Samreen Azam (sa3tnc)

9/23/19

floatingpoint.pdf

- 1. 32-bit floating point number: $-2.453125 \rightarrow$ convert to little endian binary, expressed in hexadecimal:
 - First, the sign is negative so we use a 1
 - Exponent = 1 because $2.453125 / 2^1 = 1.2265625 \rightarrow 1+127 = 128 = 10000000$
 - Mantissa:
 - o 1.2265625 -1 = 0.2265625 = 3/16 + 1/32 + 1/256
 - o This evaluates to 00111010

 - Binary to hexadecimal: starts with 0x, break into groups of 4
 - o 1100 c
 - \circ 0000 0
 - o 0001 1
 - o 1101 d
 - o Each of the 4 groups of 0's are represented by one 0
 - Hexadecimal representation: 0xc01d0000
- 2. Hexadecimal: 0x0040203f → convert to 32-bit floating point number:
 - First convert to binary:
 - o 00 = 0000 0000
 - o 40 = 0100 0000
 - o 20 = 0010 0000
 - o 3 = 0011
 - o F = 1111

 - This tells us that sign is negative and exponent is $11100110 = 230 \rightarrow 230-127 = 103$

 - Add up → 32-bit floating point number is 5.8890395 * 10^-39