Tutorat 3

Kodierung von Zahlen, Zahlensysteme, Darstellung negativer Festkommazahlen

Gruppe 9

Präsentator:
Jürgen Mattheis
(juergmatth@gmail.com)

Vorlesung von:
Prof. Dr. Scholl

Übungsgruppenbetreuung:
Tobias Seufert

10. Juni 2023

Universität Freiburg, Lehrstuhl für Rechnerarchitektur

Gliederung

Organisatorisches

Aufgabe 1

Aufgabe 2

Aufgabe 3

Appendix

Literatur

Organisatorisches

Organisatorisches

Abgaben

- schreibt bitte euren Namen und Matrikelnummer auf die Abgaben
- Vorrechnen nicht vergessen:
  1. entweder generell immer mal wieder Melden, dann zählt das irgendwann als Vorrechnen
  2. oder vor dem Tutorat ansprechen, dann werdet ihr während des Tutorats dazu gefragt, ob ihr zu einer Aufgabe vielleicht irgendetwas gehaltvolles sagen könnt
- um die Leute, die in die Tutorate kommen zu belohnen und dem Ereignis entgegenzuwirken, dass das Tutorat irgendwann leer ist, werden die Folien während des Tutorats auf einem USB-Stick verteilt
  - die Folien aller bisherigen Tutorate werden immer auch alle auf dem USB-Stick sein

Aufgabe 1

Aufgabe 1 I

Zahlensysteme und Darstellung negativer Kommazahlen

Lösung 1.1

$$\begin{aligned} 2342_{10} &= 100100100110_2 \\ &= [0100100100110]_{BV} \\ &= [0100100100110]_1 \\ &= [0100100100110]_2 \end{aligned}$$

Aufgabe 1 II

Zahlensysteme und Darstellung negativer Kommazahlen

Lösung 1.2

$$
\begin{aligned}
-BFCD_{16} &= -1011111111001101_2 \\
&= [11011111111001101]_{BV} \\
&= [10100000000110010]_1 \\
&= [10100000000110011]_2
\end{aligned}
$$

Aufgabe 1 III

Zahlensysteme und Darstellung negativer Kommazahlen

Lösung 1.3

$$
\begin{aligned}
-10_{16} &= -10000_2 \\
&= [110000]_{BV} \\
&= [101111]_1 \\
&= [110000]_2
\end{aligned}
$$

Aufgabe 1 IV

Zahlensysteme und Darstellung negativer Kommazahlen

Lösung 1.4

$$
\begin{aligned}
255_8 &= 10101101_2 \\
&= [010101101]_{BV} \\
&= [010101101]_1 \\
&= [010101101]_2
\end{aligned}
$$

Anmerkungen

- Most Significant Bits, die 0 sind fallen weg

Aufgabe 2

Aufgabe 2 I

Darstellung von Festkommazahlen - Invertieren

Voraussetzungen 2.1

- *Geometrische Summenformel:* $\sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q}$
  - *wie im Beweis auf nächster Folie:* $\sum_{k=0}^{n-1} q^k = \frac{1-q^{n-1+1}}{1-q} = \frac{1-q^n}{1-q} \cdot \frac{-1}{-1} = \frac{q^n-1}{q-1}$
  - *für $q=2$:* $\sum_{k=0}^{n} 2^k = \frac{1-2^{n+1}}{1-2} = \frac{1-2^{n+1}}{-1} = 2^{n+1}-1$

