Homework 2

(Due date: February 1st)

Presentation and clarity are very important! Show your procedure!

PROBLEM 1 (12 PTS)

- Calculate the result of the additions and subtractions for the following fixed-point numbers.

UNSIGNED		SIGNED	
1.1011010 + 0.010101	1.00101 - 0.0000111	10.001 + 1.001101	0.011 - 1.1011101
10.1101 + 1.1001	1100.1 + 0.100101	1001.101 - 111.10001	101.0001 + 1.1001001

PROBLEM 2 (18 PTS)

- Multiply the following signed fixed-point numbers:

10.011 × 0.110101	10.1101 × 01.10001	0111.111 × 10.011011

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

101.1001 ÷ 1.0101	11.011 ÷ 1.10111	0.101010 ÷ 101.0101

PROBLEM 3 (10 PTS)

- We want to represent numbers between -214.9 and 256.7. What is the fixed point format that requires the fewest number of bits for a resolution better or equal than 0.0015? (5 pts).

- Represent these numbers in Fixed Point Arithmetic (signed numbers). Select the minimum number of bits in each case.

-128.625	-231.3125	112.125

PROBLEM 4 (12 PTS)

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

Fractional bits	Integer Bits	FX Format	Range	Dynamic Range (dB)	Resolution
7	5				
12	4				
17	7				

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

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

PROBLEM 5 (16 PTS)

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

Single (32 bits)		Double (64 bits)	
✓ FDEAD360 ✓ 3DE32856	✓ 803ACBAC ✓ 7FCBEEFE	✓ FA09D3784D039800 ✓ DECAFC0FEE000000	✓ 7FFBEEFC0FFEEBEE ✓ 800ABBAF25C00000

PROBLEM 6 (32 PTS)

- Calculate the result (provide the 32-bit result) of the following operations with 32-bit floating point numbers. Truncate the results when required. When doing fixed-point division, use 8 fractional bits. Show your procedure.

✓ 40D90000 + C2EAC000 ✓ CF4A8000 + B0A90000	✓ 801A8000 – B3CEC000 ✓ FF800000 – DECAFF00	✓ FACADE80 × 7F800000 ✓ 8B092000 × 0FACE000	✓ 800C0000 ÷ 494A0000 ✓ 49744000 ÷ C0C90000