Solutions - Midterm Exam

(February 20th @ 7:30 pm)

Presentation and clarity are very important! Show your procedure!

PROBLEM 1 (20 PTS)

- Compute the result of the following operations. The operands are signed fixed-point numbers. The result must be a signed fixed-point number. For the division, use $x = 5$ fractional bits.

1.0111 + 1.101001	1.010101 − 1000.0101	01.11111 + 0.10001
10.101 × 1.01101	1.001 × 0.1011	10.1010 ÷ 0.101

```
c8=1 c7=1 c6=1 c5=1 c3=0 c2=0 c1=0 c0=0
  1.0 1 1 1 0 0 +
  1.1 0 1 0 0 1
  -------------
  1.0 0 0 1 0 1
```

```
1 1 1 1.0 1 0 1 0 1 -        c10=1 c9=1 c8=1 c7=1 c6=1 c5=1 c4=1 c3=1 c2=0 c1=0 c0=0
1 0 0 0.0 1 0 1 0 0     ➡    1 1 1 1.0 1 0 1 0 1 +
                             0 1 1 1.1 0 1 1 0 0
                             -------------------
                             0 1 1 1.0 0 0 0 0 1
```

```
c8=0 c7=0 c6=1 c5=1 c4=1 c3=1 c2=1 c1=1 c0=0
  0 0 1.1 1 1 1 1 +
  0 0 0.1 0 0 0 1
  ---------------
  0 1 0.1 0 0 0 0
```

```
10.101 x   ➡   01.011 x   ➡   1 0 0 1 1 x
1.01101        0.10011          1 0 1 1
                              ---------
                              1 0 0 1 1
                            1 0 0 1 1
                          0 0 0 0 0
                        1 0 0 1 1
                      -----------------
                      0 1 1 0 1 0 0 0 1
                             ⬇
                      0.1 1 0 1 0 0 0 1
```

```
1.001 x   ➡   0.111 x   ➡     1 0 1 1 x
0.1011        0.1011            1 1 1
                              -------
                              1 0 1 1
                            1 0 1 1
                          1 0 1 1
                        ---------------
                        0 1 0 0 1 1 0 1
                              ⬇
                        0.1 0 0 1 1 0 1
                              ⬇
                        1.0 1 1 0 0 1 1
```

✓ $\frac{10.1010}{0.101}$: To unsigned (numerator) and then alignment, $a = 3$: $\frac{01.0110}{0.101} = \frac{01.011}{00.101} \equiv \frac{1011}{101}$

```
       001000110
101 ) 101100000
      101
      -----
        01000
          101
          ---
           110
           101
           ---
            10
```

Append $x = 5$ zeros: $\frac{1011\mathbf{00000}}{101}$

Integer Division:

$Q = 1000110, R = 10$

$\rightarrow Qf = 10.00110\ (x = 5)$

Final result (2C): $\frac{10.1010}{0.101} = 2C(010.0011) = 101.1101$

PROBLEM 2 (10 PTS)

- Represent these numbers in Fixed Point Arithmetic (signed numbers). Select the minimum number of bits in each case.
  - ✓ -32.125
    +32.125 = 0100000.001 ⇒ -32.125 = 1011111.111
  - ✓ 17.75
    +17.25 = 010001.11

- Complete the table for the following fixed point formats (signed numbers): (6 pts.)

Integer bits	Fractional Bits	FX Format	Range	Resolution
8	6	[14 6]	$[-2^7, 2^7 - 2^{-6}]$	2^{-6}
6	4	[10 4]	$[-2^5, 2^5 - 2^{-4}]$	2^{-4}

PROBLEM 3 (40 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 4 fractional bits. Show your procedure.

