How does IEE 754 work?

|  |  |  |
| --- | --- | --- |
| S | Exponent | Fraction/Mantissa |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 31 | 30 | 29 | 28 | 27 | 26 | 25 | 24 | 23 | 22 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| 1 | 128 | 64 | 32 | 16 | 8 | 4 | 2 | 1 | 0.5 | 0.35 | 0.125 | 0.0625 | Etc. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**S:** Indicated signed | 1 bit

**Exponent:** Indicates what the exponent is and explained more below. Max val is bits. | 8 bits

**Fraction/Mantissa**: Tells us what the fraction is. Max val is or in this case (not ) | 23 Bits

Converting from Binary to Decimal Floating Point

* Important Formula:
* Since we are dealing with 32 bits, our Bias will be 127.
* When dealing with the exponent, it will **be signed**. However, we don’t want to represent -1 (such as in ) as 11111111 and +1 (such as ) as 00000001. Thus instead, we use what is called a **biased notation**. For single precision, we will subtract 127 from our actual value and convert that to binary to store into the exponent. When we are converting to decimal, we will obtain our exponent then turn it into decimal and subtract 127 from it. (Refer to example 211 and 212 of textbook).
  + **Note: We do deal with unsigned in the exponent but note that a lmb 1 no longer represents negative.**