**IEEE Single-Precision Floating-Point Representation (32 bits)**

Sign bit | Exponent (8 bits) | Fractional part (23 bits)

Sign bit
- 0 specifies a positive value; 1 specifies negative

Exponent (representing a base-2 exponent)
- all zeros and all ones are reserved for special values
- 8 bits → -126 to 127; 11 bits → -1,022 to 1,023

Fractional part (representing a base-2 fraction)
- Normalized to lie between 1 and 2
  all zeros = 1.0; all ones = 1.999999…

e.g., smallest positive single-precision value =
00000000100000000000000000000000 =
$1.0 \times 2^{-126} \approx 1.18\ 10^{-38}$

e.g. largest positive single-precision value =
01111111011111111111111111111111 =
$1.999\ldots \times 2^{127} \approx 3.4 \times 10^{38}$

**IEEE Double-Precision Floating-Point Representation (64 bits)**

Sign bit | Exponent (11 bits) | Fractional part (52 bits)