Tutorial 6 – Week of Oct. 25th

Question 1) (3.22)
What decimal number does the bit pattern
0×0C000000
represent if it is a floating-point number? Use the IEEE 754 standard.

Solution:

$0\times0C000000 = 0000\ 1100\ 0000\ 0000\ 0000\ 0000\ 0000\ 0000$

$= 0\ 0001\ 1000\ 0000\ 0000\ 0000\ 0000\ 0000\ 000$

sign is positive

$\exp = 0\times18 = 24 - 127 = -103$

there is a hidden 1

mantissa $= 0$

answer $= 1.0 \times 2^{-103}$

Question 2) (3.23)
Write down the binary representation of the decimal number 63.25 assuming the IEEE 754 single precision format.
Solution:

$63.25 \times 10^0 = 111111.01 \times 2^0$

normalize, move binary point five to the left

1.1111101×2^5

sign = positive, $\exp = 127 + 5 = 132$

Final bit pattern: 0 1000 0100 1111 1010 0000 0000 0000 000

= 0100 0010 0111 1101 0000 0000 0000 0000 = 0x427D0000

Question 3) (3.24)
Write down the binary representation of the decimal number 63.25 assuming the IEEE 754 double precision format.

Solution:

$63.25 \times 10^0 = 111111.01 \times 2^0$

normalize, move binary point five to the left

1.1111101×2^5

sign = positive, exp = $1023 + 5 = 1028$

Final bit pattern:

0 100 0000 0100 1111 1010 0000 0000 0000 0000 0000 0000 0000 0000 0000
0000 0000

= 0x404FA00000000000

Question 4) (3.27)
IEEE 754-2008 contains a half precision that is only 16 bits wide. The leftmost bit is still the sign bit, the exponent is 5 bits wide and has a bias of 15, and the mantissa is 10 bits long. A hidden 1 is assumed. Write down the bit pattern to represent -1.5625×10^{-1} assuming a version of this format, which uses an excess-16 format to store the exponent.

Solution:

$-1.5625 \times 10^{-1} = -0.15625 \times 10^0$

$= -0.00101 \times 2^0$

move the binary point three to the right, $= -1.01 \times 2^{-3}$

exponent $= -3 = -3 + 15 = 12$, fraction $= -0.0100000000$

answer: 1011000100000000