Aufgabe 2 II

Darstellung von Festkommazahlen - Invertieren

Lösung 2.1

$$\text{ZZ}: \; [a']_2 + 1 + [a]_2 = 0$$

$$\begin{aligned}
[a']_2 + 1 + [a]_2 &= \sum_{i=0}^{n-1} a_i' 2^i - a_n' 2^n + 1 + \sum_{i=0}^{n-1} a_i 2^i - a_n 2^n \\
&= \sum_{i=0}^{n-1} (a_i' + a_i) 2^i - (a_n' + a_n) 2^n + 1 \\
&= \sum_{i=0}^{n-1} 1 \cdot 2^i - 1 \cdot 2^n + 1 \qquad \left(a_i' + a_i = 1 \text{ da } a_i' = a_i \text{ invertiert}\right) \\
&= \frac{2^n - 1}{2 - 1} - 2^n + 1 \\
&= 2^n - 1 - 2^n + 1 \\
&= 0
\end{aligned}$$

Aufgabe 3

Aufgabe 3 I

Darstellung von Festkommazahlen - Dezimalsystem

Lösung 3.1

- *Für $n = 4$: 09999, für $n > 0 : 10^n - 1$ (größte darstellbare Zahl)*
- $x = 10^n - 1,\ [d_n d_{n-1} \dots d_0]_9 = \sum_{i=0}^{n-1} d_i \cdot 10^i - d_n \cdot\ x$
- *Für einen symmetrischen Zahlenbereich kann man sich erstmal überlegen, dass man zwei Darstellung für die 0 braucht und auch die Negation der größten darstellbaren Zahl darstellbar sein muss. D.h. für z.B. $n = 4$ muss man auch -9999 darstellen können. Und diesen Grenzfall erreicht man, indem man von 0 9999 abzieht.*

Aufgabe 3.2

- Wie das Komplementieren der Ziffern in der Neuner-KomplementDarstellung definiert werden muss, damit das folgende Lemma gilt:
- Lemma: Sei a eine Festkommazahl im Dezimalsystem, a' die Festkommazahl im Dezimalsystem, die aus a durch Komplementieren aller Ziffern hervorgeht. Dann gilt $[a']_9 + [a]_9 = 0$.

Aufgabe 3 II

Darstellung von Festkommazahlen - Dezimalsystem

Lösung 3.2

- *Das Einerkomplent im Binärsystem interpretiert Zahlen mit führenden* 0*en als nichtnegative Zahlen und Zahlen mit führenden* 1*en als negative Zahlen. Um das Einerkomplement einer negativen Zahl in BV-Repräsentation zu bekommen werden alle Bits. invertiert.*
- *Analog dazu kann man das Neunerkomplement im Dezimalsystem bilden. In der Aufgabenstellung ist definiert, dass auch hier die höchstwertigste Ziffer nur* 0 *oder* 1 *sein kann.*
- *Das Invertieren wäre hier die Ersetzung nach folgender Funktion:*

$$inv : \{0,1,\ldots,8,9\} \to \{0,1,\ldots,8,9\} : d_i \mapsto \begin{cases} 9 - d_i & 0 \leq i < n \\ 1 - d_i & i = n \text{ und } d_i \in \{0,1\} \end{cases}$$

- *Für* 01784 *ist die komplementierte Zahl:* 18215

$$[01784]_0 + [18215]_9 = 1784 + (8215 - 9999) = (1784 + 8215) - 9999 = 9999 - 9999 = 0$$

Aufgabe 3 III

Darstellung von Festkommazahlen - Dezimalsystem

Lösung 3.3

- *Beim Zweier-Komplement wird im Gegensatz zum Einer-Komplement bei der Auswertung negativer Zahlen 1 mehr abgezogen.*
- *Analog dazu unterscheidet sich das Zehner-Komplement auch vom Neuner-Komplement durch die zusätzliche Subtraktion von* 1 *bei der Auswertung negativer Zahlen.*
- *Die Argumentation ist, dass die Darstellung von* -0 *zu* -1 *geshiftet wird und somit keine doppelte Darstellung der* 0*, sondern eine durchgehende Darstellung gewährleistet ist. Es wird bei negativen Zahlen also* $y = 10^n$ *statt* $x = 10^n - 1$ *abgezogen.*

