Homework 2

(Due date: September 29th)

Presentation and clarity are very important! Show your procedure!

PROBLEM 1 (36 PTS)

- Calculate the 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.

✓ 80123000 + FACE8000	✓ CA09E378 - 80000000	✓ 80000000 × 497424FE	✓ 80000000 ÷ BEEFFACE
✓ 60A10000 + C2F97000	✓ FAD90000 – 09DECADE	✓ 7A09D300 × 7F800000	✓ FF800000 ÷ 48500000
✓ 7F90BEAD + DFEA0C98	✓ FEE32B88 – FF800000	✓ 8B092000 × 0FACE000	✓ 390D3800 ÷ C9600000

PROBLEM 2 (14 PTS)

- Complete the table for the following DFX formats:

DFX format	p_0	p_1	Number of bits of significand	Boundary value	$num0$ range	$num1$ range	Dynamic Range (dB)
8_4_2							
12_6_4							
16_8_6							
24_16_8							

PROBLEM 3 (20 PTS)

- Convert the following signed fixed point numbers in format [16 8] to the dual fixed point format 16_8_3. If more bits are required, you are allowed to use the format 17_8_3.

FX	AB.CE	0C.4F	8B.EE	8F.27	81.BE	81.E4	0A.BB	FA.09
DFX								

PROBLEM 4 (30 PTS)

- Calculate the result of the following operations where the numbers are represented in dual fixed-point arithmetic. Note that the results must be in the same format. Include an overflow bit when necessary.

DFX Format: 8_4_2	Result	overflow		Result	overflow
FA+19			2B+A9		
E2+BB			C0+C2		
FB-90			88+1A		

DFX Format 16_8_4	Result	overflow		Result	overflow
72CD+0A98			8939-09A2		
C200+B8C3			E323-7AA9		
F990-0A32			D001+F170		