

Half Precision: s10e5

Betrachten Sie die binären Gleitkommazahlen gemäß der Parameter s10e5.

1. Ermitteln Sie den Verschiebewert B (BIAS) für den Exponenten gemäß des IEEE-754 Standards, sowie den Exponentenwertebereich e_{\min} und e_{\max} .
2. Mit vielen Bits wird eine Gleitkommazahl hier gespeichert? Erklären Sie die Bedeutung der einzelnen Bits und wie daraus die Gleitkommazahl gebildet wird.
3. Bestimmen Sie die relative Maschinengenauigkeit macheps.
4. Geben Sie die Bitmuster mit zugehörigem (ungefähren) Zahlenwert an:
 - a) Signed zero und signed infinity: $\pm 0, \pm \infty$
 - b) Kleinste und größte positive normale Zahl
 - c) Kleinste und größte positive denormalisierte Zahl (*subnormals*)
 - d) Kleinste und größte positive Nichtzahl (NaN)
 - e) 001111100000000001
 - f) $\text{fl}(-4)$
 - g) $\text{fl}(\frac{2}{3})$, wobei $\frac{2}{3} = (0.\overline{10})_2 = (0.10101010\dots)_2$

Solution:

Gleitkommazahlen sind rational und enthalten Ganzzahlen

Es seien $b \geq 2$ eine Basis, $N \in \mathbb{N}$ eine Mantissenlänge und $e_{\min}, e_{\max} \in \mathbb{Z}$ der Exponentenwertebereich. Zeigen Sie, dass für die dadurch definierte Menge von Gleitkommazahlen $\mathbb{F} := \mathbb{F}(N, b, e_{\min}, e_{\max})$ folgendes gilt:

1. Gleitkommazahlen sind rational: $\mathbb{F} \subset \mathbb{Q}$.
2. Die ganzen Zahlen außer Null innerhalb des Wertebereichs sind enthalten: $([-\text{OFL}, \text{OFL}] \cap \mathbb{Z} \setminus \{0\}) \subset \mathbb{F}$, wobei OFL das Maximum der Menge \mathbb{F} bezeichnet (*Overflow-Level*) bezeichnet.