Aufgabe 3 IV

Darstellung von Festkommazahlen - Dezimalsystem

Lösung 3.3

- *Beweis:*

$$
\begin{aligned}
[a']_{10} + [a]_{10} + 1 &= \sum_{i=0}^{n-1} a'_i b^i - a'_n 10^n + \sum_{i=0}^{n-1} a_i b^i - a_n b^n + 1 \\
&= \sum_{i=0}^{n-1} (a'_i + a_i) 10^i - (a'_n + a_n) 10^n + 1 \\
&= \sum_{i=0}^{n-1} (9) \cdot 10^i - (1) \cdot 10^n + 1 \\
&= 9 \cdot \sum_{i=0}^{n-1} 10^i - 10^n + 1 \\
&= 9 \cdot \frac{10^n - 1}{10 - 1} - 10^n + 1 \\
&= 0
\end{aligned}
$$

Aufgabe 3 V

Darstellung von Festkommazahlen - Dezimalsystem

Aufgabe 3.4

- Wie muss man das Komplementieren definieren bei allgemeiner Basis b, sodass $[a']_{b-1} + [a]_{b-1} = 0$ sowie $[a']_b + [a]_b + 1 = 0$ gelten?

Aufgabe 3 I

Darstellung von Festkommazahlen - Dezimalsystem

Lösung 3.4

- *Damit die Lemmata gelten für beliebige b, muss das Komplementieren einer Ziffer* $d_i (i = 0, \dots, n-1)$ *für die Basis b als die Differenzrechnung* $(b-1) - d_i$ *definiert werden.*
- *Sowohl in Lemma 1 als auch in Lemma 2 kann man dann* $10^{i/n}$ *durch* $b^{i/n}$ *ersetzen, wie im folgenden zu sehen ist:*

Aufgabe 3 II

Darstellung von Festkommazahlen - Dezimalsystem

Lösung 3.4

- *1er-Komplement:*

$$
\begin{aligned}
[a']_{b-1} + [a]_{b-1} &= \sum_{i=0}^{n-1}((b-1) - a_i)b^i - (1 - a_n)(b^n - 1) + \sum_{i=0}^{n-1} a_i b^i - a_n(b^n - 1) \\
&= \sum_{i=0}^{n-1}((b-1) - a_i + a_i)b^i - (1 - a_n + a_n)(b^n - 1) \\
&= \sum_{i=0}^{n-1}(b-1) \cdot b^i - (1) \cdot (b^n - 1) \\
&= (b-1) \cdot \sum_{i=0}^{n-1} b^i - (b^n - 1) \\
&= (b-1) \cdot \frac{b^n - 1}{b - 1} - b^n + 1 \\
&= 0
\end{aligned}
$$

Aufgabe 3 III

Darstellung von Festkommazahlen - Dezimalsystem

Lösung 3.4

- *2er-Komplement:*

$$
\begin{aligned}
[a']_b + [a]_b + 1 &= \sum_{i=0}^{n-1}((b-1) - a_i)b^i - (1 - a_n)b^n + \sum_{i=0}^{n-1} a_i b^i - a_n b^n + 1 \\
&= \sum_{i=0}^{n-1}((b-1) - a_i + a_i)b^i - (1 - a_n + a_n)b^n + 1 \\
&= \sum_{i=0}^{n-1}(b-1) \cdot b^i - (1) \cdot b^n + 1 \\
&= (b-1) \cdot \sum_{i=0}^{n-1} b^i - b^n + 1 \\
&= (b-1) \cdot \frac{b^n - 1}{b - 1} - b^n + 1 \\
&= 0
\end{aligned}
$$

