

Rechnerstrukturen: Übungsblatt 6

Florian Ludewig (185722)

25. Juni 2020

Aufgabe 2

- 2) a) Der Exponent wird nicht im Zweierkomplement gespeichert, sondern als unsigned int. Deswegen muss vom Exponenten ein Bias abgezogen werden, dieser ist bei 32-BitFloat 127.

Bsp.: gespeicherter Exponent $1000\ 0000_2 \triangleq 128_{10}$

$$128 - \underbrace{127}_{\text{Bias}} = 1 \quad \curvearrowright \text{Exponent} = 1$$

- b) Die Darstellung ist nicht eindeutig. So ist z.Bsp.: $17 = 17 \cdot 10^0 = 0,017 \cdot 10^2$. Deswegen normalisiert man Mantisse, wobei das erste Bit vor dem Komma immer gesetzt ist. So ist z.B. $0.\underbrace{001}_{3}1011_2$ in normalisierter Form $1,011_2 \cdot 2^{-3}$.

- c) Da die normalisierte Mantisse immer eine "1" vor dem Komma steht wird diese einfach nicht gespeichert. Dieses Bit nennt man "versteckt".

Bsp.: $1.1001 \cdot 2^0$ wird als $\begin{array}{c} 1 \\ 0001 \end{array}$ gespeichert
diese 1 wird nicht gespeichert

e) $|m|=3 \quad |e|=3 \quad \curvearrowright \text{Bias}=3 \quad (2^3=8 \quad \curvearrowright \frac{8}{2}-1=3)$

$$\begin{array}{ll} \bullet 1: & 1_{10} = 1.000_2 \cdot 2^0 \\ & \hookrightarrow e = 3 - 0 = 3_{10} \triangleq 011_2 \\ & \underline{\underline{0|011|000}} \end{array} \quad \begin{array}{ll} \bullet 0: & \underline{\underline{0|000|000}} \\ & \bullet \infty, \quad \underline{\underline{0|111|000}} \end{array}$$

$$\begin{array}{ll} \bullet \frac{3}{4}: & \frac{3}{4} = 0,75_{10} \triangleq 0,11_2 \\ & \rightsquigarrow = 1,1 \cdot 2^1 \\ & e = 3 - 1 = 2 \triangleq 010_2 \\ & \underline{\underline{0|010|100}} \end{array} \quad \begin{array}{ll} \bullet -\frac{5}{4}: & \frac{5}{4} = 1,25 \triangleq 1,01_2 \cdot 2^0 \\ & e = 3 - 0 = 3 \triangleq 011_2 \\ & \underline{\underline{1|011|010}} \end{array}$$

2d)

- höchste positive Zahl: 0|11111110|11111111111111111111111111
- höchste negative Zahl: 1|11111110|11111111111111111111111111
- kleinste positive Zahl: 0|00000000|00000000000000000000000000000001

Aufgabe 3

```
1  unsigned i2f(int i) {  
2      if (i == 0) return 0;  
3      int abs = i < 0 ? -i : i;  
4  
5      int s = i > 0 ? 0b0 : 0x80000000;  
6  
7      int tmp = abs;  
8      int right_shifts = 0;  
9      while (0b01 != abs) {  
10         abs = abs >> 1;  
11         right_shifts++;  
12     }  
13     int e = (127 + right_shifts) << 23;  
14  
15     tmp = (tmp << (32 - right_shifts)) >> 9;  
16     int m = tmp & 0x007FFFFF;  
17  
18     return s | e | m;  
19 }
```