

Flyttal komplettering (32-bitars)

Format: s / c / f
Antal bitar: 1 8 23

s är teckenbit: 0 motsvarar + och 1 motsvarar −.

c är karakteristika: $c = \exp + 127$.

f kallas fraction och är normaliserade mantissan utan inledande 1.

$\pm \infty$: s / 255 / 0

± 0 : s / 0 / 0 (Obs! i detta fall sätts inledande mantissabiten = 0.)

Största belopp blir: $1.111\dots 1 \cdot 2^{+127} \approx 2^{+128}$

Minsta belopp (förutom 0) blir: $1.000\dots 0 \cdot 2^{-126}$

För att kunna representera tal med ännu mindre belopp än minsta beloppet ovan har man infört undantaget nedan, som dock ger sämre upplösning än normalfallet.

Undantag:

Man kan i uttrycket för ”0” ovan välja $f \neq 0$.

Största belopp skulle i detta fall bli:

$$0.111\dots 1 \cdot 2^{-127} = 0.0111\dots 1 \cdot 2^{-126}$$

Det uppstår då ett ”glapp” till minsta beloppet ($1.000\dots 0 \cdot 2^{-126}$) ovan. För att undvika ”glappet” bestämmer man att exponenten skall sättas till -126 istället för -127 då $c = 0$.

Man får då största beloppet = $0.111\dots 1 \cdot 2^{-126}$ i detta fall.