

## 2. Seštevanje celih števil, zapis v plavajoči vejici

- 2. Seštevanje celih števil, zapis v plavajoči vejici
  - Seštevanje celih števil
  - Zapis realnih števil s plavajočo vejico

### Seštevanje celih števil

- Prenos in pravilnost rezultata pri seštevanju **nepredznačenih števil**

Pravilnost rezultata pri seštevanju nepredznačenih števil določa prenos (*ang. carry*) na najpomembnejšem bitu  $C_{MSB}$ . Če je prenos  $C_{MSB}$  enak 0, je rezultat pravilen. Če je prenos 1, je rezultat nepravilen. V tem primeru se rezultata ne da predstaviti z danim številom bitov. Torej:

$$C_{MSB} = \begin{cases} 0, & \text{Rezultat seštevanja je pravilen} \\ 1, & \text{Rezultat seštevanja ni pravilen} \end{cases},$$

- **Primer:** Opazujte prenos pri seštevanju 190 in 20 :

Najprej pretvorimo števili v dvojiški sistem:

$$190_{(10)} = 128 + 32 + 16 + 8 + 4 + 2 \rightarrow 1011\ 1110_{(2)}$$

$$20_{(10)} = 16 + 4 \rightarrow 0001\ 0100_{(2)}$$

Seštejemo binarni števili po modulu 2 ( $1+1 = 0 \rightarrow$  prenos v naslednji bit je 1)

$$\begin{array}{r} 1011\ 1110 \\ +\ 0001\ 0100 \\ \hline C_{MSB} = 0\ 1101\ 0010 \end{array}$$

$C_{MSB}$  je enak 0, torej je rezultat pravilen.

- Prenos in pravilnost rezultata pri seštevanju **predznačenih števil**

Veljavnost rezultat pri seštevanju predznačenih števil določa bit preliva (*ang. overflow*)  $V$ . Če sta oba predznaka seštevancev enaka in je predznak rezultata različen, potem je  $V$  enak 1 in je rezultat nepravilen. Drugače pa je  $V = 0$  in je rezultat pravilen. To lahko predstavimo z naslednjo tabelo:

op1	op2	rez	$V$
+	-	+ / -	0
-	+	+ / -	0
+	+	+	0
+	+	-	1
-	-	-	0
-	-	+	1

$$V = \begin{cases} 0, & \text{Rezultat seštevanja je pravilen} \\ 1, & \text{Rezultat seštevanja ni pravilen} \end{cases},$$

- **Primer:** Opazujte prenos pri seštevanju 123 in (-123) :

Pretvorba v dvojiški sistem (Dvojiški komplement):

$$\begin{aligned} 123_{(10)} &\rightarrow 0111\ 1100_{(2)} \\ -123_{(10)} &\rightarrow 1000\ 0101_{(2)} \end{aligned}$$

Seštejemo po modulu 2

$$\begin{array}{r} 0111\ 1100 \\ +\ 1000\ 0101 \\ \hline C_{MSB} = 1\ 0000\ 0001 \end{array}$$

Seštevanca in rezultat so pozitivni  $\rightarrow V = 0 \rightarrow$  Rezultat je pravilen, čeprav je  $C_{MSB} = 1$ . Pri seštevanju predznačenih števil upoštevamo samo bit preliva  $V$ .

- **Primer:** Opazujte prenos pri seštevanju -80 in (-60) :

Pretvorba v dvojiški sistem (dvojiški komplement):

$$\begin{aligned} -80_{(10)} &\rightarrow 1011\ 0000_{(2)} \\ -60_{(10)} &\rightarrow 1100\ 0100_{(2)} \end{aligned}$$

Seštejemo po modulu 2

$$\begin{array}{r}
 1011\ 0000 \\
 +\ 1100\ 0100 \\
 \hline
 C_{MSB} = 1\ 0111\ 0100
 \end{array}$$

Seštevanca sta pozitivna, rezultat je negativen  $\rightarrow V = 1 \rightarrow$  Rezultat je nepravilen.

- **Naloge:** Opazujte prenos, preliv in pravilnost rezultata pri seštevanju naslednjih 8-bitnih števil:

- nepredznačeni števili 190 in 70 :

Pretvorba v dvojiški sistem:

$$\begin{aligned}
 190_{(10)} &\rightarrow 1011\ 1110_{(2)} \\
 70_{(10)} &\rightarrow 0100\ 0110_{(2)}
 \end{aligned}$$

Seštejemo števili v dvojiškem zapisu po modulu 2

$$\begin{array}{r}
 1011\ 1110 \\
 +\ 0100\ 0110 \\
 \hline
 C_{MSB} = 1\ 0000\ 0100
 \end{array}$$

$C_{MSB}$  je enak 1, torej rezultat ni pravilen.

