Tutorato Architettura degli Elaboratori Modulo 1 (Lezione 2)

Francesco Pelosin

23 Ottobre 2019

1 Numeri razionali in Base b

1.1 Virgola fissa

Dato un numero razionale espresso con notazione a virgola fissa in Base b su n cifre intere e m cifre razionali:

$$d_{n-1}d_{n-2}\dots d_1d_0, d_{-1}d_{-2}\dots d_{m-1}d_m$$

Possiamo ricavare il suo corrispondente valore in decimale nel seguente modo:

$$d_{n-1}\cdot b^{n-1}+\dots+d_1\cdot b^1+d_0\cdot b^0+d_{-1}\cdot b^{-1}+\dots+d_{-m+1}\cdot b^{-m+1}+d_{-m}\cdot b^{-m}$$

Volendo trasformare un numero razionale espresso in Base 10 in una generica Base b dobbiamo applicare il seguente procedimento:

- Trasformare la parte intera del numero da Base 10 a Base b (applicando l'algoritmo visto nella precedente lezione).
- Trasformare la parte frazionaria del numero in Base b. La trasformazione della parte frazionaria può essere computata procedendo per moltiplicazioni successive nel seguente modo:
  - Moltiplichiamo la parte frazionaria del numero per b.
  - Continuiamo a moltiplicare il risultato ottenuto per b tenendo presente che se il numero ottenuto è ha una parte intera maggiore di zero dobbiamo continuare con la sola parte frazionaria. La parte intera in ogni caso, diventa la nuova cifra frazionaria in base b
  - Andiamo avanti fino a quando la parte frazionaria diventa 0 oppure un risultato già ottenuto (periodicità).

1.1.1 Esercizi da Base 10 a Base b

Convertire i seguenti numeri da Base 10 a Base b:

(a) $310,63_{10} \rightarrow$ in Base 6

(b) $321,225_{10} \rightarrow$ in Base 2

(c) $519,51_{10} \rightarrow$ in Base 5

(d) $921,75_{10} \rightarrow$ in Base 9

1.1.2 Soluzioni

(a) Troviamo il valore in Base 6 della parte intera di $310,63_{10}$:

$$\begin{array}{rl} 310_{10}: & \\ 6\lfloor \underline{310} & 4 \\ 6\lfloor \underline{51} & 3 \\ 6\lfloor \underline{8} & 2 \\ 6\lfloor \underline{1} & 1 \end{array} \Bigg\} = 1234_6$$

Troviamo il valore in Base 6 della parte frazionaria del numero procedendo per moltiplcazioni successive:

	$\cdot 6$	
0,63	3,78	3
0,78	4,68	4
$\underline{0,68}$	4,08	4
0,08	0,48	0
0,48	2,88	2
0,88	5,28	5
0,28	1,68	1
$\underline{0,68}$	...	...

Avendo trovato una ripetizione sappiamo che la parte frazionaria ha una periodicità:

$$0,63_{10} = 0,34\overline{40251}_6$$

Possiamo ora scrivere il corrispondente valore in base 6 del numero di partenza:

$$310,63_{10} = 1234,34\overline{40251}_6$$

(b) Troviamo il valore in Base 2 della parte intera di $321{,}225_{10}$:

321_{10} :

$$\left.\begin{array}{r|l} 2\lfloor\underline{321} & 1 \\ 2\lfloor\underline{160} & 0 \\ 2\lfloor\underline{80} & 0 \\ 2\lfloor\underline{40} & 0 \\ 2\lfloor\underline{20} & 0 \\ 2\lfloor\underline{10} & 0 \\ 2\lfloor\underline{5} & 1 \\ 2\lfloor\underline{2} & 0 \\ 2\lfloor\underline{1} & 1 \end{array}\right\} = \texttt{101000001}_2$$

Troviamo il valore in Base 2 della parte frazionaria del numero procedendo per moltiplcazioni successive:

	$\cdot\ 2$	
0,225	0,45	0
0,45	0,9	0
0,9	1,8	1
$\underline{0,8}$	1,6	1
0,6	1,2	1
0,2	0,4	0
0,4	0,8	0
$\underline{0,8}$	...	...

Avendo trovato una ripetizione sappiamo che la parte frazionaria avrà una periodicità:

$$0,225_{10} = 0,001\overline{1100}_2$$

Possiamo ora scrivere il corrispondente valore in Base 2 del numero di partenza:

$$310,63_{10} = 101000001,001\overline{1100}_2$$

(c) Troviamo il valore in base 5 della parte intera di $519{,}51_{10}$:

$$\left.\begin{array}{r|l} 5\lfloor\underline{519} & 4 \\ 5\lfloor\underline{103} & 3 \\ 5\lfloor\underline{20} & 0 \\ 5\lfloor\underline{4} & 4 \end{array}\right\} = \texttt{4034}_5$$

Troviamo il valore in base 5 della parte frazionaria del numero procedendo per moltiplcazioni successive:

$$\begin{array}{r|c|c} & \cdot 5 & \\ 0{,}51 & 2{,}55 & 2 \\ 0{,}55 & 2{,}75 & 2 \\ \underline{0{,}75} & 3{,}75 & 3 \\ \underline{0{,}75} & \ldots & \ldots \end{array}$$

Avendo trovato una ripetizione sappiamo che la parte frazionaria avrà una periodicità:

$$0,51_{10} = 0,22\overline{3}_5$$

Possiamo ora scrivere il corrispondente valore in base 5 del numero di partenza:

$$519,51_{10} = 4034,22\overline{3}_2$$

(d) Troviamo il valore in base 9 della parte intera di $921{,}75_{10}$:

$$\begin{array}{l} 921_{10}: \\ 9\lfloor \underline{921} \\ 9\lfloor \underline{102} \\ 9\lfloor \underline{11} \\ 9\lfloor \underline{1} \end{array} \left| \begin{array}{c} 3 \\ 3 \\ 2 \\ 1 \end{array} \right\} = 1233_9$$

