Rechnerarithmetik: Fließpunktzahlen

TechGI 2 - SoSe 2014

Genauigkeit und Formate



Mini-Float 16b

VZ	е	fraction
1b	5b	10b

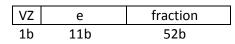
Bias = 2^{5-1} - 1 = 15

Singleprecision-Float 32b



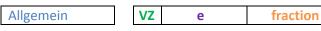
Bias =
$$2^{8-1}$$
 - 1 = 127

Doubleprecision-Float 64b



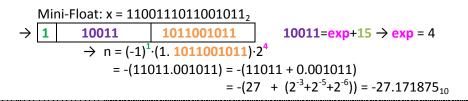
Bias =
$$2^{11-1}$$
 - 1 = 1023

Berechnung von Fließpunktzahlen



e=exp+Bias

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



Mini-Float: n = 14.125_{10} = $14_{10} + 0.125_{10} = 1110_2 + 0.001_2$ = 1110.001_2 = $= (-1)^0 \cdot 1.110001_2 \cdot 2^3$ $\Rightarrow \exp = 3 \Rightarrow \exp + 15 = 18 = 10010$ $\Rightarrow VZ = 0$ $\Rightarrow \text{fraction} = 110001_2 = 1100010000_2$

= 0100111100010000