Aufgabe 3 IV

Darstellung von Festkommazahlen - Dezimalsystem

Anmerkungen

- Man sieht das Komplementieren funktioniert unabhängig von 1er- oder 2er-Komplement gleich, da der Unterschied zwischen beiden Darstellungen nur darin liegt, dass man entweder $b^n - 1$ (1er-Komplement) oder b^n (2er-Komplement) subtrahiert um die Symmetrie bei zwei 0en (1er-Komplement) oder einer 0 (2er-Komplement) herzustellen und daher zum Ausgleich bei der einen Darstellung nichts dazuaddieren muss (1er-Komplement) oder 1 dazuaddieren (2er-Komplement) muss um bei der Addition einer Zahl mit ihrem Komplement 0 zu erhalten.
- Man muss -1 rechnen, da es zwar 10 Zeichen gibt, aber die Zahl mit der höchsten Wertigkeit, also die Zahl 9 auch nur die Wertigkeit 9 hat und nicht 10, da die 0 die Anzahl ist, wo nichts zählbares vorliegt.

Appendix

Appendix I

Zahlendarstellungen zur Basis b - Stellenwertsysteme

- Zahlensytem $S = (b, Z, \delta)$
  - Zahlensystem wird durch Anhängen der Basis als Index an die Ziffernfolge ${d_{n-1} \dots d_0}_b$ vermittelt
  - Basis $b \in \mathbb{N}, b \rangle 1$ mit welcher für jede Stelle i der Stellenwert b^i berechnet wird
  - Ziffernmenge Z
  - Ziffernwertigkeit δ ordnet jeder Ziffer bzw. Symbol ihre Wertigkeit zu

Appendix II

Zahlendarstellungen zur Basis b - Stellenwertsysteme

Anmerkungen

- bei der n-stelligen Binärdarstellung einer Zahl werden dem LSB und MSB jeweils die Stellenwerte b^0 und b^{n-1} zugeordnet
- Im Binärsystem / Dualsystem werden den beiden Ziffern 0 und 1 jeweils die Wertigkeiten 0 und 1 zugeordnet
- Im Dezimalsystem werden den zehn Ziffern $0, 1, 2, 3, 4, 5, 6, 7, 8$ und 9 jeweils die Wertigkeiten 0 bis 9 in der konventionellen Reihenfolge zugeordnet
- Im Hexadezimalsystem werden den sechzehn Ziffern $0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E$ und F jeweils die Werte der Dezimalzahlen von 0 bis 15 zugeordnet.
  - Eselsbrücken: Cwölf, Dreizehn, Fünfzehn

- positiver Wert $\langle d \rangle = \sum_{i=0}^{n-1} \delta(d_i) \cdot b^i$ einer nicht-negativen Natürlichen Zahl, wobei $d = d_{n-1} \ldots d_0$ mit $d_i \in Z$ eine Folge von n Ziffern bzw. Symbolen ist

Appendix I

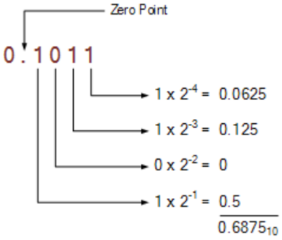
Zahlendarstellung zur Basis b - Stellenwertsysteme

- positiver Wert $\langle d \rangle = \sum_{i=-k}^{n-1} \delta(d_i) \cdot b^i$ einer nicht-negativen Festkommazahl, wobei $d = d_{n-1} \dots d_0 \dots d_{-k}$ mit $d_i \in Z$
  - die Anzahl der Nachkommastellen ist fest: $\langle d_{n-1} \dots d_0 d_{-1} \dots d_{-k} \rangle \cdot 2^{-k} = \langle d_{n-1} \dots d_0 . d_{-1} \dots d_{-k} \rangle$
  - Beispiel 3-Bit Festkommazahlen mit $n = 1$ und $k = 2$:

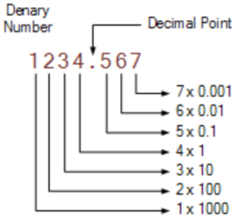
d	0.00	0.01	0.10	0.11	1.00	1.01	1.10	1.11
$\langle d \rangle$	0.0	0.25	0.5	0.75	1.0	1.25	1.5	1.75