- predznačeni števili v dvojiškem komplementu 124 in 7 Pretvorba v dvojiški sistem:

$$\begin{aligned}
 124_{(10)} &\rightarrow 0111\ 1100_{(2)} \\
 7_{(10)} &\rightarrow 0000\ 0111_{(2)}
 \end{aligned}$$

Seštejemo števili v dvojiškem zapisu po modulu 2

$$\begin{array}{r}
 0111\ 1100 \\
 +\ 0000\ 0111 \\
 \hline
 C_{MSB} = 0\ 1000\ 0011
 \end{array}$$

Seštevanca sta pozitivna, rezultat je negativen  $\rightarrow V = 1 \rightarrow$  Rezultat je nepravilen.

- predznačeni števili v dvojiškem komplementu -80 in 60

Pretvorba v dvojiški sistem:

$$\begin{aligned}
 -80_{(10)} &\rightarrow 1011\ 0000_{(2)} \\
 60_{(10)} &\rightarrow 0011\ 1100_{(2)}
 \end{aligned}$$

Seštejemo števili v dvojiškem zapisu po modulu 2 ( $1+1 = 0 \rightarrow$  prenos v naslednji bit je 1)

$$\begin{array}{r} 1011\ 0000 \\ +\ 0011\ 1100 \\ \hline C_{MSB} = 0\ 1110\ 1100 \end{array}$$

Predznaka seštevancev sta različna  $\rightarrow V = 0 \rightarrow$  Rezultat je pravilen.

◦ **Naloga iz izpita 1** Imamo podani naslednji števili:

- $A = 0x0A63$  16-bitno predznačeno celo število v zapisu z dvojiškim komplementom
- $B = 0xF4$  8-bitno predznačeno celo število v zapisu z dvojiškim komplementom

Izračunajte  $E = A + (-B)$ . Ali je pri seštevanju prišlo do prenosa ali preliva? Število  $B$  najprej pretvorite v 16-bitno predznačeno celo število. Rezultat  $E$  zapišite v šestnajstiški obliki kot 16-bitno predznačeno celo število v dvojiškem komplementu.

**Rešitev:**

Najprej predstavimo število  $B$  kot 16-bitno :

$$B = \underbrace{????\ ????}_{0xF4} 1111\ 0100$$

Ker je  $B$  predznačeno število, namesto zgornjih bitov vstavimo bit predznaka, oziroma razširimo  $B$  z bitom predznaka. Ker je  $B$  negativno število, namesto ? podamo 1. Torej:

$$B = 1111\ 1111\ 1111\ 0100 = 0xFF F4$$

Izračunamo  $-B$

$$-B = \neg(B) + 1 = 0000\ 0000\ 0000\ 1100$$

pri čemer  $\neg(\cdot)$  pomeni negacijo bitov ( $1 - > 0, 0 - > 1$ ). Na koncu seštejemo  $A + (-B)$

$$E = A + (-B) = 0000\ 1010\ 0110\ 1111_{(2)} = 0x0A6F_{(16)}$$

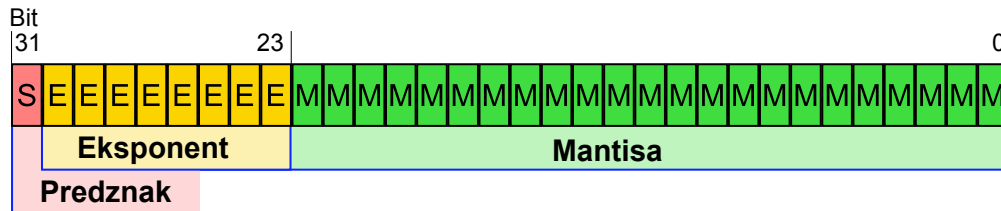
Seštevanka  $A$  in  $(-B)$  ter rezultat  $E$  so pozitivni  $\rightarrow$  Rezultat je pravilen.

