Homework 2

(Due date: February 14th @ 11:59 pm)

Presentation and clarity are very important! Show your procedure!

Problem 1 (15 pts)

a) Multiply the following signed fixed-point numbers (6 pts):

10.0111 × 0.110101	10.01111 × 10.1011	0111.111 × 10.011011

b) Get the division result (with $x = 4$ fractional bits) for the following signed fixed-point numbers:

101.1001 ÷ 1.0101	0.10101 ÷ 101.0101	0.101 ÷ 1.10111

Problem 2 (9 pts)

a) We want to represent numbers between -510.5 and 512.7. What is the fixed-point format that requires the fewest number of bits for a resolution better or equal than 0.0015? (3 pts).

b) We want to represent numbers between -127.69 and 125.42. What is the fixed-point format that requires the fewest number of bits for a resolution better or equal than 0.0025? (3 pts).

c) Represent these numbers in Fixed Point Arithmetic (signed numbers). Use the FX format [16 4].

69.25	-117.3125	-128.875

Problem 3 (8 pts)

a) Complete the table for the following fixed-point formats (signed numbers): (3 pts)

Fractional bits	Integer Bits	FX Format	Range	Dynamic Range (dB)	Resolution
12	4				
17	7				

b) Complete the table for these floating point formats (which resemble the IEEE-754 standard). Only consider ordinary numbers.

Exponent bits (E)	Significand		Min	Max	Range of e	Range of significand
	bits (p)	FX Format				
7	8					
8	15					
11	36					

Problem 4 (22 pts)

a) For the given floating-point numbers (displayed as hexadecimals), complete: bits in the fields, significand's FX format (4 pts)

✓ `7BEAD360` **(single – 32 bits)**

sign e+bias

FX format of significand: _____

significand

✓ `7A09D3784D039800` **(double – 64 bits)**

sign e+bias

FX format of significand: _____

significand

b) Calculate the decimal values of the following floating-point numbers represented as hexadecimals. Show your procedure.

Single (32 bits)		**Double (64 bits)**	
✓ 3DE32856 ✓ 003ACBAC	✓ FFCE3720	✓ BADAFC0FEE000000 ✓ FFFECE4710FACADE	✓ 800ABBAF25C00000

PROBLEM 5 (46 PTS)

- Perform the following 32-bit floating point operations. For fixed-point division, use 8 fractional bits. Truncate the result when required. Show your work: how you got the significand and the biased exponent bits of the result. Provide the 32-bit result.

✓ ECE5721A + FF800000 ✓ 40D90000 + C2EAC000	✓ 801A8000 − B3CEC000 ✓ B0AE0000 − 4F4A8000	✓ DECAFFAD × 7F800000 ✓ 0E2CE000 × 8B092000	✓ 800C0000 ÷ 494A0000 ✓ 49744000 ÷ C0C90000