Last Time in Lecture 2

- ISA is the hardware/software interface
  - Defines set of programmer visible state
  - Defines instruction format (bit encoding) and instruction semantics
  - Examples: IBM 360, MIPS, RISC-V, x86, JVM

- Many possible implementations of one ISA
  - 360 implementations: model 30 (c. 1964), z12 (c. 2012)
  - x86 implementations: 8086 (c. 1978), 80186, 286, 386, 486, Pentium, Pentium Pro, Pentium-4 (c. 2000), Core 2 Duo, Nehalem, Sandy Bridge, Ivy Bridge, Atom, AMD Athlon, Transmeta Crusoe, SoftPC
  - MIPS implementations: R2000, R4000, R10000, R18K, ...
  - JVM: HotSpot, PicoJava, ARM Jazelle, ...

- Microcoding: straightforward methodical way to implement machines with low logic gate count and complex instructions
“Iron Law” of Processor Performance

\[
\text{Time} = \frac{\text{Instructions}}{\text{Cycles}} = \frac{\text{Time}}{\text{Cycle}}
\]

- Instructions per program depends on source code, compiler technology, and ISA
- Cycles per instructions (CPI) depends on ISA and \( \mu \)architecture
- Time per cycle depends upon the \( \mu \)architecture and base technology

CPI for Microcoded Machine

7 cycles 5 cycles 10 cycles

Inst 1 Inst 2 Inst 3

Total clock cycles = 7 + 5 + 10 = 22
Total instructions = 3
CPI = 22/3 = 7.33
CPI is always an average over a large number of instructions.
Technology Influence

- When microcode appeared in 50s, different technologies for:
  - Logic: Vacuum Tubes
  - Main Memory: Magnetic cores
  - Read-Only Memory: Diode matrix, punched metal cards,…

- Logic very expensive compared to ROM or RAM
- ROM cheaper than RAM
- ROM much faster than RAM

But seventies brought advances in integrated circuit technology and semiconductor memory…

First Microprocessor
Intel 4004, 1971

- 4-bit accumulator architecture
- 8µm pMOS
- 2,300 transistors
- 3 x 4 mm2
- 750kHz clock
- 8-16 cycles/inst.

Made possible by new integrated circuit technology
Microprocessors in the Seventies

- Initial target was embedded control
  - First micro, 4-bit 4004 from Intel, designed for a desktop printing calculator
  - Constrained by what could fit on single chip
  - Accumulator architectures, similar to earliest computers
  - Hardwired state machine control

- 8-bit micros (8085, 6800, 6502) used in hobbyist personal computers
  - Micral, Altair, TRS-80, Apple-II
  - Usually had 16-bit address space (up to 64KB directly addressable)
  - Often came with simple BASIC language interpreter built into ROM or loaded from cassette tape.

VisiCalc – the first “killer” app for micros

- Microprocessors had little impact on conventional computer market until VisiCalc spreadsheet for Apple-II

- Apple-II used Mostek 6502 microprocessor running at 1MHz

Floppy disks were originally invented by IBM as a way of shipping IBM 360 microcode patches to customers!
**DRAM in the Seventies**

- Dramatic progress in semiconductor memory technology
- 1970, Intel introduces first DRAM, 1Kbit 1103
- 1979, Fujitsu introduces 64Kbit DRAM

=> *By mid-Seventies, obvious that PCs would soon have >64KBytes physical memory*

---

**Microprocessor Evolution**

- Rapid progress in 70s, fueled by advances in MOSFET technology and expanding markets
- Intel i432
  - Most ambitious seventies’ micro; started in 1975 - released 1981
  - 32-bit capability-based object-oriented architecture
  - Instructions variable number of bits long
  - Severe performance, complexity, and usability problems
- Motorola 68000 (1979, 8MHz, 68,000 transistors)
  - Heavily microcoded (and nanocoded)
  - 32-bit general-purpose register architecture (24 address pins)
  - 8 address registers, 8 data registers
- Intel 8086 (1978, 8MHz, 29,000 transistors)
  - “Stopgap” 16-bit processor, architected in 10 weeks
  - Extended accumulator architecture, assembly-compatible with 8080
  - 20-bit addressing through segmented addressing scheme
IBM PC, 1981

- **Hardware**
  - Team from IBM building PC prototypes in 1979
  - Motorola 68000 chosen initially, but 68000 was late
  - IBM builds “stopgap” prototypes using 8088 boards from Display Writer word processor
  - 8088 is 8-bit bus version of 8086 => allows cheaper system
  - Estimated sales of 250,000
  - 100,000,000s sold

- **Software**
  - Microsoft negotiates to provide OS for IBM. Later buys and modifies QDOS from Seattle Computer Products.

