IEEE 754

- It is **IEEE** Standard for **Floating**-**Point** Arithmetic.

- It is a technical standard for **floating-point** Representation which was established in 1985 by the Institute of Electrical and Electronics Engineers (**IEEE**).

Components of IEEE 754

- IEEE 754 is the most efficient in most cases. IEEE 754 has 3 basic components:

1. **Sign (+ve or –ve)**
2. **The Biased exponent**
3. **The Normalized Mantissa**

Component 1: Sign (+ve or –ve)

- 0 represents a positive number
- 1 represents a negative number.

For example: 28.017 the sign is +ve. So it will be represent by 0.

-28.017 the sign is -ve. So it will be represent by 1.

Component 2: The Biased exponent

- The exponent field needs to represent both positive and negative exponents. A bias is added to the actual exponent in order to get the stored exponent.

- The value of bias based of size of exponent component.

For example:

if size of exponent field is 8 bits then bias value is

$2^{size-1}-1 = 127$

if size of exponent field is 11 bits then bias value is

$2^{size-1}-1 = 1023$

Component 3: The Normalized Mantissa

- The mantissa is part of a number in scientific notation or a floating-point number, consisting of its significant digits.

- A normalized mantissa is one with only one 1 to the left of the decimal like 1.mmmmmmmmmm….mmmmm

Example: 8.25

Binary Representation : 1000.01

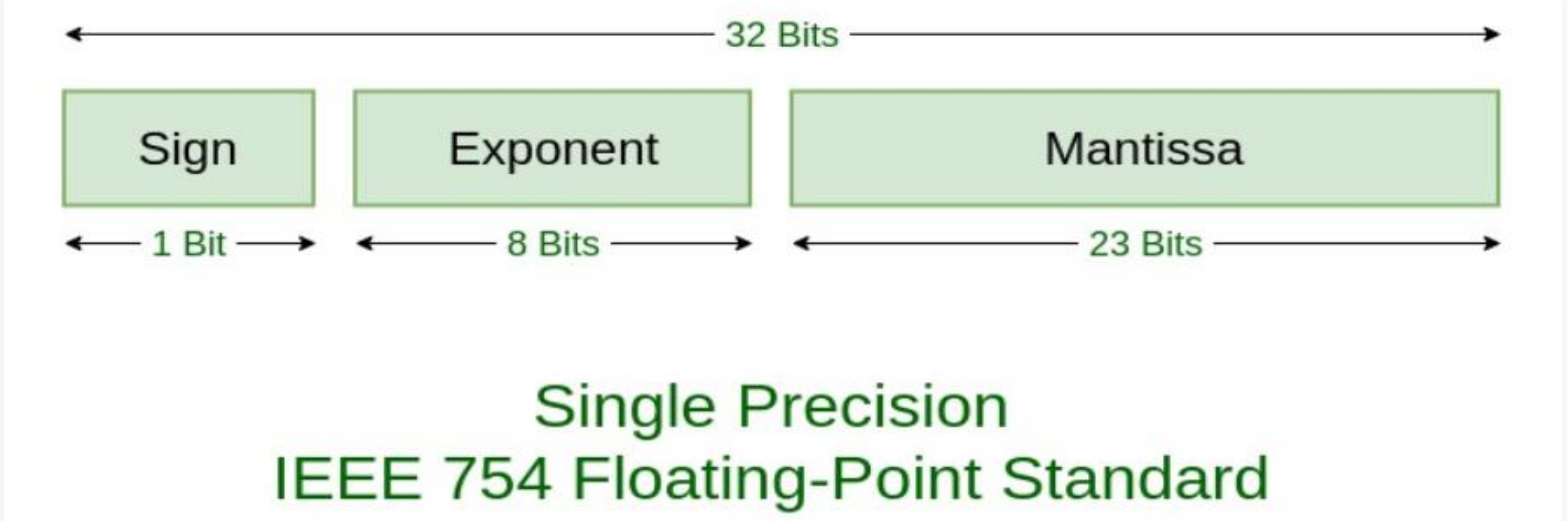
Normalized Mantissa : $1.00001 \times 2^{+3}$

Representation of IEEE 754

There are two types of representation:

1. **Single precision**

   (+ve or –ve sign) 1.mantissa x $2^{\text{Exponent-127}}$



Single Precision Example

Question: Represent 85.125 in IEEE 32 bits Format.

