Solutions - 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

UNSIGNED:

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

```
  0.1 1 0 1 0 0 0 +
  1.0 1 0 1 1 0 1
-------------------
1 0.0 0 1 0 1 0 1
```

$b_7=0$ $b_6=0$ $b_5=0$ $b_4=0$ $b_3=1$ $b_2=1$ $b_1=1$ $b_0=0$

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

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

```
  1 0.1 0 1 0 1 +
    1.1 0 0 1 0
-----------------
1 0 0.0 0 1 1 1
```

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

```
1 0 0.1 0 0 0 0 +
    0.1 0 1 0 1
-----------------
1 0 1.0 0 1 0 1
```

SIGNED:

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

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

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

```
0.0 1 0 1 0 0 0 -        0.0 1 0 1 0 0 0 +
1.0 1 0 1 1 0 1    ➡     0.1 0 1 0 0 1 1
                         -----------------
                         0.1 1 1 1 0 1 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 -        1 0 0 0.0 1 0 1 0 0 +
1 1 1 1.0 1 0 1 0 1    ➡     0 0 0 0.1 0 1 0 1 1
                             ---------------------
                             1 0 0 0.1 1 1 1 1 1
```

$c_{10}=1$ $c_9=1$ $c_8=1$ $c_7=0$ $c_6=1$ $c_5=1$ $c_4=1$ $c_3=0$ $c_2=0$ $c_1=0$ $c_0=0$

```
1 0 1.0 1 0 1 0 0 0 +
1 1 1.0 1 1 1 1 0 1
---------------------
1 0 0.1 1 0 0 1 0 1
```

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

```
01.001 x     ➡    01.001 x     ➡    1 1 0 1 1 1 x
1.001001          0.110111                1 0 0 1
                                    -------------
                                    1 1 0 1 1 1
                                  0 0 0 0 0 0
                                0 0 0 0 0 0
                              1 1 0 1 1 1
                              -------------------
                              1 1 1 1 0 1 1 1 1
                                      ⬇
                              0.1 1 1 1 0 1 1 1 1
                                      ⬇
                              1.0 0 0 0 1 0 0 0 1
```

```
10.0001 x    ➡    01.1111 x    ➡            1 0 1 0 0 1 x
01.01001          01.01001                    1 1 1 1 1
                                        -------------
                                        1 0 1 0 0 1
                                      1 0 1 0 0 1
                                    1 0 1 0 0 1
                                  1 0 1 0 0 1
                                1 0 1 0 0 1
                              -----------------------
                              1 0 0 1 1 1 1 0 1 1 1
                                        ⬇
                              0 1 0.0 1 1 1 1 0 1 1 1
                                        ⬇
                              1 0 1.1 0 0 0 0 1 0 0 1
```

```
1100.001 x   ➡  011.111 x  ➡        1 1 0 1 0 1 1 x
10.010101       01.101011                 1 1 1 1 1
                                    -------------
                                    1 1 0 1 0 1 1
                                  1 1 0 1 0 1 1
                                1 1 0 1 0 1 1
                              1 1 0 1 0 1 1
                            1 1 0 1 0 1 1
                        -----------------------
                          1 1 0 0 1 1 1 1 0 1 0 1
                                    ⬇
                         0 1 1 0.0 1 1 1 1 0 1 0 1
```

```
0.1101010 x  ➡  0.110101 x   ➡      1 1 0 1 0 1 x
11.1111011      0.0000101                   1 0 1
                                    -------------
                                    1 1 0 1 0 1
                                  0 0 0 0 0 0
                                1 1 0 1 0 1
                              -------------------
                                1 0 0 0 0 1 0 0 1
                                      ⬇
                    0.0 0 0 0 1 0 0 0 0 1 0 0 1
                                      ⬇
                    1.1 1 1 1 0 1 1 1 1 0 1 1 1
```

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

✓ $\frac{101.001}{1.001001}$: To unsigned and then alignment, $a = 6$: $\frac{101.001}{1.001001} = \frac{010.111}{0.110111} = \frac{10.111000}{00.110111} = \frac{10111000}{110111}$

Append $x = 4$ zeros: $\frac{10111000\mathbf{0000}}{110111}$

Integer Division:

```
         000000110101
110111 ) 101110000000
          110111
          -------
          1001010
           110111
          -------
            1001100
             110111
            -------
              1010100
               110111
              -------
                11101
```

