CS2323: Computer Architecture, Autumn 2025

Homework-3: Floating point arithmetic

Total: 25 marks

Show calculation steps for each.

1. Convert the following decimal numbers into IEEE-754 floating-point format (write the final answer in Hex). Show all steps [6 marks]
   a. -13.25 (single precision)
   b. 0.1 (single precision)
   c. 156.75 (double precision)
   d. -0.0078125 (double precision)

2. Convert the following hexadecimal values into their **decimal equivalents**. Show steps. [6 marks]

   a. 0xC1200000 (single precision)
   b. 0x3F800000 (single precision)
   c. 0xBFF0000000000000 (double precision)
   d. 0x4024000000000000 (double precision)

3. You are given two IEEE-754 single-precision numbers as 32-bit hex values: [4 marks]
   A = 0x41480000 (single-precision)
   B = 0xC0700000 (single-precision)

Perform the addition A + B and write the final answer in IEEE-754 **double** precision format.

4. You are given two IEEE-754 double-precision numbers as 64-bit hex values: [4 marks]
   A = 0x4039000000000000 (double precision)
   B = 0xC008000000000000 (double precision)

Perform the multiplication A x B and write the final answer in IEEE-754 **single** precision format.

5. Identify and explain one number which can be represented in a 32-bit signed integer format, but not in a 32-bit single precision floating point representation. [2 marks]

6. Show one example to prove that addition is not associative for floating point numbers i.e., (a + b) + c ≠ a + (b + c) [3 marks]

Submission instructions:

1. Create a pdf file answering the above questions.
2. The submission should be entirely your work
3. The pdf file should be named YOUR_ROLLNUM.pdf (e.g., CSYYBTECHXXXXX.pdf)
4. Submit the pdf file