Readings

This week
- Introduction to microarchitecture and single-cycle microarchitecture
  - H&H, Chapter 7.1-7.3
  - P&P, Appendices A and C
- Multi-cycle microarchitecture
  - H&H, Chapter 7.4
  - P&P, Appendices A and C

Next week
- Pipelining
  - H&H, Chapter 7.5
- Pipelining Issues
  - H&H, Chapter 7.7, 7.8.1-7.8.3
Agenda for Today & Next Few Lectures

- Instruction Set Architectures (ISA): LC-3 and MIPS
- Assembly programming: LC-3 and MIPS
- Microarchitecture (principles & single-cycle uarch)
- Multi-cycle microarchitecture
- Pipelining
- Issues in Pipelining: Control & Data Dependence Handling, State Maintenance and Recovery, ...
- Out-of-Order Execution
Recall: The von Neumann Model

1. **INPUT**
   - Keyboard
   - Mouse
   - Disk

2. **MEMORY**
   - Mem Addr Reg
   - Mem Data Reg

3. **PROCESSING UNIT**
   - ALU
   - TEMP

4. **CONTROL UNIT**
   - PC or IP
   - Inst Register

5. **OUTPUT**
   - Monitor
   - Printer
   - Disk
Recall: LC-3: A von Neumann Machine

Figure 4.3 The LC-3 as an example of the von Neumann model
Recall: The Instruction Cycle

- FETCH
- DECODE
- EVALUATE ADDRESS
- FETCH OPERANDS
- EXECUTE
- STORE RESULT
Recall: The Instruction Set Architecture

- The ISA is the **interface between** what the **software** commands and what the **hardware** carries out

- The ISA specifies
  - The **memory organization**
    - Address space (LC-3: $2^{16}$, MIPS: $2^{32}$)
    - Addressability (LC-3: 16 bits, MIPS: 8 bits)
    - Word- or Byte-addressable
  - The **register set**
    - R0 to R7 in LC-3
    - 32 registers in MIPS
  - The **instruction set**
    - Opcodes
    - Data types
    - Addressing modes
    - Semantics of instructions
Microarchitecture

- An **implementation** of the ISA

- How do we implement the ISA?
  - We will discuss this for many lectures

- There can be many implementations of the same ISA
  - MIPS R2000, R10000, ...
  - x86: Intel 80486, Pentium, Pentium Pro, Pentium 4, Kaby Lake, Coffee Lake, Comet Lake, ... AMD K5, K7, K9, Bulldozer, BobCat, ...
  - IBM POWER 4, 5, 6, 7, 8, 9, 10
  - ARM Cortex-M*, ARM Cortex-A*, NVIDIA Denver, Apple A*, M1, ...
  - Alpha 21064, 21164, 21264, 21364, ...
  - ...

(A Bit More on)
ISA Design and Tradeoffs
The von Neumann Model/Architecture

- Von Neumann model is also called *stored program computer* (instructions in memory). It has two key properties:

  - **Stored program**
    - Instructions stored in a linear memory array
    - **Memory is unified** between instructions and data
      - The interpretation of a stored value depends on the control signals

  - **Sequential instruction processing**

When is a value interpreted as an instruction?
Whether a value fetched from memory is interpreted as an instruction depends on when that value is fetched in the instruction processing cycle.
The von Neumann Model/Architecture

- Von Neumann model is also called *stored program computer* (instructions in memory). It has two key properties:

- **Stored program**
  - Instructions stored in a linear memory array
  - Memory is unified between instructions and data
    - The interpretation of a stored value depends on the control signals

- **Sequential instruction processing**
  - One instruction processed (fetched, executed, completed) at a time
  - Program counter (instruction pointer) identifies the current instruction
  - Program counter is advanced sequentially except for control transfer instructions
The von Neumann Model/Architecture

- **Recommended reading**
  - Burks, Goldstein, von Neumann, “Preliminary discussion of the logical design of an electronic computing instrument,” 1946.

- **Required reading**
  - Patt and Patel book, Chapter 4, “The von Neumann Model”

- **Stored program**

- **Sequential instruction processing**
The Von Neumann Model (of a Computer)
Q: Is this the only way that a computer can process computer programs?

A: No.

Qualified Answer: No. But, it has been the dominant way
- i.e., the dominant paradigm for computing
- for N decades

Let's examine a **completely different model** for processing computer programs
The Dataflow Execution Model of a Computer
The Dataflow Model (of a Computer)

- Von Neumann model: An instruction is fetched and executed in **control flow order**
  - As specified by the program counter (instruction pointer)
  - Sequential unless explicit control flow instruction

- Dataflow model: An instruction is fetched and executed in **data flow order**
  - i.e., when its operands are ready
  - i.e., there is no program counter (instruction pointer)
  - Instruction ordering specified by data flow dependence
    - Each instruction specifies “who” should receive the result
    - An instruction can “fire” whenever all operands are received
  - Potentially many instructions can execute at the same time
    - Inherently more parallel
Consider a Von Neumann program

- What is the significance of the program order?
- What is the significance of the storage locations?

```
v <= a + b;
w <= b * 2;
x <= v - w
y <= v + w
z <= x * y
```

Sequential

Which model is more natural to you as a programmer?
More on Dataflow

- In a dataflow machine, a program consists of dataflow nodes
  - A dataflow node fires (fetched and executed) when all its inputs are ready
    - i.e. when all inputs have tokens

- Dataflow node and its ISA representation
Example Dataflow Nodes

*Conditional

*Relational

*Barrier Synch
A Simple Example Dataflow Program
Do we need a Program Counter (PC or IP) in the ISA?

- Yes: Control-driven, sequential execution
  - An instruction is executed when the PC points to it
  - PC automatically changes sequentially (except for control flow instructions)
- No: Data-driven, parallel execution
  - An instruction is executed when all its operand values are available (dataflow)

Tradeoffs: MANY high-level ones

- Ease of programming (for average programmers)?
- Ease of compilation?
- Performance: Extraction of parallelism?
- Hardware complexity?
ISA vs. Microarchitecture Level Tradeoff

- A similar tradeoff (control vs. data-driven execution) can be made at the microarchitecture level.

- ISA: Specifies how the programmer sees the instructions to be executed.
  - Programmer sees a sequential, control-flow execution order vs. a dataflow execution order.

- Microarchitecture: How the underlying implementation actually executes instructions.
  - Microarchitecture can execute instructions in any order as long as it obeys the semantics specified by the ISA when making the instruction results visible to software.
  - Programmer should see the order specified by the ISA.
Let’s Get Back to the von Neumann Model

- But, if you want to learn more about dataflow...


- A later lecture

- If you are really impatient:
  - [http://www.youtube.com/watch?v=D2uue7izU2c](http://www.youtube.com/watch?v=D2uue7izU2c)
Lecture Video on Dataflow Model

http://www.youtube.com/watch?v=D2uue7izU2c
The von Neumann Model

- All major *instruction set architectures* today use this model
  - x86, ARM, MIPS, SPARC, Alpha, POWER, RISC-V, ...

- Underneath (at the microarchitecture level), the execution model of almost all *implementations (or, microarchitectures)* is very different
  - Pipelined instruction execution: *Intel 80486 uarch*
  - Multiple instructions at a time: *Intel Pentium uarch*
  - Out-of-order execution: *Intel Pentium Pro uarch*
  - Separate instruction and data caches

- But, what happens underneath that is *not consistent* with the von Neumann model is *not exposed* to software
  - Difference between ISA and microarchitecture
What is Computer Architecture?

- **ISA+implementation definition:** The science and art of designing, selecting, and interconnecting hardware components and designing the hardware/software interface to create a computing system that meets functional, performance, energy consumption, cost, and other specific goals.

- **Traditional (ISA-only) definition:** “The term *architecture* is used here to describe the attributes of a system as seen by the programmer, i.e., the conceptual structure and functional behavior as distinct from the organization of the dataflow and controls, the logic design, and the physical implementation.”