## Zapis realnih števil s plavajočo vejico

Zapis s plavajočo vejico v formatu IEEE 754 v računalništvu uporabljamo za predstavitev realnih števil. Obstajajo več formatov zapisov števil v plavajoči vejici. Najpogostejše se uporabljata dva zapisa:

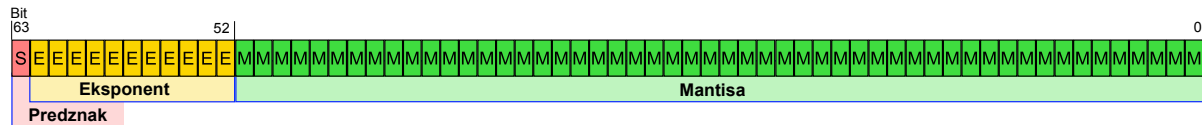
- IEEE 754 z enojno natančnostjo (podatkovni tip *float* - 32 bitov):

1. Predznak - 1 bit
2. Eksponent - 8 bitov
3. Mantisa - 23 bitov



- IEEE 754 z dvojno natančnostjo (podatkovni tip *double* - 64 bitov):

1. Predznak - 1 bit
2. Eksponent - 11 bitov
3. Mantisa - 52 bitov



- **Primer:** Število  $-210,5937510$  najprej zapišimo v dvojiški obliki s plavajočo vejico, nato pa še šestnajstiško v prestavitvi IEEE 754 z enojno natančnostjo.

$$-210,5937510_{(10)} \rightarrow ?_{(IEEE754 \text{ z enojno natančnostjo})}$$

Postopek:

1. Pretvorba v **dvojiški sistem**:

$$210,5937510_{(10)} \rightarrow 1101\,0010,1001\,1_{(2)}$$

To ni IEEE 754 zapis !!!!

2. Dvojiška eksponentna oblika  $\rightarrow$  Pretvorimo v zapis  $(1, m \cdot 2^e)$

$$1101\,0010,1001\,1 = 1,1010\,0101\,0011 \cdot 2^7$$

pri čemer:

- $m = 1010\,0101\,0011$ ,
- $e = 7$

3. Zapis v IEEE 754 z enojno natančnostjo

- Predznak: Število je negativno  $\rightarrow s = 1_{(2)}$

- Eksponent:  $E = e + 127 = 134_{(10)} \rightarrow 1000\ 0110_{(2)}$
- Mantisa:  $m = 1010\ 0101\ 0011\ 0000\ 0000\ 000_{(2)}$

4. Rešitev:

$$\underbrace{1}_{\text{Predznak}} \underbrace{100\ 0011\ 0}_{\text{Eksponent}} \underbrace{101\ 0010\ 1001\ 1000\ 0000\ 0000}_{\text{Mantisa}} = 0xC352\ 9800$$

- **Primer:** Število  $0xBF580000$  je zapisano v IEEE 754 z enojno natančnostjo. Zapišimo desetiško vrednost.

$$0xBF580000_{(IEEE754\ z\ enojno\ natančnostjo)} \rightarrow ?_{(10)}$$

Zapišimo podano število v dvojiškem sistemu:

$$0xBF580000 = \underbrace{1}_{\text{Predznak}} \underbrace{011\ 1111\ 0}_{\text{Eksponent}} \underbrace{101\ 1000\ 0000 \dots 0}_{\text{Mantisa}}$$

- Predznank
  - $s = 1 \rightarrow$  Število je negativno
- Eksponent
  - $E = 0111\ 1110_{(2)} \rightarrow 126_{(10)} \Rightarrow e = E - 127 = -1$
- Mantisa
  - $m = 1011$
- Končni rezultat
  - $(-1)^s \cdot 1, m \cdot 2^e = -1.1011_2 \cdot 2^{-1} = -0,84375$

- **Naloga:** V šestnajstiškem sestavu zapišite število  $-87,421875$  v zapis IEEE 754 z enojno natančnostjo.

**Rešitev:**

1. Pretvorba v **dvojiški sistem:**

$$87,421875_{(10)} \rightarrow 0101\ 0111, 0110\ 1100_{(2)}$$

2. Dvojiška eksponentna oblika  $\rightarrow$  Pretvorimo v zapis  $(1, m \cdot 2^e)$

$$0101\ 0111, 0110\ 11 = 1, 0101\ 1101\ 1011 \cdot 2^6$$

3. Zapis v IEEE 754 z enojno natančnostjo

- Predznak: Število je negativno  $\rightarrow s = 1_{(2)}$
- Eksponent:  $E = e + 127 = 133_{(10)} \rightarrow 1000\ 0101_{(2)}$
- Mantisa:  $m = 0101\ 1101\ 1011_{(2)}$

4. Rešitev:

$$1\ 1000\ 0101\ 0101\ 1101\ 1011\ 0000\ 0000\ 000_{IEEE754} = 0xC2AE\ D800_{IEEE754}$$

- **Naloga:** Zapišite pozitivno neskončno vrednost v šestnajstiškem zapisu v IEEE 754 z dvojno natančnostjo.

