CMPT 295 Assignment 5 Solutions (2%)

1. [7 marks] *Floating-Point Integers*

   Just to show the significance, the percentage of the code that applies to each part is included.

   (a) [1 mark] There are $106 \cdot 2^{23}$ positive integers (including $+0$), and an equal number of negative integers. (41.4% of all floating point)

   (b) [1 mark] $2^{24} - 1$ ($= 16777215$).

   (c) [1 mark] $2^{128} - 2^{104}$, and $2^{128} - 2^{105}$.

   (d) [1 mark] The consecutive integers fall in the range $[-2^{24}, 2^{24}]$ and there are $2^{25} + 1$ ($= 33554433$) of them. (0.78% of all floating point)

   (e) [1 mark] $2^{32} = 1.000\ 0000\ 0000\ 0000\ 0000\ 0000 \times 2^{32}$. The nearest neighbours would therefore be $1.000\ 0000\ 0000\ 0000\ 0000\ 0001 \times 2^{32} = 2^{32} + 2^9$ and $1.111\ 1111\ 1111\ 1111\ 1111\ 1111 \times 2^{31} = 2^{32} - 2^8$.

   (f) [2 marks] There are 10×2^{23} positive integers (again including $+0$) that are less than 2^{32}. (2.0% of all floating point, 4.7% of S)

2. [2 marks] *Floating-Point Addition*

   - These are $+1.001\ 0011\ 1000\ 0000\ 0000\ 0000_2 \times 2^8$ and $-1.010\ 1000\ 0000\ 0000\ 0000\ 0000_2 \times 2^5$. To subtract their magnitudes, the latter must be shifted right by 3 places, because the difference in the exponents is 3.

$$
\begin{array}{rll}
  & 1.0010\ 0111\ 0000\ 0000\ 0000\ 000 & \times 2^8 \ (295) \\
  - & 0.0010\ 1010\ 0000\ 0000\ 0000\ 0000\ 00 & \times 2^8 \ (\text{-}42) \\
  \hline
  = & 0.1111\ 1101\ 0000\ 0000\ 0000\ 0000\ 00 & \times 2^8 \ (253)
\end{array}
$$

   Thus we have a positive number (sign bit $= 0$), with a normalized significand of $1.111\ 1101$ and a normalized exponent of 2^7. This encodes as `0x437d0000`.

   - These are $+1.001\ 1001\ 1001\ 1001\ 1001\ 1010_2 \times 2^{-1}$ and $1.100\ 1100\ 1100\ 1100\ 1100\ 1101_2 \times 2^{-2}$. Again, alignment must occur before adding: this time it is shifted by one place.

$$
\begin{array}{rll}
  & 1.0011\ 0011\ 0011\ 0011\ 0011\ 010 & \times 2^{-1} \ (3/5) \\
  + & 0.1100\ 1100\ 1100\ 1100\ 1100\ 1101 & \times 2^{-1} \ (2/5) \\
  \hline
  = & 10.0000\ 0000\ 0000\ 0000\ 0000\ 0001 & \times 2^{-1}
\end{array}
$$

   The resulting significand is too long, so it must be truncated and rounded. Since the 25th significant digit is 0, the result is rounded down. The result is $10_2 \times 2^{-1} = 1 \times 2^0$, or `0x3f800000`.