CPE 315 Spring 2016

Homework 1 – Floating Point Exercises

Please feel free to make use of the IEEE 754 Converter reference on Polylearn, http://www.h-schmidt.net/FloatConverter/IEEE754.html

1. Extract the Sign, Exponent (less bias), and Fraction values (retrieving the "hidden 1") for the following floating point numbers, shown as decimal values. Write the resulting exponent and fraction as 2's complement numbers.

a) 0.0
b) 6.5
c) 4.0
d) -4.0

2. What is the maximum value, in decimal notation, of a single-precision and double-precision IEEE 754 floating point number? How is this determined mathematically?

3. When adding the following floating point numbers: 0.375 and 28.0:

a) Show each number in 2's complement floating point representation (exponent, fraction), where the fraction has the binary point immediately to the right of the high-order bit.

b) What must be done to place them on a common scale?

4. Normalize the following 2's complement floating point numbers, showing the updated exponent and fraction values:
a) Exponent of 4, and a fraction of 0x00200000
b) Exponent of -3 and fraction of 0xFFC00000

5. In single- and double-precision IEEE 754 numbers, how many actual bits of fraction are there after including the "hidden 1"?

6. A current-day graphics company (Nvidia) has standardized a "half precision" format for floating point, consisting of a sign, a 5-bit fraction, and a 10-bit fraction, using a bias of 15 (a total of 16 bits). What would be the decimal value of the number **0xBE00** ? Show the exponent and fraction portions in 2's complement form (as 16-bit numbers).