Ans

0100 0011 0101 0100 0000 0000 0000 0000

MT2019519
ROSHAN ALI

0 — sign bit
1000 0110 — exp bits
01010000000000000000000 — 23 bits fraction

Sign Single precision is condisered

exp. bias = 127
exp. value = 134
so, e = 134 − 127 = 7

$$m = 2^{-1} \times 1 + 2^{-2} \times 0 + 2^{-3} \times 1 + 2^{-4} \times 0 + 2^{-1} \times 1 + \cdots$$

$$(0) \times (2^{-1} + \cdots + 2^{-23})$$

$$= 0.65625$$

The decimal value

$$(-1)^0 \times (1 + 0.65625) \times 2^7 = 212$$

so, if the value of m changes of bits in fraction part change the decimal value is changing so,

fraction part defines the precision.

As, even when we look at single & double precision,

single precision has 23 Frachin bits which give $(\log_{10}(2^{23}) = 7)$ 7 decimal digit of accuracy.

double precision has 52 frachin bits gives $(\log_{10}(2^{52}) = 15.654)$ nearly 16 decimal digits of accuracy.

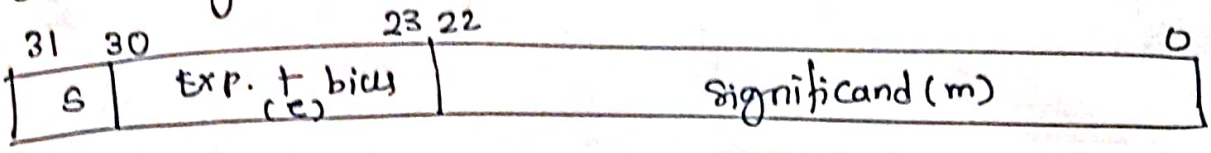
As, Smallest change that can be represented in floating point is called precision so, even from this Fraction part defining the precision.

2nd normal and subnormal values as per IEEE Standard

normal value

1. '1' before decimal point

ex:- single precision

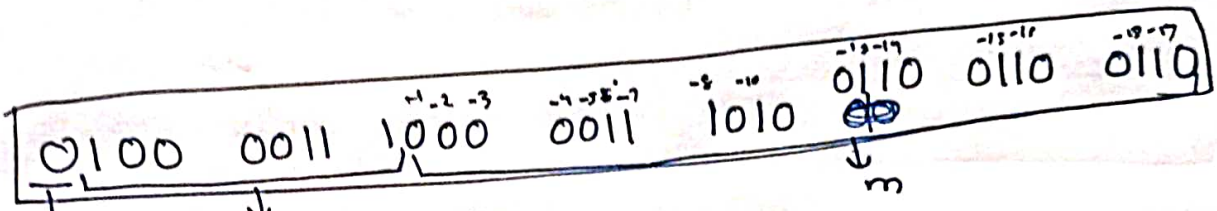


$$F = (-1)^S \times (1+m) \cdot 2^{e-127}$$

Bias = 127
e → ranged from 1 → 255

Ex:-

32 bit single precision fp.



$$(-1)^0 \times (1+m) 2^{e^{135-127}}$$

$e = 1000\,0111 = 135$

$$(-1)^0 \times (1+0.2851) \times 2^8$$

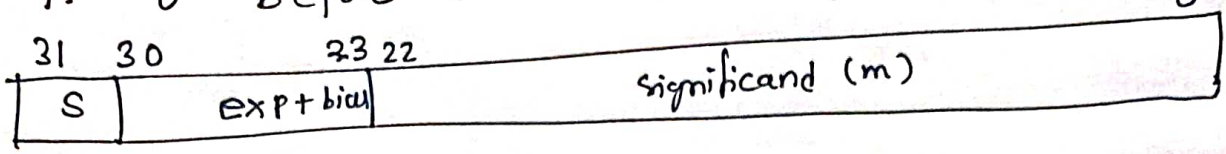
~~$m = 1 + 2^{-2} + 2^{-7} + 2^{-8} + 2^{-10} + 2^{-13} + 2^{-14} + 2^{-15} + 2^{-16} + 2^{-18} + 2^{-19}$~~

$F = 263.3$

$$1.00000111010\,0110\ldots \times 2^8$$

Subnormal value

1. `0' before decimal point



$$F = (-1)^S \times (0+m) \cdot 2^{-127+1}$$

$$= (-1)^S \times (m) \cdot 2^{-126}$$

exp = 0

m ≠ 0 (significand)

3rd

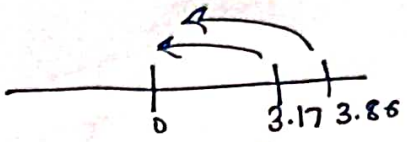
Five methods defined by IEEE for rounding floating point numbers

1. Rounding towards zero
2. Round down $(-\infty)$
3. Round up $(+\infty)$
4. Round to nearest
5. Round to even

1. Round towards zero

Ex:- 3.17 ⇒ 3

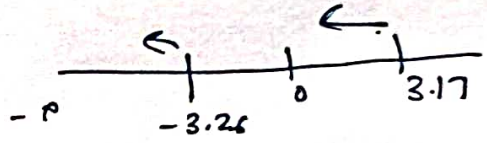
3.86 ⇒ 3



2. Round down $(-\infty)$

Ex:- 3.17 ⇒ 3

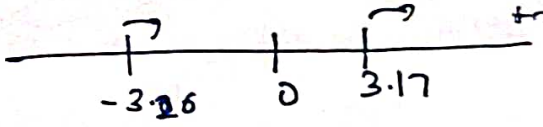
−3.26 ⇒ −4



3. Round up $(+\infty)$

Ex: 3.17 ⇒ 4

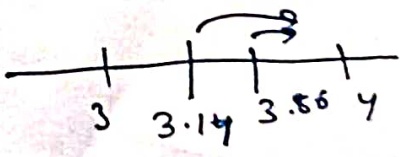
−3.26 ⇒ −3



4. Round to even

Ex:- 3.17 ⇒ 4

3.86 ⇒ 4



5. Round to nearest (ties away from zero)

$3.17 \Rightarrow 4$

$3.5 \Rightarrow 4$

$-3.6 \Rightarrow -4$