupts to state diagram?

```
How add traps and interrupts to state diagram?

IR <= MEM[PC]
PC <= PC + 4

R-type
S <= A fun B
R[rd] <= S
PC <= PC + 4

S <= A + ZX
R[rt] <= S
PC <= PC + 4

SW
MEM[S] <= B
PC <= PC + 4

LW
S <= A + SX
R[rt] <= M
PC <= PC + 4

BEQ
PC <= Next(PC)

R[rd] <= S
PC <= PC + 4

S <= PC + SX
EPC <= PC - 4

Pending INT
EPC <= PC - 4
PC <= exp_addr
cause <= 0(INT)

PC <= exp_addr
cause <= 10(RI)

overflow
other

EPC <= PC - 4
cause <= 12 (Ovf)

SW
MEM[S] <= B
PC <= PC + 4

LW
S <= A + SX
R[rt] <= M
PC <= PC + 4

BEQ
PC <= Next(PC)

S <= PC + SX
EPC <= PC - 4

Pending INT
EPC <= PC - 4
PC <= exp_addr
cause <= 0(INT)

PC <= exp_addr
cause <= 10(RI)

overflow
other

EPC <= PC - 4
cause <= 12 (Ovf)

But: What has to change in our µ-sequencer?

° Need concept of branch at micro-code level

Interrupts are precise because user-visible state committed after exceptions flagged!

Example: Can easily use with for non-ideal memory

