

Biased Exponents, why?

Floating point representations were designed for ease of comparisons.

Sign of number comes first

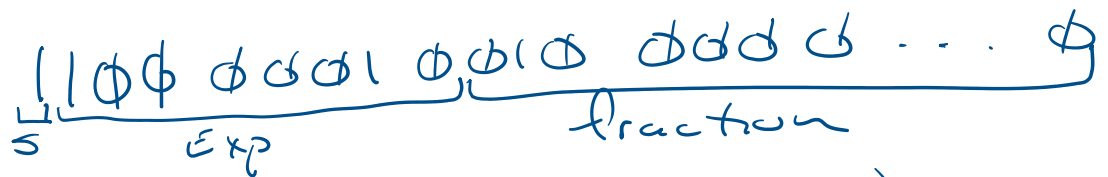
This makes it easy to determine if one is larger when signs are different. Also, compares with \emptyset are easy.

Exponent comes next

Bigger exponent means bigger number when signs are the same.

Single precision numbers use a bias of 127 ($\emptyset 111 1111$). Add exponent to 127 to get bias.
ex. $-1 + 127 = \emptyset 111 111 \emptyset$. Store 126 to represent -1 . So this forces all negative exponents to start with \emptyset .

Let's go the other way.
 Which floating point value
 is represented by:
 0120 0000₁₆?



$$(-1)^s \times (1 + \text{fraction}) \times 2^{(\text{exp})}$$

$$\begin{aligned}
 & \downarrow \\
 & 1 + \frac{0}{2} + \frac{1}{4} = 1.25 \\
 & -1 \times 1.010 \dots \times 2^{130-127} \\
 & \quad -1.25 \times 2^3 \\
 & \quad -1.25 \times 8 \\
 & \quad -10.0_{16} \\
 & \text{exp} = 10000000_2 \\
 & \quad \uparrow \quad \quad \quad \uparrow \\
 & \quad 128 \quad + \quad 2 \\
 & \quad \quad \quad 130 \\
 & \quad \quad \quad -127 \\
 & \quad \quad \quad \hline
 & \quad \quad \quad 3
 \end{aligned}$$

Given $\phi 214 0000_2$
 which base 10 floating point value?

$\underbrace{\phi \phi \phi \phi \phi \phi 1 \phi \phi \phi 1 \phi 1 \phi \phi \phi \phi \phi \phi \dots \phi \phi \phi \phi}_{\text{positive}}$

$$\text{exp} = 4 - 127 = -123$$

$$\text{frac} \quad 1.00101 \times 2^{-123}$$

$$1 + \frac{1}{8} + \frac{1}{32}$$

$$1 + 0.125 + 0.03125 = 1.15625$$

$$1.15625 \times 2^{-123}$$

$$-139.90625 \underline{\underline{?}}$$

To add $9.999 \times 10^1 + 1.616 \times 10^{-1}$

1. Convert to common exponent, usually, smaller number.

$$\begin{array}{r} 9.999 \times 10^1 \\ + 0.016 \times 10^1 \\ \hline 10.015 \times 10^1 \end{array}$$

2. Add

3. Normalize 1.0015×10^2

Note, at this time we check for over/under flow.

4. Round

$$1.002 \times 10^2$$

Since we only had 3 digits to the right in both original numbers.

Now, in binary!

$$1.000 \times 2^{-1} \\ + (-1.110) \times 2^{-2}$$

$$(0.5_{10} + -0.4375)$$

$$\begin{array}{r} 0.5000 \\ - 0.4375 \\ \hline 0.0625 \end{array}$$

1. Shift smaller and add

$$\begin{array}{r} 1.000 \times 2^{-1} \\ + -0.111 \times 2^{-1} \\ \hline 0.001_2 \times 2^{-1} \end{array}$$

2. Normalize

$$1.000_2 \times 2^{-4} \quad (\text{no under/overflow})$$

3. Round and renormalize if needed

$$1.000_2 \times 2^{-4} \\ = 0.0625$$

$$1/2 = 0.5, 1/4 = 0.25, 1/8 = 0.125$$

$$1/16 = 0.0625, 1/32 = 0.03125 \dots$$