  *Gene Amdahl*, IBM Journal of R&D, April 1964
ISA vs. Microarchitecture

- **ISA**
  - Agreed upon interface between software and hardware
    - SW/Compiler assumes, HW promises
  - What the software writer needs to know to write and debug system/user programs

- **Microarchitecture**
  - Specific implementation of an ISA
  - Not visible to the software

- **Microprocessor**
  - **ISA, uarch, circuits**
  - “Architecture” = ISA + microarchitecture

<table>
<thead>
<tr>
<th>Problem</th>
<th>Algorithm</th>
<th>Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISA</td>
<td>Microarchitecture</td>
<td>Circuits</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Electrons</td>
</tr>
</tbody>
</table>
ISA vs. Microarchitecture

- What is part of ISA vs. Uarch?
  - Gas pedal: interface for “acceleration”
  - Internals of the engine: implement “acceleration”

- Implementation (uarch) can be various as long as it satisfies the specification (ISA)
  - Add instruction vs. Adder implementation
    - Bit serial, ripple carry, carry lookahead adders are all part of microarchitecture (see H&H Chapter 5.2.1)
  - x86 ISA has many implementations:
    - Intel 80486, Pentium, Pentium Pro, Pentium 4, Kaby Lake, Coffee Lake, Comet Lake, AMD K5, K7, K9, Bulldozer, BobCat, ...

- Microarchitecture usually changes faster than ISA
  - Few ISAs (x86, ARM, SPARC, MIPS, Alpha, RISC-V) but many uarchs
  - Why?
ISA

- Instructions
  - Opcodes, Addressing Modes, Data Types
  - Instruction Types and Formats
  - Registers, Condition Codes

- Memory
  - Address space, Addressability, Alignment
  - Virtual memory management

- Call, Interrupt/Exception Handling
- Access Control, Priority/Privilege
- I/O: memory-mapped vs. instr.
- Task/thread Management
- Power and Thermal Management
- Multi-threading support, Multiprocessor support
- ...

Intel® 64 and IA-32 Architectures
Software Developer’s Manual

Volume 1: Basic Architecture
Microarchitecture

- Implementation of the ISA under specific design constraints and goals
- Anything done in hardware without exposure to software
  - Pipelining
  - In-order versus out-of-order instruction execution
  - Memory access scheduling policy
  - Speculative execution
  - Superscalar processing (multiple instruction issue?)
  - Clock gating
  - Caching? Levels, size, associativity, replacement policy
  - Prefetching?
  - Voltage/frequency scaling?
  - Error correction?
Property of ISA vs. Uarch?

- ADD instruction’s opcode
- Bit-serial adder vs. Ripple-carry adder
- Number of general purpose registers
- Number of cycles to execute the MUL instruction
- Number of ports to the register file
- Whether or not the machine employs pipelined instruction execution

Remember

- Microarchitecture: Implementation of the ISA under specific design constraints and goals
Design Point

- A set of design considerations and their importance
  - leads to tradeoffs in both ISA and uarch
- Example considerations:
  - Cost
  - Performance
  - Maximum power consumption, thermal
  - Energy consumption (battery life)
  - Availability
  - Reliability and Correctness
  - Time to Market
  - Security, safety, predictability, ...

- Design point determined by the “Problem” space (application space), the intended users/market
Application Space

Dream, and they will appear...

Other examples of the application space that continue to drive the need for unique design points are the following:

1) **scientific applications** such as those whose computations control nuclear power plants, determine where to drill for oil, and predict the weather;
2) **transaction-based applications** such as those that handle ATM transfers and e-commerce business;
3) **business data processing** applications, such as those that handle inventory control, payrolls, IRS activity, and various personnel record keeping, whether the personnel are employees, students, or voters;
4) **network applications**, such as high-speed routing of Internet packets, that enable the connection of your home system to take advantage of the Internet;
5) **guaranteed delivery (a.k.a. real time) applications** that require the result of a computation by a certain critical deadline;
6) **embedded applications** where the processor is a component of a larger system that is used to solve the (usually) dedicated application;
7) **media applications** such as those that decode video and audio files;
8) random software packages that desktop users would like to run on their PCs.

Each of these application areas has a very different set of characteristics. Each application area demands a different set of tradeoffs to be made in specifying the microprocessor to do the job.


**Many other workloads:**
- Genome analysis
- Machine learning
- Robotics
- Web search
- Graph analytics

...
Increasingly Demanding Applications

Dream

and, they will come

As applications push boundaries, computing platforms will become increasingly strained.
Tradeoffs: Soul of Computer Architecture

- ISA-level tradeoffs
- Microarchitecture-level tradeoffs
- System and Task-level tradeoffs
  - How to divide the labor between hardware and software

- Computer architecture is the science and art of making the appropriate trade-offs to meet a design point
  - Why art?
Why Is It (Somewhat) Art?

We do not (fully) know the future (applications, users, market)
And, the future is not constant (it changes)!
Analogue from Macro-Architecture

- Future is not constant in macro-architecture, either

- Example: Can a mill be later used as a theater + restaurant + conference room?
Mühle Tiefenbrunnen

- Originally built as a **brewery** in 1889, part of it was converted into a **mill** in 1913, and the other part into a **cold store**
- Today is a **center for a variety of activities**: theater, conferences, restaurants, shops, museum...

[Brewery in 1900](http://www.muehle-tiefenbrunnen.ch/)
Another Example (I)

Photo credit: Prof. Can Alkan
Another Example (II)
Implementing the ISA: Microarchitecture Basics
Now That We Have an ISA

- How do we implement it?

- i.e., how do we design a system that obeys the hardware/software interface?

- Aside: “System” can be solely hardware or a combination of hardware and software
  - “Translation of ISAs”
  - A virtual ISA can be converted by “software” into an implementation ISA

- We will assume “hardware” implementation for most lectures
How Does a Machine Process Instructions?

- What does processing an instruction mean?
- We will assume the von Neumann model (for now)

\[
\text{AS} = \text{Architectural (programmer visible) state before an instruction is processed}
\]

\[
\text{Process instruction}
\]

\[
\text{AS'} = \text{Architectural (programmer visible) state after an instruction is processed}
\]

- Processing an instruction: Transforming AS to AS' according to the ISA specification of the instruction
The Von Neumann Model/Architecture

Stored program

Sequential instruction processing
Recall: The Von Neumann Model

Recall: The Von Neumann Model

MEMORY

Mem Addr Reg
Mem Data Reg

INPUT

Keyboard, Mouse, Disk...

PROCESSING UNIT

ALU
TEMP

CONTROL UNIT

PC or IP
Inst Register

OUTPUT

Monitor, Printer, Disk...
Instructions (and programs) specify how to transform the values of programmer visible state.
The “Process Instruction” Step

- ISA specifies abstractly what AS’ should be, given an instruction and AS
  - It defines an abstract finite state machine where
    - State = programmer-visible state
    - Next-state logic = instruction execution specification
  - From ISA point of view, there are no “intermediate states” between AS and AS’ during instruction execution
    - One state transition per instruction

- Microarchitecture implements how AS is transformed to AS’
  - There are many choices in implementation
  - We can have programmer-invisible state to optimize the speed of instruction execution: **multiple** state transitions per instruction
    - Choice 1: $AS \rightarrow AS'$ (transform AS to AS’ in a single clock cycle)
    - Choice 2: $AS \rightarrow AS+MS1 \rightarrow AS+MS2 \rightarrow AS+MS3 \rightarrow AS'$ (take multiple clock cycles to transform AS to AS’)


A Very Basic Instruction Processing Engine

- Each instruction takes a single clock cycle to execute
- Only combinational logic is used to implement instruction execution
  - *No intermediate, programmer-invisible state updates*

\[
\text{AS} = \text{Architectural (programmer visible) state at the beginning of a clock cycle}
\]

Process instruction in one clock cycle

\[
\text{AS}' = \text{Architectural (programmer visible) state at the end of a clock cycle}
\]
A Very Basic Instruction Processing Engine

- Single-cycle machine

What is the *clock cycle time* determined by?
What is the *critical path* (i.e., longest delay path) of the combinational logic determined by?

