Solutions - Midterm Exam

(October 13th @ 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.

1.0001 + 1.001001	1000.0101 − 1.010101	01.11111 + 0.00001
01.011 × 1.01101	1.001 × 1.0101	01.01110 ÷ 1.011

$c_8=1, c_7=1, c_6=0, c_5=0, c_4=0, c_3=0, c_2=0, c_1=0, c_0=0$

```
  1 1.0 0 0 1 0 0 +
  1 1.0 0 1 0 0 1
  ---------------
  1 0.0 0 1 1 0 1
```

```
1 0 0 0.0 1 0 1 0 0 -
1 1 1 1.0 1 0 1 0 1
```

$c_{10}=0, c_9=0, c_8=0, c_7=0, c_6=0, c_5=0, c_4=0, c_3=0, c_2=0, c_1=0, c_0=0$

```
1 0 0 0.0 1 0 1 0 0 +
0 0 0 0.1 0 1 0 1 1
--------------------
1 0 0 0.1 1 1 1 1 1
```

$c_8=0, c_7=0, c_6=1, c_5=1, c_4=1, c_3=1, c_2=1, c_1=1, c_0=0$

```
0 0 1.1 1 1 1 1 +
0 0 0.0 0 0 0 1
---------------
0 1 0.0 0 0 0 0
```

```
 01.011 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

                          1.0 0 1 0 1 1 1 1
```

```
 1.001 x       0.111 x          1 0 1 1 x
1.0101        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
```

✓ $\frac{01.01110}{1.011}$: To unsigned (denominator) and then alignment, $a = 4$: $\frac{01.0111}{0.101} = \frac{01.0111}{00.1010} = \frac{010111}{001010} \equiv \frac{10111}{1010}$

```
       0001001001
1010 ) 1011100000
       1010
       ----
          1100
          1010
          ----
           10000
            1010
            ----
             110
```

Append $x = 5$ zeros: $\frac{10111\mathbf{00000}}{1010}$

Integer Division:

$Q = 1001001, R = 110$
$\rightarrow Qf = 10.01001\ (x = 5)$

Final result (2C): $\frac{01.01110}{1.011} = 2C(010.01001) = 101.10111$

PROBLEM 2 (30 PTS)

- Calculate the result (provide the 32-bit result) of the following operations with single floating point numbers. Truncate the results when required. When doing fixed-point division, use $x = 4$ fractional bits.

✓ 42FA8000 + C0E00000	✓ 50DAD000 - D0FAD000	✓ 01800000 × FAB80000	✓ 7B390000 ÷ C8C00000

✓ X = 42FA8000 + C0E00000:

42FA8000: 0100 0010 1111 1010 1000 0000 0000 0000

$e + bias = 10000101 = 133 \rightarrow e = 133 - 127 = 6$ $\quad$ $Mantissa = 1.11110101$

42FA8000 = 1.11110101×2^6

C0E00000: 1100 0000 1110 0000 0000 0000 0000 0000

$e + bias = 10000001 = 129 \rightarrow e = 129 - 127 = 2$ $\quad$ $Mantissa = 1.11$

C0E00000 = -1.11×2^2

$$X = 1.11110101 \times 2^6 - 1.11 \times 2^2 = 1.11110101 \times 2^6 - \frac{1.11}{2^4} \times 2^6$$
$$= (1.11110101 - 0.000111) \times 2^6$$

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

$R = 01.11110101 - 0.000111 = 01.11110101 + 1.111001$

The result in 2C is: $R = 01.11011001$

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

$\Rightarrow R(SM) = +1.11011001$

```
c10=1 c9=1 c8=1 c7=1 c6=1 c5=0 c4=0 c3=1 c2=0 c1=0 c0=0

  0 1.1 1 1 1 0 1 0 1 +
  1 1.1 1 1 0 0 1 0 0
  -------------------
  0 1.1 1 0 1 1 0 0 1
```

$X = 1.11011001 \times 2^6$, $e + bias = 6 + 127 = 133 = 10000101$

X = 0100 0010 1110 1100 1000 0000 0000 0000 = 42EC8000

✓ X = 50DAD000 - D0FAD000:

50DAD000: 0101 0000 1101 1010 1101 0000 0000 0000

$e + bias = 10100001 = 161 \rightarrow e = 161 - 127 = 34$ $\quad$ $Mantissa = 1.10110101101$

10DAD000 = $1.10110101101 \times 2^{34}$

D0FAD000: 1101 0000 1111 1010 1101 0000 0000 0000

$e + bias = 10100001 = 161 \rightarrow e = 161 - 127 = 34$ $\quad$ $Mantissa = 1.11110101101$

90FAD000 = $-1.11110101101 \times 2^{34}$

$X = 1.10110101101 \times 2^{34} + 1.11110101101 \times 2^{34}$ (unsigned addition)

```
c12=1 c11=1 c10=1 c9=1 c8=1 c7=0 c6=1 c5=0 c4=1 c3=1 c2=0 c1=1 c0=0

    1.1 0 1 1 0 1 0 1 1 0 1 +
    1.1 1 1 1 0 1 0 1 1 0 1
  -------------------------
  1 1.1 0 1 0 1 0 1 1 0 1 0
```

$X = 11.1010101101 \times 2^{34} = 1.11010101101 \times 2^{35}$

$e + bias = 35 + 127 = 162 = 10100010$

X = 0101 0001 0110 1010 1101 0000 0000 0000 = 516AD000

✓ X = 01800000 × FAB80000:

01800000: 0000 0001 1000 0000 0000 0000 0000 0000

$e + bias = 00000011 = 3 \rightarrow e = 3 - 127 = -124$ $\quad$ $Mantissa = 1.0$

