IEEE 754

IEEE 754 is a standard for representing floating-point numbers in binary form, which computers use to handle decimals efficiently. Unlike integers, which are easy to represent in binary, floating-point numbers pose a challenge due to their decimal points and varying precision.

The key concerns in choosing a standard method for representing floating-point numbers include range, precision, time efficiency, space considerations, and ensuring one-to-one relationships.

To address these challenges, IEEE 754 uses a form of scientific notation with binary numbers. Just as with base-10 scientific notation, floating-point numbers are expressed as a value multiplied by a power of 2. For example, 0.05 would be written as 1.6 × 2^-5.

The IEEE 754 format consists of three main components:

1. Sign bit: Determines if the number is positive or negative.
2. Fraction: Represents the significant digits of the number in binary form.
3. Exponent: Indicates the power of 2 by which the fraction is multiplied, adjusted by a bias.

The exponent, after adjusting for bias, is added to the IEEE 754 string along with the sign bit and fraction. This arrangement allows for efficient comparisons and arithmetic operations.

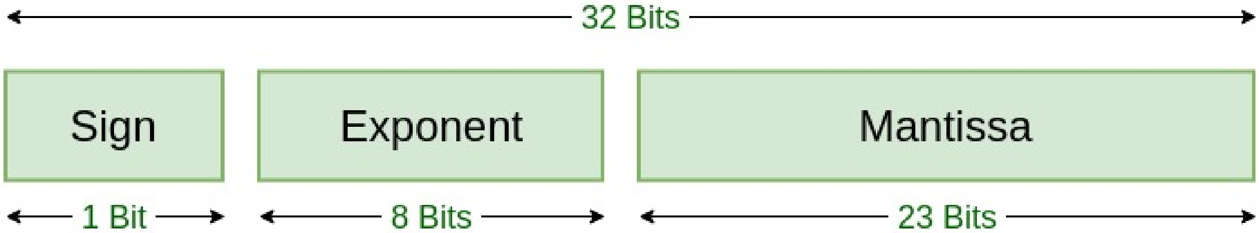
Special cases like zero, positive/negative infinity, and NaN (Not-A-Number) have specific representations in IEEE 754 format.

Converting a number to IEEE 754 format involves determining the sign, expressing the number in binary scientific notation, calculating the exponent, converting the fraction to binary, and arranging the components into the IEEE 754 string.

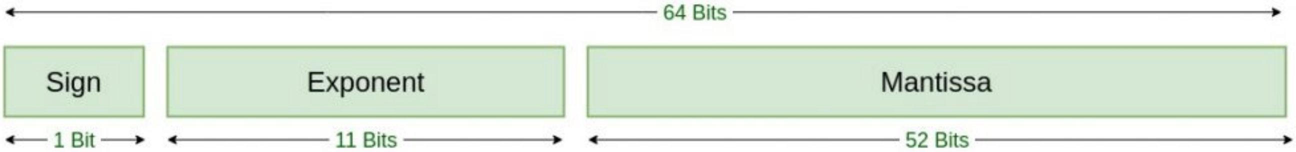
Converting from IEEE 754 format back to a decimal involves extracting the sign, exponent, and fraction, adjusting the exponent by subtracting the bias, converting the fraction back to decimal, and performing the necessary calculations to obtain the final value.

Due to the limitations of the number of bits available for representing the fraction, rounding may be necessary, leading to small errors in the representation of floating-point numbers.

IEEE 754 provides a standardized way for computers to handle floating-point numbers, ensuring consistency across different systems and platforms.



Single Precision
IEEE 754 Floating-Point Standard



Double Precision
IEEE 754 Floating-Point Standard