Homework 6

Floating point

1. Fill in the missing information in the following table:

Fractional value	Binary representation	Decimal representation
1/8		
3/4		
	10.1011	
25/16		
		3.1875

2. Given a floating-point format with a *k-bit* exponent and an *n-bit* fraction, write formulas for the exponent *E*, significand *M,* the fraction *f*, and the value *V* for the quantities that follow. In addition, describe the bit representation.

A. The number 5.0

B The largest odd integer that can be represented exactly

C The reciprocal of the smallest positive normalized value

3. Consider the following two 9-bit floating-point representations based on the IEEE floating-point format.

Format A

- There is one sign bit.
- There are *k* = 5 exponent bits. The exponent bias is 15.
- There are *n* = 3 fraction bits.

Format B

- There is one sign bit.
- There are *k* = 4 exponent bits. The exponent bias is 7.
- There are *n* = 4 fraction bits.

Below, you are given some bit patterns in Format A, and your task is to convert them to the closest value in Format B. If rounding is necessary, you should ***round toward*** +∞. In addition, give the values of numbers given by the Format A and Format B bit patterns. Given these as whole numbers(eg.,17) or as fractions(eg.,17/64 or $17/2^6$).

Format A		Format B	
Bits	Value	Bits	Value
1 01110 001	$\frac{-9}{16}$	1 0110 0010	$\frac{-9}{16}$
0 10110 101			
1 00111 110			
0 00000 101			
1 11011 000			
0 11000 100			

X86-64

4. For each of the following structure declarations, determine the offset of each field, the total size of the structure, and its alignment requirement under x86-64.

A. struct P1 { int I; char c; long j; char d;};

B. struct P2 { long I; char c; char d; int j;};

C. struct P3 { short w[3]; char c*[3]};

D. struct P4 { struct P1 a[2]; struct P2 *p};

E. struct P5 { short w[3]; char c[3]}.

	Offset of each field				Total size	Alignment
A	i:0	c:4				
B						
C						
D						
E						