Appendix II

Zahlendarstellung zur Basis b - Stellenwertsysteme



(a) *Binärerbrüche*



(b) *Dezimalbrüche*

- Darstellung negativer Festkommazahlen, wobei $d = d_n d_{n-1} \dots d_0 \dots d_{-k}$ mit $\forall i \langle n : d_i \in Z$ und $d_n \in \{0, 1\}$:

Appendix III

Zahlendarstellung zur Basis b - Stellenwertsysteme

- potentiell negativer Wert $[d]_{BV} = (-1)^{d_n} \sum_{i=0}^{n-1} \delta(d_i) 2^i$ in Darstellung durch Betrag und Vorzeichen
  - Beispiel 3-Bit Festkommazahlen mit Vorzeichenbit, $n = 1$ und $k = 1$:

d	11.1	11.0	10.1	10.0	00.0	00.1	01.0	01.1
$[d]_{BV}$	-1.5	-1.0	-0.5	0.0	0.0	0.5	1.0	1.5

- potentiell negativer Wert $[d]_1 = \sum_{i=0}^{n-1} \delta(d_i) 2^i - \delta(d_n)(2^n - 2^{-k})$ in Einerkomplement-Darstellung
  - Beispiel 3-Bit Festkommazahlen mit Vorzeichenbit, $n = 1$ und $k = 1$:

d	10.0	10.1	11.0	11.1	00.0	00.1	01.0	01.1
$[d]_1$	-1.5	-1.0	-0.5	0.0	0.0	0.5	1.0	1.5

Appendix IV

Zahlendarstellung zur Basis b - Stellenwertsysteme

- im Negativen hat eine Folge mit mehr 1en anders als im Positiven betragsmäßig eine kleineren Wert, denn umso mehr 1en da sind, umso größer wird die Summe $\sum_{i=0}^{n-1} \delta(d_i)2^i$ und umso mehr kann von der subtrahierten größten positiven Zahl $2^n - 1$ ausgeglichen werden
- um den Wert einer negativen Dezimalzahl binär darzustellen gibt es 2 Methoden
  1. mit größtmöglichen positven Zahl anfangen $-(2^n - 1)$ und überlegen, welche 2er Potenzen (Stellenwerte) man draufaddieren muss, um die gewünschte Zahl zu erhalten und an den entsprechenden Bits, die diesen 2er Potenzen entsprechen 1en setzen
  2. die passende Ziffernfolge wie gewohnt erstellen, aber mit vertauschten 1en und 0en und das Vorzeichenbit ist eine 1

- potentiell negativer Wert $[d]_2 = \sum_{i=0}^{n-1} \delta(d_i)2^i - \delta(d_n)2^n$ in Zweierkomplement-Darstellung

Appendix V

Zahlendarstellung zur Basis b - Stellenwertsysteme

- Beispiel 3-Bit Festkommazahlen mit Vorzeichenbit, $n = 1$ und $k = 1$:

d	10.0	10.1	11.0	11.1	00.0	00.1	01.0	01.1
$[d]_2$	-2	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5

- genauso wie beim Einerkomplment bedeuten mehr 1en im Negativen einen betragsmäßig kleineren Wert
- um den Wert einer negativen Dezimalzahl binär darzustellen gibt es 2 Methoden
  1. mit der größtmöglichen positven Zahl + kleinmöglichen positven Zahl ungleich 0 anfangen $-(2^n)$ und überlegen, welche 2er Potenzen (Stellenwerte) man draufaddieren muss, um die gewünschte Zahl zu erhalten und an den entsprechenden Bits, die diesen 2er Potenzen entsprechen 1en setzen

Appendix VI

Zahlendarstellung zur Basis b - Stellenwertsysteme

2. die passende Ziffernfolge wie gewohnt erstellen, aber die kleinstmögliche positve Zahl ungleich 0 wird als Startwert genommen und die 1en und 0en sind vertauscht, sowie das Vorzeichenbit ist eine 1