Troviamo il valore in base 9 della parte frazionaria del numero procedendo per moltiplcazioni successive:

$$\begin{array}{r|c|c} & \cdot 9 & \\ \underline{0{,}75} & 6{,}75 & 6 \\ \underline{0{,}75} & \ldots & \ldots \end{array}$$

Avendo trovato una ripetizione sappiamo che la parte frazionaria avrà una periodicità:

$$0,75_{10} = 0,\overline{6}_9$$

Possiamo ora scrivere il corrispondente valore in base 5 del numero di partenza:

$$921,75_{10} = 1233,\overline{6}_9$$

1.1.3 Esercizi da Base b a Base 10

Tradurre i seguenti numeri in Base 10:

(a) $11310{,}32_5$

(b) $147{,}75_{12}$

(c) $3A7{,}68_{16}$

(d) $1703{,}12_8$

1.1.4 Soluzioni

(a) $11310,32_5 = 1 \cdot 5^4 + 1 \cdot 5^3 + 3 \cdot 5^2 + 1 \cdot 5^1 + 0 \cdot 5^0 + 3 \cdot 5^{-1} + 2 \cdot 5^{-2} = 625 + 125 + 75 + 5 + 0,6 + 0,08 = 830,68_{10}$

(b) $147,75_{12} = 1 \cdot 12^2 + 4 \cdot 12^1 + 7 \cdot 12^0 + 7 \cdot 12^{-1} + 5 \cdot 12^{-2} = 144 + 48 + 7 + 0,584 + 0,03473 = 199,61873_{10}$

(c) $3A7,68_{16} = 3 \cdot 16^2 + 10 \cdot 16^1 + 7 \cdot 16^0 + 6 \cdot 16^{-1} + 8 \cdot 16^{-2} = 768 + 160 + 7 + 0,375 + 0,03125 = 935,40625_{10}$

(d) $1703,12_8 = 1 \cdot 8^3 + 7 \cdot 8^2 + 0 \cdot 8^1 + 3 \cdot 8^0 + 1 \cdot 8^{-1} + 2 \cdot 8^{-2} = 512 + 448 + 3 + 0,125 + 0,03125 = 963,15625_{10}$

1.1.5 Esercizi da Base 10 a Base 2

Convertire i seguenti numeri da Base 10 a Base 2:

(a) $35{,}75_{10}$

(b) $44{,}35_{10}$

(c) $63{,}875_{10}$

(d) $136{,}5625_{10}$

(e) $192{,}625_{10}$

(f) $255{,}90625_{10}$

1.1.6 Soluzioni

(a) Troviamo il valore in Base 2 della parte intera di $35{,}75_{10}$:

$$
\begin{array}{l}
35_{10}: \\
\left.\begin{array}{rl}
2\lfloor \underline{35} & 1 \\
2\lfloor \underline{17} & 1 \\
2\lfloor \underline{8} & 0 \\
2\lfloor \underline{4} & 0 \\
2\lfloor \underline{2} & 0 \\
2\lfloor \underline{1} & 1
\end{array}\right\} = \texttt{100011}_2
\end{array}
$$

Troviamo il valore in base 6 della parte frazionaria del numero procedendo per moltiplcazioni successive:

$$
\begin{array}{r|c|l}
 & \cdot 2 & \\
0{,}75 & 1{,}5 & 1 \\
0{,}5 & \underline{1{,}0} & 1 \\
0{,}0 & &
\end{array}
$$

Avendo ottenuto un risultato intero posso fermarmi e affermare che:

$$0,75_{10} = 0,11_2$$

Possiamo ora scrivere il corrispondente valore in base 6 del numero di partenza:

$$35,75_{10} = 100011,11_2$$

(b) Troviamo il valore in Base 2 della parte intera di $44{,}35_{10}$:

$$44_{10}:$$

$2 \lfloor \underline{44}$	0
$2 \lfloor \underline{22}$	0
$2 \lfloor \underline{11}$	1
$2 \lfloor \underline{5}$	1
$2 \lfloor \underline{2}$	0
$2 \lfloor \underline{1}$	1

$\Big\} = 101100_2$

Troviamo il valore in Base 2 della parte frazionaria del numero procedendo per moltiplcazioni successive:

	$\cdot 2$	
0,35	0,7	0
0,7	1,4	1
$\underline{0,4}$	0,8	0
$\overline{0,8}$	1,6	1
0,6	1,2	1
0,2	0,4	0
$\underline{0,4}$	0,8	0
$\overline{0,8}$	...	...

Avendo trovato una ripetizione sappiamo che la parte frazionaria avrà una periodicità:

$$0,35_{10} = 0,01\overline{0110}_2$$

Possiamo ora scrivere il corrispondente valore in Base 2 del numero di partenza:

$$44,35_{10} = 101100,01\overline{0110}_2$$

(c) Troviamo il valore in Base 2 della parte intera di $63{,}875_{10}$:

$$63_{10}:$$

$2 \lfloor \underline{63}$	1
$2 \lfloor \underline{31}$	1
$2 \lfloor \underline{15}$	1
$2 \lfloor \underline{7}$	1
$2 \lfloor \underline{3}$	1
$2 \lfloor \underline{1}$	1

$\Big\} = 111111_2$

Troviamo il valore in Base 2 della parte frazionaria del numero procedendo per moltiplcazioni successive:

	$\cdot 2$	
0,875	1,75	1
0,75	1,5	1
0,5	1,0	1
0,0		

Avendo ottenuto un risultato intero posso fermarmi e affermare che:

$$0,875_{10} = 0,111_2$$

Possiamo ora scrivere il corrispondente valore in Base 2 del numero di partenza:

$$63,875_{10} = 111111,111_2$$

(d) Troviamo il valore in Base 2 della parte intera di $136{,}5625_{10}$:

136_{10} :

$2\lfloor \underline{136}$	0
$2\lfloor \underline{68}$	0
$2\lfloor \underline{34}$	0
$2\lfloor \underline{17}$	1
$2\lfloor \underline{8}$	0
$2\lfloor \underline{4}$	0
$2\lfloor \underline{2}$	0
$2\lfloor \underline{1}$	1

$\Big\} = \texttt{10001000}_2$

Troviamo il valore in Base 2 della parte frazionaria del numero procedendo per moltiplcazioni successive:

	$\cdot 2$	
0,5625	1,125	1
0,125	0,25	0
0,25	0,5	0
0,5	1,0	1
0,0		

Avendo ottenuto un risultato intero posso fermarmi e affermare che:

$$0,5625_{10} = 0,1001_2$$

Possiamo ora scrivere il corrispondente valore in Base 2 del numero di partenza:

$$136,5625_{10} = 10001000,1001_2$$

(e) Troviamo il valore in Base 2 della parte intera di $192{,}625_{10}$:

$$
\begin{array}{r|l}
192_{10}: & \\
2\lfloor \underline{192} & 0 \\
2\lfloor \underline{96} & 0 \\
2\lfloor \underline{48} & 0 \\
2\lfloor \underline{24} & 0 \\
2\lfloor \underline{12} & 0 \\
2\lfloor \underline{6} & 0 \\
2\lfloor \underline{3} & 1 \\
2\lfloor \underline{1} & 1
\end{array}
\Bigg\} = \texttt{11000000}_2
$$

Troviamo il valore in Base 2 della parte frazionaria del numero procedendo per moltiplcazioni successive:

$$
\begin{array}{r|c|l}
 & \cdot 2 & \\
0{,}625 & 1{,}25 & 1 \\
0{,}25 & 0{,}5 & 0 \\
0{,}5 & \underline{1{,}0} & 1 \\
0{,}0 & &
\end{array}
$$

Avendo ottenuto un risultato intero posso fermarmi e affermare che:

$$0,625_{10} = 0,101_2$$

Possiamo ora scrivere il corrispondente valore in Base 2 del numero di partenza:

$$192,625_{10} = 11000000,101_2$$

(f) Troviamo il valore in Base 2 della parte intera di $255{,}90625_{10}$:

$$
\begin{array}{r|l}
255_{10}: & \\
2\lfloor \underline{255} & 1 \\
2\lfloor \underline{127} & 1 \\
2\lfloor \underline{63} & 1 \\
2\lfloor \underline{31} & 1 \\
2\lfloor \underline{15} & 1 \\
2\lfloor \underline{7} & 1 \\
2\lfloor \underline{3} & 1 \\
2\lfloor \underline{1} & 1
\end{array}
\Bigg\} = \texttt{11111111}_2
$$

Troviamo il valore in Base 2 della parte frazionaria del numero procedendo per moltiplcazioni successive:

$$
\begin{array}{r|c|l}
 & \cdot 2 & \\
0{,}90625 & 1{,}8125 & 1 \\
0{,}8125 & 1{,}625 & 1 \\
0{,}625 & 1{,}25 & 1 \\
0{,}25 & 0{,}5 & 0 \\
0{,}5 & \underline{1{,}0} & 1 \\
0{,}0 & &
\end{array}
$$

Avendo ottenuto un risultato intero posso fermarmi e affermare che:

$$0,90625_{10} = 0,11101_2$$

Possiamo ora scrivere il corrispondente valore in Base 2 del numero di partenza:

$$255,90625_{10} = 11111111,11101_2$$

1.1.7 Esercizi da Base 2 a Base 10

Tradurre i seguenti numeri in Base 10:

(a) $100011{,}0111_2$

(b) $101100{,}101_2$

(c) $111111{,}0011_2$

(d) $10001000{,}1111_2$

(e) $11000000{,}11001_2$

(f) $11111111{,}001_2$

1.1.8 Soluzioni

(a) $100011,0111_2 = 1 \cdot 2^5 + 0 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 + 0 \cdot 2^{-1} + 1 \cdot 2^{-2} + 1 \cdot 2^{-3} + 1 \cdot 2^{-4} = 32 + 2 + 1 + 0,25 + 0,125 + 0,0625 = 35,4375_{10}$

(b) $101100,101_2 = 1{\cdot}2^5{+}0{\cdot}2^4{+}1{\cdot}2^3{+}1{\cdot}2^2{+}0{\cdot}2^1{+}0{\cdot}2^0{+}1{\cdot}2^{-1}{+}0{\cdot}2^{-2}{+}1{\cdot}2^{-3} = 32 + 8 + 4 + 0,5 + 0,125 = 44,625_{10}$

(c) $111111,0011_2 = 1 \cdot 2^5 + 1 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 + 0 \cdot 2^{-1} + 0 \cdot 2^{-2} + 1 \cdot 2^{-3} + 1 \cdot 2^{-4} = 32 + 16 + 8 + 4 + 2 + 1 + 0,125 + 0,0625 = 63,1875_{10}$

(d) $10001000,1111_2 = 1{\cdot}2^7{+}0{\cdot}2^6{+}0{\cdot}2^5{+}0{\cdot}2^4{+}1{\cdot}2^3{+}0{\cdot}2^2{+}0{\cdot}2^1{+}0{\cdot}2^0{+}1{\cdot}2^{-1}{+}1{\cdot}2^{-2}{+}1{\cdot}2^{-3}{+}1{\cdot}2^{-4} = 128{+}8{+}0,5{+}0,25{+}0,125{+}0,0625 = 136,9375_{10}$

