Aufgabe 45

a) Bei der *scientific notation* wird die Zahl derart umgeformt, dass sie das Produkt einer Zehnerpotenz und einer Zahl ist, die nur eine Ziffer vor dem Komma hat.
Eine normalisierte Gleitkommazahl entspricht im wesentlichen der scientific notation, lediglich darf die einzelne Ziffer vor dem Komma keine Null sein.

b) *Hinweis:* Ich verwende die englische Zahlendarstellung, d.h. ein Punkt entspricht dem deutschen Komma in einer Zahl.

dezimal: $2^{17} = 1.31072 \cdot 10^5$ $\qquad \frac{1}{20} = 5.0 \cdot 10^{-2}$

binär: $2^{17_{10}} = 1.0 \cdot 2^{10001_2}$ $\qquad \frac{1}{20} \approx \overline{1.100} \cdot 2^{-5}$

c) $-8.75 = -\frac{35}{4} = -1000.11_2 = (-1)^1 \cdot 1.00011_2 \cdot 2^{130-127} =$

1	10000010	00011000000000000000000

$\frac{1}{20} \approx \overline{1.100} \cdot 2^{-5} = (-1)^0 \cdot \overline{1.100} \cdot 2^{122-127} =$

0	01111010	10011001100110011001101

d) Die Null wird als Gleitkommazahl in der Form sign=exponent=fraction=0 dargestellt, d.h. besteht auf Bitebene nur aus Nullen.

e) Bei der Gegenüberstellung der Eigenschaften muss man noch ergänzen, dass die IEEE-Norm den Zahlenbereich der Exponenten einschränkt, um noch Ausnahmen darstellen und mit nicht-darstellbare Zahlen, wie z.B. ∞, arbeiten zu können. Ich gebe hier aber trotzdem die theoretischen Grenzen an (die kleinste bzw. größte negative Zahl ergibt sich durch Änderung von *sign* in 1, sie entsprechen in ihrem Absolutbetrag jedoch den hier angegebenen Werten):

Typ	verwendete Bits	kleinste positive Zahl	größte positive Zahl
einfach genau	32	$2^{-128} \approx 2.9 \cdot 10^{-39}$	$(2-2^{-23}) \cdot 2^{127} \approx 1.7 \cdot 10^{39}$
doppelt genau	64	$2^{-1024} \approx 5.5 \cdot 10^{-309}$	$(2-2^{-52}) \cdot 2^{1023} \approx 1.8 \cdot 10^{308}$

Aufgabe 46

Die Addition von Gleitkommazahlen ist *nicht immer* assoziativ. Zwar ist generell die Addition zweier ganzer Zahlen (solange der Wertebereich nicht verlassen wird) assoziativ, jedoch müssen Gleitkommazahlen zur Darstellung im Computer Rundungsprozessen unterworfen werden. Diesen Einfluss zeige ich anhand eines Beispiels:

Es seien zwei Gleitkommaahlen *a* und *b* gegeben. Dabei sei *b* eine derart kleine Zahl, dass sie, wenn ihre Zehnerpotenz an die von *a* angepasst wird, der *fraction*-Anteil zu Null wird. Dann gilt: $a+b=a$. Für die Gleichung $-a+a+b$ ergibt sich dann ein Widerspruch:

$$-a + (a+b) \neq (-a+a) + b$$
$$-a + a \neq 0 + b$$
$$0 \neq b$$

Da *b* zwar klein ist, aber immer noch deutlich von Null verschieden, kann die Addition in diesem Fall nicht assoziativ sein.
Die anfangs verwendete Formulierung nicht immer bedeutet, dass es Beispiele gibt, wo die Addition tatsächlich assoziativ ist, allerdings gibt es deutlich mehre Gegenbeispiele.

Aufgabe 47

Hinweis: Ich verwende öfters das Wort *fraction*, welche ein Synonym für Signifikant ist. Mir ist ersteres lieber, da sonst Verwechslungsgefahr mit dem Vorzeichen *sign* besteht.

a) Gleitkommaaddition:
1. Exponentenanpassung (die Zahl mit dem kleineren Exponenten wird der Zahl mit dem größeren durch Shiften des Dezimalpunktes angepaßt, bis beide Exponenten übereinstimmen)
2. stellenrichtige Addition der *fraction*-Anteile
3. Normalisieren (Exponenteninkrementation bzw. -dekrementation, so dass nur noch eine Ziffer links vom Komma steht)
4. Fehlermeldung bei nicht darstellbarer Zahl
5. Runden des *fraction*-Anteils (Anpassung an die tatsächlich zur Verfügung stehenden Bits)
6. falls nicht normalisiert, dann zurück zu Schritt 3

Gleitkommamultiplikation:
1. Addition der Exponenten
2. Multiplikation der *fraction*-Anteile
3. Normalisieren
4. Fehlermeldung bei nicht darstellbarer Zahl
5. Runden des *fraction*-Anteils (Anpassung an die tatsächlich zur Verfügung stehenden Bits)
6. falls nicht normalisiert, dann zurück zu Schritt 3
7. Ermittlung des Vorzeichens (bei Faktoren identisch $\rightarrow$ -, unterschiedlich $\rightarrow$ +)

b) Zur einfacheren Implementation der Hardware entschloss man sich, nur positive Zahlen für die Exponenten zu verwenden. Dies erreicht man, indem eine Konstante addiert, die *bias* genannt wird.

c) einfach: *bias* = 127
doppelt: *bias* = 1023

d) Rechnung 1:

1. Exponentenanpassung	$6.666 \cdot 10^{09} = 0.06666 \cdot 10^{11}$
2. Addition von *fraction*	9.933+0.06666 = 9.99966
3. Normalisieren	$9.99966 \cdot 10^{11} = 9.99966 \cdot 10^{11}$
4. kein Fehler	
5. Runden von *fraction*	$10.000 \cdot 10^{11}$
6. erneut ab Schritt 3 !	
3. Normalisieren	$1.0000 \cdot 10^{12}$
4. kein Fehler	
5. Runden von *fraction*	$1.000 \cdot 10^{12}$
6. fertig	

Rechnung 2:

1. Exponentenaddition	-03 + 01 = -2
2. Multiplikation von *fraction*	$4.200 \cdot 1.111 = 4.6662$
3. Normalisieren	$4.6662 \cdot 10^{-2} = 4.6662 \cdot 10^{-2}$
4. kein Fehler	
5. Runden von *fraction*	$4.666 \cdot 10^{-2}$
6. Zahl ist normalisiert	
6. Vorzeichenbestimmung	$+4.666 \cdot 10^{-2}$

Aufgabe 48

Gemäß Aufgabe 45b und 45c ist $\frac{1}{20} \approx \overline{1.100} \cdot 2^{-5}$. Zur Darstellung der Zahlen verwende ich *nicht* den IEEE-Standard (nicht verwechseln mit dem Rundungsverfahren !), d.h. in *fraction* taucht jeweils die führende 1 auf.

Rundungsverfahren	*sign*	*exponent*	*fraction*
rounding down	0	01111010	11001100
rounding up	0	01111010	11001101
rounding off	0	01111010	11001101
IEEE-Standard	0	01111010	11001101
von-Neumann	0	01111010	11001101