IEEEfp Notes

* Notes for IEEEfp (ranges and precision)
  + Ranges[[1]](#footnote-1)

IEEE single precision floating points use normalized representations where possible, and can extend it’s range (at the expense of normalization) with denormalized numbers. This extension of representation’s tradeoff additionally is in precision for small numbers. “Denormals” have an exponent field of zero, and represent:

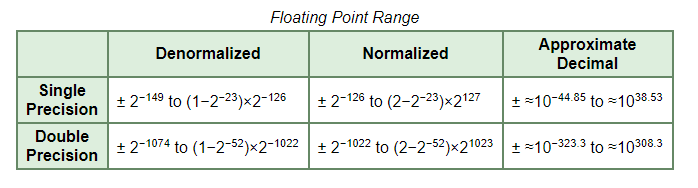
(-1)^s \* 0.f \* 2^-126

Important Considerations:

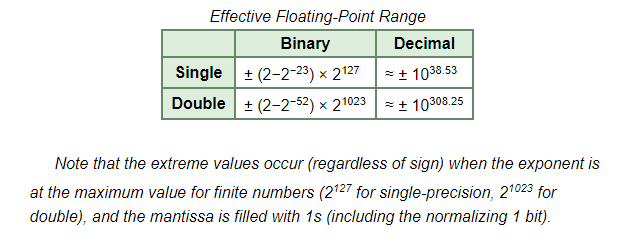
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sign** | **Exponent** | **Fraction** | **Represents** | **Notes** |
| 1 | ff | !=0 | NaN |  |
| 1 | ff | 0 | -Inf |  |
| 1 | 01-fe | anything | -1.f \* 2^(exp-127) |  |
| 1 | 00 | !=0 | -0.f \* 2^-126 |  |
| 1 | 00 | 0 | -0 | (special case of last line) |
| 0 | 00 | 0 | 0 | (special case of next line) |
| 0 | 00 | !=0 | 0.f \* 2^-126 |  |
| 0 | 01-fe | anything | 1.f \* 2^(exp–127) |  |
| 0 | ff | 0 | Inf |  |
| 0 | ff | !=0 | NaN |  |

|  |  |
| --- | --- |
| **Operation** | **Result** |
| n / +-Inf | 0 |
| +-Inf \* +-Inf | +-Inf |
| +-nonzero / 0 | +-Inf |
| Inf + Inf | Inf |
| +-0 / +-0 | NaN |
| Inf – Inf | NaN |
| +-Inf / +-Inf | NaN |
| +-Inf \* 0 | NaN |

Ranges:

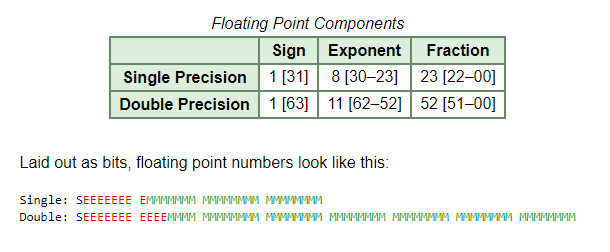


Effective Floating-Point Range

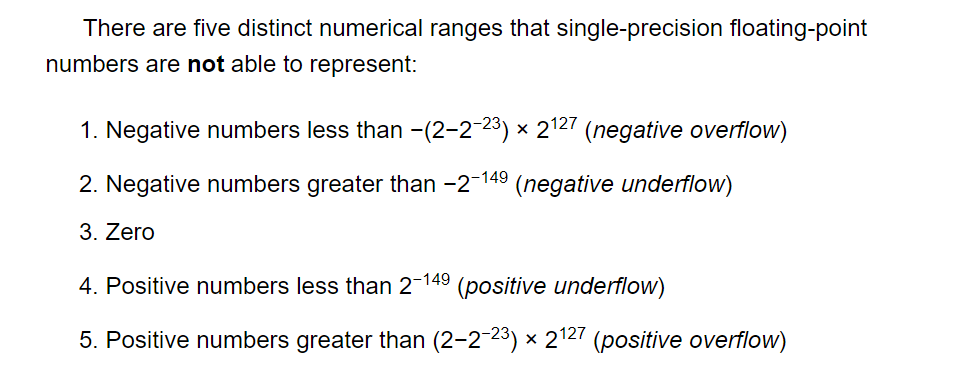


* + Precision[[2]](#footnote-2)

Precision refers to a given floating point format’s storage capacity:



There are three basic components to the floating point: the sign, the exponent, and the mantissa. The mantissa refers to the fraction portion of a floating point and an implicit leading digit.



1. Source: http://www.oneonta.edu/faculty/zhangs/csci201/IEEE%20Floating%20Point%20Format.htm [↑](#footnote-ref-1)
2. Source: http://steve.hollasch.net/cgindex/coding/ieeefloat.html [↑](#footnote-ref-2)