AS: Architectural State
Single-cycle vs. Multi-cycle Machines

- **Single-cycle machines**
  - Each instruction takes a single clock cycle
  - All state updates made at the end of an instruction’s execution
  - **Big disadvantage:** The slowest instruction determines cycle time → long clock cycle time

- **Multi-cycle machines**
  - Instruction processing broken into multiple cycles/stages
  - State updates can be made during an instruction’s execution
  - Architectural state updates made at the end of an instruction’s execution
  - **Advantage over single-cycle:** The slowest “stage” determines cycle time

- Both single-cycle and multi-cycle machines literally follow the von Neumann model at the microarchitecture level
Instruction Processing “Cycle”

- Instructions are processed under the direction of a “control unit” step by step.
- Instruction cycle: Sequence of steps to process an instruction
- Fundamentally, there are six steps:
  - Fetch
  - Decode
  - Evaluate Address
  - Fetch Operands
  - Execute
  - Store Result

- Not all instructions require all six steps (see P&P Ch. 4)
Recall: The Instruction Processing “Cycle”

- FETCH
- DECODE
- EVALUATE ADDRESS
- FETCH OPERANDS
- EXECUTE
- STORE RESULT
Instruction Processing “Cycle” vs. Machine Clock Cycle

- **Single-cycle machine:**
  - All six phases of the instruction processing cycle take a single machine clock cycle to complete

- **Multi-cycle machine:**
  - All six phases of the instruction processing cycle can take multiple machine clock cycles to complete
  - In fact, each phase can take multiple clock cycles to complete
Instruction Processing Viewed Another Way

- Instructions transform Data (AS) to Data’ (AS’)
- This transformation is done by functional units
  - Units that “operate” on data
- These units need to be told what to do to the data

- An instruction processing engine consists of two components
  - **Datapath**: Consists of hardware elements that deal with and transform data signals
    - **functional units** that operate on data
    - **hardware structures** (e.g., wires, muxes, decoders, tri-state bufs) that enable the flow of data into the functional units and registers
    - **storage units** that store data (e.g., registers)
  - **Control logic**: Consists of hardware elements that determine control signals, i.e., signals that specify what the datapath elements should do to the data
Recall: LC-3: A von Neumann Machine

Figure 4.3 The LC-3 as an example of the von Neumann model
Single-cycle vs. Multi-cycle: Control & Data

- **Single-cycle machine:**
  - Control signals are generated in the same clock cycle as the one during which data signals are operated on.
  - Everything related to an instruction happens in one clock cycle (serialized processing).

- **Multi-cycle machine:**
  - Control signals needed in the next cycle can be generated in the current cycle.
  - Latency of control processing can be overlapped with latency of datapath operation (more parallelism).

- See P&P Appendix C for more (microprogrammed multi-cycle microarchitecture).
Many Ways of Datapath and Control Design

- There are many ways of designing the datapath and control logic

- Example ways
  - Single-cycle, multi-cycle, pipelined datapath and control
  - Single-bus vs. multi-bus datapaths
  - Hardwired/combinational vs. microcoded/microprogrammed control
    - Control signals generated by combinational logic versus
    - Control signals stored in a memory structure

- Control signals and structure depend on the datapath design
Flash-Forward: Performance Analysis

- Execution time of a single instruction
  - \(\{\text{CPI}\} \times \{\text{clock cycle time}\}\)  
  CPI: Cycles Per Instruction

- Execution time of an entire program
  - Sum over all instructions \([\{\text{CPI}\} \times \{\text{clock cycle time}\}]\)
  - \(\{\# \text{ of instructions}\} \times \{\text{Average CPI}\} \times \{\text{clock cycle time}\}\)

- Single-cycle microarchitecture performance
  - CPI = 1
  - Clock cycle time = long

- Multi-cycle microarchitecture performance
  - CPI = different for each instruction
    - Average CPI \(\rightarrow\) hopefully small
  - Clock cycle time = short

In multi-cycle, we have two degrees of freedom to optimize independently.
A Single-Cycle Microarchitecture
A Closer Look
Remember…

- Single-cycle machine

AS: Architectural State
Let’s Start with the State Elements

- Data and control inputs
MIPS State Elements

- **Program counter:**
  32-bit register

- **Instruction memory:**
  Takes input 32-bit address A and reads the 32-bit data (i.e., instruction) from that address to the read data output RD.

- **Register file:**
  The 32-element, 32-bit register file has 2 read ports and 1 write port

- **Data memory:**
  If the write enable, WE, is 1, it writes 32-bit data WD into memory location at 32-bit address A on the rising edge of the clock.
  If the write enable is 0, it reads 32-bit data from address A onto RD.

This notation is used in H&H single-cycle MIPS implementation (H&H Chapter 7.3)
For Now, We Will Assume

- “Magic” memory and register file
- Combinational read
  - output of the read data port is a combinational function of the register file contents and the corresponding read select port
- Synchronous write
  - the selected register is updated on the positive edge clock transition when write enable is asserted
  - Cannot affect read output in between clock edges
- Single-cycle, synchronous memory
  - Contrast this with memory that tells when the data is ready
    - i.e., Ready signal: indicating the read or write is done
    - See P&P Appendix C (LC3-b) for multi-cycle memory
Instruction Processing

- 5 generic steps (P&H book)
  - Instruction fetch (IF)
  - Instruction decode and register operand fetch (ID/RF)
  - Execute/Evaluate memory address (EX/AG)
  - Memory operand fetch (MEM)
  - Store/writeback result (WB)

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]**
We Need to Provide the Datapath + Control Logic to Execute All ISA Instructions
**What Is To Come: The Full MIPS Datapath**

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]

JAL, JR, JALR omitted
Another Complete Single-Cycle Processor

Single-cycle processor. Harris and Harris, Chapter 7.3.
Single-Cycle Datapath for Arithmetic and Logical Instructions
R-Type ALU Instructions

- R-type: 3 register operands

MIPS assembly (e.g., register-register signed addition)

```
add $s0, $s1, $s2  #$s0=rd, $s1=rs, $s2=rt
```

Machine Encoding

<table>
<thead>
<tr>
<th></th>
<th>rs</th>
<th>rt</th>
<th>rd</th>
<th>0</th>
<th>add (32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>6 bits</td>
</tr>
</tbody>
</table>

- Semantics

if MEM[PC] == add rd rs rt

```
GPR[rd] ← GPR[rs] + GPR[rt]
PC ← PC + 4
```
(R-Type) ALU Datapath

```
if MEM[PC] == ADD rd rs rt
  GPR[rd] ← GPR[rs] + GPR[rt]
  PC ← PC + 4
```

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]

**Combinational state update logic**

![ALU Datapath Diagram]
Example: ALU Design

- ALU operation ($F_{2:0}$) comes from the control logic

<table>
<thead>
<tr>
<th>$F_{2:0}$</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>$A &amp; B$</td>
</tr>
<tr>
<td>001</td>
<td>$A \mid B$</td>
</tr>
<tr>
<td>010</td>
<td>$A + B$</td>
</tr>
<tr>
<td>011</td>
<td>not used</td>
</tr>
<tr>
<td>100</td>
<td>$A &amp; \sim B$</td>
</tr>
<tr>
<td>101</td>
<td>$A \mid \sim B$</td>
</tr>
<tr>
<td>110</td>
<td>$A - B$</td>
</tr>
<tr>
<td>111</td>
<td>SLT</td>
</tr>
</tbody>
</table>
## I-Type ALU Instructions

- **I-type**: 2 register operands and 1 immediate

### MIPS assembly (e.g., register-immediate signed addition)

```
addi $s0, $s1, 5  #$s0=rs, $s1=rt
```

### Machine Encoding

<table>
<thead>
<tr>
<th>addi (0)</th>
<th>rs</th>
<th>rt</th>
<th>immediate</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>

### Semantics

- if MEM[PC] == addi rs rt immediate
- PC ← PC + 4
- GPR[rt] ← GPR[rs] + sign-extend(immediate)
if MEM[PC] == ADDI rt rs immediate
GPR[rt] ← GPR[rs] + sign-extend (immediate)
PC ← PC + 4

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]**
Recall: ADD with one Literal in LC-3

- ADD assembly and machine code

LC-3 assembly

ADD R1, R4, #-2

Field Values

<table>
<thead>
<tr>
<th>OP</th>
<th>DR</th>
<th>SR</th>
<th>imm5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-2</td>
</tr>
</tbody>
</table>

