More ISA

Property of ISA vs. Uarch?

- ADD instruction’s opcode
- Number of general purpose registers
- Number of cycles to execute the MUL instruction
- 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
- Considerations
  - Cost
  - Performance
  - Maximum power consumption
  - Energy consumption (battery life)
  - Availability
  - Reliability and Correctness
  - Time to Market

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

Many Different ISAs Over Decades

- x86
- PDP-x: Programmed Data Processor (PDP-11)
- VAX
- IBM 360
- CDC 6600
- SIMD ISAs: CRAY-1, Connection Machine
- VLIW ISAs: Multiflow, Cydrome, IA-64 (EPIC)
- PowerPC, POWER
- RISC ISAs: Alpha, MIPS, SPARC, ARM

- What are the fundamental differences?
  - E.g., how instructions are specified and what they do
  - E.g., how complex are the instructions
Instruction

- Basic element of the HW/SW interface
- Consists of
  - opcode: what the instruction does
  - operands: who it is to do it to

- Example from Alpha ISA:
Set of Instructions, Encoding, and Spec

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Encoding</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD*</td>
<td>0001 DR SR A op spec</td>
<td></td>
</tr>
<tr>
<td>AND*</td>
<td>0101 DR SR A op spec</td>
<td></td>
</tr>
<tr>
<td>BR</td>
<td>0000 m p PCofflag</td>
<td></td>
</tr>
<tr>
<td>JMP</td>
<td>1100 000 BaseR 000000</td>
<td></td>
</tr>
<tr>
<td>JSR(R)</td>
<td>0100 A operand specifier</td>
<td></td>
</tr>
<tr>
<td>LDB*</td>
<td>0010 DR BaseL offsetPC</td>
<td></td>
</tr>
<tr>
<td>LDW*</td>
<td>0110 DR BaseL offsetPC</td>
<td></td>
</tr>
<tr>
<td>LEA*</td>
<td>1110 DR PCoffset</td>
<td></td>
</tr>
<tr>
<td>RTI</td>
<td>1000 000000000000</td>
<td></td>
</tr>
<tr>
<td>SHF*</td>
<td>1101 DR SR A D amount</td>
<td></td>
</tr>
<tr>
<td>STB</td>
<td>0011 SR BaseL offsetPC</td>
<td></td>
</tr>
<tr>
<td>SWW</td>
<td>0111 SR BaseL offsetPC</td>
<td></td>
</tr>
<tr>
<td>TRAP</td>
<td>1111 0000 trapspec8</td>
<td></td>
</tr>
<tr>
<td>XOR*</td>
<td>1011 DR SR A op spec</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1011</td>
<td></td>
</tr>
</tbody>
</table>

- Example from LC-3b ISA
- x86 Manual
- Aside: concept of “bit steering”
  - A bit in the instruction determines the interpretation of other bits
- Why unused instructions?

Bit Steering in Alpha

**Figure 3-4: Operate Instruction Format**

```
  31  26  21  16  15  12  11  5  4  0
 Opcode  Rb  Rb  SBZ 0    Function   Rc
```

If bit <12> of the instruction is 0, the Rb field specifies a source register operand.

If bit <12> of the instruction is 1, an 8-bit zero-extended literal constant is formed by bits <20:13> of the instruction. The literal is interpreted as a positive integer between 0 and 255 and is zero-extended to 64 bits. Symbolically, the integer Rb operand is formed as follows:
What Are the Elements of An ISA?

- **Instruction processing style**
  - Specifies the number of “operands” an instruction “operates” on and how it does so
  - 0, 1, 2, 3 address machines
    - 0-address: stack machine (push A, pop A, op)
    - 1-address: accumulator machine (ld A, st A, op A)
    - 2-address: 2-operand machine (one is both source and dest)
    - 3-address: 3-operand machine (source and dest are separate)

- **Tradeoffs?**
  - Larger operate instructions vs. more executed operations
  - Code size vs. execution time vs. on-chip memory space

An Example: Stack Machine

+ Small instruction size (no operands needed for operate instructions)
  - Simpler logic
  - Compact code

+ Efficient procedure calls: all parameters on stack
  - No additional cycles for parameter passing

-- Computations that are not easily expressible with “postfix notation” are difficult to map to stack machines
  - Cannot perform operations on many values at the same time (only top N values on the stack at the same time)
  - Not flexible
An Example: Stack Machine (II)

http://www.ece.cmu.edu/~koopman/stack_computers/sec3_2.html

Figure 3.1 -- The canonical stack machine.