(e) $11000000,11001_2 = 1{\cdot}2^7{+}1{\cdot}2^6{+}0{\cdot}2^5{+}0{\cdot}2^4{+}0{\cdot}2^3{+}0{\cdot}2^2{+}0{\cdot}2^1{+}0{\cdot}2^0{+}1{\cdot}2^{-1}{+}1{\cdot}2^{-2}{+}0{\cdot}2^{-3}{+}0{\cdot}2^{-4}{+}1{\cdot}2^{-5} = 128{+}64{+}0,5{+}0,25{+}0,03125 = 192,78125_{10}$

(f) $11111111,001_2 = 1{\cdot}2^7{+}1{\cdot}2^6{+}1{\cdot}2^5{+}1{\cdot}2^4{+}1{\cdot}2^3{+}1{\cdot}2^2{+}1{\cdot}2^1{+}1{\cdot}2^0{+}0{\cdot}2^{-1}{+}0{\cdot}2^{-2}{+}1{\cdot}2^{-3} = 128{+}64{+}32{+}16{+}8{+}4{+}2{+}1{+}0,125 = 255,125_{10}$

1.2 Virgola mobile

Nella notazione in virgola mobile (floating point) si usa la notazione scientifica normalizzata:

$$\pm 1,\underbrace{mmmmmm}_{Mantissa} \times 2^{\overbrace{eeee}^{Esponente}}$$

Su Base 2 i bit disponibili per la rappresentazione di un numero in virgola mobile, vengono suddivisi in:

- Un bit per il segno S
- Un blocco di bit per l'esponente E
- Un blocco di bit per la mantissa M

In particolare, per la rappresentazione dei numeri in Base 2 in virgola mobile, si usa lo standard IEEE754, nello specifico:

- IEEE754 a Singola Precisione (32 bit):
  - 1 bit per il segno S
  - 8 bit per l'esponente E
  - 23 bit per la mantissa M
- IEEE754 a Doppia Precisione (64 bit):
  - 1 bit per il segno S
  - 11 bit per l'esponente E
  - 20+32 bit per la mantissa M

Ricordiamo, che lo standard prevede la polarizzazione dell'esponente (per accelerare le operazioni di ordinamento):

- IEEE754 a Singola Precisione (32 bit) $\rightarrow$ polarizziamo aggiungendo 127 all'esponente del numero
- IEEE754 a Doppia Precisione (64 bit) $\rightarrow$ polarizziamo aggiungendo 1023 ad all'esponente del numero

La presenza di un esponente polarizzato implica che il valore rappresentato da un numero in virgola mobile nella realtà sia dato da:

$$(-1)^S \times (1 + Mantissa) \times 2^{E-127}$$

Ricordiamo, in oltre, che lo standard considera implicito il primo bit (sempre uguale a 1) dei numeri binari normalizzati.

1.2.1 Esercizi

Rappresentare i seguenti numeri in Base 2 secondo lo standard IEEE754 Singola Precisione (32 bit):

(a) $35{,}75_{10}$

(b) $-63{,}875_{10}$

(c) $136{,}5625_{10}$

(d) $-192{,}625_{10}$

1.2.2 Soluzioni

(a) Essendo $35,75_{10} = 100011,11_2$ riscriviamo il numero in notazione scientifica normalizzata Base 2:

$$+1,0001111 \cdot 2^5$$

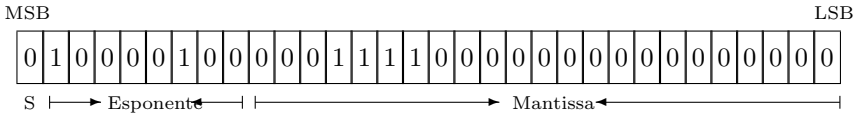
Identifichiamo segno esponente e mantissa:

$S = 0$

$M = 1 + 0,0001111 = 1,0001111$

$E = 5 + 127 = 132_{10} = 10000100_2$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



(b) Essendo $63,875_{10} = 111111,111_2$ riscriviamo il numero in notazione scientifica normalizzata Base 2:

$$-1,11111111 \cdot 2^5$$

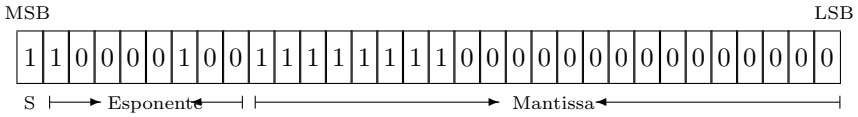
Identifichiamo segno esponente e mantissa:

$S = 1$

$M = 1 + 0,11111111 = 1,11111111$

$E = 5 + 127 = 132_{10} = 10000100_2$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



(c) Essendo $136,5625_{10} = 10001000,1001_2$ riscriviamo il numero in notazione scientifica normalizzata Base 2:

$$+1,00010001001 \cdot 2^7$$

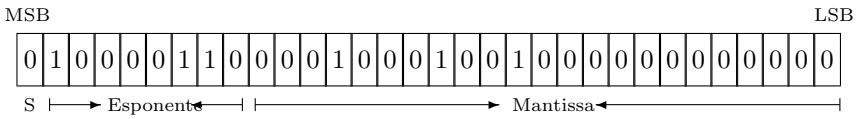
Identifichiamo segno esponente e mantissa:

$S = 0$

$M = 1 + 0,00010001001 = 1,00010001001$

$E = 7 + 127 = 134_{10} = 10000110_2$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



(d) Essendo $192,625_{10} = 11000000,101_2$ riscriviamo il numero in notazione scientifica normalizzata Base 2:

$$-1,1000000101 \cdot 2^7$$

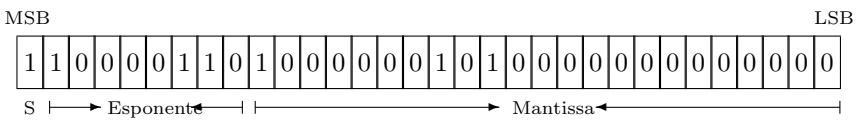
Identifichiamo segno esponente e mantissa:

$S = 1$

$M = 1 + 0,1000000101 = 1,1000000101$

$E = 7 + 127 = 134_{10} = 10000110_2$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



1.3 Somme di numeri Floating Point

L'algoritmo per sommare numeri floating point (FP) espressi in Base 2 è il seguente:

1. Confronto dell'esponente dei due numeri; scorrimento del numero più piccolo a destra finché il suo esponente non è uguale a quello del numero maggiore
2. Somma dei significandi (nel caso sia una somma tra due numeri con segno diverso usare l'algoritmo di cambio di segno ed aggiungere un bit di segno per passare alla rappresentazione in complemento a due)
3. Normalizzazione della somma, facendo scorrere la virgola verso sinistra e aumentando l'esponente, oppure farla scorrere verso destra e diminuendo l'esponente.
4. Controllare se è avvenuto overflow o underflow dell'esponente. Nel caso positivo generare un'eccezione oppure, nel caso negativo, procedere con il prossimo step.
5. Arrotondamento del significando al numero adeguato di bit.
6. Controllare se il numero è normalizzato. In caso positivo terminare, in caso positivo tornare allo step 3.

1.3.1 Esercizi

Considerando i numeri convertiti in floating point precedentemente:

A=$35{,}75_{10}$

B=-$63{,}875_{10}$

C=$136{,}5625_{10}$

D=-$192{,}625_{10}$

Svolgere le seguenti somme ed esprimere il risultato secondo lo standard IEEE754 a Singola Precisione (32 bit):

(a) $A + B$

(b) $A + D$

(c) $B + C$

1.3.2 Soluzioni

(a) La rappresentazione di A nello standard IEEE754 Singola Precisione è:

$$S_{\mathrm{A}} = 0$$
$$M_{\mathrm{A}} = 1 + 0,0001111 = 1,0001111$$
$$E_{\mathrm{A}} = 5 + 127 = 132_{10} = 10000100_2$$

La rappresentazione di B nello standard IEEE754 Singola Precisione è:

$$S_{\mathrm{B}} = 1$$
$$M_{\mathrm{B}} = 1 + 0,11111111 = 1,11111111$$
$$E_{\mathrm{B}} = 5 + 127 = 132_{10} = 10000100_2$$

Svolgiamo la somma A + B nel seguente modo:

- Allineamento degli esponenti: poiché gli esponenti sono uguali, non serve allinearli.
- Essendo B un numero FP negativo (si guardi il bit di segno) usiamo l'algoritmo di cambio di segno ed esprimiamo i numeri in complemento a due:

$$M_{\mathrm{A}} = 01,0001111$$
$$M_{\mathrm{B}} = 01,11111111 \rightarrow 10,00000001$$

- Sommiamo:

$$\begin{array}{cccccccccccc} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & \\ & 0 & 1 & , & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ + & 1 & 0 & , & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline & 1 & 1 & , & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{array}$$

Osserviamo che il bit di segno del risultato è 1. Utilizziamo l'algoritmo di cambio di segno per trovare il valore assoluto del numero:

$$M_A + M_B = 11,00011111 \rightarrow 00,11100001$$

Scriviamo il risutato della somma in notazione scientifica normalizzata Base 2:

$$A + B = -0,11100001 \cdot 2^5 = -1,1100001 \cdot 2^4$$

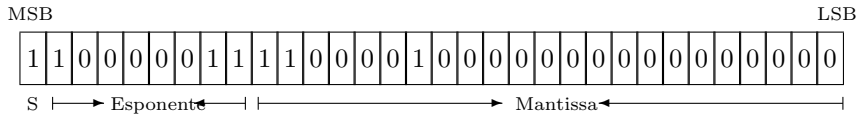
Disponiamo ora di tutti i dati per rappresentare il risultato della somma secondo lo standard IEEE754 Singola Precisione:

$$S = 1$$

$M = 1 + 0,1100001 = 1,1100001$

$E = 4 + 127 = 131_{10} = 10000011_2$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



(b) La rappresentazione di A nello standard IEEE754 Singola Precisione è:

$S_{\mathrm{A}} = 0$

$M_{\mathrm{A}} = 1 + 0,0001111 = 1,0001111$

$E_{\mathrm{A}} = 5 + 127 = 132_{10} = 10000100_2$

La rappresentazione di D nello standard IEEE754 Singola Precisione è:

$S_{\mathrm{D}} = 1$

$M_{\mathrm{D}} = 1 + 0,1000000101 = 1,1000000101$

$E_{\mathrm{D}} = 7 + 127 = 134_{10} = 10000110_2$

Svolgiamo la somma di $A + D$ nel seguente modo:

- Allineiamo l'esponente del numero più piccolo a quello del numero più grande (cioè allineiamo l'esponente di A):

$$A = 1,0001111 \cdot 2^5 = 0,010001111 \cdot 2^7$$

- Essendo D un numero FP negativo (si guardi il bit di segno) usiamo l'algoritmo di cambio di segno ed esprimiamo i numeri in complemento a due:

  $M_{\mathrm{A}} = 00,010001111$

  $M_{\mathrm{D}} = 01,1000000101 \rightarrow 10,0111111011$

- Sommiamo:

$$\begin{array}{cccccccccccccc} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 0} \\ & 0 & 0 & , & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ + & 1 & 0 & , & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ \hline & 1 & 0 & , & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{array}$$

