

Ans 1: Yes all the components play a role in defining precision of the number. For example take number 5.2. So if we convert this number in binary

5.2 in binary is 101.0011..... . For this number same pattern of 0011 (after decimal) will repeat again and again , so its representation in binary format will exceed th 32 bit format. So IEEE has given some standard to represent such numbers which are very large.We need to divide such numbers in three parts (in single precision format)

- 1.Sign bit : 1 bit
- 2.Exponent : 8 bits
- 3.Mantissa : 23 bits

Now in the above example we will shift decimal by two places to left to get the sign bit. So the bit is 1.Now we add $127 + 2 = 129$ (depends upon how many decimal places we shift to left till we get a sign bit).So 129 in binary is 100000001.

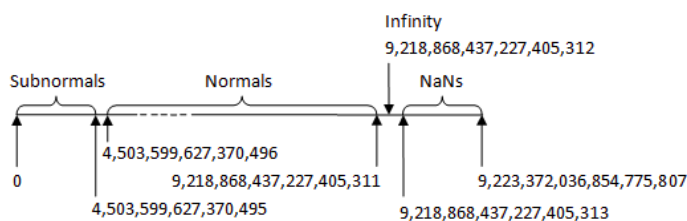
sign bit(1 bit)	Exponent(8 bit)	mantissa (23 bit)
1	10000001	00000000000000000000000

Ans 2:

1.Sub-normal numbers : When all the exponent bits are 0 and the leading hidden bit of the significand is 0, then the floating point number is called subnormal number. Thus, one logical representation of a subnormal number is :

$$(-1)^s \cdot (1.x) \cdot (2^{(\text{Exponent}-\text{bias})})$$

2.Normal numbers :Normal values are those which are represented as $(-1)^s \cdot (1.x) \cdot (2^{(\text{Exponent}-\text{bias})})$. For normal numbers the value of exponent should not be zero.



Ans 3. The five methods for rounding floating point numbers:

1. Round to nearest, ties to even – rounds to the nearest value
:rounded to nearest value that has 0 as least significant bit

2. Round to nearest, ties away from zero : rounded to nearest value
above given number for positive number and below for negative
numbers

3. Round toward 0 : rounding is directed towards zero

4. Round toward $+\infty$: rounding is directed towards +ve infinity

5. Round toward $-\infty$: rounding is directed towards -ve infinity

ex: +13.3 -13.3 14.4

Round to nearest, ties to even : +14.0 -14.0 +14.0

Round to nearest, ties away from zero : +14.0 -14.0 +15.0

toward 0 : +13.0 -13.0 +14.0

toward $+\infty$: +14.0 -13.0 +15.0

toward $-\infty$: +13.0 -14.0 +14.0
Ans1: The fractional part plays main role
in defining the precision. The length of the fractional part determines
the precision. For ex: 15675.656 This no. is represented as
 1.5675656×10^4 if significand is 8 bits and 1.56756 if significand is 5
bits. former has more precision.