An Example: Stack Machine Operation

http://www.ece.cmu.edu/~koopman/stack_computers/sec3_2.html

Figure 3.2 -- An example stack machine.
Other Examples

- PDP-11: A 2-address machine
  - PDP-11 ADD: 4-bit opcode, 2 6-bit operand specifiers
  - Why? Limited bits to specify an instruction
  - Disadvantage: One source operand is always clobbered with the result of the instruction
    - How do you ensure you preserve the old value of the source?

- X86: A 2-address (memory/memory) machine
- Alpha: A 3-address (load/store) machine
- MIPS?

What Are the Elements of An ISA?

- Instructions
  - Opcode
  - Operand specifiers (addressing modes)
    - How to obtain the operand? Why are there different addressing modes?

- Data types
  - Definition: Representation of information for which there are instructions that operate on the representation
  - Integer, floating point, character, binary, decimal, BCD
  - Doubly linked list, queue, string, bit vector, stack
    - VAX: INSQUEUE and REMQUEUE instructions on a doubly linked list or queue; FINDFIRST
    - X86: SCAN opcode operates on character strings; PUSH/POP
Data Type Tradeoffs

- What is the benefit of having more or high-level data types in the ISA?
- What is the disadvantage?

- Think compiler/programmer vs. microarchitect

- Concept of semantic gap
  - Data types coupled tightly to the semantic level, or complexity of instructions

- Example: Early RISC architectures vs. Intel 432
  - Early RISC: Only integer data type
  - Intel 432: Object data type, capability based machine

What Are the Elements of An ISA?

- Memory organization
  - Address space: How many uniquely identifiable locations in memory
  - Addressability: How much data does each uniquely identifiable location store
    - Byte addressable: most ISAs, characters are 8 bits
    - Bit addressable: Burroughs 1700. Why?
    - 64-bit addressable: Some supercomputers. Why?
    - 32-bit addressable: First Alpha
  - Food for thought
    - How do you add 2 32-bit numbers with only byte addressability?
    - How do you add 2 8-bit numbers with only 32-bit addressability?
    - Big endian vs. little endian? MSB at low or high byte.

- Support for virtual memory
Big-endian vs. Little-endian

- **Big-endian** systems store the most significant byte of a word in the smallest address.
- **Little-endian** systems, in contrast, store the least significant byte in the smallest address.

What Are the Elements of An ISA?

- **Registers**
  - How many
  - Size of each register

- Why is having registers a good idea?
  - Because programs exhibit a characteristic called **data locality**
  - A recently produced/accessed value is likely to be used more than once (temporal locality)
    - Storing that value in a register eliminates the need to go to memory each time that value is needed
Programmer Visible (Architectural) State

- Memory array of storage locations indexed by an address
- Registers
  - given special names in the ISA (as opposed to addresses)
  - general vs. special purpose

Instructions (and programs) specify how to transform the values of programmer visible state

Aside: Programmer Invisible State

- Microarchitectural state
- Programmer cannot access this directly

- E.g. cache state
- E.g. pipeline registers
Evolution of Register Architecture

- **Accumulator**
  - a legacy from the “adding” machine days

- **Accumulator + address registers**
  - need register indirection
  - initially address registers were special-purpose, i.e., can only be loaded with an address for indirection
  - eventually arithmetic on addresses became supported

- **General purpose registers (GPR)**
  - all registers good for all purposes
  - grew from a few registers to 32 (common for RISC) to 128 in Intel IA-64

Instruction Classes

- **Operate instructions**
  - Process data: arithmetic and logical operations
  - Fetch operands, compute result, store result
  - Implicit sequential control flow

- **Data movement instructions**
  - Move data between memory, registers, I/O devices
  - Implicit sequential control flow

- **Control flow instructions**
  - Change the sequence of instructions that are executed
What Are the Elements of An ISA?

