COM1006/COM1090 Devices and Networks (Autumn)

Tutorial Sheet #3: Computer Arithmetic II

1. Explain why floating point numbers have normalized mantissas.

2. Express the following numbers in IEEE 754 single precision (32-bit) format:

   a) 0.5
   b) -17.375
   c) 0

3. Convert the following IEEE 754 single precision numbers to decimal. You may write down results as an expression (e.g. "5/16" or "2^{-8}"), provided that it cannot be further simplified.

   a) 1 01111101 10000000000000000000000
   b) 0 10000101 00100100000000000000000
   c) 0 00000001 00000000000000000000000

4. Perform the following arithmetic operations on normalised floating point numbers. Normalise the result if necessary and truncate the mantissa to using 6 bits for its fractional part. In which cases is the result accurate?

   a) $1.011001 \times 2^6 + 1.001000 \times 2^4$
   b) $1.010000 \times 2^{-4} + 1.110001 \times 2^{-3}$
   c) $1.001100 \times 2^8 - 1.001000 \times 2^8$
   d) $1.100010 \times 2^{11} + 1.000000 \times 2^3$

5. Explain the following terms in relation to floating point arithmetic:

   biased exponent
   fractional part
   range
   precision
   normalization
   exponent overflow
   exponent underflow
   gradual underflow