Lab Exercise #7 -- Floating Point Representation

A. The file named "lab07.viewdp" contains an executable program which allows the user to enter any real constant and displays the corresponding 64-bit internal representation of that constant.

1. Use "lab07.viewdp" to complete the following table.

Real constant	Internal representation (in hexadecimal)
1.0	_____
2.0	_____
4.0	_____
4.5	_____
4.25	_____
4.125	_____

2. Use "lab07.viewdp" to complete the following table.

Real constant	Internal representation (in hexadecimal)
3.14	_____
3.142	_____
3.1416	_____
3.14159265359	_____

3. What is the internal representation (in hexadecimal) of 0.0 (zero)?

4. What is the internal representation (in hexadecimal) of 1.0e+500 (a value which is too large to represent)?

5. The largest real number which can be represented is a value between 1.0e+308 and 2.0e+308. Use the program to find the largest 4-digit decimal number (in the form X.XXXe+308) which can be represented.

B. The file named "lab07.makedp" contains an executable program which allows the user to enter the internal representation of a 64-bit floating point number and displays the corresponding real constant.

1. Use "lab07.makedp" to complete the following table.

Internal representation	Real constant
4069 0000 0000 0000	_____
4069 0400 0000 0000	_____
4520 8B2A 2C28 0291	_____
39B4 484B FEEB C2A0	_____

2. Use "lab07.makedp" to complete the following table.

Internal representation	Real constant
400A 6666 6666 6664	_____
400A 6666 6666 6665	_____
400A 6666 6666 6666	_____
400A 6666 6666 6667	_____

3. What is the decimal value of the largest and most precise real constant that can be represented as a double-precision floating point number?

4. What is the decimal value of the smallest positive real constant that can be represented as a normalized double-precision floating point number?

5. A "denormal" floating point number is one in which the biased exponent is zero, but the fraction field is non-zero. The smallest denormal floating point number which can be represented in 64 bits is 0000 0000 0000 0001. What is the value of this number as a decimal real constant?

C. Write a C program which demonstrates overflow, underflow, and loss of precision on floating point numbers. Use type "float" and type "double" in the program.