- **Load/store vs. memory/memory architectures**

  - Load/store architecture: operate instructions operate only on registers
    - E.g., MIPS, ARM and many RISC ISAs
  
  - Memory/memory architecture: operate instructions can operate on memory locations
    - E.g., x86, VAX and many CISC ISAs

---

What Are the Elements of An ISA?

- **Addressing modes** specify how to obtain the operands

  - Absolute: \( \text{LW rt, 10000} \)
    - use immediate value as address
  
  - Register Indirect: \( \text{LW rt, } (r_{\text{base}}) \)
    - use GPR[\( r_{\text{base}} \)] as address
  
  - Displaced or based: \( \text{LW rt, offset(}r_{\text{base}}) \)
    - use offset+GPR[\( r_{\text{base}} \)] as address
  
  - Indexed: \( \text{LW rt, } (r_{\text{base}}, r_{\text{index}}) \)
    - use GPR[\( r_{\text{base}} \)]+GPR[\( r_{\text{index}} \)] as address
  
  - Memory Indirect: \( \text{LW rt, (}r_{\text{base}}) \)
    - use value at M[ GPR[ \( r_{\text{base}} \) ] ] as address
  
  - Auto inc/decrement: \( \text{LW Rt, } (r_{\text{base}}) \)
    - use GPR[\( r_{\text{base}} \)] as address, but inc. or dec. GPR[\( r_{\text{base}} \)] each time
What Are the Benefits of Different Addressing Modes?

- Another example of programmer vs. microarchitect tradeoff

- Advantage of more addressing modes:
  - Enables better mapping of high-level constructs to the machine: some accesses are better expressed with a different mode → reduced number of instructions and code size
    - Think array accesses (autoincrement mode)
    - Think indirection (pointer chasing)
  
- Disadvantage:
  - More work for the compiler
  - More work for the microarchitect

ISA Orthogonality

- Orthogonal ISA:
  - All addressing modes can be used with all instruction types
  - Example: VAX
    - (~13 addressing modes) x (>300 opcodes) x (integer and FP formats)

- Who is this good for?
- Who is this bad for?
### Is the LC-3b ISA Orthogonal?

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Opcode</th>
<th>Use</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD</td>
<td>0001</td>
<td>DR</td>
<td>SR1</td>
<td>A</td>
<td>op specifier</td>
</tr>
<tr>
<td>AND</td>
<td>0101</td>
<td>DR</td>
<td>SR1</td>
<td>A</td>
<td>op specifier</td>
</tr>
<tr>
<td>BR</td>
<td>0000</td>
<td>n</td>
<td>p</td>
<td>PC+off</td>
<td>0</td>
</tr>
<tr>
<td>JMP</td>
<td>1100</td>
<td>000</td>
<td>BaseR</td>
<td>000000</td>
<td></td>
</tr>
<tr>
<td>JSR(R)</td>
<td>0100</td>
<td>A</td>
<td>operand specifier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDB</td>
<td>0010</td>
<td>DR</td>
<td>BaseR</td>
<td>offset</td>
<td></td>
</tr>
<tr>
<td>LDW</td>
<td>0110</td>
<td>DR</td>
<td>BaseR</td>
<td>offset</td>
<td></td>
</tr>
<tr>
<td>LEA</td>
<td>1110</td>
<td>DR</td>
<td>PC+offset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RTI</td>
<td>1000</td>
<td>00000000000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHF</td>
<td>1101</td>
<td>DR</td>
<td>SR</td>
<td>A</td>
<td>D amount</td>
</tr>
<tr>
<td>STB</td>
<td>0011</td>
<td>DR</td>
<td>BaseR</td>
<td>offset</td>
<td></td>
</tr>
<tr>
<td>STW</td>
<td>0111</td>
<td>SR</td>
<td>BaseR</td>
<td>offset</td>
<td></td>
</tr>
<tr>
<td>TRAP</td>
<td>1111</td>
<td>0000</td>
<td>trapspec</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XOR</td>
<td>1011</td>
<td>DR</td>
<td>SR1</td>
<td>A</td>
<td>op specifier</td>
</tr>
<tr>
<td>not used</td>
<td>1010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not used</td>
<td>1011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### LC-3b: Addressing Modes of ADD

