

PRECISIÓN SIMPLE: $1 + 8 + 23 = 32$ bits = 1 palabra

s (1 bit) E (8 bits) m (23 bits)

Normal: si $E \neq 0^{(8)}0$ y $E \neq 1^{(8)}1$; la interpretación es $(-1)^s \times 1.m \times 2^{E_d-127}$

Subnormal: si $E = 0^{(8)}0$; la interpretación es $(-1)^s \times 0.m \times 2^{-126}$

Desbordamiento (*overflow*): si $E = 1^{(8)}1$; la interpretación es *NaN* (*not a number*)

¿Por qué -127 ? Porque $(1^{(8)}1)_2 = 2^7 + 2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 2^0 = 2^8 - 1 = 255$ y trasladamos el intervalo $[0, 255]$ al intervalo $[-127, 128]$

PRECISIÓN DOBLE: $1 + 11 + 52 = 64$ bits = 2 palabras

s (1 bit) E (11 bits) m (52 bits)

Normal: si $E \neq 0^{(11)}0$ y $E \neq 1^{(11)}1$; la interpretación es $(-1)^s \times 1.m \times 2^{E_d-1023}$

Subnormal: si $E = 0^{(11)}0$; la interpretación es $(-1)^s \times 0.m \times 2^{-1022}$

Desbordamiento (*overflow*): si $E = 1^{(11)}1$; la interpretación es *NaN* (*not a number*)

¿Por qué -1023 ? Porque $(1^{(11)}1)_2 = 2^{10} + 2^9 + \dots + 2^0 = 2^{11} - 1 = 2047$ y trasladamos el intervalo $[0, 2047]$ al intervalo $[-1023, 1024]$