# CMPT 295 Assignment 4 Solutions (2%)

1. [5 marks] *Floating-Point Conversion*

   (a) [3 marks]

   - $-255_{10} = -1.1111\ 111_2 \times 2^7$ (exactly). Thus the sign is 1, the exponent is 7 ($\mapsto$ 134 on the bias), and the significand is 1111 111 followed by 16 zeros.

     $$1\ 10000110\ 11111110000000000000000 = \texttt{0xc37f0000}.$$

   - 2.55 has an integer part of 2 ($10_2$). Continual multiplication by 2 yields the repeating binary: $-2.55_{10} = -1.010\overline{0011}_2 \times 2^1$. Since the $25^{\text{th}}$ digit of the significand is 0, the $24^{\text{th}}$ digit will not be rounded up. The exponent will be encoded as 128 on the bias.

     $$1\ 10000000\ 01000110011001100110011 = \texttt{0xc0233333}.$$

   - $1/3$ also yields a repeating binary pattern, but this time a shorter one: $1/3 = 1.\overline{01} \times 2^{-2}$. This time there will be a rounding up to achieve:

     $$0\ 01111101\ 01010101010101010101011 = \texttt{0x3eaaaaab}.$$

   (b) [2 marks]

   - `0x3e970a3d` has a sign bit of 0, an exponent field of 0111 1101, which is 125 unsigned, or $-2$ on the bias, and a significand of $1.00101110000101000111101_2$. Thus the scientific notation is $1.00101110000101000111101_2 \times 2^{-2}$.
     The exact rational fraction represented here is 9898557/33554432 which is 0.2949999869.

   - `0x3f7fffff` $= 1.11111111111111111111111_2 \times 2^{-1}$, the largest number less than 1.0. The fraction would be $\dfrac{2^{24}-1}{2^{24}} = \dfrac{16777215}{16777216} = 0.9999999404$.

2. [6 marks] *Half-Precision Floating-Point*

   (a) [1 mark] $[-6, 7]$. The *bias* would be 7.

   (b) [1 mark] All exponent fields except for 1111 are valid. Thus there are 15/16 of the $2^{16}$ possible values, i.e., 61440 different values, including the two encodings for $\pm 0$.

   (c) [1 mark] There are half positive and half negative values. The median must be $\pm 0$.

   (d) [1 mark] There are $2^{11}$ possible values, in the range $[0, 2^{-6} - 2^{-17}]$. Each pair of values is separated by $2^{-17}$.

   (e) [1 mark] $[2^{-6}, 2^8 - 2^{-4}]$.

   (f) [1 mark] Because the set is ordered, we look for the halfway point between the smallest normalized value (`0x0800`) and the largest (`0x77ff`). Taking the average of the two numbers yields two medians: `0x3fff` and `0x4000`, which represent $2 - 2^{-11}$ and 2, respectively.