**Encodings**

<table>
<thead>
<tr>
<th>Bit Positions</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>12, 11</td>
<td>DR</td>
</tr>
<tr>
<td>9</td>
<td>SR1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>15, 7</td>
<td>001</td>
</tr>
<tr>
<td>12, 11</td>
<td>DR</td>
</tr>
<tr>
<td>6</td>
<td>SR1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Imm5</td>
</tr>
</tbody>
</table>

**Operation**

- if (bit[5] == 0)
  - DR = SR1 + SR2;
  - else
  - DR = SR1 + SEXT(mm5);
  - seteQ;
LC-3b: Addressing Modes of JSR(R)

**Encodings**

<table>
<thead>
<tr>
<th>JSR</th>
<th>JSRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0100</td>
<td>0100</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>PCoffset11</td>
<td>BaseR</td>
</tr>
</tbody>
</table>

**Operation**

\[
R7 = PC^2;
\]
| if (bit[11] == 0) |
| PC = BaseR; |
| else |
| PC = PC^2 + LSHIFT(SEXT(PCoffset11), 1); |

**Description**

First, the incremented PC is saved in R7. This is the linkage back to the calling routine. Then, the PC is loaded with the address of the first instruction of the subroutine, causing an unconditional jump to that address. The address of the subroutine is obtained from the base register (if bit[11] is 0), or the address is computed by sign-extending bits [10:0] to 16 bits, left-shifting the result one bit, and then adding this value to the incremented PC (if bit[11] is 1).

---

Another Question

- Does the LC-3b ISA contain complex instructions?
Complex vs. Simple Instructions

- **Complex instruction**: An instruction does a lot of work, e.g. many operations
  - Insert in a doubly linked list
  - Compute FFT
  - String copy

- **Simple instruction**: An instruction does small amount of work, it is a primitive using which complex operations can be built
  - Add
  - XOR
  - Multiply

**Advantages of Complex instructions**
- Denser encoding $\rightarrow$ smaller code size $\rightarrow$ better memory utilization, saves off-chip bandwidth, better cache hit rate (better packing of instructions)
- Simpler compiler: no need to optimize small instructions as much

**Disadvantages of Complex Instructions**
- Larger chunks of work $\rightarrow$ compiler has less opportunity to optimize (limited in fine-grained optimizations it can do)
- More complex hardware $\rightarrow$ translation from a high level to control signals and optimization needs to be done by hardware
ISA-level Tradeoffs: Semantic Gap

**Where to place the ISA?** Semantic gap
- Closer to high-level language (HLL) \( \rightarrow \) Small semantic gap, complex instructions
- Closer to hardware control signals? \( \rightarrow \) Large semantic gap, simple instructions

**RISC vs. CISC machines**
- RISC: Reduced instruction set computer
- CISC: Complex instruction set computer
  - FFT, QUICKSORT, POLY, FP instructions?
  - VAX INDEX instruction (array access with bounds checking)

---

ISA-level Tradeoffs: Semantic Gap

**Some tradeoffs (for you to think about)**

- Simple compiler, complex hardware vs. complex compiler, simple hardware
  - Caveat: Translation (indirection) can change the tradeoff!