- wenn man die kleinste negative Zahl komplementiert: $[10.0]_2' = [01.1]_1 + 0.5 = [10.0]_2$, dann erhält man erneut die kleinste negative Zahl. Das passt auch ganz gut, da die kleinste negative Zahl im Zweierkomplement keine komplementäre positive Zahl hat. Folglich auch hier: $[a]_2 + [a']_2 + 2^{-k} = [1.00]_2 + [0.11]_2 + 0.5 = -2 + 1.5 + 0.5 = 0$ für a kleinste Zweierkomplement-Zahl

Appendix VII

Zahlendarstellung zur Basis b - Stellenwertsysteme

Anmerkungen

- ein weiterer Vorteil ist, dass die Einerkomplement- und Zweierkomplement-Darstellung zyklisch sind:

$$[01.1]_2 + 0.5 = [10.0]_2 \qquad \text{(nach 1.5 geht es mit } -2 \text{ weiter)}$$
$$[01.1]_1 + 0.5 = [10.0]_1 \qquad \text{(nach 1.5 geht es mit } -1.5 \text{ weiter)}$$

- Darstellung negativer Gleitkommazahlen, wobei $\underbrace{d_n}_{sign}\ \underbrace{d_{n-1}\dots d_0}_{exponent}\ \underbrace{d_{-1}\dots d_{-k}}_{fraction\,/\,mantissa}$ mit exponent, fraction bits und sign bit aus $\{0,1\}$:
  - varierende Anzahl von Nachkommastellen im Gegensatz zu Festkommazahlen

Appendix VIII

Zahlendarstellung zur Basis b - Stellenwertsysteme

- Normalisierte Zahlen:
  $(-1)^{d_n} \cdot \langle 1d_{-1} \dots d_{-k} \rangle \cdot 2^{-k+\langle d_{n-1} \dots d_0 \rangle - (2^{n-1}-1)} = (-1)^{sign} \times (1 + fraction) \times 2^{exponent-bias}$
  - Bias: $2^{n-1} - 1 = \frac{2^2}{2} - 1 = 2 - 1 = \boxed{01_2}$, größter, kleinster Exponent: $11_2 - 01_2 = \boxed{10_2}$, $01_2 - 01_2 = \boxed{00_2}$
  - Betragsmäßig kleinste, größte Zahl: $\pm 1.0 \times 2^{1-1} = \pm 1.0 \; (\boxed{\frac{0}{1} \mid 01 \mid 0})$, $\pm 1.5 \times 2^{2-1} = \pm 3 \; (\boxed{\frac{0}{1} \mid 10 \mid 1})$
- Denormalisierte Zahlen:
  $(-1)^{d_n} \cdot \langle 0d_{-1} \dots d_{-k} \rangle \cdot 2^{-k+\langle d_{n-1} \dots d_0 \rangle - (2^{n-1}-2)} = (-1)^{sign} \times (0 + fraction) \times 2^{-bias}$

Appendix IX

Zahlendarstellung zur Basis b - Stellenwertsysteme

- Bias: $2^{n-1} - 2 = \frac{2^2}{2} - 2 = 2 - 2 = \boxed{00_2}$, einziger Exponent: $00_2 - 00_2 = \boxed{00_2}$
- Betragsmäßig kleinste, größte Zahl: $0.0 \times 2^{0-0} = 0$ $\left(\boxed{\frac{0}{1} \mid 00 \mid 0}\right)$, $\pm 0.5 \times 2^{0-0} = \pm 0.5$ $\left(\boxed{\frac{0}{1} \mid 00 \mid 1}\right)$

Anmerkungen

- $1.d_{-1} \ldots d_{-k}$ wird als normalized significand bezeichnet
- $0.d_{-1} \ldots d_{-k}$ wird als normed significand bezeichnet

