

3. Dodatne vježbe

1. Odredite najveću moguću relativnu i najveću moguću apsolutnu pogrešku koja se može očekivati pri pohrani broja $2 \cdot 10^{22}$ u IEEE 754 formatu jednostruke preciznosti.
2. Za pohranu realnih brojeva koristi se registar u kojem mantisa ima ukupno (zajedno sa skrivenim bitom) 15 bitova, karakteristika ima 12 bitova, te se jedan bit koristi za predznak.
 - a. odredite najveću moguću relativnu pogrešku
 - b. odredite najveću moguću apsolutnu pogrešku ako se pohranjuje broj 1350
3. Koliko puta bi se smanjila najveća moguća relativna pogreška u odnosu na IEEE 754 format jednostruke preciznosti, ako bi se koristio format prikaza u kojem bi mantisa imala ukupno (zajedno sa skrivenim bitom) 27 bitova.
4. Koliko puta bi se povećala najveća moguća apsolutna pogreška ako bi se umjesto IEEE 754 formata jednostruke preciznosti koristio format u kojem bi mantisa imala ukupno (zajedno sa skrivenim bitom) 20 bitova.
5. Koliko bi se smanjila najveća moguća relativna pogreška u IEEE 754 formatu jednostruke preciznosti, ako bi se za karakteristiku koristilo 15 umjesto 8 bitova, a veličina mantise ostala ista.
6. Koliko bi približno iznosio najveći broj kojeg je moguće prikazati u IEEE 754 formatu jednostruke preciznosti, ako bi se za karakteristiku umjesto 8 bitova, koristilo 9 bitova.

Rješenja: NE GLEDATI prije nego sami pokušate riješiti zadatke

1. Najveća moguće relativna pogreška ovisi isključivo o broju bitova mantise m . Vodite računa o tome da parametar m uključuje i skriveni bit. Kod prikaza prema IEEE 754 standardu jednostruke preciznosti $m = 24$.

Najveća moguća relativna pogreška iznosi $2^{-m} = 2^{-24} \approx 6 \cdot 10^{-8}$

Najveća moguća apsolutna pogreška ovisi o parametru m i konkretnom broju x koji se prikazuje:

Najveća moguća apsolutna pogreška iznosi $x \cdot 2^{-m} \approx 2 \cdot 10^{22} \cdot 6 \cdot 10^{-8} \approx 1.2 \cdot 10^{15}$

2. Najveća moguća relativna pogreška iznosi $2^{-m} = 2^{-15} \approx 3.05 \cdot 10^{-5}$

Najveća moguća apsolutna pogreška iznosi $1350 \cdot 3.05 \cdot 10^{-5} \approx 4.12 \cdot 10^{-2}$

3. Najveća moguća relativna pogreška u IEEE 754 iznosi 2^{-24}

Najveća moguća relativna pogreška u "novom" formatu iznosi 2^{-27}

$$2^{-24} \div 2^{-27} = 2^3$$

Najveća moguća relativna pogreška u "novom" formatu manja je za 8 puta.

4. Neka je x broj koji se prikazuje.

Najveća moguća apsolutna pogreška u IEEE 754 iznosi $x \cdot 2^{-24}$

Najveća moguća apsolutna pogreška u "novom" formatu iznosi $x \cdot 2^{-20}$

$$x \cdot 2^{-20} \div x \cdot 2^{-24} = 2^4$$

Najveća moguća apsolutna pogreška u "novom" formatu veća je za 16 puta.

5. Broj bitova karakteristike NE utječe na preciznost, stoga najveća moguća relativna pogreška ostaje ista.

6. $K \in [0, 511]$, $K = 0$ i $K = 511$ se koriste za posebne slučajeve

$$BE = K - 255$$

Najveći mogući binarni eksponent je 255.

Najveći broj kojeg je moguće prikazati je $1.1111_2 \dots \cdot 2^{255} \approx 1_2 \cdot 2^{256} \approx 1.16 \cdot 10^{77}$