Machine Code

<table>
<thead>
<tr>
<th>OP</th>
<th>DR</th>
<th>SR</th>
<th>imm5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>001</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>11110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>12</td>
<td>11</td>
<td>9</td>
</tr>
</tbody>
</table>

Diagram showing the register file, instruction register, and ALU operations.
Single-Cycle Datapath for Data Movement Instructions
Load Instructions

- Load 4-byte word

MIPS assembly

```
lw $s3, 8($s0) #$s0=rs, $s3=rt
```

Machine Encoding

<table>
<thead>
<tr>
<th>op</th>
<th>rs=base</th>
<th>rt</th>
<th>imm=offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>lw (35)</td>
<td>base</td>
<td>rt</td>
<td>offset</td>
</tr>
</tbody>
</table>

I-Type

- Semantics

if MEM[PC] == lw rt offset_{16} (base)

PC ← PC + 4

EA = sign-extend(offset) + GPR(base)

GPR[rt] ← MEM[ translate(EA) ]
if MEM[PC]==LW rt offset\textsubscript{16} (base)
EA = sign-extend(offset) + GPR[base]
GPR[rt] \leftarrow MEM[\text{translate}(EA)]
PC \leftarrow PC + 4
Store Instructions

- Store 4-byte word

MIPS assembly

\[
\text{sw} ~ \$s3, \ 8(\$s0) \quad \#\$s0=rs, \ \$s3=rt
\]

Machine Encoding

<table>
<thead>
<tr>
<th>op</th>
<th>rs=base</th>
<th>rt</th>
<th>imm=offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>sw (43)</td>
<td>base</td>
<td>rt</td>
<td>offset</td>
</tr>
</tbody>
</table>

I-Type

- Semantics

if Mem[PC] == sw rt offset_{16} (base)

PC ← PC + 4

EA = sign-extend(offset) + GPR(base)

MEM[ translate(EA) ] ← GPR[rt]
if \text{MEM}[\text{PC}] == \text{SW rt offset}_{16} (\text{base})
\[ \text{EA} = \text{sign-extend} (\text{offset}) + \text{GPR} [\text{base}] \]
\[ \text{MEM} [ \text{translate} (\text{EA}) ] \leftarrow \text{GPR} [\text{rt}] \]
\[ \text{PC} \leftarrow \text{PC} + 4 \]
Load-Store Datapath

Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]
Datapath for Non-Control-Flow Insts.

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]**
We Did Not Cover Later Slides. They Will Be Covered in Later Lectures.
Single-Cycle Datapath for Control Flow Instructions
Unconditional branch or jump

\[ j \ target \]

- 2 = opcode
- immediate (target) = target address

Semantics

If \( \text{MEM}[\text{PC}] == j \ \text{immediate}_{26} \)

\[ \text{target} = \{ \text{PC}^{\dagger}[31:28], \text{immediate}_{26}, 2'b00 \} \]

\( \text{PC} \leftarrow \text{target} \)

\( ^{\dagger} \) This is the incremented PC
Unconditional Jump Datapath

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]

** What about JR, JAL, JALR?**

if MEM[PC]==J immediate26
PC = { PC[31:28], immediate26, 2’b00 }
Other Jumps in MIPS

- **jal**: jump and link (function calls)
  - Semantics
    
    \[
    \text{if MEM[PC] == jal immediate}_{26} \\
    \quad \text{ra} \leftarrow \text{PC} + 4 \\
    \quad \text{target} = \{ \text{PC}^\dagger[31:28], \text{immediate}_{26}, \text{2'b00} \} \\
    \quad \text{PC} \leftarrow \text{target}
    \]

- **jr**: jump register
  - Semantics
    
    \[
    \text{if MEM[PC] == jr rs} \\
    \quad \text{PC} \leftarrow \text{GPR(rs)}
    \]

- **jalr**: jump and link register
  - Semantics
    
    \[
    \text{if MEM[PC] == jalr rs} \\
    \quad \text{ra} \leftarrow \text{PC} + 4 \\
    \quad \text{PC} \leftarrow \text{GPR(rs)}
    \]

\(^\dagger\text{This is the incremented PC}\)
Aside: MIPS Cheat Sheet


- On the course website
Conditional Branch Instructions

- beq (Branch if Equal)

```plaintext
beq $s0, $s1, offset #$s0=rs,$s1=rt
```

<table>
<thead>
<tr>
<th>beq (4)</th>
<th>rs</th>
<th>rt</th>
<th>immediate=offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>

- Semantics (assuming no branch delay slot)

  if MEM[PC] == beq rs rt immediate
  target = PC + sign-extend(immediate) x 4
  if GPR[rs] == GPR[rt] then PC ← target
  else PC ← PC + 4

- Variations: beq, bne, blez, bgtz

† This is the incremented PC
Conditional Branch Datapath (for you to finish)

How to uphold the delayed branch semantics?
Putting It All Together

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]

JAL, JR, JALR omitted
Single-Cycle Control Logic
### Single-Cycle Hardwired Control

- As combinational function of $\text{Inst} = \text{MEM}[\text{PC}]$

<table>
<thead>
<tr>
<th>Field</th>
<th>Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>rs</td>
<td>5</td>
</tr>
<tr>
<td>rt</td>
<td>5</td>
</tr>
<tr>
<td>rd</td>
<td>5</td>
</tr>
<tr>
<td>shamt</td>
<td>5</td>
</tr>
<tr>
<td>funct</td>
<td>6</td>
</tr>
</tbody>
</table>

**R-Type**

<table>
<thead>
<tr>
<th>Field</th>
<th>Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>opcode</td>
<td>6</td>
</tr>
<tr>
<td>rs</td>
<td>5</td>
</tr>
<tr>
<td>rt</td>
<td>5</td>
</tr>
<tr>
<td>immediate</td>
<td>16</td>
</tr>
</tbody>
</table>

**I-Type**

<table>
<thead>
<tr>
<th>Field</th>
<th>Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>opcode</td>
<td>6</td>
</tr>
<tr>
<td>immediate</td>
<td>26</td>
</tr>
</tbody>
</table>

**J-Type**

<table>
<thead>
<tr>
<th>Field</th>
<th>Bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>opcode</td>
<td>6</td>
</tr>
<tr>
<td>immediate</td>
<td>26</td>
</tr>
</tbody>
</table>

- Consider
  - All R-type and I-type ALU instructions
  - `lw` and `sw`
  - `beq`, `bne`, `blez`, `bgtz`
  - `j`, `jr`, `jal`, `jalr`
Generate Control Signals (in Orange Color)

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]

JAL, JR, JALR omitted
### Single-Bit Control Signals (I)

<table>
<thead>
<tr>
<th></th>
<th>When De-asserted</th>
<th>When asserted</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RegDest</td>
<td>GPR write select according to (rt), i.e., (inst[20:16])</td>
<td>GPR write select according to (rd), i.e., (inst[15:11])</td>
<td>(\text{opcode} == 0)</td>
</tr>
<tr>
<td>ALUSrc</td>
<td>2(^{nd}) ALU input from 2(^{nd}) GPR read port</td>
<td>2(^{nd}) ALU input from sign-extended 16-bit immediate</td>
<td>((\text{opcode}! = 0) &amp;&amp; (\text{opcode}! = \text{BEQ}) &amp;&amp; (\text{opcode}! = \text{BNE}))</td>
</tr>
<tr>
<td>MemtoReg</td>
<td>Steer ALU result to GPR write port</td>
<td>steer memory load to GPR write port</td>
<td>(\text{opcode} == \text{LW})</td>
</tr>
<tr>
<td>RegWrite</td>
<td>GPR write disabled</td>
<td>GPR write enabled</td>
<td>((\text{opcode}! = \text{SW}) &amp;&amp; (\text{opcode}! = \text{Bxx}) &amp;&amp; (\text{opcode}! = \text{J}) &amp;&amp; (\text{opcode}! = \text{JR}))</td>
</tr>
</tbody>
</table>
### Single-Bit Control Signals (II)

<table>
<thead>
<tr>
<th></th>
<th>When De-asserted</th>
<th>When asserted</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MemRead</strong></td>
<td>Memory read disabled</td>
<td>Memory read port return load value</td>
<td>( \text{opcode}==\text{LW} )</td>
</tr>
<tr>
<td><strong>MemWrite</strong></td>
<td>Memory write disabled</td>
<td>Memory write enabled</td>
<td>( \text{opcode}==\text{SW} )</td>
</tr>
<tr>
<td><strong>PCSrc\textsubscript{1}</strong></td>
<td>According to <strong>PCSrc\textsubscript{2}</strong></td>
<td>next PC is based on 26-bit immediate jump target</td>
<td>((\text{opcode}==\text{J}) \</td>
</tr>
<tr>
<td><strong>PCSrc\textsubscript{2}</strong></td>
<td>next PC = PC + 4</td>
<td>next PC is based on 16-bit immediate branch target</td>
<td>((\text{opcode}==\text{Bxx}) \ &amp;&amp; \ \text{“bcond is satisfied”})</td>
</tr>
</tbody>
</table>

JR and JALR require additional PCSrc\textsubscript{c} options
R-Type ALU

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]**
**I-Type ALU**

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]**
**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]**
**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]**
Some control signals are dependent on the processing of data.
Some control signals are dependent on the processing of data
Jump

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]**
What is in That Control Box?

