Team #:____ 8/31/06
Absent:

IEEE 754 Standard Floating Point Representation

	Sign bit	8-bit Exponent (bias 127)	23-bit Mantissa (for normalized values, leading 1 not stored)
0 ≡ +			
1 ≡ -			

	Sign bit	11-bit Exponent (bias 1023)	52-bit Mantissa (for normalized values, leading 1 not stored)
0 ≡ +			
1 ≡ -			

Single Precision		Double Precision		Object
Exponent	**Mantissa**	**Exponent**	**Mantissa**	**Represented**
1-254	any value	1-2046	any value	normalized #
0	0	0	0	0
0	nonzero	0	nonzero	denormalized #
255	0	2,047	0	infinity
255	nonzero	2,047	nonzere	NaN (not a #)

1) Convert the value 23.625_{10} to its binary representation.

64 32 16 8 4 2 1		.5 .25 .125 .0625
	.	

2) Normalize the above value so that the most significant 1 is immediately to the left of the radix point. Include the corresponding exponent value to indicate the motion of the radix point.

$1.\boxed{} \times 2^{\boxed{}}$

3) Write the corresponding 32-bit IEEE 754 floating point representation for 23.625_{10}.

Team #:____ 8/31/06
Absent:

4) Write the corresponding 64-bit IEEE 754 floating point representation for 23.625_{10}.

5) What would be the smallest positive normalized 32-bit IEEE 754 floating point value?

6) The smallest positive denormalized 32-bit IEEE 754 floating point value has representation of

	Sign bit	8-bit Exponent (bias 127)	23-bit Mantissa (for denormalized values, leading 0 not stored)
0 ≡ + 1 ≡ -	0	0 0 0 0 0 0 0 0	0 0 0 ... 0 1

What value would it represent?

$$2^{\square} \times 2^{\square}$$

7) What would be the representation for the largest positive denormalized 32-bit IEEE 754 floating point?

	Sign bit	8-bit Exponent (bias 127)	23-bit Mantissa (for denormalized values, leading 0 not stored)
0 ≡ + 1 ≡ -	0	0 0 0 0 0 0 0 0	

Team #:____ 8/31/06
Absent:

8) How would you add two IEEE 754 floating point numbers?

9) How would you multiply two IEEE 754 floating point numbers?

10) Consider adding 1.011×2^{40} and 1.01×2^{5}.
a) How many places does the second number's mantissa get shifted?

b) After we add these two numbers and store the results back into a 32-bit IEEE 754 value, what would be the result?