Selinie Wang
September 24, 2018
floatingpoint.pdf

Floating Point

Your magic (32 bit) floating point number is 39.0625.
This is the number that needs to be converted to (little endian) binary, and expressed in hexadecimal.

1. CONVERT TO LITTLE ENDIAN BINARY (HEX): 39.0625
    a. *Sign bit* is 0 because it is positive.
    b. *Exponent* is 5; $2^5 = 32$ and 39.0625/32 = 1.2207
       5 + 127 = 132 which is $1000\ 0100_b$
    c. *Mantissa* is 1.220703125 – 1 = 0.220703125 = 113/512.
       113/512 – 64/512 = 49/512 ($1/8 = 1/2^3$)
       49/512 – 32/512 = 17/512 ($1/16 = 1/2^4$)
       17/512 – 16/512 = 1/512 ($1/32 = 1/2^5$)
       1/512 – 1/512 = 0/512 ($1/512 = 1/2^9$)
       0011 1000 1000 0000 0000 000
    d. Binary Representation: 0100 0010 0001 1100 0100 0000 0000 0000
       Little Endian Binary: 0000 0000 0100 0000 0001 1100 0100 0010
       Big Endian: 0x421c4000
       Little Endian: 0x00401c42

Your other magic floating point number is, in hex, 0x00401ec3.
This is the number that needs to be converted to a (32 bit) floating point number.
Note that the hexadecimal printed above is in little-endian format!

1. CONVERT TO FLOATING POINT NUMBER
    a. Big Endian: 0xc31c4000
       Big Endian Binary: 1100 0011 0001 1100 0100 0000 0000 0000
    b. Sign bit is 1 so it is negative.
    c. Exponent is 1000 0110; 134 – 127 = 7
    d. Mantissa is 0011 1000 1000 0000 0000 000
       $(1/2)^3 + (1/2)^4 + (1/2)^5 + (1/2)^9 = 0.220703125$
       0.220703125 + 1 = 1.220703125
    e. $-1.220703125 * 2^7 =$ **-156.25**