Machine-Level Representation of Programs: Instruction set architectures

Yipeng Huang

Rutgers University

March 6, 2023
Table of contents

Announcements
  Quizzes and programming assignments
  Reading assignments

Floats: Understanding its design
  Deep understanding 1: Why is exp field encoded using bias?
  Deep understanding 2: Why have denormalized numbers?
  Deep understanding 3: Why is bias chosen to be $2^{k-1} - 1$?

Floats: Properties
  Floating point multiplication
  Properties of floating point

Computer organization: A primer

Assembly code: Human readable representation of machine code

Instruction set architectures

swap.s: Assembly implementation of function that swaps memory contents

Data size and IA32, x86, and x86-64 registers
Quizzes and programming assignments

Short quiz 5
- Due Friday. All about floats.

Programming assignment 3
- Has been out, due Friday before spring break.
Reading assignments

CS:APP Chapters 3.1-3.4
<table>
<thead>
<tr>
<th>Table of contents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Announcements</strong></td>
</tr>
<tr>
<td>Quizzes and programming assignments</td>
</tr>
<tr>
<td>Reading assignments</td>
</tr>
<tr>
<td><strong>Floats: Understanding its design</strong></td>
</tr>
<tr>
<td>Deep understanding 1: Why is exp field encoded using bias?</td>
</tr>
<tr>
<td>Deep understanding 2: Why have denormalized numbers?</td>
</tr>
<tr>
<td>Deep understanding 3: Why is bias chosen to be $2^{k-1} - 1$?</td>
</tr>
<tr>
<td><strong>Floats: Properties</strong></td>
</tr>
<tr>
<td>Floating point multiplication</td>
</tr>
<tr>
<td>Properties of floating point</td>
</tr>
<tr>
<td><strong>Computer organization: A primer</strong></td>
</tr>
<tr>
<td><strong>Assembly code: Human readable representation of machine code</strong></td>
</tr>
<tr>
<td>Instruction set architectures</td>
</tr>
<tr>
<td><strong>swap.s</strong>: Assembly implementation of function that swaps memory contents</td>
</tr>
<tr>
<td>Data size and IA32, x86, and x86-64 registers</td>
</tr>
</tbody>
</table>
The IEEE 754 number line

Figure: Full picture of number line for floating point values. Image credit CS:APP

Figure: Zoomed in number line for floating point values. Image credit CS:APP
## Floats: Summary

<table>
<thead>
<tr>
<th>normalized</th>
<th>denormalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(-1)^s \times M \times 2^E$</td>
<td>$(-1)^s \times M \times 2^E$</td>
</tr>
<tr>
<td>$E = \text{exp-bias}$</td>
<td>$E = \text{-bias} + 1$</td>
</tr>
<tr>
<td>$2^{k-1} - 1$</td>
<td>$2^{k-1} - 1$</td>
</tr>
<tr>
<td>$0 &lt; \text{exp} &lt; (2^k - 1)$</td>
<td>$\text{exp} = 0$</td>
</tr>
<tr>
<td>$M = 1.\text{frac}$</td>
<td>$M = 0.\text{frac}$</td>
</tr>
<tr>
<td>M has implied leading 1</td>
<td>M has leading 0</td>
</tr>
</tbody>
</table>

| greater range | greater precision |
| large magnitude numbers | small magnitude numbers |
| denser near origin | evenly spaced |

**Table:** Summary of normalized and denormalized numbers
Deep understanding 1: Why is exp field encoded using bias?

exp field needs to encode both positive and negative exponents. Why not just use one of the signed integer formats? 2’s complement, 1s’ complement, signed magnitude?
Deep understanding 1: Why is exp field encoded using bias?

exp field needs to encode both positive and negative exponents. Why not just use one of the signed integer formats? 2’s complement, 1s’ complement, signed magnitude?

Answer: allows easy comparison of magnitudes by simply comparing bits.
Deep understanding 1: Why is exp field encoded using bias?

exp field needs to encode both positive and negative exponents. Why not just use one of the signed integer formats? 2’s complement, 1s’ complement, signed magnitude?