✓ 40D00000 + C2EA0000	✓ 50A90000 – 4F480000	✓ 80200000 × 7AB80000	✓ FB380000 ÷ 48C00000

✓ X = 40D00000 + C2EA0000:

40D00000: 0100 0000 1101 0000 1000 0000 0000 0000

$e + bias = 10000001 = 129 \rightarrow e = 129 - 127 = 2$ $\qquad Significand = 1.101$

40D00000 = 1.101×2^2

C2EA0000: 1100 0010 1110 1010 0000 0000 0000 0000

$e + bias = 10000101 = 133 \rightarrow e = 133 - 127 = 6$ $\qquad Significand = 1.110101$

C2EA0000 = -1.110101×2^6

$$X = 1.101 \times 2^2 - 1.110101 \times 2^6 = +\frac{1.101}{2^4} \times 2^6 - 1.110101 \times 2^6 = (0.0001101 - 1.110101) \times 2^6$$

To subtract these numbers, we first convert to 2C:

$R = 0.0001101 - 01.110101 = 0.0001101 + 10.001011$ (2C addition)

The result in 2C is: $R = 10.0100011$, $-R = 01.1011101$

* *Note that you can also do unsigned subtraction:* $X = -(1.110101 - 0.0001101 -) \times 2^6$

c_9=0	c_8=0	c_7=0	c_6=0	c_5=1	c_4=1	c_3=1	c_2=0	c_1=0	c_0=0	
	0	0.	0	0	0	1	1	0	1	+
	1	0.	0	0	1	0	1	1	0	
	1	0.	0	1	0	0	0	1	1	

For floating point, we need to convert to sign-and-magnitude:

$\Rightarrow R(SM) = -1.1011101$

$X = -1.1011101 \times 2^6$, $e + bias = 6 + 127 = 133 = 10000101$

X = 1100 0010 1101 1101 0000 0000 0000 0000 = C2DD0000

✓ X = 50A90000 – 4F480000:

50A90000: 0101 0000 1010 1001 0000 0000 0000 0000

$e + bias = 10100001 = 161 \rightarrow e = 161 - 127 = 34$ $\qquad Significand = 1.0101001$

50A90000 = 1.0101001×2^{34}

4F480000: 0100 1111 0100 1000 0000 0000 0000 0000

$e + bias = 10011110 = 158 \rightarrow e = 158 - 127 = 31$ $\qquad Significand = 1.1001$

4F480000 = 1.1001×2^{31}

$$X = 1.0101001 \times 2^{34} - 1.1001 \times 2^{31} = 1.0101001 \times 2^{34} - \frac{1.1001}{2^3} \times 2^{34}$$

$X = (1.0101001 - 0.0011001) \times 2^{34}$ (unsigned subtraction)

b_8=0	b_7=0	b_6=0	b_5=1	b_4=0	b_3=0	b_2=0	b_1=0	b_0=0	
	1.	0	1	0	1	0	0	1	–
	0.	0	0	1	1	0	0	1	
	1.	0	0	1	0	0	0	0	

$X = 1.001 \times 2^{34}$, $e + bias = 34 + 127 = 161 = 10100001$

X = 0101 0000 1001 0000 0000 0000 0000 0000 = 50900000

✓ X = 80200000 × 7AB80000:

80200000: 1000 0000 0010 0000 0000 0000 0000 0000

$e + bias = 00000000 = 0 \rightarrow Denormal\ number \rightarrow e = -126$ $\qquad Significand = 0.01$

80200000 = -0.01×2^{-126}

7AB80000: 0111 1010 1011 1000 0000 0000 0000 0000

$e + bias = 11110101 = 245 \rightarrow e = 245 - 127 = 118$ $\qquad Significand = 1.0111$

7AB80000 = 1.0111×2^{118}

$X = (-0.01 \times 2^{-126}) \times (1.0111 \times 2^{118}) = -0.010111 \times 2^{-8} = -1.0111 \times 2^{-10}$

$e + bias = -10 + 127 = 117 = 01110101$

X = 1011 1010 1011 1000 0000 0000 0000 0000 = BAB80000

✓ X = FB380000 ÷ 48C00000:

FB380000: 1111 1011 0011 1000 0000 0000 0000 0000

$e + bias = 11110110 = 246 \rightarrow e = 246 - 127 = 119$ $\qquad Significand = 1.0111$

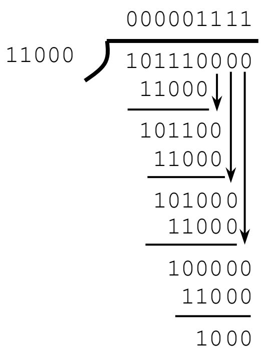
FB380000 = -1.0111×2^{119}

48C00000: 0100 1000 1100 0000 0000 0000 0000 0000

$e + bias = 10010001 = 145 \rightarrow e = 145 - 127 = 18$ $\qquad Significand = 1.1$