- der Exponent ist immer als nicht-negative Natürliche Zahl zu interpretieren, die Mantissa ist immer als nicht-negativer Bruch zu interpretieren, also mit 2^{-k} zu multiplizieren

Appendix X

Zahlendarstellung zur Basis b - Stellenwertsysteme

- der Bias macht den *encoded_exponent* immer positiv
  - $encoded\_exponent = real\_exponent + bias$
  - $real\_exponent = encoded\_exponent - bias$
- Zero: $\boxed{\frac{0}{1} \mid 00 \mid 0}$, Infinity: $\boxed{\frac{0}{1} \mid 11 \mid 0}$, NaN: $\boxed{\frac{0}{1} \mid 11 \mid 1}$
- Beispiel 4-Bit Gleitkommazahl mit Vorzeichenbit, 2 Exponentbits und Mantissabit:

d	1 | 00 | 0	1 | 00 | 1	1 | 01 | 0	1 | 01 | 1	1 | 10 | 0	1 | 10 | 1	1 | 11 | 0	1 | 11 | 1
$[d]_{GK}$	0.0	-0.5	-1.0	-1.5	-2.0	-3.0	$-\infty$	NaN
d als BV	0.0	-0.1	-1.0	-1.1	-10.0	-11.0	-	-

Appendix XI

Zahlendarstellung zur Basis b - Stellenwertsysteme

d	0 | 00 | 0	0 | 00 | 1	0 | 01 | 0	0 | 01 | 1	0 | 10 | 0	0 | 10 | 1	0 | 11 | 0	0 | 11 | 1
$[d]_{GK}$	0.0	0.5	1.0	1.5	2.0	3.0	∞	NaN
d als BV	0.0	0.1	1.0	1.1	10.0	11.0	-	-

- Beispiel ohne angepassten Bias (-2) für Denormalisierte Zahlen mit 4-Bit Gleitkommazahl mit Vorzeichenbit, 2 Exponentbits und Mantissabit:

d	0 | 00 | 0	0 | 00 | 1	0 | 01 | 0	0 | 01 | 1	0 | 10 | 0	0 | 10 | 1	0 | 11 | 0	0 | 11 | 1
$[d]_{GK}$	0.0	0.25	1.0	1.5	2.0	3.0	∞	NaN
d als BV	0.0	0.01	1.0	1.1	10.0	11.0	-	-

- Beispiel ohne um -1 geshifteten Bias mit 4-Bit Gleitkommazahl mit Vorzeichenbit, 2 Exponentbits und Mantissabit:

d	0 | 00 | 0	0 | 00 | 1	0 | 01 | 0	0 | 01 | 1	0 | 10 | 0	0 | 10 | 1	0 | 11 | 0	0 | 11 | 1
$[d]_{GK}$	0.0	0.25	0.5	0.75	1.0	1.5	∞	NaN
d als BV	0.0	0.01	0.1	0.11	1.0	1.1	-	-

Appendix XII

Zahlendarstellung zur Basis b - Stellenwertsysteme

- es gibt mehr positive Exponenten als negative Exponenten
- die Zahl 1.0 und damit viele Festkommazahlen mit so vielen Mantissa-Bits, wie für die Gleitkommazahl zu Verfügung stehen sind sehr einfach zu kodieren, da $(1 + fraction) \cdot 2^{\sum_{i=0}^{n-2} 1 \cdot 2^i - (2^{n-1} - 1)} = (1 + fraction) \cdot 2^{2^{n-1} - 1 - 2^{n-1} + 1} = (1 + fraction) \cdot 2^0 = (1 + fraction)$

- es gibt verschiedene Standards, wie Bfloat16 (1 sign bit, 8 exponent bits, 7 fraction bits), Single-precision (1 sign bit, 8 exponent bits, 23 fraction bits) und Double-precision (1 sign bit, 11 exponent bits, 52 fraction bits), die sich in der Anzahl der Bits für Exponent und Mantissa unterscheiden
- Runden mit GRS-Bits:
  - mit GRS braucht man nur 3 zusätzliche Bits und braucht so keinen großen und langsamen Addierer mit dem man die Berechnungen erstellt. Ohne GRS müsste man die Berechnungen mit allen Bits machen und am Ende runden