Binary of 85 = 1010101
Binary of 0.125 = 001
Floating Point Number 85.125 = 1010101.001 =1.010101001 x 2^{+6}

So, Sign component: 0 because of +ve sign.
biased exponent: 127+6=133
Binary of 133 = 10000101
Normalized mantissa = 010101001

we will add 0's to complete the 23 bits in mantissa

The IEEE 754 Single precision is: =

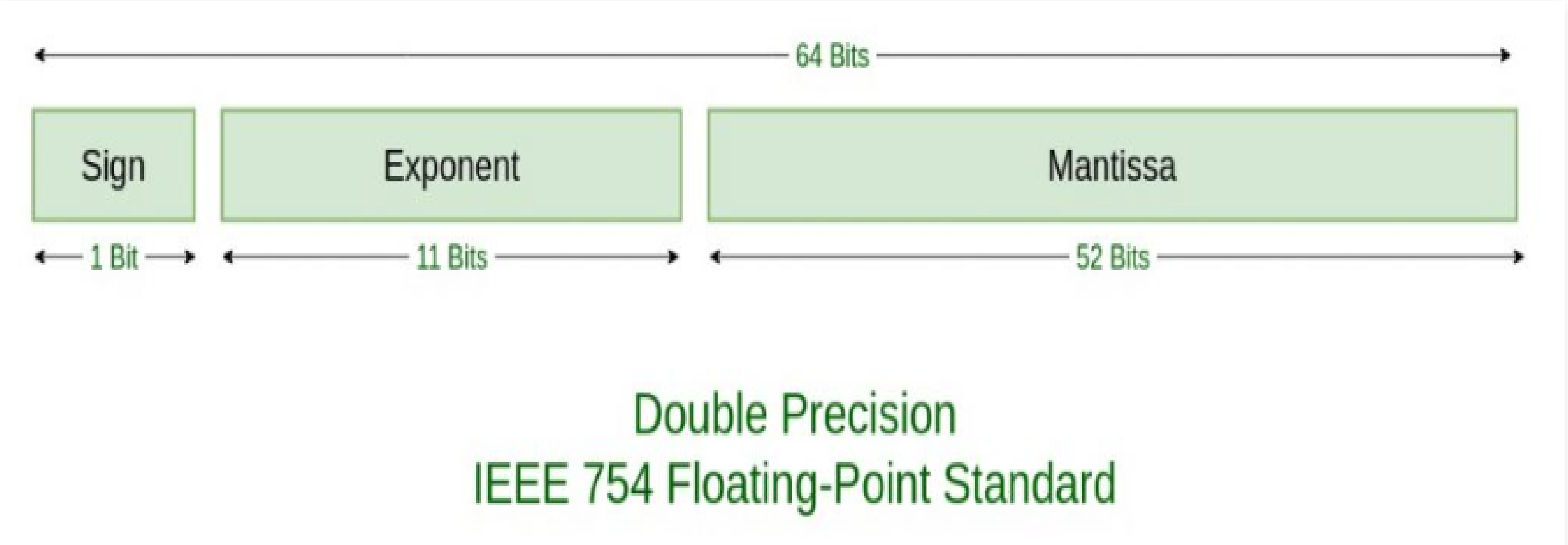
32bits :
0 10000101 01010100100000000000000

Sign	Exponent	Mantissa
0	10000101	01010100100000000000000
1bit	8bits	23bits

This can be written in hexadecimal form **42AA4000** in Computer.

2. Double precision

(+ve or –ve sign) 1.mantissa x $2^{\text{Exponent-1023}}$



Double Precision Example

Question: Represent 85.125 in IEEE 64 bits Format.

Binary of 85 = 1010101
Binary of 0.125 = 001
Floating Point Number 85.125 = 1010101.001 =1.010101001 x 2^{+6}

So, Sign component: 0 because of +ve sign.
Biased exponent: 1023+6=1029
Binary of 1029 = 10000000101
Normalized mantissa = 010101001

we will add 0's to complete the 52 bits in mantissa

The IEEE 754 Double precision is

64 bits :
010000000101
0101010010000000000000000000000000000000000000000000

Sign	Exponent	Mantissa
0	10000000101	0101010010000000000000000000000000000000000 000000000
1bit	11bits	52bits

This can be written in hexadecimal form **4055480000000000** in Computer.