Answer: allows easy comparison of magnitudes by simply comparing bits.

Consider hypothetical 8-bit floating point format (from the textbook)
1-bit sign, $k = 4$-bit exp, 3-bit frac.

What is the decimal value of $0b1_0110_111$? What is the decimal value of $0b1_0111_000$?
Deep understanding 1: Why is exp field encoded using bias?

exp field needs to encode both positive and negative exponents. Why not just use one of the signed integer formats? 2’s complement, 1s’ complement, signed magnitude?

Answer: allows easy comparison of magnitudes by simply comparing bits.

Consider hypothetical 8-bit floating point format (from the textbook)
1-bit sign, $k = 4$-bit exp, 3-bit frac.

What is the decimal value of $0b1_0110_111$?
$-1.875 \times 2^{-1}$

What is the decimal value of $0b1_0111_000$?
$-2.000 \times 2^{-1}$
Deep understanding 2: Why have denormalized numbers?

Why not just continue normalized number scheme down to smallest numbers around zero?
Answer: makes sure that smallest increments available are maintained around zero.

Suppose denormalized numbers NOT used.

What is the decimal value of 0b0_0000_001?
1.125 × 2⁻⁷

What is the decimal value of 0b0_0000_111?
1.875 × 2⁻⁷

What is the decimal value of 0b0_0001_000?
2.000 × 2⁻⁷
Deep understanding 2: Why have denormalized numbers?

Why not just continue normalized number scheme down to smallest numbers around zero?
Answer: makes sure that smallest increments available are maintained around zero.

Suppose denormalized numbers ARE used.

What is the decimal value of 0b0_0000_001?
0.125 \times 2^{-6}

What is the decimal value of 0b0_0000_111?
0.875 \times 2^{-6}

What is the decimal value of 0b0_0001_000?
1.000 \times 2^{-6}
## Floats: Special cases

<table>
<thead>
<tr>
<th>number class</th>
<th>when it arises</th>
<th>exp field</th>
<th>frac field</th>
</tr>
</thead>
<tbody>
<tr>
<td>+0 / -0</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>+infinity / -infinity</td>
<td>overflow or division by 0</td>
<td>$2^k - 1$</td>
<td>0</td>
</tr>
<tr>
<td>NaN not-a-number</td>
<td>illegal ops. such as $\sqrt{-1}$, inf-inf, inf*0</td>
<td>$2^k - 1$</td>
<td>non-0</td>
</tr>
</tbody>
</table>

*Table: Summary of special cases*
Table of contents

Announcements
  Quizzes and programming assignments
  Reading assignments

Floats: Understanding its design
  Deep understanding 1: Why is exp field encoded using bias?
  Deep understanding 2: Why have denormalized numbers?
  Deep understanding 3: Why is bias chosen to be $2^{k-1} - 1$?

Floats: Properties
  Floating point multiplication
  Properties of floating point

Computer organization: A primer

Assembly code: Human readable representation of machine code

Instruction set architectures

swap.s: Assembly implementation of function that swaps memory contents

Data size and IA32, x86, and x86-64 registers
How to multiply scientific notation?

Recall: $\log(x \times y) = \log(x) + \log(y)$
Floating point multiplication

**FP Multiplication**

- $(-1)^{s_1} M_1 \times 2^{E_1} \times (-1)^{s_2} M_2 \times 2^{E_2}$
- Exact Result: $(-1)^s M \times 2^E$
  - Sign $s$: $s_1 \ ^\wedge s_2$
  - Significand $M$: $M_1 \times M_2$
  - Exponent $E$: $E_1 + E_2$

**Fixing**
- If $M \geq 2$, shift $M$ right, increment $E$
- If $E$ out of range, overflow
- Round $M$ to fit frac precision

**Implementation**
- Biggest chore is multiplying significands
### Mathematical Properties of FP Add

- **Compare to those of Abelian Group**
  - Closed under addition?  
    - Yes  
    - But may generate infinity or NaN  
  - Commutative?  
    - Yes  
  - Associative?  
    - No  
    - Overflow and inexactness of rounding  
      - \((3.14+1e10)-1e10 = 0, 3.14+(1e10-1e10) = 3.14\)  
  - 0 is additive identity?  
    - Yes  
  - Every element has additive inverse?  
    - Yes  
    - Yes, except for infinities & NaNs  
  - Monotonicity  
    - \(a \geq b \Rightarrow a+c \geq b+c?\)  
      - Almost  
      - Except for infinities & NaNs
Mathematical Properties of FP Mult

