Homework 2

(Due date: February 13th @ 11:59 pm)

Presentation and clarity are very important! Show your procedure!

Problem 1 (20 pts)

a) Compute the result of the additions and subtractions for the following fixed-point numbers (5 pts):

UNSIGNED		SIGNED		
1101.1 + 0.1100101	1.01011 - 0.0011011	1010.01 - 101.01001	10.1100 + 01.01100101	0.0011011 - 1.01011

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

01.1001 × 1.001011	10.0101 × 11.011	1010.001 × 01.100101

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

100.1001 ÷ 1.0101	1.011 ÷ 0.01001	1.01011 ÷ 010.1011

Problem 2 (9 pts)

a) We want to represent numbers between -256 and 255.97. What is the fixed-point format that requires the fewest number of bits for a resolution better or equal than 0.001? (3 pts).

b) We want to represent numbers between -63.42 and 65.69. What is the fixed-point format that requires the fewest number of bits for a resolution better or equal than 0.0015? (3 pts).

c) Represent these numbers in Fixed Point Arithmetic (signed numbers). Use the FX format [16 4]. Truncate (the LSB) and perform Saturation when required.

2048.25	-117.53125	-129.375

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
11	5				
15	9				

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				
8	6					
10	13					
12	32					

PROBLEM 4 (19 PTS)

a) For the given IEEE-754 floating-point numbers (displayed as hexadecimals), complete: bits in the fields (sign, biased exponent, significand) and significand's FX format (4 pts)

✓ **`90DECADE` (single – 32 bits)**

sign e+bias

FX format of significand: _____

significand

✓ **`3DECAFC0FFEE5000` (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)**	
✓ `803AD0BE` ✓ `BEA7BEEF`	✓ `7FCE4710`	✓ `7FFCABFEEBA5ED00` ✓ `000C0FFEEFAD0000`	✓ `ECE4710A96B80C60`

PROBLEM 5 (44 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: <u>how you got the significand and the biased exponent bits of the result.</u> Provide the 32-bit result.

✓ `7F800000 + C512290A` ✓ `C0D90000 + 42EAC000`	✓ `B3BEE000 – 8037C000` ✓ `80123000 – 004E8000`	✓ `5A09C000 × CD080000` ✓ `7CDA0000 × 80200000`	✓ `C9744000 ÷ 81C90000` ✓ `000C0000 ÷ BACA0000`