- Burden of backward compatibility

- Performance?
  - Optimization opportunity: Example of VAX INDEX instruction: who (compiler vs. hardware) puts more effort into optimization?
  - Instruction size, code size
X86: Small Semantic Gap: String Operations

- An instruction operates on a string
  - Move one string of arbitrary length to another location
  - Compare two strings

- Enabled by the ability to specify repeated execution of an instruction (in the ISA)
  - Using a "prefix" called REP prefix

- Example: REP MOVXS instruction
  - Only two bytes: REP prefix byte and MOVXS opcode byte (F2 A4)
  - Implicit source and destination registers pointing to the two strings (ESI, EDI)
  - Implicit count register (ECX) specifies how long the string is

---

X86: Small Semantic Gap: String Operations

**REP MOVS** (DEST SRC)

IF AddressSize = 16
THEN
  Use CX for CountReg;
ELSE IF AddressSize = 64 and REX.W used
  THEN Use REX for CountReg; F1;
ELSE
  Use ECX for CountReg;
F1;
WHILE CountReg ≠ 0
  DO
    Service pending interrupts (if any);
    Execute associated string instruction;
    CountReg ← (CountReg − 1);
    IF CountReg = 0
      THEN exit WHILE loop; F1;
    IF (Repeat prefix is REPZ or REPE) and (ZF = 0)
      or (Repeat prefix is REPNZ or REPNE) and (ZF = 1)
      THEN exit WHILE loop; F1;
  OD;

**How many instructions does this take in MIPS?**
Small Semantic Gap Examples in VAX

- FIND FIRST
  - Find the first set bit in a bit field
  - Helps OS resource allocation operations
- SAVE CONTEXT, LOAD CONTEXT
  - Special context switching instructions
- INSQUEUE, REMQUEUE
  - Operations on doubly linked list
- INDEX
  - Array access with bounds checking
- STRING Operations
  - Compare strings, find substrings, ...
- Cyclic Redundancy Check Instruction
- EDITPC
  - Implements editing functions to display fixed format output


Small versus Large Semantic Gap

- CISC vs. RISC
  - Complex instruction set computer → complex instructions
    - Initially motivated by “not good enough” code generation
  - Reduced instruction set computer → simple instructions
    - John Cocke, mid 1970s, IBM 801
      - Goal: enable better compiler control and optimization

- RISC motivated by
  - Memory stalls (no work done in a complex instruction when there is a memory stall?)
    - When is this correct?
  - Simplifying the hardware → lower cost, higher frequency
  - Enabling the compiler to optimize the code better
    - Find fine-grained parallelism to reduce stalls
How High or Low Can You Go?

- **Very large semantic gap**
  - Each instruction specifies the complete set of control signals in the machine
  - Compiler generates control signals
  - Open microcode (John Cocke, circa 1970s)
    - Gave way to optimizing compilers

- **Very small semantic gap**
  - ISA is (almost) the same as high-level language
  - Java machines, LISP machines, object-oriented machines, capability-based machines

A Note on ISA Evolution

- ISAs have evolved to reflect/satisfy the concerns of the day

- **Examples:**
  - Limited on-chip and off-chip memory size
  - Limited compiler optimization technology
  - Limited memory bandwidth
  - Need for specialization in important applications (e.g., MMX)

- Use of translation (in HW and SW) enabled underlying implementations to be similar, regardless of the ISA
  - Concept of dynamic/static interface
  - Contrast it with hardware/software interface
Effect of Translation

- One can translate from one ISA to another ISA to change the semantic gap tradeoffs

Examples

- Intel’s and AMD’s x86 implementations translate x86 instructions into programmer-invisible microoperations (simple instructions) in hardware
- Transmeta’s x86 implementations translated x86 instructions into “secret” VLIW instructions in software (code morphing software)

Think about the tradeoffs

ISA-level Tradeoffs: Instruction Length

- Fixed length: Length of all instructions the same
  - Easier to decode single instruction in hardware
  - Easier to decode multiple instructions concurrently
  - Wasted bits in instructions (Why is this bad?)
  - Harder-to-extend ISA (how to add new instructions?)