$Q = 110101, R = 11101$
$\rightarrow Qf = 11.0101\ (x = 4)$

Final result (2C): $\frac{101.001}{1.001001} = 011.0101$

✓ $\frac{10.011001}{1.0101}$: To unsigned and then alignment, $a = 6$: $\frac{01.100111}{0.10011} = \frac{1.100111}{0.100110} = \frac{1100111}{100110}$

Append $x = 4$ zeros: $\frac{1100111\mathbf{0000}}{100110}$

Integer Division:

```
         00000101011
100110 ) 11001110000
         100110
         ------
           110110
           100110
           ------
            1000000
             100110
            -------
              110100
              100110
              ------
                1110
```

$Q = 101011, R = 1110$
$\rightarrow Qf = 10.1011 (x = 4)$

Final result (2C): $\frac{101.001}{1.001001} = 010.1011$

✓ $\frac{01001.001}{10.101}$: To unsigned (denominator) and then alignment, $a = 3$: $\frac{01001.001}{01.011} = \frac{1001001}{1011}$

Append $x = 4$ zeros: $\frac{1001001\mathbf{0000}}{1011}$

Integer Division:

```
       00001101010
1011 ) 10010010000
        1011
        ----
          1110
          1011
          ----
            1110
            1011
            ----
              1100
              1011
              ----
                10
```

$Q = 1101010, R = 10$
$\rightarrow Qf = 110.1010 (x = 4)$

Final result (2C): $\frac{01001.001}{10.101} = 2C(0110.101) = 1001.0110$

✓ $\frac{0.110101}{010.110111}$: To unsigned (denominator) and then alignment, $a = 6$: $\frac{0.110101}{10.110101} = \frac{110101}{10110111}$

Append $x = 4$ zeros: $\frac{110101\mathbf{0000}}{10110111}$

Integer Division:

```
          0000000100
10110111 ) 1101010000
           10110111
           --------
              1110100
```

$Q = 100, R = 1110100$
$\rightarrow Qf = 0.0100\ (x = 4)$

Final result (2C): $\frac{01001.001}{10.101} = 0.01$

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?

  2C representation for integers: -2^{n-1} to $2^{n-1}-1$. For $2^{n-1}-1 \geq 179$, we have that $n \geq 9$, so we pick $n = 9$.

  For the fractional part, we select the number of fractional bits p that make the resolution better or equal than 0.0005:

  $$2^{-p} \leq 0.0005 \rightarrow p \geq 10.96 \rightarrow p = 11$$

  Then the Fixed Point format required in $[20\ 11]$.

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	9.3132×10^{-10}	4.2907×10^{9}	$[-30,31]$	$[1,1.998046875]$
7	16	2.1684×10^{-19}	1.8447×10^{19}	$[-62,63]$	$[1,1.9999847512]$
10	37	2.9833×10^{-154}	1.3408×10^{154}	$[-510,511]$	$[1,1.9999999999927]$

PROBLEM 6 (20)

- 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

✓ F8000378: 1111 1000 0000 0000 0000 0011 0111 1000
$e + bias = 11110000 = 240 \rightarrow e = 240 - 127 = 113$
$Mantissa\ ([24\ 23]) = 1.00000000000001101111000 = 1.000105857849121$
$X = -1.000105857849121 \times 2^{113} = -1.0385693 \times 10^{34}$

✓ 80DECADE: 1000 0000 1101 1110 1100 1010 1101 1110
$e + bias = 00000001 = 1 \rightarrow e = 1 - 127 = -126$
$Mantissa = 1.10111011001010111011110 = 1.740566$
$X = -1.740566 \times 2^{-126} = -2.046 \times 10^{-38}$

✓ FDEAD378: 1111 1101 1110 1010 1101 0011 0111 1000
$e + bias = 11111011 = 251 \rightarrow e = 251 - 127 = 124$
$Mantissa = 1.11010101101001101111000 = 1.834578$
$X = -1.834578 \times 2^{124} = -3.9017 \times 10^{37}$

✓ 3DE38866: 0011 1101 1110 0011 1000 1000 0110 0110
$e + bias = 01111011 = 123 \rightarrow e = 123 - 127 = -4$
$Mantissa = 1.11000111000100001100110 = 1.7776$
$X = 1.7776 \times 2^{-4} = 0.111100003123283$