- **Open System**
  - Standard processor, Intel 8088
  - Standard interfaces
  - Standard OS, MS-DOS
  - IBM permits cloning and third-party software
Microprogramming: early Eighties

- Evolution bred more complex micro-machines
  - Complex instruction sets led to need for subroutine and call stacks in microcode
  - Need for fixing bugs in control programs was in conflict with read-only nature of μROM
  - Writable Control Store (WCS) (B1700, QMachine, Intel i432, ...)
- With the advent of VLSI technology assumptions about ROM & RAM speed became invalid → more complexity
- Better compilers made complex instructions less important.
- Use of numerous micro-architectural innovations, e.g., pipelining, caches and buffers, made multiple-cycle execution of reg-reg instructions unattractive.

Analyzing Microcoded Machines

- John Cocke and group at IBM
  - Working on a simple pipelined processor, 801, and advanced compilers inside IBM
  - Ported experimental PL8 compiler to IBM 370, and only used simple register-register and load/store instructions similar to 801
  - Code ran faster than other existing compilers that used all 370 instructions! (up to 6MIPS whereas 2MIPS considered good before)
- Emer, Clark, at DEC
  - Measured VAX-11/780 using external hardware
  - Found it was actually a 0.5MIPS machine, although usually assumed to be a 1MIPS machine
  - Found 20% of VAX instructions responsible for 60% of microcode, but only account for 0.2% of execution time!
- VAX8800
  - Control Store: 16K*147b RAM, Unified Cache: 64K*8b RAM
  - 4.5x more microstore RAM than cache RAM!
IC Technology Changes Tradeoffs

- Logic, RAM, ROM all implemented using MOS transistors
- Semiconductor RAM ~ same speed as ROM

Nanocoding

Exploits recurring control signal patterns in \( \mu \)code, e.g.,

\[
\begin{align*}
ALU_0 & \quad A \leftarrow \text{Reg}[rs1] \\
\cdots & \\
ALU_i & \quad A \leftarrow \text{Reg}[rs1] \\
\cdots
\end{align*}
\]

- MC68000 had 17-bit \( \mu \)code containing either 10-bit \( \mu \)jump or 9-bit nanoinstruction pointer
  - Nanoinstructions were 68 bits wide, decoded to give 196 control signals
From CISC to RISC

- Use fast RAM to build fast instruction cache of user-visible instructions, not fixed hardware microroutines
  - Contents of fast instruction memory change to fit what application needs right now
- Use simple ISA to enable hardwired pipelined implementation
  - Most compiled code only used a few of the available CISC instructions
  - Simpler encoding allowed pipelined implementations
- Further benefit with integration
  - In early '80s, could finally fit 32-bit datapath + small caches on a single chip
  - No chip crossings in common case allows faster operation

Berkeley RISC Chips

RISC-I (1982) Contains 44,420 transistors, fabbed in 5 µm NMOS, with a die area of 77 mm², ran at 1 MHz. This chip is probably the first VLSI RISC.

RISC-II (1983) contains 40,760 transistors, was fabbed in 3 µm NMOS, ran at 3 MHz, and the size is 60 mm².

Stanford built some too…
CS152 Administrivia

- PS1 and Lab 1 available on website
  - Due Thursday February 14 at start of class
- Quiz 1 on Tuesday February 19 in class
- Class moving to ETCHEVERRY 3108 starting Thursday!

“Iron Law” of Processor Performance

\[
\text{Time} = \frac{\text{Instructions}}{\text{Program}} \times \frac{\text{Cycles}}{\text{Instruction}} \times \frac{\text{Time}}{\text{Cycle}}
\]

- Instructions per program depends on source code, compiler technology, and ISA
- Cycles per instructions (CPI) depends on ISA and \( \mu \)architecture
- Time per cycle depends upon the \( \mu \)architecture and base technology

<table>
<thead>
<tr>
<th>Microarchitecture</th>
<th>CPI</th>
<th>cycle time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microcoded</td>
<td>&gt;1</td>
<td>short</td>
</tr>
<tr>
<td>Single-cycle unpipelined</td>
<td>1</td>
<td>long</td>
</tr>
<tr>
<td>Pipelined</td>
<td>1</td>
<td>short</td>
</tr>
</tbody>
</table>
Hardware Elements

- **Combinational circuits**
  - Mux, Decoder, ALU, ...