Per prima cosa osserviamo che il bit del segno del risultato è 1. Utilizziamo l'algoritmo di cambio di segno per trovare il valore assoluto del numero:

$$M_{\mathrm{A}} + M_{\mathrm{D}} = 10,1100011001 \rightarrow 01,0011100111$$

Scriviamo il risultato della somma in notazione scientifica normalizzata Base2:

$$A + D = -1,0011100111 \cdot 2^7$$

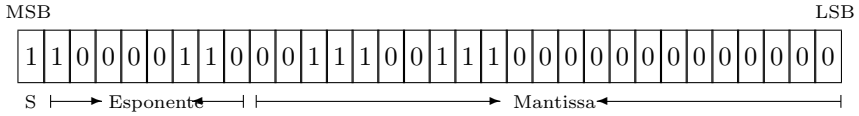
Disponiamo ora di tutti i dati per rappresentare il risultato della somma secondo lo standard IEEE754 Singola Precisione:

$S = 1$

$M = 1 + 0,0011100111 = 1,0011100111$

$E = 7 + 127 = 134_{10} = 10000110_2$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



(c) La rappresentazione di B nello standard IEEE754 Singola Precisione è:

$S_{\mathrm{B}} = 1$

$M_{\mathrm{B}} = 1 + 0,11111111 = 1,11111111$

$E_{\mathrm{B}} = 5 + 127 = 132_{10} = 10000100_2$

La rappresentazione di C nello standard IEEE754 Singola Precisione è:

$S_{\mathrm{C}} = 0$

$M_{\mathrm{C}} = 1 + 0,00010001001 = 1,00010001001$

$E_{\mathrm{C}} = 7 + 127 = 134_{10} = 10000110_2$

Svolgiamo la somma B + C nel seguente modo:

- Allineiamo l'esponente del numero più piccolo a quello del numero più grande (cioè allineiamo l'esponente di B):

$$B = -1,11111111 \cdot 2^5 = -0,0111111111 \cdot 2^7$$

- Essendo B un numero FP negativo (si guardi il bit di segno) usiamo l'algoritmo di cambio di segno ed esprimiamo i numeri in complemento a due:

$M_{\mathrm{B}} = 00,0111111111 \rightarrow 11,1000000001$
$M_{\mathrm{C}} = 01,00010001001$

- Ora possiamo sommare le due mantisse:

$$\begin{array}{cccccccccccccc} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 0} & & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} \\ & 1 & 1 & , & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ + & 0 & 1 & , & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ \hline & 0 & 0 & , & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array}$$

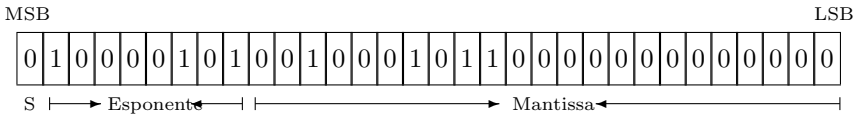
Per prima cosa osserviamo che il bit del segno del risultato è 0. Scriviamo il risultato della somma in notazione scientifica normalizzata Base2:

$$B + C = +0,10010001011 \cdot 2^7 = +1,0010001011 \cdot 2^6$$

Disponiamo ora di tutti i dati per rappresentare il risultato della somma secondo lo standard IEEE754:

$S = 0$
$M = 1 + 0,0010001011 = 1,0010001011$
$E = 6 + 127 = 133_{10} = 10000101_2$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



1.3.3 Esercizi

Date le seguenti sequenze binarie espresse su 32 bit:

$A = 01111011100110101111000000000000$

$B = 11111010011010100100000000000000$

$C = 01111100010100101111100000000000$

$D = 11111010100011011100000000000000$

Interpretarli come numeri FP espressi secondo lo standard IEEE754 Sine svolgere le seguenti operazioni:

(a) $A - B$

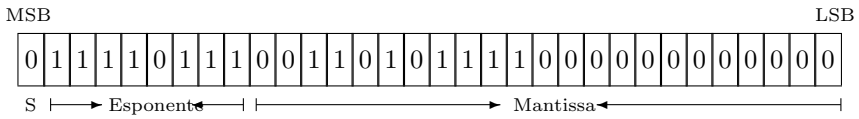
(b) $B + C$

(c) $C - B$

(d) $D + B$

1.3.4 Soluzioni

- Riscriviamo A come:



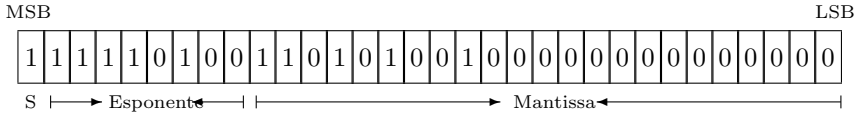
  Identifichiamo segno esponente e mantissa::

  $S_{\mathrm{A}} = 0$
  $M_{\mathrm{A}} = 1 + 0,00110101111 = 1,00110101111$
  $E_{\mathrm{A}} = 11110111_2 = 247_{10} = 120 + 127$

  Infine esprimiamo A in notazione scientifica normalizzata Base 2:

$$A = +1,00110101111 \cdot 2^{120}$$

- Riscriviamo B come:



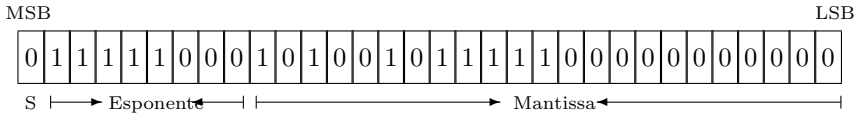
  Identifichiamo segno esponente e mantissa::

  $S_{\mathrm{B}} = 1$
  $M_{\mathrm{B}} = 1 + 0,110101001 = 1,110101001$
  $E_{\mathrm{B}} = 11110100_2 = 244_{10} = 117 + 127$