**Rešitev:** Poglejte prosojnice iz predavanj

$$+\infty \rightarrow 0x7FF0\ 0000\ 0000\ 0000_{IEEE754} \text{ z dvojno natančnostjo}$$

$$-\infty \rightarrow 0xFFF0\ 0000\ 0000\ 0000_{IEEE754} \text{ z dvojno natančnostjo}$$

- **Naloga:** Katere desetiško vrednost predstavlja  $0xFFF1\ 0000\ 0000\ 0000$  v IEEE 754 z dvojno natančnostjo?

**Rešitev:** Poglejte prosojnice iz predavanja

$$0xFFF1\ 0000\ 0000\ 0000_{IEEE754} \text{ z dvojno natančnostjo} \rightarrow NaN$$

- **Naloga iz izpita 2:** Imamo podani naslednji števili:

- $C = 0x3F58\ 0000$  32-bitno število, zapisano po standardu IEEE 754
- $D = 0x425C\ 4000$  32-bitno število, zapisano po standardu IEEE 754

Izračunajte  $F = C + D$ . Zapišite števili  $C$  in  $D$  v dvojiški eksponentni obliki (primer:  $1,1001011 \cdot 2^{-10}$ ). Število  $F$  mora biti zapisano na oba načina – v dvojiški eksponentni obliki in šestnajstiško po formatu IEEE 754.

**Rešitev:**

- Zapišemo podani števili v dvojiški eksponentni obliki

$$C = 0x3F58\ 0000 \rightarrow 1,1011 \cdot 2^{-1}$$

$$D = 0x425C\ 4000 \rightarrow 1,101110001 \cdot 2^5$$

- Števili prestavimo na isto potenco

$$C = 1,1011 \cdot 2^{-1}$$

$$D = 1101110,0010 \cdot 2^{-1}$$

- Števili seštejemo in premaknemo vejico

$$F = C + D = 1,101111101 \cdot 2^5$$

- Zapišemo v IEEE 754

$$F = 0x425F\ A000$$

- **Naloga iz izpita 3:** Izračunajte produkt  $P = M \cdot N$  dveh realnih števil, zapisanih v 32-bitnem zapisu v plavajoči vejici po standardu IEEE 754. Števila pred množenjem zapišite v dvojiški eksponentni obliki brez odmika (primer:  $+1,011 \cdot 2^{-32}$ ). Množenje izvedite v dvojiški obliki. Rezultat  $P$  zapišite po standardu IEEE 754 v šestnajstiškem zapisu.

- $M = 0xABCD\ 0000$

- $N = 0x4EB0\ 0000$

**Rešitev:**

- Zapišemo podani števili v dvojiški eksponentni obliki

$$M = 0xABCD\ 0000 \rightarrow -1,1001101 \cdot 2^{-40}$$

$$N = 0x4EB0\ 0000 \rightarrow 1,1011 \cdot 2^{30}$$

- Pomnožimo  $M$  in  $N$

$$P = M \cdot N = (-1,1001101) \cdot 2^{-40} \cdot 1,1011 \cdot 2^{30}$$

$$= -(1,1001101 \cdot 1,1011) \cdot 2^{-40+30}$$

$$= -(1,1001101 \cdot 1,1011) \cdot 2^{-10}$$

$$= -(10,0011001111) \cdot 2^{-10}$$

$$= -(1,00011001111) \cdot 2^{-9}$$

1,1001101 · 1,011	
11001101	(1 · 11001101)
11001101	(1 · 11001101 pomaknjeno za 1 mesto)
00000000	(0 · 11001101 pomaknjeno za 2 mesti)
+ 11001101	(1 · 11001101 pomaknjeno za 3 mesta)
10,0011001111	Rezultat ( 7 + 3 = 10 decimalnih mest)

- Zapišemo v IEEE 754

$$P = 0xBB0C\ F000$$