- **Synchronous state elements**
  - Flipflop, Register, Register file, SRAM, DRAM

Edge-triggered: Data is sampled at the rising edge

Register Files

- **Reads are combinational**
- RISC-V integer instructions have at most 2 register source operands

A Simple Memory Model

- Reads and writes are always completed in one cycle
  - a Read can be done any time (i.e. combinational)
  - a Write is performed at the rising clock edge if it is enabled
    \[ \Rightarrow \text{the write address and data must be stable at the clock edge} \]

Later in the course we will present a more realistic model of memory
Implementing RISC-V

Single-cycle per instruction datapath & control logic
(Should be review of CS61C)

Instruction Execution

Execution of an instruction involves

1. Instruction fetch
2. Decode and register fetch
3. ALU operation
4. Memory operation (optional)
5. Write back (optional)

and compute address of next instruction
Datapath: Reg-Reg ALU Instructions

```
<table>
<thead>
<tr>
<th>rd</th>
<th>rs1</th>
<th>rs2</th>
<th>func</th>
<th>opcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>27</td>
<td>26</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

RegWrite Timing?

```
rd ← (rs1) func (rs2)
```

Datapath: Reg-Imm ALU Instructions

```
<table>
<thead>
<tr>
<th>rd</th>
<th>rs1</th>
<th>immediate12</th>
<th>func</th>
<th>opcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>27</td>
<td>26</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

```
rd ← (rs1) op immediate
```
Conflicts in Merging Datapath

Datapath for ALU Instructions

1/29/2013  CS152, Spring 2013
Load/Store Instructions

rs1 is the base register
rd is the destination of a Load, rs2 is the data source for a Store

RISC-V Conditional Branches

- Compare two integer registers for equality (BEQ/BNE) or signed magnitude (BLT/BGE) or unsigned magnitude (BLTU/BGEU)
- 12-bit immediate encodes branch target address as a signed offset from PC, in units of 16-bits (i.e., shift left by 1 then add to PC).
Conditional Branches (BEQ/BNE/BLT/BGE/BLTU/BGEU)

RISC-V Unconditional Jumps

- 25-bit immediate encodes jump target address as a signed offset from PC, in units of 16-bits (i.e., shift left by 1 then add to PC). (+/- 16MB)
- JAL is a subroutine call that also saves return address (PC +4) in register x1
RISC-V Register Indirect Jumps

- Jumps to target address given by adding 12-bit offset (not shifted by 1 bit) to register rs1
- The return address (PC+4) is written to rd (can be x0 if value not needed)
- The RDNPC instruction simply writes return address to register rd without jumping (used for dynamic linking)
Hardwired Control is pure Combinational Logic

ALU Control & Immediate Extension
Single-Cycle Hardwired Control

We will assume clock period is sufficiently long for all of the following steps to be “completed”:

1. Instruction fetch
2. Decode and register fetch
3. ALU operation
4. Data fetch if required
5. Register write-back setup time

\[ t_C > t_{IF} + t_{RF} + t_{ALU} + t_{DM} + t_{RWB} \]

At the rising edge of the following clock, the PC, register file and memory are updated
Summary

- Microcoding became less attractive as gap between RAM and ROM speeds reduced, and logic implemented in same technology as memory.
- Complex instruction sets difficult to pipeline, so difficult to increase performance as gate count grew.
- Iron Law explains architecture design space.
- Trade instruction/program, cycles/instruction, and time/cycle.
- Load-Store RISC ISAs designed for efficient pipelined implementations.
- Very similar to vertical microcode.
- Inspired by earlier Cray machines (CDC 6600/7600).
- RISC-V ISA will be used in lectures, problems, and labs.
- Berkeley RISC chips: RISC-I, RISC-II, SOAR (RISC-III), SPUR (RISC-IV).

Acknowledgements

- These slides contain material developed and copyright by:
  - Arvind (MIT)
  - Krste Asanovic (MIT/UCB)
  - Joel Emer (Intel/MIT)
  - James Hoe (CMU)
  - John Kubiatowicz (UCB)
  - David Patterson (UCB)
- MIT material derived from course 6.823
- UCB material derived from course CS252