- ✓ `800ABBAA`: 1000 0000 0000 1010 1011 1011 1010 1010
  $e + bias = 00000000 = 0 \rightarrow Denormal\ number \rightarrow e = -126$
  $Mantissa = 0.00010101011101110101010 = 0.083852$
  $X = -0.083852 \times 2^{-126} = -9.8567 \times 10^{-40}$

- ✓ `FACEB0E8`: 1111 1010 1100 1110 1011 0000 1110 1000
  $e + bias = 11110101 = 245 \rightarrow e = 245 - 127 = 118$
  $Mantissa = 1.10011101011000011101000 = 1.61477$
  $X = -1.61477 \times 2^{118} = -5.366 \times 10^{35}$

- ✓ `7FF32B5A`: 0111 1111 1111 0011 0010 1011 0101 1010
  $e + bias = 11111111 = 255, f \neq 0$
  $X = NaN$

- ✓ `ACCEDE78`: 1010 1100 1100 1110 1101 1110 0111 1000
  $e + bias = 01011001 = 89 \rightarrow e = 89 - 127 = -38$
  $Mantissa = 1.10011101101111001111000 = 1.61616$
  $X = -1.61616 \times 2^{-38} = -5.8796 \times 10^{-12}$

- ✓ `FA09D3784D089B7D`: 1111 1010 0000 1001 1101 0011 0111 1000 0100 1101 0000 1000 1001 1011 0111 1101
  $e + bias = 11110100000 = 1952 \rightarrow e = 1952 - 1023 = 929$
  $Mantissa\ ([53\ 52]) = 1.1001110100110111100001001101000010001001101101111101 = 1.61412$
  $X = -1.6142 \times 2^{929} = -7.3249 \times 10^{279}$

- ✓ `80DEADBEE9742400`: 1000 0000 1101 1110 1010 1101 1011 1110 1110 1001 0111 0100 0010 0100 0000 0000
  $e + bias = 00000001101 = 13 \rightarrow e = 13 - 1023 = -1010$
  $Mantissa = 1.1110101011011011111011101001011101000010010000000000 = 1.9174$
  $X = -1.9174 \times 2^{-1010} = -1.7475 \times 10^{-304}$

- ✓ `8009D3787F888800`: 1000 0000 0000 1001 1101 0011 0111 1000 0111 1111 1000 1000 1000 1000 0000 0000
  $e + bias = 00000000000 = 0 \rightarrow Denormal\ number \rightarrow e = -1022$
  $Mantissa = 0.1001110100110111100001111111100010001000100000000000 = 0.61412$
  $X = -0.61412 \times 2^{-1022} = -1.3665 \times 10^{-308}$

- ✓ `FA0BEBE80BEEF0A0`: 1111 1010 0000 1011 1110 1011 1110 1000 0000 1011 1110 1110 1111 0000 1010 0000
  $e + bias = 11110100000 = 1952 \rightarrow e = 1952 - 1023 = 929$
  $Mantissa = 1.1011111010111110100000001011111011101111000010100000 = 1.7451$
  $X = -1.7451 \times 2^{929} = -7.9193 \times 10^{279}$

- ✓ `4974240040490FDB`: 0100 1001 0111 0100 0010 0100 0000 0000 0100 0000 0100 1001 0000 1111 1101 1011
  $e + bias = 10010010111 = 1175 \rightarrow e = 1175 - 1023 = 152$
  $Mantissa = 1.0100001001000000000001000000010010010000111111011011 = 1.2587$
  $X = 1.2587 \times 2^{152} = 7.1864 \times 10^{45}$

- ✓ `FA09D37809ABC0DE`: 1111 1010 0000 1001 1101 0011 0111 1000 0000 1001 1010 1011 1100 0000 1101 1110
  $e + bias = 11110100000 = 1952 \rightarrow e = 1952 - 1023 = 929$
  $Mantissa = 1.1001110100110111100000001001101010111100000011011110 = 1.61413$
  $X = -1.61413 \times 2^{929} = -7.3249 \times 10^{279}$

- ✓ `FF80000009ABC0DE`: 1111 1111 1000 0000 0000 0000 0000 0000 0000 1001 1010 1011 1100 0000 1101 1110
  $e + bias = 11111111000 = 2040 \rightarrow e = 2040 - 1023 = 1017$
  $Mantissa = 1.0000000000000000000000001001101010111100000011011110 = 1.00000003602$
  $X = -1.00000003602 \times 2^{1017} = -1.4044 \times 10^{306}$