```
Example: Can easily use with for non-ideal memory

```

Question #1: Why do microcoding?
° If simple instruction could execute at very high clock rate...
° If you could even write compilers to produce microinstructions...
° If most programs use simple instructions and addressing modes...
° If microcode is kept in RAM instead of ROM so as to fix bugs ...
° If same memory used for control memory could be used instead as cache for “macroinstructions”...
° Then why not skip instruction interpretation by a microprogram and simply compile directly into lowest language of machine? (microprogramming is overkill when ISA matches datapath 1-1)
Legacy Software and Microprogramming

- IBM bet company on 360 Instruction Set Architecture (ISA): single instruction set for many classes of machines
  - (8-bit to 64-bit)
- Stewart Tucker stuck with job of what to do about software compatibility
  - If microprogramming could easily do same instruction set on many different microarchitectures, then why couldn’t multiple microprograms do multiple instruction sets on the same microarchitecture?
  - Coined term “emulation”: instruction set interpreter in microcode for non-native instruction set
  - Very successful: in early years of IBM 360 it was hard to know whether old instruction set or new instruction set was more frequently used

Microprogramming Pros and Cons

- Ease of design
- Flexibility
  - Easy to adapt to changes in organization, timing, technology
  - Can make changes late in design cycle, or even in the field
- Can implement very powerful instruction sets (just more control memory)
- Generality
  - Can implement multiple instruction sets on same machine.
  - Can tailor instruction set to application.
- Compatibility
  - Many organizations, same instruction set
- Costly to implement
- Slow

Recall: Performance Evaluation

- What is the average CPI?
  - state diagram gives CPI for each instruction type
  - workload gives frequency of each type

<table>
<thead>
<tr>
<th>Type</th>
<th>CPI, for type</th>
<th>Frequency</th>
<th>CPI x freq,</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arith/Logic</td>
<td>4</td>
<td>40%</td>
<td>1.6</td>
</tr>
<tr>
<td>Load</td>
<td>5</td>
<td>30%</td>
<td>1.5</td>
</tr>
<tr>
<td>Store</td>
<td>4</td>
<td>10%</td>
<td>0.4</td>
</tr>
<tr>
<td>branch</td>
<td>3</td>
<td>20%</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Average CPI: 4.1

Question #2: Can we get CPI < 4.1?

- Seems to be lots of “idle” hardware
  - Why not overlap instructions???
The Big Picture: Where are We Now?

° The Five Classic Components of a Computer

- Processor
  - Control
- Memory
- Datapath
- Input
- Output

° Next Topics:
  - Pipelining by Analogy
  - Pipeline hazards

Pipelining is 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

6 PM 7 8 9 10 11 Midnight

- Task Order

○ Sequential laundry takes 6 hours for 4 loads

○ If they learned pipelining, how long would laundry take?

Pipelined Laundry: Start work ASAP

6 PM 7 8 9 10 11 Midnight

- Task Order

○ Pipelined laundry takes 3.5 hours for 4 loads
Pipelining Lessons

- Pipelining doesn’t help latency of single task, it helps throughput of entire workload.
- Pipeline rate limited by slowest pipeline stage.
- Multiple tasks operating simultaneously using different resources.
- Potential speedup = Number pipe stages.
- Unbalanced lengths of pipe stages reduces speedup.
- Time to “fill” pipeline and time to “drain” it reduces speedup.
- Stall for Dependences.

The Five Stages of Load

- Ifetch: Instruction Fetch
  - Fetch the instruction from the Instruction Memory.
- Reg/Dec: Registers Fetch and Instruction Decode
- Exec: Calculate the memory address.
- Mem: Read the data from the Data Memory.
- Wr: Write the data back to the register file.

Note: These 5 stages were there all along!

- If IR <= MEM[PC], PC <= PC + 4
- If R-type
  - ALUout <= A fun B
  - ALUout <= A op ZX
  - ALUout <= A + SX
  - If A = B then PC <= ALUout
- If ORI
  - ALUout <= PC + SX
- If LW
  - ALUout <= A + SX
- If SW
  - ALUout <= ALUout
- If BEQ
  - PC <= ALUout

Pipelining

- Improve performance by increasing throughput.
  - Ideal speedup is number of stages in the pipeline. Do we achieve this?
Basic Idea

What do we need to add to split the datapath into stages?

Graphically Representing Pipelines

Can help with answering questions like:
- how many cycles does it take to execute this code?
- what is the ALU doing during cycle 4?
- use this representation to help understand datapaths

Conventional Pipelined Execution Representation

Single Cycle, Multiple Cycle, vs. Pipeline
Why Pipeline?

- Suppose we execute 100 instructions
  - Single Cycle Machine
    - 45 ns/cycle \times 1 \text{ CPI} \times 100 \text{ inst} = 4500 \text{ ns}
  - Multicycle Machine
    - 10 ns/cycle \times 4.1 \text{ CPI} (due to inst mix) \times 100 \text{ inst} = 4100 \text{ ns}
  - Ideal pipelined machine
    - 10 ns/cycle \times (1 \text{ CPI} \times 100 \text{ inst} + 4 \text{ cycle drain}) = 1040 \text{ ns}

Why Pipeline? Because we can!

Can pipelining get us into trouble?

- Yes: Pipeline Hazards
  - structural hazards: attempt to use the same resource two different ways at the same time
    - E.g., combined washer/dryer would be a structural hazard or folder busy doing something else (watching TV)
  - control hazards: attempt to make a decision before condition is evaluated
    - E.g., washing football uniforms and need to get proper detergent level; need to see after dryer before next load in
      - branch instructions
  - data hazards: attempt to use item before it is ready
    - E.g., one sock of pair in dryer and one in washer; can’t fold until get sock from washer through dryer
    - instruction depends on result of prior instruction still in the pipeline

- Can always resolve hazards by waiting
  - pipeline control must detect the hazard
  - take action (or delay action) to resolve hazards

Single Memory is a Structural Hazard

Detection is easy in this case! (right half highlight means read, left half write)
**Structural Hazards limit performance**

- Example: if 1.3 memory accesses per instruction and only one memory access per cycle then
  - average CPI ≥ 1.3
  - otherwise resource is more than 100% utilized

Control Hazard Solution #1: Stall

- **Stall**: wait until decision is clear
- **Impact**: 2 lost cycles (i.e. 3 clock cycles per branch instruction) => slow
- **Move decision to end of decode**
  - save 1 cycle per branch

Control Hazard Solution #2: Predict

- **Predict**: guess one direction then back up if wrong
- **Impact**: 0 lost cycles per branch instruction if right, 1 if wrong (right - 50% of time)
  - Need to “Squash” and restart following instruction if wrong
  - Produce CPI on branch of (1 * .5 + 2 * .5) = 1.5
  - Total CPI might then be: 1.5 * .2 + 1 * .8 = 1.1 (20% branch)
- **More dynamic scheme: history of 1 branch (~ 90%)**

Control Hazard Solution #3: Delayed Branch

- **Delayed Branch**: Redefine branch behavior (takes place after next instruction)
- **Impact**: 0 clock cycles per branch instruction if can find instruction to put in “slot” (~ 50% of time)
- **As launch more instruction per clock cycle, less useful**
Data Hazard on r1:

- Dependencies backwards in time are hazards