- **Combinational Logic → Hardwired Control**
  - Idea: Most control signals generated combinationally based on bits in instruction encoding

- **Sequential Logic → Sequential Control**
  - Idea: A memory structure contains the control signals associated with an instruction
    - Called Control Store

- Both types of control structure can be used in single-cycle processors
  - Choice depends on latency of each structure + how much on the critical path control signal generation is, etc.
Review: Complete Single-Cycle Processor

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]

JAL, JR, JALR omitted
Another Single-Cycle MIPS Processor (from H&H)

See backup slides to reinforce the concepts we have covered. They are to complement your reading:
H&H, Chapter 7.1-7.3, 7.6
Another Complete Single-Cycle Processor

Single-cycle processor. Harris and Harris, Chapter 7.3.
Example: Single-Cycle Datapath: `lw` fetch

**STEP 1:** Fetch instruction

```
lw $s3, 1($0)  # read memory word 1 into $s3
```

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>

I-Type
Single-Cycle Datapath: lw register read

**STEP 2:** Read source operands from register file

```
lw $s3, 1($0)  # read memory word 1 into $s3
```

**I-Type**

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: `lw` immediate

- **STEP 3:** Sign-extend the immediate

```
lw $s3, 1($0)  # read memory word 1 into $s3
```

**I-Type**

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: `lw` address

**STEP 4:** Compute the memory address

| `lw $s3, 1($0)` | # read memory word 1 into $s3 |

**I-Type**

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: lw memory read

**STEP 5:** Read from memory and write back to register file

```
lw $s3, 1($0)  # read memory word 1 into $s3
```

**I-Type**

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: \texttt{lw} PC increment

**STEP 6:** Determine address of next instruction

\texttt{lw} $\text{\$s3, 1($0)}$  \# read memory word 1 into $\text{\$s3}

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Similarly, We Need to Design the Control Unit

- **Control signals** are generated by the decoder in control unit

<table>
<thead>
<tr>
<th>Instruction</th>
<th>( \text{Op}_{5:0} )</th>
<th>RegWrite</th>
<th>RegDst</th>
<th>AluSrc</th>
<th>Branch</th>
<th>MemWrite</th>
<th>MemtoReg</th>
<th>( \text{ALUOp}_{1:0} )</th>
<th>Jump</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-type</td>
<td>000000</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>lw</td>
<td>100011</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>sw</td>
<td>101011</td>
<td>0</td>
<td>( X )</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>( X )</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>beq</td>
<td>000100</td>
<td>0</td>
<td>( X )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>( X )</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>addi</td>
<td>001000</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>j</td>
<td>000010</td>
<td>0</td>
<td>( X )</td>
<td>( X )</td>
<td>( X )</td>
<td>0</td>
<td>( X )</td>
<td>( XX )</td>
<td>1</td>
</tr>
</tbody>
</table>
Another Complete Single-Cycle Processor (H&H)
Your Reading Assignment

- Please read the Lecture Slides and the Backup Slides

- Please do your readings from the H&H Book
  - H&H, Chapter 7.1-7.3, 7.6
Single-Cycle Uarch I (We Developed in Lectures)

**Based on original figure from [P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]

JAL, JR, JALR omitted
Single-Cycle Uarch II (In Your Readings)
Evaluating the Single-Cycle Microarchitecture
A Single-Cycle Microarchitecture

- Is *this* a good idea/design?

- When is this a good design?

- When is this a bad design?

- How can we design a better microarchitecture?
Performance Analysis Basics
Processor Performance

- How fast is my program?
  - Every program consists of a series of instructions
  - Each instruction needs to be executed
Processor Performance

How fast is my program?
- Every program consists of a series of instructions
- Each instruction needs to be executed

So how fast are my instructions?
- Instructions are realized on the hardware
- They can take one or more clock cycles to complete
- \( \text{Cycles per Instruction} = \text{CPI} \)
Processor Performance

- How fast is my program?
  - Every program consists of a series of instructions
  - Each instruction needs to be executed.

- So how fast are my instructions?
  - Instructions are realized on the hardware
  - They can take one or more clock cycles to complete
  - Cycles per Instruction = CPI

- How much time is one clock cycle?
  - The critical path determines how much time one cycle requires = clock period.
  - 1/clock period = clock frequency = how many cycles can be done each second.
Processor Performance

Now as a general formula

- Our program consists of executing $N$ instructions
- Our processor needs $CPI$ cycles for each instruction
- The maximum clock speed of the processor is $f$, and the clock period is therefore $T=1/f$
Processor Performance

■ Now as a general formula
  ▪ Our program consists of executing \( N \) instructions
  ▪ Our processor needs \( CPI \) cycles for each instruction
  ▪ The maximum clock speed of the processor is \( f \), and the clock period is therefore \( T=1/f \)

■ Our program executes in

\[
N \times CPI \times (1/f) = N \times CPI \times T \text{ seconds}
\]
Performance Analysis Basics

- Execution time of a single instruction
  - \( \text{CPI} \times \text{clock cycle time} \)
    - CPI: Number of cycles it takes to execute an instruction

- Execution time of an entire program
  - Sum over all instructions \([\text{CPI} \times \text{clock cycle time}]\)
  - \( \# \text{ of instructions} \times \text{Average CPI} \times \text{clock cycle time} \)
Performance Analysis of Our Single-Cycle Design
A Single-Cycle Microarchitecture: Analysis

- Every instruction takes 1 cycle to execute
  - CPI (Cycles per instruction) is strictly 1

- How long each instruction takes is determined by how long the slowest instruction takes to execute
  - Even though many instructions do not need that long to execute

- Clock cycle time of the microarchitecture is determined by how long it takes to complete the slowest instruction
  - Critical path of the design is determined by the processing time of the slowest instruction
What is the Slowest Instruction to Process?

- Let’s go back to the basics

- All six phases of the instruction processing cycle take a single machine clock cycle to complete

  - Fetch
  - Decode
  - Evaluate Address
  - Fetch Operands
  - Execute
  - Store Result

  1. Instruction fetch (IF)
  2. Instruction decode and register operand fetch (ID/RF)
  3. Execute/Evaluate memory address (EX/AG)
  4. Memory operand fetch (MEM)
  5. Store/writeback result (WB)

- Do each of the above phases take the same time (latency) for all instructions?
Let’s Find the Critical Path
Example Single-Cycle Datapath Analysis

- Assume (for the design in the previous slide)
  - memory units (read or write): 200 ps
  - ALU and adders: 100 ps
  - register file (read or write): 50 ps
  - other combinational logic: 0 ps

<table>
<thead>
<tr>
<th>steps</th>
<th>IF</th>
<th>ID</th>
<th>EX</th>
<th>MEM</th>
<th>WB</th>
<th>Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>resources</td>
<td>mem</td>
<td>RF</td>
<td>ALU</td>
<td>mem</td>
<td>RF</td>
<td></td>
</tr>
<tr>
<td>R-type</td>
<td>200</td>
<td>50</td>
<td>100</td>
<td>50</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>I-type</td>
<td>200</td>
<td>50</td>
<td>100</td>
<td>50</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>LW</td>
<td>200</td>
<td>50</td>
<td>100</td>
<td>200</td>
<td>50</td>
<td>600</td>
</tr>
<tr>
<td>SW</td>
<td>200</td>
<td>50</td>
<td>100</td>
<td>200</td>
<td></td>
<td>550</td>
</tr>
<tr>
<td>Branch</td>
<td>200</td>
<td>50</td>
<td>100</td>
<td></td>
<td></td>
<td>350</td>
</tr>
<tr>
<td>Jump</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200</td>
</tr>
</tbody>
</table>
Let’s Find the Critical Path

[Based on original figure from P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]
R-Type and I-Type ALU

[Based on original figure from P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]
Branch Taken

[Based on original figure from P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]
Jump

[Based on original figure from P&H CO&D, COPYRIGHT 2004 Elsevier. ALL RIGHTS RESERVED.]
What About Control Logic?