- **Compare to Commutative Ring**
  - Closed under multiplication? Yes
  - But may generate infinity or NaN
  - Multiplication Commutative? Yes
  - Multiplication is Associative? No
    - Possibility of overflow, inexactness of rounding
      - Ex: \((1e20 * 1e20) * 1e-20 = \text{inf}, 1e20 * (1e20 * 1e-20) = 1e20\)
  - 1 is multiplicative identity? Yes
  - Multiplication distributes over addition? No
    - Possibility of overflow, inexactness of rounding
      - \(1e20 * (1e20 - 1e20) = 0.0, 1e20 * 1e20 - 1e20 * 1e20 = \text{NaN}\)

- **Monotonicity**
  - \(a \geq b \& c \geq 0 \Rightarrow a \cdot c \geq b \cdot c?\)
    - Except for infinities & NaNs
    - Almost
Table of contents

Announcements
  Quizzes and programming assignments
  Reading assignments

Floats: Understanding its design
  Deep understanding 1: Why is exp field encoded using bias?
  Deep understanding 2: Why have denormalized numbers?
  Deep understanding 3: Why is bias chosen to be $2^{k-1} - 1$?

Floats: Properties
  Floating point multiplication
  Properties of floating point

Computer organization: A primer

Assembly code: Human readable representation of machine code

Instruction set architectures

swap.s: Assembly implementation of function that swaps memory contents

Data size and IA32, x86, and x86-64 registers
Computer organization
Layer cake: remember the first day of class, we discussed what are parts of a computer?

- Society
- Human beings
- Applications
- Algorithms
- High-level programming languages
- Interpreters
- Low-level programming languages
- Compilers
- Architectures
- Microarchitectures
- Sequential/combinational logic
- Transistors
- Semiconductors
- Materials science
- Physics
- Mathematics
Stored program computer

Stored program: Instructions reside in memory, loaded as needed.

von Neumann architecture: Data and instructions share same connection to memory.

Programmer-Visible State
- PC: Program counter
  - Address of next instruction
  - Called “RIP” (x86-64)
- Register file
  - Heavily used program data
- Condition codes
  - Store status information about most recent arithmetic or logical operation
  - Used for conditional branching

Memory
- Byte addressable array
- Code and user data
- Stack to support procedures
# Memory hierarchy

<table>
<thead>
<tr>
<th></th>
<th>Capacity</th>
<th>Access speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tape</td>
<td>250Pb</td>
<td></td>
</tr>
<tr>
<td>Hard drives</td>
<td>16TB</td>
<td>2Mb/s</td>
</tr>
<tr>
<td>Solid state drives</td>
<td>4TB</td>
<td>2Gb/s</td>
</tr>
<tr>
<td>DRAM</td>
<td>8Gb - 1Tb+</td>
<td>8Gb/s</td>
</tr>
<tr>
<td>Last-level cache</td>
<td>64Mb</td>
<td></td>
</tr>
<tr>
<td>Level-1 cache</td>
<td>1Mb</td>
<td></td>
</tr>
<tr>
<td>Registers</td>
<td>1Kb</td>
<td></td>
</tr>
</tbody>
</table>

- Registers (.25ns; 4GHz => .25e-9s)
Table of contents

Announcements
  Quizzes and programming assignments
  Reading assignments

Floats: Understanding its design
  Deep understanding 1: Why is exp field encoded using bias?
  Deep understanding 2: Why have denormalized numbers?
  Deep understanding 3: Why is bias chosen to be $2^{k-1} - 1$?

Floats: Properties
  Floating point multiplication
  Properties of floating point

Computer organization: A primer

Assembly code: Human readable representation of machine code

Instruction set architectures

swap.s: Assembly implementation of function that swaps memory contents

Data size and IA32, x86, and x86-64 registers
Unraveling the compilation chain

**Turning C into Object Code**

- Code in files `p1.c` `p2.c`
- Compile with command: `gcc -Og p1.c p2.c -o p`
  - Use basic optimizations (`-Og`) [New to recent versions of GCC]
  - Put resulting binary in file `p`

![Diagram showing the stages of compilation](image-url)

- **Text**
  - C program (`p1.c` `p2.c`)
  - Compiler (`gcc -Og`)
- **Text**
  - Asm program (`p1.s` `p2.s`)
  - Assembler (`gcc` or `as`)
- **Binary**
  - Object program (`p1.o` `p2.o`)
  - Linker (`gcc` or `ld`)
- **Binary**
  - Executable program (`p`)
  - Static libraries (`.a`)
Assembly

Human readable machine code

- Very limited
- Not much control flow
- Any more complex functionality is built up
- for loops, while loops, turn into assembly sequence

Choice of what assembly to experiment with

- MIPS
- ARM
- x86 / x86-64 (not ideal for teaching, but it allows us to experiment on ilab)
Assembly instructions

Instructions for the microarchitecture

- Binary streams that tell an electronic circuit what to do
- Fetch, decode, execute, memory, writeback
A preview of microarchitecture

Figure: Stages of compilation. Image credit Wikimedia
Table of contents

Announcements
  Quizzes and programming assignments
  Reading assignments

Floats: Understanding its design
  Deep understanding 1: Why is exp field encoded using bias?
  Deep understanding 2: Why have denormalized numbers?
  Deep understanding 3: Why is bias chosen to be $2^{k-1} - 1$?

Floats: Properties
  Floating point multiplication
  Properties of floating point

Computer organization: A primer

Assembly code: Human readable representation of machine code

Instruction set architectures

swap.s: Assembly implementation of function that swaps memory contents

Data size and IA32, x86, and x86-64 registers
Why are instruction set architectures important

Interface between computer science and electrical and computer engineering

- Software is varied, changes
- Hardware is standardized, static

Computer architect Fred Brooks and the IBM 360

- IBM was selling computers with different capacities,
- Compile once, and can run software on all IBM machines.
- Backward compatibility.
- An influential idea.
CISC vs. RISC

Complex instruction set computer

- Intel and AMD
- Have an extensive and complex set of instructions
- For example: x86’s extensions: x87, IA-32, x86-64, MMX, 3DNow!, SSE, SSE2, SSE3, SSSE3, SSE4, SSE4.2, SSE5, AES-NI, CLMUL, RDRAND, SHA, MPX, SGX, XOP, F16C, ADX, BMI, FMA, AVX, AVX2, AVX512, VT-x, VT-d, AMD-V, AMD-Vi, TSX, ASF
- Can license Intel’s compilers to extract performance
- Secret: inside the processor, they break it down to more elementary instructions
CISC vs. RISC

Reduced instruction set computer

- MIPS, ARM, RISC-V (can find Patterson and Hennessy Computer Organization and Design textbook in each of these versions), and PowerPC
- Have a relatively simple set of instructions
- For example: ARM’s extensions: SVE;SVE2;TME; All mandatory: Thumb-2, Neon, VFPv4-D16, VFPv4
- Obsolete: Jazelle
- ARM: smartphones, Apple ARM M1 Mac
Into the future: Post-ISA world

Post-ISA world

- Increasingly, the CPU is not the only character
- It orchestrates among many pieces of hardware
- Smartphone die shot
- GPU, TPU, FPGA, ASIC

Figure: Apple A13 (2019 Apple iPhone 11 CPU). Image credit AnandTech
Table of contents

Announcements
   Quizzes and programming assignments
   Reading assignments

Floats: Understanding its design
   Deep understanding 1: Why is exp field encoded using bias?
   Deep understanding 2: Why have denormalized numbers?
   Deep understanding 3: Why is bias chosen to be $2^{k-1} - 1$?