```
add r1, r2, r3
sub r4, r1, r3
and r6, r1, r7
or r8, r1, r9
xor r10, r1, r11
```

Data Hazard Solution:

- “Forward” result from one stage to another

```
add r1, r2, r3
sub r4, r1, r3
and r6, r1, r7
or r8, r1, r9
xor r10, r1, r11
```

Forwarding (or Bypassing): What about Loads?

- Dependencies backwards in time are hazards

```
lw r1, 0(r2)
sub r4, r1, r3
```

```
• Can’t solve with forwarding:
• Must delay/stall instruction dependent on loads
```
**Forwarding (or Bypassing): What about Loads**

- Dependencies backwards in time are hazards

  *Time (clock cycles)*

  ![Diagram](image)

  - Cannot solve with forwarding:
  - Must delay/stall instruction dependent on loads

**Designing a Pipelined Processor**

- Go back and examine your datapath and control diagram
- Associated resources with states
- Ensure that flows do not conflict, or figure out how to resolve
- Assert control in appropriate stage

**Control and Datapath: Split state diag into 5 pieces**

- IR <- Mem[PC]; PC <- PC+4;
- A <- R[rs]; B <- R[rt]
- S <- A + B;
- S <- A or ZX;
- S <- A + SX;
- If Cond PC <- PC+SX;
- M <- Mem[S]; Mem[S] <- B
- R[rd] <- S;
- R[rt] <- S;
- R[rd] <- M;
- Next PC
- PC
- Inst. Mem
- IR
- Reg File
- A
- B
- Exec
- Mem Access
- M
- Reg File

**Summary: Exceptions**

- Microprogramming is a fundamental concept
  - Implement an instruction set by building a very simple processor and interpreting the instructions
  - Essential for very complex instructions and when few register transfers are possible
  - Control design reduces to Microprogramming

- Exceptions are the hard part of control
  - Need to find convenient place to detect exceptions and to branch to state or microinstruction that saves PC and invokes the operating system
  - Providing clean interrupt model gets hard with pipelining!

- Precise Exception ⇒ state of the machine is preserved as if program executed up to the offending instruction
  - All previous instructions completed
  - Offending instruction and all following instructions act as if they have not even started
Summary: Pipelining

- Reduce CPI by overlapping many instructions
  - Average throughput of approximately 1 CPI with fast clock
- Utilize capabilities of the Datapath
  - start next instruction while working on the current one
  - limited by length of longest stage (plus fill/flush)
  - detect and resolve hazards
- What makes it easy
  - all instructions are the same length
  - just a few instruction formats
  - memory operands appear only in loads and stores
- What makes it hard?
  - structural hazards: suppose we had only one memory
  - control hazards: need to worry about branch instructions
  - data hazards: an instruction depends on a previous instruction

Summary: Where this class is going

- We'll build a simple pipeline and look at these issues
  - Lab 5 ⇒ Pipelined Processor
  - Lab 6 ⇒ With caches
- We'll talk about modern processors and what's really hard:
  - Branches (control hazards) are really hard!
  - Exception handling
  - Trying to improve performance with out-of-order execution, etc.
  - Trying to get CPI < 1 (Superscalar execution)