01800000 = 1.0×2^{-124}

FAB80000: 1111 1010 1011 1000 0000 0000 0000 0000

$e + bias = 11110101 = 245 \rightarrow e = 245 - 127 = 118$ $\quad$ $Mantissa = 1.0111$

FAB80000 = -1.0111×2^{118}

$X = (-1.0 \times 2^{-124}) \times (1.0111 \times 2^{118}) = -1.0111 \times 2^{-6}$

$e + bias = -6 + 127 = 121 = 01111001$

X = 1011 1100 1011 1000 0000 0000 0000 0000 = BCB800000

✓ X = 7B390000 ÷ C8C00000:

7B390000: 0111 1011 0011 1001 0000 0000 0000 0000

$e + bias = 11110110 = 246 \rightarrow e = 246 - 127 = 119$ $\quad$ $Mantissa = 1.0111$

7B390000 = 1.0111001×2^{119}

C8C00000: 1100 1000 1100 0000 0000 0000 0000 0000

$e + bias = 10010001 = 145 \rightarrow e = 145 - 127 = 18$ $\quad$ $Mantissa = 1.1$

C8C00000 = -1.1×2^{18}

$$X = -\frac{1.0111001 \times 2^{119}}{1.1 \times 2^{18}} = -\frac{1.0111001}{1.1} \times 2^{101}$$

```
            000000001111
11000000 ) 101110010000
            11000000
            101100100
             11000000
             101001000
              11000000
              100010000
               11000000
                1010000
```

Alignment:
$$\frac{1.0111001}{1.1} = \frac{1.0111001}{1.1000000} = \frac{10111001}{11000000}$$

Append $x = 4$ zeros: $\frac{10111001\mathbf{0000}}{11000000}$

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

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 3 (15 PTS)

- Convert the following signed fixed point numbers in format [12 8] to the dual fixed point format 12_8_4.

FX	A.CE	F.EE	C.0B	8.BF	1.0A
DFX					

- ✓ A.CE:
  1010.1100 1110 ⇒ To DFX 12_8_4 ($num0$): **0**01011001110 = not a $num0$!
  ⇒ To DFX 12_8_4 ($num1$): **1**11110101100 = FAC

- ✓ F.EE:
  1111.1110 1110 ⇒ To DFX 12_8_4 ($num0$): **0**11111101110 = 7EE

- ✓ C.0B:
  1100.0000 1011 ⇒ To DFX 12_8_4 ($num0$): **0**10000001011 = 40B

- ✓ 8.BF:
  1000.1011 1111 ⇒ To DFX 12_8_4 ($num0$): **0**00010111111 = not a $num0$!
  ⇒ To DFX 12_8_4 ($num1$): **1**11110001011 = F8B

- ✓ 1.0A:
  0001.0000 1010 ⇒ To DFX 12_8_4 ($num0$): **0**00100001010 = 10A

PROBLEM 4 (20 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 12_6_4	Result	Overflow		Result	overflow
C0A + C2B			FB9-072		
2CD + 398			F33-CBF		

- ✓ C0A + C2B:

```
                          1 1 0 0 0 0 0 0 1 0 1 0 0
**1**10000001010 +          1 1 0 0 0 0 0 0.1 0 1 0 +
**1**10000101011            1 1 0 0 0 0 1 0.1 0 1 1
                           --------------------------
                            1 0 0 0 0 0 1 1.0 1 0 1
```

10000011.0101 ⇒ To DFX 12_6_4 ($num0$): **0**00011.010100 = not a $num0$!
⇒ To DFX 12_6_4 ($num1$): **1**0000011.0101 = 835 → not a $num1$! Overflow = 1

✓ 2CD + 398:

```
                                0 0 1 1 1 0 0 1 1 0 0 0 0
   001011001101 +          ➡      0¦0 1 0 1 1.0 0 1 1 0 1 +
   001110011000                   0¦0 1 1 1 0.0 1 1 0 0 0
                                  -------------------------
                                  0 1 1 0 0 1.1 0 0 1 0 1
```

011001.100101 ⇒ To DFX 12_6_4 ($num0$): 011001.100101 = not a $num0$!
⇒ To DFX 12_6_4 ($num1$): 10011001.1001 = 999 Overflow = 0

✓ FB9 - 072:

```
                                                      1 1 1 1 1 1 0 0 0 1 0 0
 111110111001 -      ➡    1111011.1001 -          ➡     1 1 1 1 0 1 1.1 0 0 1 +
 000001110010             0000001.1100~~10~~              1 1 1 1 1 1 0.0 1 0 0
                                optional to keep        ---------------------
                                                        1 1 1 1 0 0 1.1 1 0 1
```

1111001.1101 ⇒ To DFX 12_6_4 ($num0$): 011001.110100 = 674 Overflow = 0

✓ F33 - CBF:

```
                                                      1 1 1 0 0 0 0 0 0 1 1 0
 111100110011 +      ➡    1110011.0011 -          ➡     1 1 1 0 0 1 1.0 0 1 1 +
 110010111111             1001011.1111                  0 1 1 0 1 0 0.0 0 0 1
                                                        ---------------------
                                                        0 1 0 0 1 1 1.0 1 0 0
```

0100111.0100 ⇒ To DFX 12_6_4 ($num0$): 000111.010000 = not a $num0$!
⇒ To DFX 12_6_4 ($num1$): 10100111.0100 = A74 Overflow = 0

PROBLEM 5 (15 PTS)

- Complete the timing diagram of the following iterative unsigned multiplier ($N = 4, M = 4$).
  Register: $sclr$: synchronous clear. Here, if $sclr = E = 1$, the register contents are initialized to 0.
  Parallel access shift register: If $E = 1$: $s\_l = 1 \rightarrow$ Load, $s\_l = 0 \rightarrow$ Shift