#### **CMSC 313 Lecture 03**

- **•Multiple-byte data**
	- **big-endian vs little-endian**
	- **sign extension**
- **•Multiplication and division**
- **• Floating point formats**
- **•Character Codes**

For the following questions, *show all of your work*. It is not sufficient to provide the answers.

Exercise 1. Convert the following numbers.

- a.  $62347_{10}$  to unsigned binary
- b. 8DF $6_{16}$  to base 2
- c.  $41.375_{10}$  to base 4
- d. 10011101.0101<sub>2</sub> to base 10

Exercise 2. Convert each of the following numbers to 8-bit signed magnitude, 8-bit one's complement, 8-bit two's complement and 8-bit excess 128 formats.

- a.  $(-122)_{10}$
- b.  $(-31)_{10}$
- c.  $(-16)_{10}$
- d.  $127_{10}$

Exercise 3. Find the decimal equivalents for the following 8-bit two's complement numbers.

- a. 1000 0001
- b. 0111 1011
- c. 1111 0001
- d. 0010 1010

Exercise 4. Perform two's complement addition on the following pairs of numbers. In each case, indicate whether an overflow has occured.

- a.  $1110\ 1011 + 0110\ 1001$
- b. 1110 1011 + 1111 1111
- c.  $1000\ 1100 + 1100\ 0001$
- d. 0111  $1001 + 0000$  1001

#### **Common Sizes for Data Types**

- **A byte is composed of 8 bits. Two nibbles make up a byte.**
- **Halfwords, words, doublewords, and quadwords are composed of bytes as shown below:**



Principles of Computer Architecture by M. Murdocca and V. Heuring **Computer Architecture by M. Murdocca** and V. Heuring **Computer Architecture by M. Murdocca** and V. Heuring

#### **Big-Endian and Little-Endian Formats**

- **In a byte-addressable machine, the smallest datum that can be referenced in memory is the byte. Multi-byte words are stored as a sequence of bytes, in which the address of the multi-byte word is the same as the byte of the word that has the lowest address.**
- **When multi-byte words are used, two choices for the order in which the bytes are stored in memory are: most significant byte at lowest address, referred to as big-endian, or least significant byte stored at lowest address, referred to as little-endian.**



Word address is *x* for both big-endian and little-endian formats.

Principles of Computer Architecture by M. Murdocca and V. Heuring **Computer Architecture by M. Murdocca** and V. Heuring **Computer Architecture by M. Murdocca** and V. Heuring

#### **Two's Complement Sign Extension**



**•Why does sign extension work?**

\n- $$
-x
$$
 is represented as  $2^8 - x$  in 8-bit
\n- $-x$  is represented as  $2^{16} - x$  in 16-bit
\n- $2^8 - x + ??? = 2^{16} - x$
\n- $??? = 2^{16} - 2^8$
\n- $1\ 0000\ 0000\ 0000\ 0000 = 65536$
\n

 **1111 1111 0000 0000 = 65280**

UMBC, CMSC313, Richard Chang <chang@umbc.edu>

## **Multiplication Example**

**• Multiplication of two 4-bit unsigned binary integers produces an 8-bit result.**



**• Multiplication of two 4-bit signed binary integers produces only a 7-bit result (each operand reduces to a sign bit and a 3-bit magnitude for each operand, producing a sign-bit and a 6-bit result).**

# **Multiplication of Signed Integers**

- **Sign extension to the target word size is needed for the negative operand(s).**
- **A target word size of 8 bits is used here for two 4-bit signed operands, but only a 7-bit target word size is needed for the result.**



#### **Example of Base 2 Division**

- $(7/3 = 2)_{10}$  with a remainder R of 1.
- Equivalently,  $(0111/ 11 = 10)$ <sub>2</sub> with a remainder R of 1.

#### 1 1 0111 0010 R 1 1 1 0 1

Principles of Computer Architecture by M. Murdocca and V. Heuring **Computer Architecture by M. Murdocca and V. Heuring** 

#### **Base 10 Floating Point Numbers**

- **Floating point numbers allow very large and very small numbers to be represented using only a few digits, at the expense of precision. The precision is primarily determined by the number of digits in the fraction (or significand, which has integer and fractional parts), and the range is primarily determined by the number of digits in the exponent.**
- **Example (+6.023** × **1023):**



## **Normalization**

**• The base 10 number 254 can be represented in floating point form** as  $254 \times 10^0$ , or equivalently as:

```
25.4 \times 10^1, or
2.54 \times 10^2, or
.254 × 103, or
```
**.0254** × **104, or**

**infinitely many other ways, which creates problems when making comparisons, with so many representations of the same number.**

- **Floating point numbers are usually normalized, in which the radix point is located in only one possible position for a given number.**
- **Usually, but not always, the normalized representation places the radix point immediately to the left of the leftmost, nonzero digit in** the fraction, as in:  $.254 \times 10^3$ .

**2-20 Chapter 2: Data Representation**

# **Floating Point Example**

- **Represent .254** × **103 in a normalized base 8 floating point format with a sign bit, followed by a 3-bit excess 4 exponent, followed by four base 8 digits.**
- **Step #1: Convert to the target base.**

**.254**  $\times$  10<sup>3</sup> = 254<sub>10</sub>. Using the remainder method, we find that 254<sub>10</sub>  $= 376 \times 8^{0}$ :

```
254/8 = 31 R 6
```

```
31/8 = 3 R 7
```

```
3/8 = 0 R 3
```
- **Step #2: Normalize: 376** × **80 = .376** × **83.**
- **Step #3: Fill in the bit fields, with a positive sign (sign bit = 0), an exponent of 3 + 4 = 7 (excess 4), and 4-digit fraction = .3760:**

#### **0 111 . 011 111 110 000**

#### **Error, Range, and Precision**

- **In the previous example, we have the base b = 8, the number of significant digits (not bits!) in the fraction s = 4, the largest exponent value (not bit pattern) M = 3, and the smallest exponent value**  $m = -4$ .
- **In the previous example, there is no explicit representation of 0, but there needs to be a special bit pattern reserved for 0 otherwise there would be no way to represent 0 without violating the normalization rule. We will assume a bit pattern of 0 000 000 000 000 000 represents 0.**
- **Using b, s, M, and m, we would like to characterize this floating point representation in terms of the largest positive representable number, the smallest (nonzero) positive representable number, the smallest gap between two successive numbers, the largest gap between two successive numbers, and the total number of numbers that can be represented.**

#### **Error, Range, and Precision (cont')**

- Largest representable number:  $b^M \times (1 b^{-s}) = 8^3 \times (1 8^{-4})$
- Smallest representable number:  $b^m \times b^{-1} = 8^{-4-1} = 8^{-5}$

• Largest gap: 
$$
b^M \times b^{-s} = 8^{3-4} = 8^{-1}
$$

• Smallest gap: 
$$
b^m \times b^{-s} = 8^{-4-4} = 8^{-8}
$$

# **Error, Range, and Precision (cont')**



**• Number of representable numbers: There are 5 components: (A) sign bit; for each number except 0 for this case, there is both a positive and negative version; (B) (M - m) + 1 exponents; (C) b - 1 values for the first digit (0 is disallowed for the first normalized digit); (D) b<sup>s</sup>-1 values for each of the s-1 remaining digits, plus (E) a special representation for 0. For this example, the 5 components result in: 2 × ((3 - 4) + 1) × (8 - 1) × 8<sup>4-1</sup> + 1 numbers that can be represented. Notice this number must be no greater than the number of possible bit patterns that can be generated, which is 216.**

Principles of Computer Architecture by M. Murdocca and V. Heuring **Computer Architecture by M. Murdocca and V. Heuring** 

#### **Example Floating Point Format**



- **Smallest number is 1/8**
- **Largest number is 7/4**
- **Smallest gap is 1/32**
- **Largest gap is 1/4**
- **Number of representable numbers is 33.**

## **Gap Size Follows Exponent Size**

- **The relative error is approximately the same for all numbers.**
- **If we take the ratio of a large gap to a large number, and compare that to the ratio of a small gap to a small number, then the ratios are the same:**



### **Conversion Example**

- **• Example: Convert (9.375** × **10-2) 10 to base 2 scientific notation**
- **Start by converting from base 10 floating point to base 10 fixed point by moving the decimal point two positions to the left, which corresponds to the -2 exponent: .09375.**
- **Next, convert from base 10 fixed point to base 2 fixed point:**