48C00000 = 1.1×2^{18}

$$X = -\frac{1.0111 \times 2^{119}}{1.1 \times 2^{18}} = -\frac{1.0111}{1.1} \times 2^{101}$$



Alignment:
$$\frac{1.0111}{1.1} = \frac{1.0111}{1.1000} = \frac{10111}{11000}$$

Append $x = 4$ zeros: $\frac{10111\mathbf{0000}}{11000}$

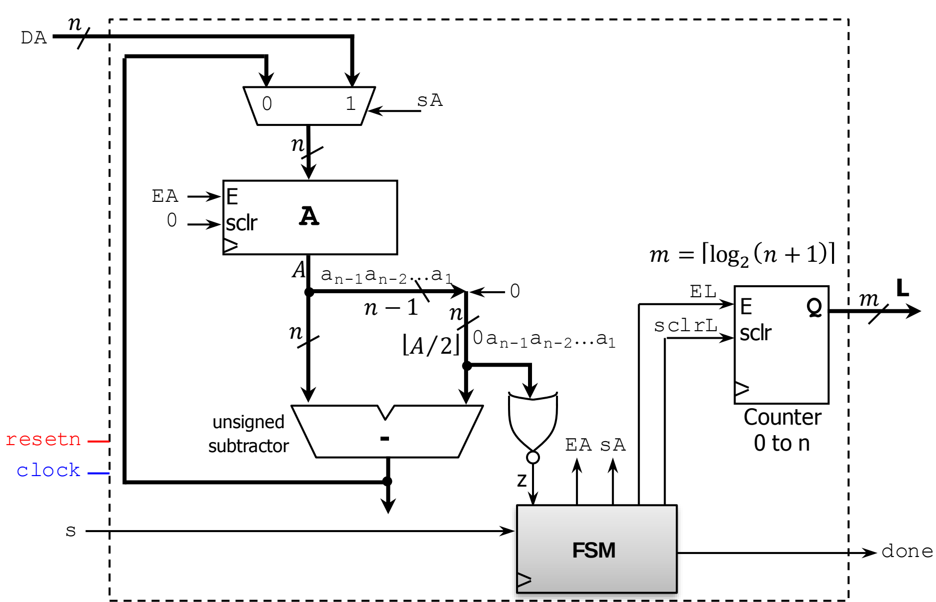
Integer division
$$Q = 1111 \to Qf = 0.1111$$

Thus: $X = -0.1111 \times 2^{101} = -1.111 \times 2^{100}$
$e + bias = 100 + 127 = 227 = 11100011$

X = 1111 0001 1111 0000 0000 0000 0000 0000 = F1F00000

PROBLEM 4 (30 PTS)

- "Integer Base-2 Logarithm": This circuit computes $L = \lceil \log_2 A \rceil$, where A is an unsigned n-bit number stored in register A. The digital system is depicted below: FSM + Datapath. Example: For $n = 8$: if $A = 00110110$, then $L = 0110$.
  - ✓ Note that for $n = 8$, if $A = a_7a_6a_5a_4a_3a_2a_1a_0$, then $\lfloor A/2 \rfloor = 0a_7a_6a_5a_4a_3a_2a_1$. Also: $z = 1$ if $\lfloor A/2 \rfloor = 0$.
  - ✓ m-bit counter: $sclr$. If $E = sclr = 1$, the count is initialized to zero. If $E = 1, sclr = 0$, the count is increased by 1.
  - ✓ Register: If $E = 1$: $sclr = 1 \to$ Clear, $sclr = 0 \to$ Load.
- Sketch the Finite State Machine diagram (in ASM form) given the algorithm (for $n = 8$, $m = 4$). (18 pts.)
  - ✓ The process begins when s is asserted, at this moment we capture DA on register A. Then the process continues by updating A and it is concluded when $A = 1$. The signal $done$ is asserted when we finish computing $\lceil \log_2 A \rceil$.
- Complete the timing diagram where $n = 8, m = 4$. (12 pts.)



ALGORITHM

```
L ← 0
while A ≠ 1
    A ← A - ⌊A/2⌋
    L ←L + 1
end while
```

Finite State Machine: