

Übung 4

Armin Kleinert und Ruth Höner zu Siederdisen

Aufgabe 1: Zahlendarstellung und Rechnen

1. *Warum werden i. d. R. Fließkommazahlen und nicht Festkommazahlen zur Darstellung von reellen Zahlen verwendet? Anders formuliert: Welche Vorteile bringt diese Art der Darstellung?*

Mit Hilfe von Fließkommazahlen kann ein größerer Wertebereich dargestellt werden. Das bedeutet es sind sehr große positive und negative Zahlen, sowie sehr kleine positive und negative Zahlen darstellbar. Zusätzlich ist die Durchführung der Division mit Fließkommazahlen schneller. Das Fließkomma-Format kennt zudem die Werte NaN (not a number, erzeugt wenn die Teilung $0 / 0$ stattfindet) sowie +Infinity und -Infinity, wenn eine Teilung durch 0 stattfindet oder ein Overflow geschieht. Bei einem Underflow kann außerdem das „Sticky Bit“ gesetzt werden, das dann für die Fehlerkontrolle genutzt werden kann.

2. *Für eine Fließkommazahl-Darstellung werden die gegebenen Bits in drei Abschnitte unterteilt. Wie heißen diese Abschnitte?*

Die Bits werden aufgeteilt in Sign (Vorzeichen), in Characteristic (Exponent) und Mantisse (Zahl/Wert). Je nachdem wie viele Bits für die Darstellung zur Verfügung stehen, werden diese auf die drei Abschnitte verteilt. Das Sign ist dabei immer das erste Bit und besteht auch immer nur aus einem Bit. Darauf folgen die Bits der Characteristic und dann die der Mantisse.

3. *Welchen Vorteil bringt es wenn dem Exponenten mehr Bits zugeteilt werden bzw. welchen Vorteil bringt es wenn der Mantisse mehr Bits zugeteilt werden?*

Wenn dem Exponenten mehr Bits zugeteilt werden, wird der Wertebereich größer, den man mit der Anzahl an Bits darstellen kann. Werden der Mantisse mehr Bits zugeteilt, erhöht sich die Genauigkeit eines dargestellten Wertes, da mehr Nachkommastellen zur Verfügung stehen.

4. *Erklären Sie Unterlauf (Underflow) und Überlauf (Overflow) für Fließkommazahlen.*

Bei einem Unterlauf (Underflow) kommt die Zahl bei einer Berechnung in einen Bereich zwischen der 0 und der kleinsten darstellbaren positiven oder negativen Zahl. Das bedeutet die Zahl kann nicht dargestellt werden und es würde eine 0 zurückgegeben werden.

Bei einem Überlauf (Overflow) kommt die Zahl bei einer Berechnung in den Bereich, der größer ist als die größte darstellbare positive oder negative Zahl. Hier wird ein Overflow Flag gesetzt und als Ergebnis positiv oder negativ \pm „Unendlich“ zurückgegeben.

5. *Nennen Sie Beispiele für Festlegungen, die der IEEE-754 Standard mitbringt. Warum ist eine solche Standardisierung sinnvoll?*

Der IEEE-754 Standard gibt vor, wie viele Bits für Charakteristik und Mantisse verwendet werden je nachdem, wie viele Bits zur Verfügung stehen. Bei 32 Bits wird zum Beispiel für das Sign ein Bit verwendet, für die Charakteristik 8 Bits und für die Mantisse 23 Bits.

Laut dem Standard, soll die Mantisse der Fließkommazahl immer das Format 1,... haben. Dadurch ist das erste Bit der Mantisse implizit und muss nicht mehr dargestellt werden (Hidden Bit). Diese

Darstellung wird als normalisiert bezeichnet. Der Standard legt damit auch fest wo genau sich das Komma befindet. Nämlich links vom Most Significant Bit der Mantisse.

Wenn die Charakteristik 0 ist, werden die Fließkommazahlen denormalisiert dargestellt. Das bedeutet, das Hidden Bit muss ebenfalls explizit dargestellt werden. So ist auch die Zahl 0 im Fließkommaformat darstellbar.

Zusätzlich bietet der Standard Regeln für Sonderfälle. Wenn zum Beispiel alle Bits der Charakteristik 1 sind und die der Mantisse alle 0 wird ein Overflow signalisiert.

Bei Fließkommazahlen hat man normalerweise verschiedene Möglichkeiten ein und dieselbe Zahl darzustellen. Durch diese Festlegungen werden Fließkommazahlen immer im selben Format dargestellt, sodass die gleiche Folge von Bits auf jedem Rechner dieselbe Zahl repräsentiert. Zusätzlich wird dadurch das Rechnen mit Fließkommazahlen vereinfacht und eine Umsetzung direkt in der Hardware ermöglicht. Durch das Hidden Bit hat man ein Bit mehr für die Nachkommastellen zur Verfügung und kann Zahlen genauer darstellen. Durch die Sonderregelungen, wie die Charakteristik und die Mantisse aussehen sollen, kann man Exceptions signalisieren und die 0 ebenfalls darstellen.

6. *Was ist die betragsmäßig größte bzw. kleinste darstellbare Zahl im IEEE-754 32 bit Standard?*

Die betragsmäßig größte darstellbare Zahl ist $\pm 3.4 \times 10^{38}$. Im normalisierten Format ist die betragsmäßig kleinste darstellbare Zahl $\pm 1.18 \times 10^{-38}$ und im denormalisierten Format ist die kleinste darstellbare Zahl $\pm 1,4 \times 10^{-45}$.

Quellen

- Vorlesungsfolien
- https://en.wikipedia.org/wiki/Arithmetic_underflow
- https://de.wikipedia.org/wiki/IEEE_754
- <https://youtu.be/o4-CwDo2zpg> (Hier wird die Geschwindigkeit der Division von Floating Point Zahlen erwähnt)