Floats: Properties
   Floating point multiplication
   Properties of floating point

Computer organization: A primer

Assembly code: Human readable representation of machine code

Instruction set architectures

swap.s: Assembly implementation of function that swaps memory contents

Data size and IA32, x86, and x86-64 registers
Unraveling the compilation chain

## Turning C into Object Code

- Code in files `p1.c` `p2.c`
- Compile with command: `gcc -Og p1.c p2.c -o p`
  - Use basic optimizations (`-Og`) [New to recent versions of GCC]
  - Put resulting binary in file `p`

```
method
---

C program (p1.c p2.c)

Compiler (gcc -Og -S)

Asm program (p1.s p2.s)

Assembler (gcc or as)

Object program (p1.o p2.o)

Static libraries (.a)

Linker (gcc or ld)

Executable program (p)
```

Let’s go to CS:APP textbook lecture slides (05-machine-basics.pdf) slide 28
Data movement instructions

Does unsigned / signed matter?

1. void swap_uc ( unsigned char*a, unsigned char*b );
2. void swap_sc ( signed char*a, signed char*b );

Swapping different data sizes

1. void swap_c ( char*a, char*b );
2. void swap_s ( short*a, short*b );
3. void swap_i ( int*a, int*b );
4. void swap_l ( long*a, long*b );
Table of contents

Announcements
   Quizzes and programming assignments
   Reading assignments

Floats: Understanding its design
   Deep understanding 1: Why is exp field encoded using bias?
   Deep understanding 2: Why have denormalized numbers?
   Deep understanding 3: Why is bias chosen to be $2^{k-1} - 1$?

Floats: Properties
   Floating point multiplication
   Properties of floating point

Computer organization: A primer

Assembly code: Human readable representation of machine code

Instruction set architectures

swap.s: Assembly implementation of function that swaps memory contents

Data size and IA32, x86, and x86-64 registers
Assembly syntax
Instruction Source, Dest

\[
\text{swap}_l: \\
\text{movq} \ (\%rhi), \ %rax \\
\text{movq} \ (\%rdi), \ %rdx \\
\text{movq} \ %rdx, \ (\%rhi) \\
\text{movq} \ %rax, \ (\%rdi) \\
\text{return}
\]

<table>
<thead>
<tr>
<th>swap</th>
<th>data type</th>
<th>mov operation</th>
<th>registers</th>
</tr>
</thead>
<tbody>
<tr>
<td>swap_uc</td>
<td>unsigned char</td>
<td>movb (move byte)</td>
<td>%al, %dl</td>
</tr>
<tr>
<td>swap_sc</td>
<td>signed char</td>
<td>movb (move byte)</td>
<td>%al, %dl</td>
</tr>
<tr>
<td>swap_c</td>
<td>char</td>
<td>movb (move byte)</td>
<td>%al, %dl</td>
</tr>
<tr>
<td>swap_s</td>
<td>short</td>
<td>movw (move word)</td>
<td>%ax, %dx</td>
</tr>
<tr>
<td>swap_i</td>
<td>int</td>
<td>movl</td>
<td>%eax, %edx</td>
</tr>
<tr>
<td>swap_l</td>
<td>long</td>
<td>movq</td>
<td>%rax, %rdx</td>
</tr>
</tbody>
</table>
Data size and IA32, x86, and x86-64 registers

Some History: IA32 Registers

<table>
<thead>
<tr>
<th>data type</th>
<th>registers</th>
</tr>
</thead>
<tbody>
<tr>
<td>char</td>
<td>%al, %dl</td>
</tr>
<tr>
<td>short</td>
<td>%ax, %dx</td>
</tr>
<tr>
<td>int</td>
<td>%eax, %edx</td>
</tr>
<tr>
<td>long</td>
<td>%rax, %rdx</td>
</tr>
</tbody>
</table>

Note the backward compatibility.
Data size and IA32, x86, and x86-64 registers

<table>
<thead>
<tr>
<th>data type</th>
<th>registers</th>
</tr>
</thead>
<tbody>
<tr>
<td>char</td>
<td>%al, %dl</td>
</tr>
<tr>
<td>short</td>
<td>%ax, %dx</td>
</tr>
<tr>
<td>int</td>
<td>%eax, %edx</td>
</tr>
<tr>
<td>long</td>
<td>%rax, %rdx</td>
</tr>
</tbody>
</table>

Note the backward compatibility.