- How does that affect the critical path?

- Food for thought for you:
  - Can control logic be on the critical path?
  - Historical example:
    - CDC 5600: control store access too long...
What is the Slowest Instruction to Process?

- Real world: **Memory is slow (not magic)**

- What if memory *sometimes* takes 100ms to access?

- Does it make sense to have a simple register to register add or jump to take \{100ms+all else to do a memory operation\}?  

- And, what if you need to access memory more than once to process an instruction?  
  - Which instructions need this?  
  - Do you provide multiple ports to memory?
Single Cycle uArch: Complexity

- Contrived
  - All instructions run as slow as the slowest instruction

- Inefficient
  - All instructions run as slow as the slowest instruction
  - Must provide worst-case combinational resources in parallel as required by any instruction
  - Need to replicate a resource if it is needed more than once by an instruction during different parts of the instruction processing cycle

- Not necessarily the simplest way to implement an ISA
  - Single-cycle implementation of REP MOVS (x86) or INDEX (VAX)?

- Not easy to optimize/improve performance
  - Optimizing the common case does not work (e.g. common instructions)
  - Need to optimize the worst case all the time
(Micro)architecture Design Principles

- **Critical path design**
  - Find and decrease the maximum combinational logic delay
  - Break a path into multiple cycles if it takes too long

- **Bread and butter (common case) design**
  - Spend time and resources on where it matters most
    - i.e., improve what the machine is really designed to do
  - Common case vs. uncommon case

- **Balanced design**
  - Balance instruction/data flow through hardware components
  - Design to eliminate bottlenecks: balance the hardware for the work
Single-Cycle Design vs. Design Principles

- Critical path design
- Bread and butter (common case) design
- Balanced design

How does a single-cycle microarchitecture fare with respect to these principles?
Aside: System Design Principles

- When designing computer systems/architectures, it is important to follow good principles
  - Actually, this is true for *any* system design
    - Real architectures, buildings, bridges, ...
    - Good consumer products
    - ...

- Remember: “principled design” from our second lecture
  - Frank Lloyd Wright: “architecture [...] based upon principle, and not upon precedent”
Aside: From Lecture 2

- “architecture [...] based upon principle, and not upon precedent”
This
Recall: Takeaways

- It all starts from the **basic building blocks and design principles**

- And, **knowledge of how to use, apply, enhance them**

- **Underlying technology might change** (e.g., steel vs. wood)
  - but **methods of taking advantage of technology bear resemblance**
  - **methods used for design depend on the principles employed**
Aside: System Design Principles

- We will continue to cover key principles in this course
- Here are some references where you can learn more

- Yale Patt, "Requirements, Bottlenecks, and Good Fortune: Agents for Microprocessor Evolution," Proc. of IEEE, 2001. (Levels of transformation, design point, etc)
- Gene M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities," AFIPS Conference, April 1967. (Amdahl’s Law → Common-case design)
A Key System Design Principle

- Keep it simple

- “Everything should be made as simple as possible, but no simpler.”
  - Albert Einstein

- And, keep it low cost: “An engineer is a person who can do for a dime what any fool can do for a dollar.”

- For more, see:
Multi-Cycle Microarchitectures
Backup Slides on Single-Cycle Uarch for Your Own Study

Please study these to reinforce the concepts we covered in lectures.

Please do the readings together with these slides:
H&H, Chapter 7.1-7.3, 7.6
Another Single-Cycle MIPS Processor (from H&H)

These are slides for your own study. They are to complement your reading H&H, Chapter 7.1-7.3, 7.6
What to do with the Program Counter?

- The PC needs to be incremented by 4 during each cycle (for the time being).
- Initial PC value (after reset) is 0x00400000

```
reg [31:0] PC_p, PC_n;       // Present and next state of PC

// [...]
assign PC_n <= PC_p + 4;    // Increment by 4;

always @(posedge clk, negedge rst)
begin
    if (rst == '0') PC_p <= 32'h00400000; // default
    else          PC_p <= PC_n;          // when clk
end
```
We Need a Register File

- **Store 32 registers, each 32-bit**
  - $2^5 = 32$, we need 5 bits to address each

- **Every R-type instruction uses 3 register**
  - Two for reading (RS, RT)
  - One for writing (RD)

- **We need a special memory with:**
  - 2 read ports (address x2, data out x2)
  - 1 write port (address, data in)
Register File

```verilog
input [4:0]  a_rs, a_rt, a_rd;
input [31:0]  di_rd;
input    we_rd;
output [31:0] do_rs, do_rt;

reg [31:0] R_arr [31:0]; // Array that stores regs

// Circuit description
assign do_rs = R_arr[a_rs]; // Read RS

assign do_rt = R_arr[a_rt]; // Read RT

always @ (posedge clk)
    if (we_rd) R_arr[a_rd] <= di_rd; // write RD
```
Register File

input [4:0] a_rs, a_rt, a_rd;
input [31:0] di_rd;
input we_rd;
output [31:0] do_rs, do_rt;

reg [31:0] R_arr [31:0]; // Array that stores regs

// Circuit description; add the trick with $0
assign do_rs = (a_rs != 5’b00000)? // is address 0?
R_arr[a_rs] : 0; // Read RS or 0

assign do_rt = (a_rt != 5’b00000)? // is address 0?
R_arr[a_rt] : 0; // Read RT or 0

always @ (posedge clk)
if (we_rd) R_arr[a_rd] <= di_rd; // write RD
Data Memory Example

- Will be used to store the bulk of data

```vhdl
input [15:0] addr; // Only 16 bits in this example
input [31:0] di;
input we;
output [31:0] do;

reg [31:0] M_arr [0:65535]; // Array for Memory

// Circuit description
assign do = M_arr[addr]; // Read memory

always @(posedge clk)
  if (we) M_arr[addr] <= di; // write memory
```
Single-Cycle Datapath: \texttt{lw} fetch

\textbf{STEP 1:} Fetch instruction

\texttt{lw} \ $s3, \ 1(\$0) \ # \ read \ memory \ word \ 1 \ into \ $s3

\begin{itemize}
  \item \texttt{op} \ 6 \ bits
  \item \texttt{rs} \ 5 \ bits
  \item \texttt{rt} \ 5 \ bits
  \item \texttt{imm} \ 16 \ bits
\end{itemize}
Single-Cycle Datapath: `lw` register read

**STEP 2:** Read source operands from register file

```
lw $s3, 1($0)  # read memory word 1 into $s3
```

**I-Type**

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: lw immediate

- **STEP 3:** Sign-extend the immediate

```plaintext
lw $s3, 1($0)  # read memory word 1 into $s3
```

**I-Type**

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: `lw` address

**STEP 4:** Compute the memory address

```assembly
lw $s3, 1($0)  # read memory word 1 into $s3
```

**I-Type**

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: `lw` memory read

**STEP 5:** Read from memory and write back to register file

- `lw $s3, 1($0) # read memory word 1 into $s3`

**I-Type**

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
**Single-Cycle Datapath: lw PC increment**

- **STEP 6**: Determine address of next instruction

```plaintext
lw $s3, 1($0)  # read memory word 1 into $s3
```

**I-Type**

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>imm</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: sw

- Write data in rt to memory

```
sw $t7, 44($0)  # write t7 into memory address 44
```

**I-Type**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>op</td>
<td>rs</td>
<td>rt</td>
<td>imm</td>
</tr>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>16 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: R-type Instructions

