Gleitkommazahlen

Um die Einschränkung der Festkommazahlen aufzuheben, dass dort nur Brüche mit 2er-Potenmzen im Nenner exakt dargestellt werden können benötigt man eine weitere Zahlendarstellung. Diese funktioniert allerdings grundlegend anders als die bisherige binäre Zahldarstellung (Binär, Zweierkomplement, Festkommazahlen). Damit ein Computer mit solchen "neuen" Zahlen rechnen kann, benötigt man eine angepasste elektronische Schaltung, die sogenannte FPU [1)].

Vorüberlegungen im Dezimalsystem

Jede endliche Dezimalzahl kann man in der Form $M \cdot 10^{E}$ schreiben.

dabei heißt **M "Mantisse"** und E **"Exponent"**.

- Die Mantisse ist eine Dezimalzahl, die stets eine Ziffer zwischen 1 und 9 vor dem Komma besitzt.
- Der Exponent ist eine Ganze Zahl

Beispiele:

- $435{,}861 = 4{,}35861 * 10^{2}$ (Mantisse 4,25861, Exponent 2)
- $0{,}0003562 = 3{,}562 * 10^{4}$ (Mantisse 3,562, Exponent 4)

Im Binärsystem (IEEE 754)

Die Norm IEEE 754 (ANSI/IEEE Std 754-1985; IEC-60559:1989 – International version) definiert Standarddarstellungen für binäre Gleitkommazahlen in Computern und legt genaue Verfahren für die Durchführung mathematischer Operationen, insbesondere für Rundungen, fest.

Die Darstellung einer Gleitkommazahl $\mathbf{x = s \cdot m \cdot b^{e}}$

besteht aus:

- Vorzeichen s (1 Bit)
- Mantisse m (p Bits)
- Basis b (bei normalisierten Gleitkommazahlen nach IEEE 754 ist b = 2)
- Exponent e (r Bits)

Ein wesentlicher Unterschied zur Dezimadarstellung ist also, dass Mantisse, Exponent binär dargestellt werden, und die Basis der Potenz 2 ist. Das Vorzeichenbit ist so definiert, dass s=0 positive Zahlen darstellt, s=1 negative Zahlen.

An das **Mantissenbitmuster wird links stets eine 1 angehängt: 1,m**. Das Bitmuster der Mantisse m steht also nur für negative 2er-Potenzen. Damit spart man sich ein Bit Speicherplatz (und gewinnt damit ein Bit an Genauigkeit) weil man sich diese erste 1 immer "denkt".

Durch die Normalisierung kann man wie bei den dezimalen Fließkommazahlen diesen Zustand immer erreichen, so dass die Mantisse mit der vorangestellten 1 auf der Stelle 2^0 immer zwischen 1 und 2 ist - so wie sie bei den Dezimalzahlen immer zwischen 1 und 9 sein kann.

Der Exponent wird aus technischen Gründen mit einem **Bias** versehen, so dass das Exponentenbitmuster immer eine positive Zahl darstellt.

Beispiel: Wenn wer Exponent 8 Bit lang ist, kann man damit als Zweierkomplement Werte zwischen -126 und 127 für den Exponenten darstellen. Damit man sich die Komplementdartstellung spart, wird festgelegt, dass das Bitmuster für den Exponenten ermittelt wird, indem zum echten Wert von e der **Bias** 127 addiert wird.

Man arbeitet dann mit Bitmustern für Werte zwischen 1 und 254 und kann auf das Zweierkomplement verzichten, wenn man weiß, dass man umgekehrt wieder 127 von einem errechneten Bitmuster subtrahieren muss, um den "echten" Wert des Exponenten zu erhalten. Diese Vorgehen vereinfacht den Entwurf der elektronischen Schaltungen für diese Rechenvorgänge.

Typ	Größe (1+r+p)	Exponent (r)	Mantisse (p)	Werte des Exponenten (e)	Biaswert (B)
single	32 bit	8 bit	23 bit	$-126 \leq e \leq 127$	127
double	64 bit	11 bit	52 bit	$-1022 \leq e \leq 1023$	1023
quadruple	128 bit	15 bit	112 bit	$-16382 \leq e \leq 16383$	16383

2)

1)

Floating Point Unit, siehe auch https://de.wikipedia.org/wiki/Gleitkommaeinheit

2)

Quelle: https://de.wikipedia.org/wiki/IEEE_754

From:
https://www.info-bw.de/ -

Permanent link:
https://www.info-bw.de/faecher:informatik:oberstufe:codierung:zahlendarstellungen:gleitkommazahlen:start?rev=1663169324

Last update: **14.09.2022 15:28**