- Thus, (.09375)<sub>10</sub> = (.00011)<sub>2</sub>.
- **Finally, convert to normalized base 2 floating point:**

$$
.00011=.00011\times 2^0=1.1\times 2^{-4}
$$

#### **IEEE-754 32-bit Floating Point Format**

- **• sign bit, 8-bit exponent, 23-bit mantissa**
- **• normalized as 1.xxxxx**
- **• leading 1 is hidden**
- **•8-bit exponent in excess 127 format**
	- **NOT excess 128**
	- **0000 0000 and 1111 1111 are reserved**
- **•+0 and -0 is zero exponent and zero mantissa**
- **• 1111 1111 exponent and zero mantissa is infinity**

**2-27 Chapter 2: Data Representation**

# **IEEE-754 Floating Point Formats**



Principles of Computer Architecture by M. Murdocca and V. Heuring **Computer Architecture by M. Murdocca** and V. Heuring

#### **IEEE-754 Examples**

#### **Value**

#### **Bit Pattern**



Principles of Computer Architecture by M. Murdocca and V. Heuring **Computer Architecture by M. Murdocca** and V. Heuring

#### **IEEE-754 Conversion Example**

- Represent -12.625<sub>10</sub> in single precision IEEE-754 format.
- Step #1: Convert to target base.  $-12.625_{10} = -1100.101_{2}$
- Step #2: Normalize.  $-1100.101$ <sub>2</sub> =  $-1.100101$ <sub>2</sub>  $\times$  2<sup>3</sup>
- **Step #3: Fill in bit fields. Sign is negative, so sign bit is 1. Exponent is in excess 127 (not excess 128!), so exponent is represented as the unsigned integer 3 + 127 = 130. Leading 1 of significand is hidden, so final bit pattern is:**

#### **1 1000 0010 . 1001 0100 0000 0000 0000 000**

# **Floating Point Arithmetic**

- **Floating point arithmetic differs from integer arithmetic in that exponents must be handled as well as the magnitudes of the operands.**
- **The exponents of the operands must be made equal for addition and subtraction. The fractions are then added or subtracted as appropriate, and the result is normalized.**
- **Ex: Perform the floating point operation: (.101** × **23 + .111** × **24) 2**
- **Start by adjusting the smaller exponent to be equal to the larger exponent, and adjust the fraction accordingly. Thus we have .101**  $\times$  2<sup>3</sup> = .010  $\times$  2<sup>4</sup>, losing .001  $\times$  2<sup>3</sup> of precision in the process.
- **The resulting sum is (.010 + .111)** × **24 = 1.001** × **24 = .1001** × **25, and rounding to three significant digits, .100** × **25, and we have lost another 0.001**  $\times$  2<sup>4</sup> in the rounding process.

# **Floating Point Multiplication/Division**

- **Floating point multiplication/division are performed in a manner similar to floating point addition/subtraction, except that the sign, exponent, and fraction of the result can be computed separately.**
- **Like/unlike signs produce positive/negative results, respectively. Exponent of result is obtained by adding exponents for multiplication, or by subtracting exponents for division. Fractions are multiplied or divided according to the operation, and then normalized.**
- Ex: Perform the floating point operation: (+.110  $\times$  2<sup>5</sup>) / (+.100  $\times$  2<sup>4</sup>)<sub>2</sub>
- **The source operand signs are the same, which means that the result will have a positive sign. We subtract exponents for division, and so the exponent of the result is 5 – 4 = 1.**
- **We divide fractions, producing the result: 110/100 = 1.10.**
- **Putting it all together, the result of dividing (+.110** × **25) by (+.100** × **24) produces (+1.10** × **21). After normalization, the final result is (+.110** × **22).**

#### **ASCII Character Code**

- **ASCII is a 7-bit code, commonly stored in 8-bit bytes.**
- "A" is at 41<sub>16</sub>. To convert **upper case letters to lower case letters, add 20<sub>16</sub>. Thus "a" is at 41<sub>16</sub> +**  $20_{16} = 61_{16}$ .
- **The character "5" at position 3516 is different than the number 5. To convert character-numbers into number-numbers, sub**tract  $30_{16}$ :  $35_{16}$  -  $30_{16}$  = 5.



**2-32 Chapter 2: Data Representation**



Principles of Computer Architecture by M. Murdocca and V. Heuring **Computer Architecture by M. Murdocca** and V. Heuring



**• Unicode is a 16 bit code.**



#### **Next Time**

- **•Basic Architecture of Intel Pentium Chip**
- **• "Hello World" in Linux Assembly**
- **•Addressing Modes**