- Read from rs and rt, write ALUResult to register file

```
add t, b, c  # t = b + c
```

### R-Type

<table>
<thead>
<tr>
<th>op</th>
<th>rs</th>
<th>rt</th>
<th>rd</th>
<th>shamt</th>
<th>funct</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>5 bits</td>
<td>6 bits</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath: beq

- Determine whether values in rs and rt are equal
- Calculate BTA = (sign-extended immediate << 2) + (PC+4)

beq $s0, $s1, target  # branch is taken
Complete Single-Cycle Processor
Our MIPS Datapath has Several Options

- **ALU inputs**
  - Either RT or Immediate \( (MUX) \)

- **Write Address of Register File**
  - Either RD or RT \( (MUX) \)

- **Write Data In of Register File**
  - Either ALU out or Data Memory Out \( (MUX) \)

- **Write enable of Register File**
  - Not always a register write \( (MUX) \)

- **Write enable of Memory**
  - Only when writing to memory (sw) \( (MUX) \)

*All these options are our control signals*
Control Unit

- **Control Unit**
  - **Opcode**
    - Main Decoder
      - MemtoReg
      - MemWrite
      - Branch
      - ALUSrc
      - RegDst
      - RegWrite
      - ALUOp
      - ALUControl

- **Funct**
  - ALUOp
    - 00: add
    - 01: subtract
    - 10: look at funct field
    - 11: n/a
ALU Does the Real Work in a Processor

<table>
<thead>
<tr>
<th>$F_{2:0}$</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>$A &amp; B$</td>
</tr>
<tr>
<td>001</td>
<td>$A \mid B$</td>
</tr>
<tr>
<td>010</td>
<td>$A + B$</td>
</tr>
<tr>
<td>011</td>
<td>not used</td>
</tr>
<tr>
<td>100</td>
<td>$A &amp; \sim B$</td>
</tr>
<tr>
<td>101</td>
<td>$A \mid \sim B$</td>
</tr>
<tr>
<td>110</td>
<td>$A - B$</td>
</tr>
<tr>
<td>111</td>
<td>SLT</td>
</tr>
</tbody>
</table>
**ALU Internals**

![ALU Circuit Diagram]

<table>
<thead>
<tr>
<th>$F_{2:0}$</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>$A &amp; B$</td>
</tr>
<tr>
<td>001</td>
<td>$A \mid B$</td>
</tr>
<tr>
<td>010</td>
<td>$A + B$</td>
</tr>
<tr>
<td>011</td>
<td>not used</td>
</tr>
<tr>
<td>100</td>
<td>$A &amp; \sim B$</td>
</tr>
<tr>
<td>101</td>
<td>$A \mid \sim B$</td>
</tr>
<tr>
<td>110</td>
<td>$A - B$</td>
</tr>
<tr>
<td>111</td>
<td>SLT</td>
</tr>
</tbody>
</table>
Control Unit: ALU Decoder

<table>
<thead>
<tr>
<th>ALUOp&lt;sub&gt;1:0&lt;/sub&gt;</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>Add</td>
</tr>
<tr>
<td>01</td>
<td>Subtract</td>
</tr>
<tr>
<td>10</td>
<td>Look at Funct</td>
</tr>
<tr>
<td>11</td>
<td>Not Used</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ALUOp&lt;sub&gt;1:0&lt;/sub&gt;</th>
<th>Funct</th>
<th>ALUControl&lt;sub&gt;2:0&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>X</td>
<td>010 (Add)</td>
</tr>
<tr>
<td>X1</td>
<td>X</td>
<td>110 (Subtract)</td>
</tr>
<tr>
<td>1X</td>
<td>100000 (add)</td>
<td>010 (Add)</td>
</tr>
<tr>
<td>1X</td>
<td>100010 (sub)</td>
<td>110 (Subtract)</td>
</tr>
<tr>
<td>1X</td>
<td>100100 (and)</td>
<td>000 (And)</td>
</tr>
<tr>
<td>1X</td>
<td>100101 (or)</td>
<td>001 (Or)</td>
</tr>
<tr>
<td>1X</td>
<td>101010 (slt)</td>
<td>111 (SLT)</td>
</tr>
</tbody>
</table>
Let us Develop our Control Table

<table>
<thead>
<tr>
<th>Instruction</th>
<th>$\text{Op}_{5:0}$</th>
<th>RegWrite</th>
<th>RegDst</th>
<th>AluSrc</th>
<th>MemWrite</th>
<th>MemtoReg</th>
<th>ALUOp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **RegWrite**: Write enable for the register file
- **RegDst**: Write to register RD or RT
- **AluSrc**: ALU input RT or immediate
- **MemWrite**: Write Enable
- **MemtoReg**: Register data in from Memory or ALU
- **ALUOp**: What operation does ALU do
Let us Develop our Control Table

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Op_{5:0}</th>
<th>RegWrite</th>
<th>RegDst</th>
<th>AluSrc</th>
<th>MemWrite</th>
<th>MemtoReg</th>
<th>ALUOp</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-type</td>
<td>0000000</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>funct</td>
</tr>
</tbody>
</table>

- **RegWrite**: Write enable for the register file
- **RegDst**: Write to register RD or RT
- **AluSrc**: ALU input RT or immediate
- **MemWrite**: Write Enable
- **MemtoReg**: Register data in from Memory or ALU
- **ALUOp**: What operation does ALU do
Let us Develop our Control Table

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Op_{5:0}</th>
<th>RegWrite</th>
<th>RegDst</th>
<th>AluSrc</th>
<th>MemWrite</th>
<th>MemtoReg</th>
<th>ALUOp</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-type</td>
<td>000000</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>funct</td>
</tr>
<tr>
<td>lw</td>
<td>100011</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>add</td>
</tr>
</tbody>
</table>

- **RegWrite:** Write enable for the register file
- **RegDst:** Write to register RD or RT
- **AluSrc:** ALU input RT or immediate
- **MemWrite:** Write Enable
- **MemtoReg:** Register data in from Memory or ALU
- **ALUOp:** What operation does ALU do
Let us Develop our Control Table

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Op&lt;sub&gt;5:0&lt;/sub&gt;</th>
<th>RegWrite</th>
<th>RegDst</th>
<th>AluSrc</th>
<th>MemWrite</th>
<th>MemtoReg</th>
<th>ALUOp</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-type</td>
<td>000000</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>funct</td>
</tr>
<tr>
<td>lw</td>
<td>100011</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>add</td>
</tr>
<tr>
<td>sw</td>
<td>101011</td>
<td>0</td>
<td>X</td>
<td>1</td>
<td>1</td>
<td>X</td>
<td>add</td>
</tr>
</tbody>
</table>

- **RegWrite:** Write enable for the register file
- **RegDst:** Write to register RD or RT
- **AluSrc:** ALU input RT or immediate
- **MemWrite:** Write Enable
- **MemtoReg:** Register data in from Memory or ALU
- **ALUOp:** What operation does ALU do
# More Control Signals

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Op&lt;sub&gt;5:0&lt;/sub&gt;</th>
<th>RegWrite</th>
<th>RegDst</th>
<th>AluSrc</th>
<th>Branch</th>
<th>MemWrite</th>
<th>MemtoReg</th>
<th>ALUOp</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-type</td>
<td>000000</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>funct</td>
</tr>
<tr>
<td>lw</td>
<td>100011</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>add</td>
</tr>
<tr>
<td>sw</td>
<td>101011</td>
<td>0</td>
<td>X</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>X</td>
<td>add</td>
</tr>
<tr>
<td>beq</td>
<td>000100</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>X</td>
<td>sub</td>
</tr>
</tbody>
</table>

- **New Control Signal**
  - **Branch**: Are we jumping or not?
Control Unit: Main Decoder

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Op&lt;sub&gt;5:0&lt;/sub&gt;</th>
<th>RegWrite</th>
<th>RegDst</th>
<th>AluSrc</th>
<th>Branch</th>
<th>MemWrite</th>
<th>MemtoReg</th>
<th>ALUOp&lt;sub&gt;1:0&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-type</td>
<td>000000</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>lw</td>
<td>100011</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1,00</td>
</tr>
<tr>
<td>sw</td>
<td>101011</td>
<td>0</td>
<td>X</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>X</td>
<td>00</td>
</tr>
<tr>
<td>beq</td>
<td>000100</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>X</td>
<td>01</td>
</tr>
</tbody>
</table>
Single-Cycle Datapath Example: or
Extended Functionality: addi