- ✓ `DECAFC0FFEE00800`: 1101 1110 1100 1010 1111 1100 0000 1111 1111 1110 1110 0000 0000 1000 0000 0000
  $e + bias = 10111101100 = 1516 \rightarrow e = 1516 - 1023 = 493$
  $Mantissa = 1.1010111111000000111111111110111000000000100000000000 = 1.68653$
  $X = -1.68653 \times 2^{493} = -4.313 \times 10^{148}$

PROBLEM 7 (30)

- 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

✓ X = FA000378 + FF800FAD:

FF800FAD: 1111 1111 1000 0000 0000 1111 1010 1101

$e + bias = 11111111 = 255 \rightarrow, f \neq 0$

FF800FAD = NaN

$X = \# + NaN = NaN$

✓ X = 7F800FEA + 09ABC0DE:

7F800FEA: 0111 1111 1000 0000 0000 1111 1110 1010

$e + bias = 11111111 = 255, f \neq 0$

7F800FEA = NaN

$X = NaN + \# = NaN$

✓ X = FC09D378 + 7F800000:

7F800000: 0111 1111 1000 0000 0000 0000 0000 0000

$e + bias = 11111111 = 255, f = 0$

7F800000 = $+\infty$

$X = \# + \infty = +\infty$

✓ X = 3DE38866 + 3300D959:

3DE38866: 0011 1101 1110 0011 1000 1000 0110 0110

$e + bias = 01111011 = 123 \rightarrow e = 123 - 127 = -4$ $\quad Mantissa = 1.11000111000100001100110$

3DE38866 = $1.11000111000100001100110 \times 2^{-4}$

3300D959: 0011 0011 0000 0000 1101 1001 0101 1001

$e + bias = 01100110 = 102 \rightarrow e = 102 - 127 = -25$ $\quad Mantissa = 1.00000001101100101011001$

3300D959 = $1.00000001101100101011001 \times 2^{-25}$

$$X = 1.11000111000100001100110 \times 2^{-4} + 1.00000001101100101011001 \times 2^{-25}$$
$$X = 1.11000111000100001100110 \times 2^{-4} + \frac{1.00000001101100101011001}{2^{21}} \times 2^{-4}$$
$$X = 1.11000111000100001100110 \times 2^{-4} + 0.00000000000000000000001 \times 2^{-4}$$

$$X = 1.11000111000100001101010 \times 2^{-4} = 1.7776 \times 2^{-4} = 0.111100003123282$$

X = 0011 1101 1110 0011 1000 1000 0110 1010 = 3DE3886A

✓ X = CA09D378 - 80000000:

CA98D378: 1100 1010 0000 1001 1101 0011 0111 1000

$e + bias = 10010100 = 148 \rightarrow e = 148 - 127 = 21$ $\quad Mantissa = 1.00010011101001101111000$

3DE38866 = $-1.00010011101001101111000 \times 2^{21}$

80000000: 1000 0000 0000 0000 0000 0000 0000 0000

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

80000000 = 0

$$X = -1.00010011101001101111 \times 2^{21} = -2258142$$

X = 1100 1010 0000 1001 1101 0011 0111 1000 = CA09D378

✓ X = 5A09D378 - 40490FDB:

5A09D378: 0101 1010 0000 1001 1101 0011 0111 1000

$e + bias = 10110100 = 180 \rightarrow e = 180 - 127 = 53$ $\quad Mantissa = 1.00010011101001101111000$

5A09D378 = $1.00010011101001101111000 \times 2^{53}$

40490FDB: 0100 0000 0100 1001 0000 1111 1101 1011

$e + bias = 10000000 =\to e = 128 - 127 = 1$ $Mantissa = 1.10010010000111111011011$

40490FDB = $1.10010010000111111011011 \times 2^1$

$X = 1.00010011101001101111000 \times 2^{53} - 1.10010010000111111011011 \times 2^1$

$X = 1.00010011101001101111000 \times 2^{53} - \frac{1.10010010000111111011011}{2^{52}} \times 2^{53}$

The division by 2^{52} requires more than $p + 1 = 24$ bits for proper representation. Thus, we approximate the second operand by 0.