- Variable length: Length of instructions different (determined by opcode and sub-opcode)
  - Compact encoding (Why is this good?)
    - Intel 432: Huffman encoding (sort of). 6 to 321 bit instructions. How?
  - More logic to decode a single instruction
  - Harder to decode multiple instructions concurrently

Tradeoffs

- Code size (memory space, bandwidth, latency) vs. hardware complexity
- ISA extensibility and expressiveness
- Performance? Smaller code vs. imperfect decode
**ISA-level Tradeoffs: Uniform Decode**

- **Uniform decode**: Same bits in each instruction correspond to the same meaning
  - Opcode is always in the same location
  - Immediate values, ...
  - Many “RISC” ISAs: Alpha, MIPS, SPARC
    + Easier decode, simpler hardware
    + Enables parallelism: generate target address before knowing the instruction is a branch
    -- Restricts instruction format (fewer instructions?) or wastes space

- **Non-uniform decode**
  - E.g., opcode can be the 1st-7th byte in x86
    + More compact and powerful instruction format
    -- More complex decode logic

---

**x86 vs. Alpha Instruction Formats**

- **x86**:

  ![x86 Instruction Format Diagram](image)

- **Alpha**:

  ![Alpha Instruction Format Diagram](image)
MIPS Instruction Format

- **R-type**, 3 register operands
  
<table>
<thead>
<tr>
<th>opcode</th>
<th>rs</th>
<th>rt</th>
<th>rd</th>
<th>shamt</th>
<th>funct</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-bit</td>
<td>5-bit</td>
<td>5-bit</td>
<td>5-bit</td>
<td>5-bit</td>
<td>6-bit</td>
</tr>
</tbody>
</table>

- **I-type**, 2 register operands and 16-bit immediate operand
  
<table>
<thead>
<tr>
<th>opcode</th>
<th>rs</th>
<th>rt</th>
<th>immediate</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-bit</td>
<td>5-bit</td>
<td>5-bit</td>
<td>16-bit</td>
</tr>
</tbody>
</table>

- **J-type**, 26-bit immediate operand
  
<table>
<thead>
<tr>
<th>opcode</th>
<th>immediate</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-bit</td>
<td>26-bit</td>
</tr>
</tbody>
</table>

**Simple Decoding**

- 4 bytes per instruction, regardless of format
- must be 4-byte aligned \(2 \text{lsb of PC must be } 2b'00\)
- format and fields easy to extract in hardware

---

A Note on Length and Uniformity

- Uniform decode usually goes with fixed length

- In a variable length ISA, uniform decode can be a property of instructions of the same length
  
  - It is hard to think of it as a property of instructions of different lengths
A Note on RISC vs. CISC

- Usually, ...

- RISC
  - Simple instructions
  - Fixed length
  - Uniform decode
  - Few addressing modes

- CISC
  - Complex instructions
  - Variable length
  - Non-uniform decode
  - Many addressing modes

ISA-level Tradeoffs: Number of Registers

- Affects:
  - Number of bits used for encoding register address
  - Number of values kept in fast storage (register file)
  - (uarch) Size, access time, power consumption of register file

- Large number of registers:
  + Enables better register allocation (and optimizations) by compiler → fewer saves/restores
  -- Larger instruction size
  -- Larger register file size
What Are the Elements of An ISA?

- **How to interface with I/O devices**
  - Memory mapped I/O
    - A region of memory is mapped to I/O devices
    - I/O operations are loads and stores to those locations
  - Special I/O instructions
    - IN and OUT instructions in x86 deal with ports of the chip
  - Tradeoffs?
    - Which one is more general purpose?

---

What Are the Elements of An ISA?

- **Privilege modes**
  - User vs supervisor
  - Who can execute what instructions?

- **Exception and interrupt handling**
  - What procedure is followed when something goes wrong with an instruction?
  - What procedure is followed when an external device requests the processor?
  - Vectored vs. non-vectored interrupts (early MIPS)

- **Virtual memory**
  - Each program has the illusion of the entire memory space, which is greater than physical memory
x86 vs. Alpha Instruction Formats

- x86:

<table>
<thead>
<tr>
<th>Instruction Prefixes</th>
<th>Opcode</th>
<th>Mod/M</th>
<th>SIB</th>
<th>Displacement</th>
<th>Immediate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to four prefixes 1-, 2-, or 3-byte opcode</td>
<td>1 byte if required</td>
<td>1 byte if required</td>
<td>Address displacement of 1, 2, or 4 bytes or none</td>
<td>Immediate data of 1, 2, or 4 bytes or none</td>
<td></td>
</tr>
</tbody>
</table>

