

Homework 1

(Due date: September 25th)

Presentation and clarity are very important! Show your procedure!

PROBLEM 1 (10 PTS)

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

UNSIGNED		SIGNED	
0.11010 + 1.0101101	1.00111 - 0.0000111	1.0001 + 1.001001	0.0101 - 1.0101101
10.10101 + 1.1001	100.1 + 0.10101	1000.0101 - 11.010101	101.0101 + 1.0111101

PROBLEM 2 (10 PTS)

- Multiply the following signed fixed-point numbers:

01.001 × 1.001001	10.0001 × 01.01001	1100.001 × 10.010101	0.1101010 × 11.1111011
----------------------	-----------------------	-------------------------	---------------------------

PROBLEM 3 (15 PTS)

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

101.001 ÷ 1.001001	10.011001 ÷ 1.01101	01001.001 ÷ 10.101	0.1101010 ÷ 010.110111
-----------------------	------------------------	-----------------------	---------------------------

PROBLEM 4 (5 PTS)

- We want to represent numbers between -128.7 and 179 . What is the fixed point format that requires the fewest number of bits for a resolution better or equal than 0.0005 ?

PROBLEM 5 (10 PTS)

- Complete the table for the following floating point formats (which resemble the IEEE-754 standard) with 16, 24, 48 bits. Only consider ordinary numbers.

Exponent bits (E)	Significant bits (p)	Min	Max	Range of e	Range of significand
6	9				
7	16				
10	37				

PROBLEM 6 (20 PTS)

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

Single (32 bits)		Double (64 bits)	
✓ F8000378	✓ 800ABBAA	✓ FA09D3784D089B7D	✓ 4974240040490FDB
✓ 80DECADE	✓ FACEB0E8	✓ 80DEADBEE9742400	✓ FA09D37809ABC0DE
✓ FDEAD378	✓ 7FF32B5A	✓ 8009D3787F888800	✓ FF80000009ABC0DE
✓ 3DE38866	✓ ACCEDE78	✓ FA0BEBE80BEEF0A0	✓ DECAFC0FFEE00800

PROBLEM 7 (30 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.

✓ FA000378 + FF800FAD	✓ CA09D378 - 80000000	✓ FA09D300 × 4D080000	✓ 49742000 ÷ 40490000
✓ 7F800FEA + 09ABC0DE	✓ 5A09D378 - 40490FDB	✓ 80000000 × 497424FE	✓ 80000000 ÷ 09ABC0DE
✓ FC09D378 + 7F800000	✓ 7DE32B5A - FF800000	✓ FA09DF00 × 7F800000	✓ FF800000 ÷ 09FE0000
✓ 3DE38866 + 3300D959	✓ FA09D378 - 09ABC0DE	✓ 7A09D300 × 0BEEF000	✓ FA09D300 ÷ 48500000