  Infine esprimiamo B in notazione scientifica normalizzata Base 2:

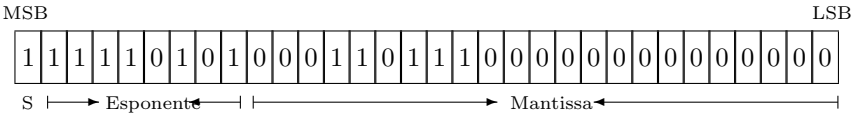
$$B = -1,110101001 \cdot 2^{117}$$

- Riscriviamo C come:



  Identifichiamo segno esponente e mantissa::

  $S_{\mathrm{C}} = 0$
  $M_{\mathrm{C}} = 1 + 0,101001011111 = 1,101001011111$
  $E_{\mathrm{C}} = 11111000_2 = 248_{10} = 121 + 127$



Infine esprimiamo C in notazione scientifica normalizzata Base 2:

$$C = +1,101001011111 \cdot 2^{121}$$

- Riscriviamo D come:
  Identifichiamo segno esponente e mantissa::

  $S_{\mathrm{D}} = 1$
  $M_{\mathrm{D}} = 1 + 0,000110111 = 1,000110111$
  $E_{\mathrm{D}} = 11110101_2 = 245_{10} = 118 + 127$

  Infine esprimiamo D in notazione scientifica normalizzata Base 2:

$$D = -1,000110111 \cdot 2^{118}$$

Ora svolgiamo le operazioni:

(a) La rappresentazione di A nello standard IEEE754 Singola Precisione è:

$S_{\mathrm{A}} = 0$
$M_{\mathrm{A}} = 1 + 0,00110101111 = 1,00110101111$
$E_{\mathrm{A}} = 11110111_2 = 247_{10} = 120 + 127$

La rappresentazione di B nello standard IEEE754 Singola Precisione è:

$S_{\mathrm{B}} = 1$
$M_{\mathrm{B}} = 1 + 0,110101001 = 1,110101001$
$E_{\mathrm{B}} = 11110100_2 = 244_{10} = 117 + 127$

Sapendo che B è un numero negativo possiamo riscrivere la sottrazione come segue $A+(-B)$. Cambiamo dunque il bit di segno di B e procediamo con l'algoritmo di somma:

- Allineiamo l'esponente del numero più piccolo a quello del numero più grande (cioè allineiamo l'esponente di B):

$$B = +1,110101001 \cdot 2^{117} = +0,001110101001 \cdot 2^{120}$$

- Essendo entrambi i numeri positivi possiamo direttamente fare la somma:

$$M_{\mathrm{A}} = +1,00110101111$$
$$M_{\mathrm{B'}} = +0,001110101001$$

Eseguiamo la somma:

$$
\begin{array}{cccccccccccccc}
 & {\scriptstyle 0} & {\scriptstyle 0} & & {\scriptstyle 0} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} \\
 & 1 & , & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 \\
+ & 0 & , & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
\hline
 & 1 & , & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{array}
$$

Avendo sommato due numeri positivi il risultato è sicuramente positivo quindi possiamo scrivere:

$$M_{\mathrm{A}} + M_{\mathrm{B}} = +1,011100000111$$

Riscriviamo il risultato in notazione scientifica normalizzata Base 2 (osserviamo che la rappresentazione è già normalizzata):

$$A + B = +1,011100000111 \cdot 2^{120}$$

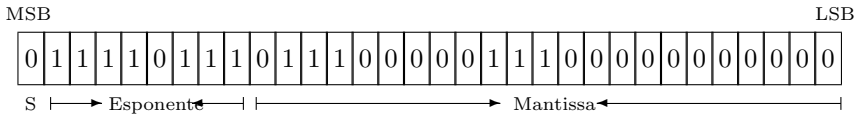
Disponiamo ora di tutti i dati per rappresentare il risultato della somma secondo lo standard IEEE754 Singola Precisione:

$S = 0$

$M = 1 + 0,011100000111 = 1,011100000111$

$E = 120 + 127 = 247_{10} = 11110111_2$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



(b) La rappresentazione di B nello standard IEEE754 Singola Precisione è:

$S_{\mathrm{B}} = 1$

$M_{\mathrm{B}} = 1 + 0,110101001 = 1,110101001$

$E_{\mathrm{B}} = 11110100_2 = 244_{10} = 117 + 127$

La rappresentazione di C nello standard IEEE754 Singola Precisione è::

$S_{\mathrm{C}} = 0$

$M_{\mathrm{C}} = 1 + 0,101001011111 = 1,101001011111$

$E_{\mathrm{C}} = 11111000_2 = 248_{10} = 121 + 127$

Svolgiamo la somma B + C nel seguente modo:

- Allineiamo l'esponente del numero più piccolo a quello del numero più grande (cioè allineiamo l'esponente di B):

$$B = -1,110101001 \cdot 2^{117} = -0,0001110101001 \cdot 2^{121}$$

- Essendo B un numero FP negativo (si guardi il bit di segno) usiamo l'algoritmo di cambio di segno ed esprimiamo i numero in complemento a due:

$$M_{\mathrm{B}} = 00,0001110101001 \rightarrow 11,1110001010111$$

$$M_{\mathrm{C}} = 01,101001011111$$

- Eseguiamo la somma:

$$\begin{array}{cccccccccccccccccc}
 & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 0} \\
 & 1 & 1 & , & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\
+ & 0 & 1 & , & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\
\hline
 & 0 & 1 & , & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1
\end{array}$$