- Alpha:

<table>
<thead>
<tr>
<th>Opcode</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PALcode Format</td>
</tr>
<tr>
<td></td>
<td>Branch Format</td>
</tr>
<tr>
<td></td>
<td>Memory Format</td>
</tr>
<tr>
<td></td>
<td>Operate Format</td>
</tr>
</tbody>
</table>

Other Example ISA-level Tradeoffs

- Condition codes vs. not
- VLIW vs. single instruction
- Precise vs. imprecise exceptions
- Virtual memory vs. not
- Unaligned access vs. not
- Hardware interlocks vs. software-guaranteed interlocking
- Software vs. hardware managed page fault handling
- Cache coherence (hardware vs. software)
- ...
Back to Programmer vs. (Micro)architect

- Many ISA features designed to aid programmers
- But, complicate the hardware designer’s job

- Virtual memory
  - vs. overlay programming
  - Should the programmer be concerned about the size of code blocks fitting physical memory?
- Addressing modes
- Unaligned memory access
  - Compile/programmer needs to align data

MIPS: Aligned Access

- LW/SW alignment restriction: 4-byte word-alignment
  - not designed to fetch memory bytes not within a word boundary
  - not designed to rotate unaligned bytes into registers
- Provide separate opcodes for the “infrequent” case

<table>
<thead>
<tr>
<th>LWL</th>
<th>rd 6(r0)</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>byte-6</td>
<td>byte-5</td>
<td>byte-4</td>
<td>D</td>
</tr>
</tbody>
</table>

| LWR | rd 3(r0) | byte-6 | byte-5 | byte-4 | byte-3 |

- LWL/LWR is slower
- Note LWL and LWR still fetch within word boundary
X86: Unaligned Access

Aligned vs. Unaligned Access

- Pros of having no restrictions on alignment

- Cons of having no restrictions on alignment

- Filling in the above: an exercise for you...
MIPS machine language code for a routine to compute and print the sum of the squares of integers between 0 and 100.

```
MIPS machine language code for a routine to compute and print the sum of the squares of integers between 0 and 100.

```

MIPS

```
The same routine written in assembly language. However, the code for the routine does not label registers or memory locations nor include comments.

```

sw $4, 32($29)
sw $5, 36($29)
sw $0, 24($29)
sw $0, 28($29)
lw $14, 28($29)
lw $24, 24($29)
multi $14, $14
addiu $8, 14, 1
slt $1, $8, 101
sw $8, 28($29)
mflo $15
addu $25, $24, $15
bne $1, 0, -9
sw $25, 24($29)
lui $4, 4096
lw $5, 24($29)
jal 1048812
addiu $4, 4, 1072
lw $31, 20($29)
adu $29, $29, 32
```
MIPS

```mips
.text
.align 2
.globl main

subu $sp, $sp, 32
sw $ra, 20($sp)
sd $a0, 32($sp)
sw $0, 24($sp)
sw $0, 28($sp)

lw $t6, 28($sp)
mul $t7, $t6, $t6
lw $t8, 24($sp)
addu $t9, $t8, $t7
sw $t9, 24($sp)
addu $t0, $t6, 1
sw $t0, 28($sp)
ble $t0, 100, loop
la $a0, str
lw $a1, 24($sp)
jal printf
move $v0, $0
lw $ra, 20($sp)
addu $sp, $sp, 32
jr $ra
```

The same routine written in assembly language with labels, but no comments.

The routine written in the C programming

```c
#include <stdio.h>

int main (int argc, char *argv[])
{
    int i;
    int sum = 0;

    for (i = 0; i <= 100; i = i + 1) sum = sum + i;
    printf ("The sum from 0 .. 100 is %d\n", sum);
}
```
The routine written in Ruby

(1..100).inject{|s,n| s + n * n}