# 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.

7. Gdje se (i zašto) u sljedećem odsječku programa nalaze sintaktičke pogreške:

```
int thin, tall, short;
float which, while, when, why, who;
char single, double, triple;
signed long a777, 7b, _19;
```

8. Pronađite koje su konstante ispravno, a koje neispravno napisane. Za ispravno napisane konstante odredite kojeg su tipa i koliko okteta zauzimaju u memoriji:

```
2   4u   7f   9.1   14.5U   0101u   12.1L   12.1e+22F   12.1e22
12.1Fe-22   12.1E11L   12.1E11u   0x22L  0xABC   0x2f    2F
0x2F.1F    021.1f
```

9. Napisati sadržaj registra u kojem je, prema IEEE 754 standardu za prikaz brojeva u **dvostrukoj preciznosti** pohranjen broj $-0.25_{10}$. Sadržaj registra napisati u heksadekadskom obliku.
10. U registru od 64 bita upisan je broj $C0\ 3D\ 80\ 00\ 00\ 00\ 00\ 00_{16}$. Napisati koji je broj predstavljen u tom registru, ukoliko registar služi za pohranu varijable `double x`. Rezultat napisati u dekadskom brojevnom sustavu.
11. Napisati sadržaj registra u kojem je, prema IEEE 754 standardu za prikaz brojeva u **dvostrukoj preciznosti** pohranjen broj $-\infty$. Sadržaj registra napisati u heksadekadskom obliku.
12. Napisati sadržaj registra u kojem je, prema IEEE 754 standardu za prikaz brojeva u **dvostrukoj preciznosti** pohranjena vrijednost NaN. Sadržaj registra napisati u heksadekadskom obliku.
13. Odredite najveću moguću relativnu i najveću moguću apsolutnu pogrešku koja se može očekivati pri pohrani broja $2 \cdot 10^{22}$ u IEEE 754 formatu dvostruke preciznosti.

## Rješenja: <u>NE GLEDATI prije nego sami pokušate riješiti zadatke</u>

**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 $2^{-m} = 2^{-15} \approx 3.05 \approx 10^{-5}$

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

**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 $1.1111_2... \cdot 2^{255} \approx 1_2 \cdot 2^{256} \approx 1.16 \cdot 10^{77}$

**7.**
```
int thin, tall, short;
float which, while, when, why, who;
char single, double, triple;
signed long a777, 7b, _19;
```

U prvom retku se za ime varijable koristi ključna riječ `short`;

U drugom retku se za ime varijable koristi ključna riječ `while`;

U trećem retku ime varijable `7b` započinje znamenkom (nije dopušteno)

**8.**

| | |
|---|---|
| `2` | signed int - 4 okteta |
| `4u` | unsigned int - 4 okteta |
| `7f` | pogreška: nedostaje točka |
| `9.1` | double - 8 okteta |
| `14.5U` | pogreška: ne postoji tip unsigned double |
| `0101u` | unsigned int u oktalnom obliku - 4 okteta |
| `12.1L` | long double - 8 okteta |
| `12.1e+22F` | float - 4 okteta |
| `12.1e22` | double - 8 okteta |
| `12.1Fe-22` | pogreška: F na pogrešnom mjestu |
| `12.1E11L` | long double - 8 okteta |
| `12.1E11u` | pogreška: ne postoji tip unsigned double |
| `0x22L` | long int u heksadekadskom obliku - 4 okteta |
| `0xABC` | int u heksadekadskom obliku - 4 okteta |
| `0x2f` | int u heksadekadskom obliku - 4 okteta |
| `2F` | pogreška: nedostaje točka |
| `0x2F.1F` | pogreška: ne može se realni broj zapisati u heksadekadskom obliku |
| `021.1F` | float, 0 na početku nema nikakvo značenje - 4 okteta |

**9.** BFD0000000000000

**10.** -29.5

**11.** FFF0000000000000

**12.** Prikazano je jedno od mogućih rješenja. Bitno je da su svi bitovi karakteristike postavljeni na 1, te da je barem jedan bit mantise postavljen na 1

FFF8000000000000

**13.** Najveća moguća relativna pogreška iznosi $2^{-m} = 2^{-53} \approx 1.1 \cdot 10^{-16}$

Najveća moguća apsolutna pogreška iznosi $x \cdot 2^{-m} \approx 2 \cdot 10^{22} \cdot 1.1 \cdot 10^{-16} \approx 2.2 \cdot 10^{6}$