  Osserviamo che il bit del segno del risultato è 0. Possiamo ora scrivere il risutato della somma secondo la notazione scientifica normalizzata in Base 2 (osserviamo che la rappresentazione è già normalizzata):

$$B + C = +1,1000100010101 \cdot 2^{121}$$

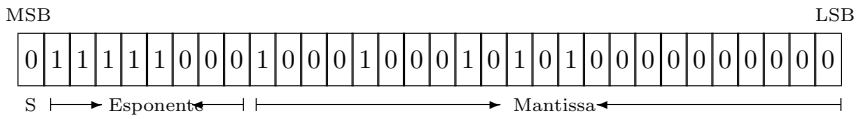
Disponiamo ora di tutti i dati per rappresentare il risultato della somma secondo lo standard IEEE754 Singola Precisione:

$S = 0$

$M = 1 + 0,1000100010101 = 1,1000100010101$

$E = 121 + 127 = 248_{10} = 11111000_2$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



(c) La rappresentazione di C nello standard IEEE754 Singola Precisione è:

$S_C = 0$

$M_C = 1 + 0,101001011111 = 1,101001011111$

$E_C = 11111000_2 = 248_{10} = 121 + 127$

La rappresentazione di B nello standard IEEE754 Singola Precisione è::

$S_B = 1$

$M_B = 1 + 0,110101001 = 1,110101001$

$$E_B = 11110100_2 = 244_{10} = 117 + 127$$

Sapendo che B è un numero negativo possiamo riscrivere la sottrazione come segue $C+(-B)$. Cambiamo dunque il bit di segno di B e procediamo con l'algoritmo di somma:

- Allineiamo l'esponente del numero più piccolo a quello del numero più grande (cioè allineiamo l'esponente di B):

$$B = +1,110101001 \cdot 2^{117} = +0,0001110101001 \cdot 2^{121}$$

- Essendo entrambi i numeri positivi possiamo direttamente fare la somma:

$$M_{\mathrm{C}} = 1,101001011111$$
$$M_{\mathrm{B}} = 0,0001110101001$$

Eseguiamo la somma come segue:

$$\begin{array}{cccccccccccccccc} & {\scriptstyle 0} & {\scriptstyle 0} & & {\scriptstyle 0} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 0} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & \\ & 1 & , & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ + & 0 & , & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ \hline & 1 & , & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \end{array}$$

Avendo sommato due numeri positivi il risultato è sicuramente positivo quindi possiamo scrivere:

$$M_C + M_B = 1,1100001100111$$

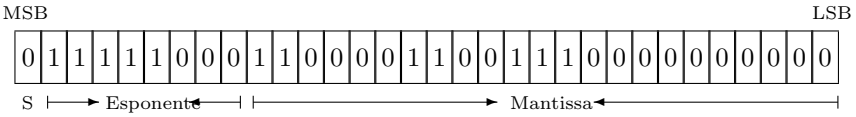
Scriviamo il risutato della somma secondo la notazione scientifica normalizzata in Base 2 (osserviamo che la rappresentazione è già normalizzata):

$$C + B = +1,1100001100111 \cdot 2^{121}$$

Disponiamo ora di tutti i dati per rappresentare il risultato della somma secondo lo standard IEEE754 Precisione Singola:

$$S = 0$$
$$M = 1 + 0,1100001100111 = 1,1100001100111$$
$$E = 121 + 127 = 248_{10} = 11111000_2$$

Ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



(d) La rappresentazione di D nello standard IEEE754 Singola Precisione è:

$S_D = 1$

$M_D = 1 + 0,000110111 = 1,000110111$

$E_D = 11110101_2 = 245_{10} = 118 + 127$

La rappresentazione di B nello standard IEEE754 Singola Precisione è::

$S_B = 1$

$M_B = 1 + 0,110101001 = 1,110101001$

$E_B = 11110100_2 = 244_{10} = 117 + 127$

Svolgiamo la somma D + B nel seguente modo:

- Allineiamo l'esponente del numero più piccolo a quello del numero più grande (cioè allineiamo l'esponente di B):

$$B = -1,110101001 \cdot 2^{117} = -0,1110101001 \cdot 2^{118}$$

- Essendo D e B numeri FP negativi (si guardi il bit di segno) possiamo sommare il loro valore assoluto tenendoci a mente che il risultato sarà sicuramente negativo. Procediamo dunque con la somma:

$$\begin{array}{cccccccccccccc} & {\scriptstyle 1} & {\scriptstyle 1} & & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 1} & {\scriptstyle 0} & {\scriptstyle 1} & {\scriptstyle 0} & {\scriptstyle 0} & {\scriptstyle 0} & \\ & & 1 & , & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ + & & 0 & , & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ \hline & 1 & 0 & , & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{array}$$

N.B: è importante non interpretare l'uno nella posizione più significativa come il segno del risultato. Questo perchè stiamo sommando valori assoluti.

Notiamo che è necessario eseguire la normalizzazione per rappresentare il risultato in notazione scientifica normalizzata Base 2:

$$D + B = -10,0000010111 \cdot 2^{118} = -1,00000010111 \cdot 2^{119}$$

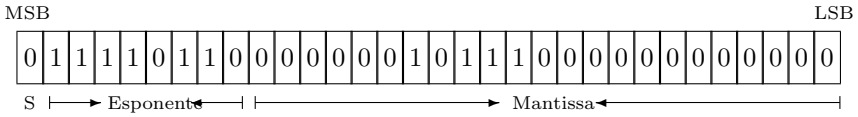
Disponiamo ora di tutti i dati per rappresentare il risultato della somma secondo lo standard IEEE754 Precisione Singola:

$S = 1$

$M = 1 + 0,00000010111 = 1,00000010111$

$$E = 119 + 127 = 246_{10} = 11110110_2$$

Ora ricordando che in singola precisione disponiamo di 32 bit rispettivamente 1 per il segno 8 per l'esponente e 23 per la mantissa scriviamo:



2 Risorse Esterne

- Floating Point Numbers - Computerphile
  https://www.youtube.com/watch?v=PZRI1IfStY0