Rechnerarithmetik: Fließpunktzahlen

TechGI 2 - SoSe 2014

Genauigkeit und Formate

Bias = $2^{(\#e)-1} - 1$

Mini-Float 16b	VZ 1b	e 5b	fraction 10b	Bias = $2^{5-1} - 1 = 15$
Singleprecision-Float 32b	VZ 1b	e 8b	fraction 23b	Bias = $2^{8-1} - 1 = 127$
Doubleprecision-Float 64b	VZ 1b	e 11b	fraction 52b	Bias = $2^{11-1} - 1 = 1023$

Berechnung von Fließpunktzahlen

Allgemein	VZ	e	fraction	e=exp+Bias

$n = (-1)^{VZ} \cdot (1.fraction) \cdot 2^{exp}$

Mini-Float: $x = 1100111011001011_2$

→	1	10011	1011001011	10011=exp+15 → exp = 4

→ $n = (-1)^{1} \cdot (1.\ 1011001011) \cdot 2^{4}$

$= -(11011.001011) = -(11011 + 0.001011)$

$= -(27 + (2^{-3}+2^{-5}+2^{-6})) = -27.171875_{10}$

Mini-Float: $n = 14.125_{10}$

$= 14_{10} + 0.125_{10} = 1110_2 + 0.001_2$

$= 1110.001_2$

$= = (-1)^{0} \cdot 1.110001_2 \cdot 2^{3}$

→ exp = 3 → exp+15 = 18 = 10010

→ VZ = 0

→ fraction = 110001_2 = 1100010000_2

→	0	10010	1100010000

= 0100111100010000