Appendix XIII

Zahlendarstellung zur Basis b - Stellenwertsysteme

- Guard und Round bit: Zwei zusätzliche Bits die bei Berechnungen mit Gleitkommazahlen rechts angehängt werden, um die Rundungsgenauigkeit zu erhöhen
- Sticky bit: Zusätzliches Bits rechts der Bits Guard und Round, dass gesetzt wird, wann immer es Bits rechts davon gibt, die nicht 0 sind
- Ties to even: Numbers exactly in the middle between two integer numbers („ties") are rounded towards the even number

$$0.5 \rightarrow 0,$$
$$1.5 \rightarrow 2,$$
$$2.5 \rightarrow 2$$

G	R	S	Ergebnis
		1/8	0.125
	1/4		0.25
	1/4	1/8	0.375
1/2			0.5
1/2		1/8	0.625
1/2	1/4		0.75
1/2	1/4	1/8	0.875

Appendix XIV

Zahlendarstellung zur Basis b - Stellenwertsysteme

Anmerkungen

- the Anordnung, dass der Exponent immer vor der Mantissa steht und der Fakt, dass der kodierte Exponent positiv ist, ist so gewählt, damit man Gleitkommazahlen möglichst einfach vergleichen kann
- Festkommazahlen haben eine kleinere Repräsentationsspanne, da sie nur eine feste Anzahl an Nachkommastellen haben. Gleitkommazahlen sind nicht gleichmäßig verteilt, man hat eine sehr hohe Dichte kleiner Zahlen nahe der 0, neben einem schmallen Bereich nahe der 0 mit gar keinen Zahlen.

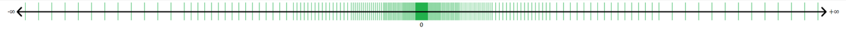


Abbildung 2: Verteilung von Gleitkommazahlen

$d[i]$	0.101	1.01	10.1	101
$\langle d[i] \rangle$	0.625	1.25	2.5	5
$\langle d[i] \rangle - \langle d[i-1] \rangle$	0.3125	0.625	1.25	2.5

Tabelle 1: Abstände erhöhen sich exponentiel

Appendix I

Hexadezimalsystem zu Binärsystem

$$
\begin{aligned}
ba_{16} &= b_{16} \cdot 16^1 + a_{16} \cdot 16^0 \\
&= 13_8 \cdot 16^1 + 12_8 \cdot 16^0 \\
&= (1_8 \cdot 8^1 + 3_8) \cdot 16^1 + (1_8 \cdot 8^1 + 2_8) \cdot 16^0 \\
&= (1_8 \cdot 8^1 + 3_8) \cdot (8 \cdot 2)^1 + (1_8 \cdot 8^1 + 2_8) \cdot (8 \cdot 2)^0 \\
&= 2_8 \cdot 8^2 + 6_8 \cdot 8^1 + 1_8 \cdot 8^1 \cdot (8 \cdot 2)^0 + 2_8 \cdot (8 \cdot 2)^0 \\
&= 2_8 \cdot 8^2 + 7_8 \cdot 8^1 + 2_8 \cdot 1 = 272_8 \\
ba_{16} &= b_{16} \cdot 16^1 + a_{16} \cdot 16^0 \\
&= 1011_2 \cdot (2^4)^1 + 1010_2 \cdot (2^4)^0 = 10111010_2
\end{aligned}
$$

Literatur

Online

[1] *Zahlendarstellung Zur Basis*. URL: https://www.inf.hs-flensburg.de/lang/informatik/zahlendarstellung.htm (besucht am 26. 05. 2023).