$X = 1.00010011101001101111000 \times 2^{53} = 9.6986 \times 10^{15}$

X = 0101 1010 0000 1001 1101 0011 0111 1000 = 5A09D378

✓ X = 7DE32B5A – FF800000:

FF800000: 1111 1111 1000 0000 0000 0000 0000 0000

$e + bias = 11111111 = 255, f = 0$

FF800000 = $-\infty$

$X = \# - (-\infty) = +\infty$

X = 7F800000

✓ X = FA09D378 – 09ABC0DE:

FA09D378: 1111 1010 0000 1001 1101 0011 0111 1000

$e + bias = 11110100 = 244 \to e = 244 - 127 = 117$ $Mantissa = 1.00010011101001101111000$

FA09D378 = $-1.00010011101001101111000 \times 2^{117}$

09ABC0DE: 0000 1001 1010 1011 1100 0000 1101 1110

$e + bias = 00010011 = 19 \to e = 19 - 127 = -108$ $Mantissa = 1.01010111100000011011110$

09ABC0DE = $1.01010111100000011011110 \times 2^{-108}$

$X = -1.00010011101001101111000 \times 2^{117} - 1.01010111100000011011110 \times 2^{-108}$

$X = -1.00010011101001101111000 \times 2^{117} - \frac{1.01010111100000011011110}{2^{225}} \times 2^{117}$

The division by 2^{225} requires more than $p + 1 = 24$ bits for proper representation. Thus, we approximate the second operand by 0.

$X = -1.00010011101001101111000 \times 2^{117} = -1.7891 \times 10^{35}$

X = 1111 1010 0000 1001 1101 0011 0111 1000 = FA09D378

✓ X = FA09D300 × 4D080000:

FA09D300: 1111 1010 0000 1001 1101 0011 0000 0000

$e + bias = 11110100 = 244 \to e = 244 - 127 = 117$ $Mantissa = 1.00010011101001100000000$

FA09D300 = $-1.000100111010011 \times 2^{117}$

4D080000: 0100 1101 0000 1000 0000 0000 0000 0000

$e + bias = 10011010 = 154 \to e = 154 - 127 = 27$ $Mantissa = 1.00010000000000000000000$

4D080000 = 1.0001×2^{27}

$X = -1.000100111010011 \times 2^{117} \times 1.0001 \times 2^{27}$

$X = -1.0010010011100000011 \times 2^{144}$

$e + bias = 144 + 127 = 271 > 254$

In this case, there is an overflow. The value X is assigned to $-\infty$.

X = 1111 1111 1000 0000 0000 0000 0000 0000 = FF800000

✓ X = 80000000 × 497424FE:

80000000: 1000 0000 0000 0000 0000 0000 0000 0000

$e + bias = 00000000 = 0 \to Denormal\ number \to e = -126$

80000000 = 0

$X = 0 \times \# = 0$

X = 80000000

✓ X = FA09DF00 × 7F800000:

7F800000: 0111 1111 1000 0000 0000 0000 0000 0000

$e + bias = 11111111 = 255, f = 0$

7F800000 = $+\infty$

$X = (-\#) \times \infty = -\infty$

X = 1111 1111 1000 0000 0000 0000 0000 0000 = FF800000

✓ X = 7A09D300 × 0BEEF000:

7A09D300: 0111 1010 0000 1001 1101 0011 0000 0000

$e + bias = 11110100 = 244 \rightarrow e = 244 - 127 = 117$ $\quad Mantissa = 1.00010011101001100000000$

7A09D300 = $1.000100111010011 \times 2^{117}$

0BEEF000: 0000 1011 1110 1110 1111 0000 0000 0000

$e + bias = 00010111 = 23 \rightarrow e = 23 - 127 = -104$ $\quad Mantissa = 1.11011101111000000000000$

0BEEF000 = $1.11011101111 \times 2^{-104}$

$X = 1.000100111010011 \times 2^{117} \times 1.11011101111 \times 2^{-104}$

$X = 10.0000000101000110101111111101 \times 2^{13} = 1.000000010100011010111111101 \times 2^{14} = 1.6466 \times 10^{4}$

$e + bias = 14 + 127 = 141 = 10001101$

X = 0100 0110 1000 0000 1010 0011 0101 1111 = 4680A35F (four bits were truncated)

✓ X = 49742000 ÷ 40490000:

49742000: 0100 1001 0111 0100 0010 0000 0000 0000

$e + bias = 10010010 = 146 \rightarrow e = 146 - 127 = 19$ $\quad Mantissa = 1.11101000010000000000000$

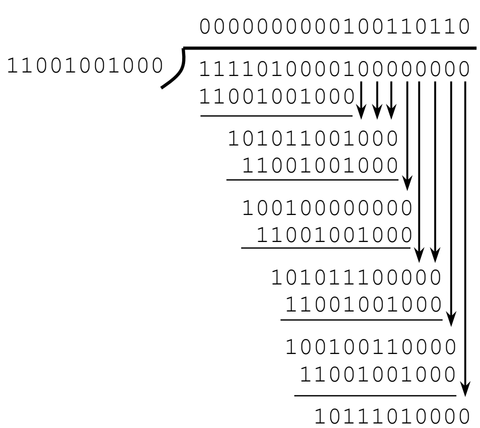
497420000 = $1.1110100001 \times 2^{19}$

40490000: 0100 0000 0100 1001 0000 0000 0000 0000

$e + bias = 10000000 = 128 \rightarrow e = 128 - 127 = 1$ $\quad Mantissa = 1.10010010000000000000000$

0BEEF000 = 1.1001001×2^{1}

$$X = \frac{1.1110100001 \times 2^{19}}{1.1001001 \times 2^{1}}$$



Alignment:

$$\frac{1.1110100001}{1.1001001} = \frac{1.1110100001}{1.1001001000} = \frac{11110100001}{11001001000}$$

Append $x = 8$ zeros: $\frac{11110100001\mathbf{00000000}}{11001001000}$

Integer division

$Q = 100110110, R = 1011101000 \rightarrow Qf = 1.00110110$

Thus: $X = \frac{1.1110100001 \times 2^{19}}{1.1001001 \times 2^{1}} = 1.0011011 \times 2^{18} = 1.2109375 \times 2^{18} = 317440$

$e + bias = 18 + 127 = 145 = 10010001$

X = 0100 1000 1001 1011 0000 0000 0000 0000 = 489B0000

✓ X = 80000000 ÷ 09ABC0DE:

80000000: 1000 0000 0000 0000 0000 0000 0000 0000

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

80000000 = 0

$X = 0 \div \# = 0$

X = 80000000

✓ X = FF800000 ÷ 09FE0000:

FF800000: 1111 1111 1000 0000 0000 0000 0000 0000

$e + bias = 11111111 = 255, f = 0$

FF800000 = $-\infty$

$X = -\infty \div \# = -\infty$

X = FF800000

✓ X = FA09D300 ÷ 48500000

FA09D300: 1111 1010 0000 1001 1101 0011 0000 0000

$e + bias = 11110100 = 244 \rightarrow e = 244 - 127 = 117$ $Mantissa = 1.00010011101001100000000$

FA09D300 = $-1.000100111010011 \times 2^{117}$

48500000: 0100 1000 0101 0000 0000 0000 0000 0000

$e + bias = 10010000 = 144 \rightarrow e = 144 - 127 = 17$ $Mantissa = 1.10100000000000000000000$

48500000 = 1.101×2^{17}

$$X = \frac{-1.000100111010011 \times 2^{117}}{1.101 \times 2^1}$$

```
                  000000000000000010101001
1101000000000000 )100010011101001100000000
                   1101000000000000
                    10000111010011000
                     1101000000000000
                       1111101001100000
                       1101000000000000
                         10101001100000000
                          1101000000000000
                          1000001100000000
```

Alignment:

$$\frac{1.00010011101001}{1.101} = \frac{1.000100111010011}{1.101000000000000} = \frac{1000100111010011}{1101000000000000}$$

Append $x = 8$ zeros: $\frac{1000100111010011\mathbf{00000000}}{1101000000000000}$

Integer division

$Q = 10101001, R = 1000001100000000$

$\rightarrow Qf = 0.10101001$

Thus:

$$X = \frac{-1.000100111010011 \times 2^{117}}{1.101 \times 2^1} = -0.10101001 \times 2^{100} = -1.0101001 \times 2^{99} = -1.3203125 \times 2^{99} = -8.368 \times 10^{29}$$

$e + bias = 99 + 127 = 226 = 11100010$

X = 1111 0001 0010 1001 0000 0000 0000 0000 = F1290000