- No change to datapath
## Control Unit: `addi`

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Op_{5:0}</th>
<th>RegWrite</th>
<th>RegDst</th>
<th>AluSrc</th>
<th>Branch</th>
<th>MemWrite</th>
<th>MemtoReg</th>
<th>ALUOp_{1:0}</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-type</td>
<td>000000</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>lw</td>
<td>100011</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>00</td>
</tr>
<tr>
<td>sw</td>
<td>101011</td>
<td>0</td>
<td>X</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>X</td>
<td>00</td>
</tr>
<tr>
<td>beq</td>
<td>000100</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>X</td>
<td>01</td>
</tr>
<tr>
<td>addi</td>
<td>001000</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>00</td>
</tr>
</tbody>
</table>
Extended Functionality: \( j \)
## Control Unit: Main Decoder

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Op&lt;sub&gt;5:0&lt;/sub&gt;</th>
<th>RegWrite</th>
<th>RegDst</th>
<th>AluSrc</th>
<th>Branch</th>
<th>MemWrite</th>
<th>MemtoReg</th>
<th>ALUOP&lt;sub&gt;1:0&lt;/sub&gt;</th>
<th>Jump</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-type</td>
<td>0000000</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>lw</td>
<td>100011</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>sw</td>
<td>101011</td>
<td>0</td>
<td>X</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>X</td>
<td>00</td>
<td>0</td>
</tr>
<tr>
<td>beq</td>
<td>000100</td>
<td>0</td>
<td>X</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>X</td>
<td>01</td>
<td>0</td>
</tr>
<tr>
<td>j</td>
<td>000100</td>
<td>0</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>0</td>
<td>X</td>
<td>XX</td>
</tr>
</tbody>
</table>
Review: Complete Single-Cycle Processor (H&H)
A Bit More on Performance Analysis
Processor Performance

- How fast is my program?
  - Every program consists of a series of instructions
  - Each instruction needs to be executed.
Processor Performance

- How fast is my program?
  - Every program consists of a series of instructions
  - Each instruction needs to be executed.

- So how fast are my instructions?
  - Instructions are realized on the hardware
  - They can take one or more clock cycles to complete
  - *Cycles per Instruction* = *CPI*
Processor Performance

- How fast is my program?
  - Every program consists of a series of instructions
  - Each instruction needs to be executed.

- So how fast are my instructions?
  - Instructions are realized on the hardware
  - They can take one or more clock cycles to complete
  - \textit{Cycles per Instruction} = CPI

- How much time is one clock cycle?
  - The critical path determines how much time one cycle requires = \textit{clock period}.
  - 1/clock period = \textit{clock frequency} = how many cycles can be done each second.
Performance Analysis

- Execution time of an instruction
  - \( \text{CPI} \times \text{clock cycle time} \)

- Execution time of a program
  - Sum over all instructions \([\text{CPI} \times \text{clock cycle time}]\)
  - \( \text{# of instructions} \times \text{Average CPI} \times \text{clock cycle time} \)
Processor Performance

Now as a general formula
- Our program consists of executing \( N \) instructions.
- Our processor needs CPI cycles for each instruction.
- The maximum clock speed of the processor is \( f \), and the clock period is therefore \( T=1/f \)
Processor Performance

- Now as a general formula
  - Our program consists of executing $N$ instructions.
  - Our processor needs $CPI$ cycles for each instruction.
  - The maximum clock speed of the processor is $f$, and the clock period is therefore $T=1/f$

- Our program will execute in

$$N \times CPI \times (1/f) = N \times CPI \times T \text{ seconds}$$
How can I Make the Program Run Faster?

\[ N \times \text{CPI} \times \left( \frac{1}{f} \right) \]
How can I Make the Program Run Faster?

$N \times CPI \times (1/f)$

- Reduce the number of instructions
  - Make instructions that ‘do’ more (CISC)
  - Use better compilers
How can I Make the Program Run Faster?

\[ N \times CPI \times (1/f) \]

- **Reduce the number of instructions**
  - Make instructions that ‘do’ more (CISC)
  - Use better compilers

- **Use less cycles to perform the instruction**
  - Simpler instructions (RISC)
  - Use multiple units/ALUs/cores in parallel
How can I Make the Program Run Faster?

\[ N \times CPI \times (1/f) \]

- **Reduce the number of instructions**
  - Make instructions that ‘do’ more (CISC)
  - Use better compilers

- **Use less cycles to perform the instruction**
  - Simpler instructions (RISC)
  - Use multiple units/ALUs/cores in parallel

- **Increase the clock frequency**
  - Find a ‘newer’ technology to manufacture
  - Redesign time critical components
  - Adopt pipelining
Single-Cycle Performance

- $T_c$ is limited by the critical path ($1w$)
Single-Cycle Performance

- Single-cycle critical path:
  \[ T_c = t_{pcq\_PC} + t_{mem} + \max(t_{RF\_read}, t_{sext} + t_{mux}) + t_{ALU} + t_{mem} + t_{mux} + t_{RF\_setup} \]

- In most implementations, limiting paths are:
  - memory, ALU, register file.
  \[ T_c = t_{pcq\_PC} + 2t_{mem} + t_{RF\_read} + t_{mux} + t_{ALU} + t_{RF\_setup} \]
## Single-Cycle Performance Example

<table>
<thead>
<tr>
<th>Element</th>
<th>Parameter</th>
<th>Delay (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Register clock-to-Q</td>
<td>( t_{pcq_PC} )</td>
<td>30</td>
</tr>
<tr>
<td>Register setup</td>
<td>( t_{setup} )</td>
<td>20</td>
</tr>
<tr>
<td>Multiplexer</td>
<td>( t_{mux} )</td>
<td>25</td>
</tr>
<tr>
<td>ALU</td>
<td>( t_{ALU} )</td>
<td>200</td>
</tr>
<tr>
<td>Memory read</td>
<td>( t_{mem} )</td>
<td>250</td>
</tr>
<tr>
<td>Register file read</td>
<td>( t_{RFread} )</td>
<td>150</td>
</tr>
<tr>
<td>Register file setup</td>
<td>( t_{RFsetup} )</td>
<td>20</td>
</tr>
</tbody>
</table>

\[ T_c = \]
## Single-Cycle Performance Example

<table>
<thead>
<tr>
<th>Element</th>
<th>Parameter</th>
<th>Delay (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Register clock-to-Q</td>
<td>$t_{pcq_{PC}}$</td>
<td>30</td>
</tr>
<tr>
<td>Register setup</td>
<td>$t_{setup}$</td>
<td>20</td>
</tr>
<tr>
<td>Multiplexer</td>
<td>$t_{mux}$</td>
<td>25</td>
</tr>
<tr>
<td>ALU</td>
<td>$t_{ALU}$</td>
<td>200</td>
</tr>
<tr>
<td>Memory read</td>
<td>$t_{mem}$</td>
<td>250</td>
</tr>
<tr>
<td>Register file read</td>
<td>$t_{RFread}$</td>
<td>150</td>
</tr>
<tr>
<td>Register file setup</td>
<td>$t_{RFsetup}$</td>
<td>20</td>
</tr>
</tbody>
</table>

\[
T_c = t_{pcq_{PC}} + 2t_{mem} + t_{RFread} + t_{mux} + t_{ALU} + t_{RFsetup} \\
= [30 + 2(250) + 150 + 25 + 200 + 20] \text{ ps} \\
= 925 \text{ ps}
\]
Single-Cycle Performance Example

- Example:
  For a program with 100 billion instructions executing on a single-cycle MIPS processor:
Single-Cycle Performance Example

Example:

For a program with 100 billion instructions executing on a single-cycle MIPS processor:

\[
\text{Execution Time} = \# \text{ instructions} \times \text{CPI} \times \text{TC} \\
= (100 \times 10^9)(1)(925 \times 10^{-12} \text{ s}) \\
= 92.5 \text{ seconds}
\]