- Can reference low-order 4 bytes (also low-order 1 & 2 bytes)
Data size and IA32, x86, and x86-64 registers

<table>
<thead>
<tr>
<th>ZMM0</th>
<th>YMM0</th>
<th>XMM0</th>
<th>ZMM1</th>
<th>YMM1</th>
<th>XMM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZMM2</td>
<td>YMM2</td>
<td>XMM2</td>
<td>ZMM3</td>
<td>YMM3</td>
<td>XMM3</td>
</tr>
<tr>
<td>ZMM4</td>
<td>YMM4</td>
<td>XMM4</td>
<td>ZMM5</td>
<td>YMM5</td>
<td>XMM5</td>
</tr>
<tr>
<td>ZMM6</td>
<td>YMM6</td>
<td>XMM6</td>
<td>ZMM7</td>
<td>YMM7</td>
<td>XMM7</td>
</tr>
<tr>
<td>ZMM8</td>
<td>YMM8</td>
<td>XMM8</td>
<td>ZMM9</td>
<td>YMM9</td>
<td>XMM9</td>
</tr>
<tr>
<td>ZMM10</td>
<td>YMM10</td>
<td>XMM10</td>
<td>ZMM11</td>
<td>YMM11</td>
<td>XMM11</td>
</tr>
<tr>
<td>ZMM12</td>
<td>YMM12</td>
<td>XMM12</td>
<td>ZMM13</td>
<td>YMM13</td>
<td>XMM13</td>
</tr>
<tr>
<td>ZMM14</td>
<td>YMM14</td>
<td>XMM14</td>
<td>ZMM15</td>
<td>YMM15</td>
<td>XMM15</td>
</tr>
<tr>
<td>ZMM16</td>
<td>ZMM17</td>
<td>ZMM10</td>
<td>ZMM20</td>
<td>ZMM21</td>
<td>ZMM22</td>
</tr>
<tr>
<td>ZMM24</td>
<td>ZMM25</td>
<td>ZMM26</td>
<td>ZMM27</td>
<td>ZMM28</td>
<td>ZMM29</td>
</tr>
</tbody>
</table>

ST(0) MM0 MM1 MM2 MM3 MM4 MM5 MM6 MM7 MM8 MM9 MM10 MM11 MM12 MM13 MM14 MM15 MM16 MM17 MM18 MM19 MM20 MM21 MM22 MM23 MM24 MM25 MM26 MM27 MM28 MM29 MM30 MM31


ST(17) MM17 MM18 MM19 MM20 MM21 MM22 MM23 MM24 MM25 MM26 MM27 MM28 MM29 MM30 MM31

ST(18) MM18 MM19 MM20 MM21 MM22 MM23 MM24 MM25 MM26 MM27 MM28 MM29 MM30 MM31

ST(19) MM19 MM20 MM21 MM22 MM23 MM24 MM25 MM26 MM27 MM28 MM29 MM30 MM31

ST(20) MM20 MM21 MM22 MM23 MM24 MM25 MM26 MM27 MM28 MM29 MM30 MM31

ST(21) MM21 MM22 MM23 MM24 MM25 MM26 MM27 MM28 MM29 MM30 MM31

ST(22) MM22 MM23 MM24 MM25 MM26 MM27 MM28 MM29 MM30 MM31

ST(23) MM23 MM24 MM25 MM26 MM27 MM28 MM29 MM30 MM31

ST(24) MM24 MM25 MM26 MM27 MM28 MM29 MM30 MM31

ST(25) MM25 MM26 MM27 MM28 MM29 MM30 MM31

ST(26) MM26 MM27 MM28 MM29 MM30 MM31

ST(27) MM27 MM28 MM29 MM30 MM31

ST(28) MM28 MM29 MM30 MM31

ST(29) MM29 MM30 MM31

ST(30) MM30 MM31

Figure: x86-64 with SIMD extensions registers. Image credit: https://commons.wikimedia.org/wiki/File:Table_of_x86_Registers_svg.svg