Rappresentazione di un reale in virgola mobile

Sia dato il numero reale a = -7.3. Rappresentarlo in virgola mobile con precisione dimezzata (half precision), cioè con 1 bit di segno, 5 di esponente, e 10 di parte frazionaria.

Soluzione

Prima di tutto occorre esprimere il numero in cifre binarie. Con il metodo del div&mod troviamo la rappresentazione della parte intera (7):

$q_0 = 7$

$q_1 = q_0 \text{ div } 2 = 3$ $\mathbf{a_0} = q_0 \text{ mod } 2 = \mathbf{1}$

$q_2 = q_1 \text{ div } 2 = 1$ $\mathbf{a_2} = q_1 \text{ mod } 2 = \mathbf{1}$

$q_3 = q_2 \text{ div } 2 = \mathbf{0}$ $\mathbf{a_3} = q_2 \text{ mod } 2 = \mathbf{1}$

Quindi 7 = $(\mathbf{111})_2$.

Andiamo a rappresentare la parte frazionaria, cioè 0.3, con il metodo "parte frazionaria – parte intera":

$f_0 = 0.3$

$f_{-1} = F(f_0 \cdot 2 = 0.3 \cdot 2 = \mathbf{0.6}) = 0.6$ $a_{-1} = I(0.6) = \mathbf{0}$

$f_{-2} = F(f_{-1} \cdot 2 = 0.6 \cdot 2 = \mathbf{1.2}) = 0.2$ $a_{-2} = I(1.2) = \mathbf{1}$

$f_{-3} = F(f_{-2} \cdot 2 = 0.2 \cdot 2 = \mathbf{0.4}) = 0.4$ $a_{-3} = I(0.4) = \mathbf{0}$

$f_{-4} = F(f_{-3} \cdot 2 = 0.4 \cdot 2 = \mathbf{0.8}) = 0.8$ $a_{-4} = I(0.8) = \mathbf{0}$

$f_{-5} = F(f_{-4} \cdot 2 = 0.8 \cdot 2 = \mathbf{1.6}) = 0.6$ $a_{-5} = I(1.6) = \mathbf{1}$ f_{-5} è uguale a f_{-1}, quindi il metodo si ripete uguale a partire dal passo successivo

$a_{-6} = a_{-2} = \mathbf{1}$

$a_{-7} = a_{-3} = \mathbf{0}$

$a_{-8} = a_{-4} = \mathbf{0}$

$a_{-9} = a_{-5} = \mathbf{1}$

$a_{-10} = a_{-2} = \mathbf{1}$

E così via all'infinito. Il numero in cifre binarie sarà dunque:

$a = (-111.0\underline{1001})_2$,

dove le cifre sottolineate indicano il periodo. Normalizzando tale numero in modo da avere la forma a = -1.[F]·2^e, dove F è la parte frazionaria ed e è l'esponente, si ottiene:

$a = (-1.110\underline{1001})_2 \cdot 2^2$.

La parte frazionaria sarà dunque: F = **1101001100**, su 10 bit. Notare che sono state troncate le cifre binarie dalla decima in poi, quindi la rappresentazione in virgola mobile di -7.3 non avrà precisione infinita. Il segno sarà S = **1**, in quanto il numero a è negativo. Andiamo adesso a calcolare la rappresentazione dell'esponente e = 2. Tale rappresentazione è con bias, il cui valore dipende dai bit di rappresentazione. In generale: bias = 2^{p-1}-1. In half precision abbiamo p = 5, quindi bias = **15**. La rappresentazione dell'esponente e = 2 sarà dunque E = 15+2 = 17 = (mediante metodo div&mod omesso) $\mathbf{(10001)_2}$.

Infine, la rappresentazione del numero a = -7.3 sarà A = **1|10001|1101001100**.

Da rappresentazione in virgola mobile a numero

Sia data la rappresentazione in virgola mobile in half precision A = 0|01110|1100000000. Si calcoli il corrispondente numero reale rappresentato.

Soluzione

Il segno S = 0, quindi il reale è **positivo**.

La rappresentazione con bias del suo esponente è E = 01110. Il bias per gli half precision è sempre bias = 15, quindi l'esponente vale e = $(01110)_2$ - 15 = (mediante metodo della sommatoria omesso) 14 – 15 = **-1**. La forma normalizzata del numero sarà quindi a = $+(1.11)_2 \cdot 2^{-1}$ = **$+(0.111)_2$**.

Applicando il metodo della sommatoria generalizzata si ottiene: a = $+(0.111)_2 = 1 \cdot 2^{-1} + 1 \cdot 2^{-2} + 1 \cdot 2^{-3}$ = 0.5 + 0.25 + 0.125 = **0.875**.

Moltiplicazione per costante in virgola mobile

Sia data la rappresentazione in virgola mobile in half precision A = 0|10010|1101000000. Si calcoli la rappresentazione A' del suo quadruplo.

Soluzione

Sfruttando il fatto che i numeri in virgola mobile sono normalizzati nella forma la forma a = $-1.[F] \cdot 2^e$, le moltiplicazioni e le divisioni per potenze di due sono realizzabili operando solo sull'esponente. In questo caso, a' = $4 \cdot a = 4 \cdot (+1.[F]) \cdot 2^e = 2^2 \cdot (+1.[F]) \cdot 2^e = (+1.[F]) \cdot 2^{e+2}$. Quindi la rappresentazione A' avrà esattamente lo stesso bit di segno e gli stessi bit di parte frazionaria di A, ma con un esponente incrementato di due: e' = e+2. Per incrementare di due un numero rappresentato con bias basta incrementare di due la rappresentazione stessa. Svolgendo il calcolo in colonna si ottiene:

```
E = 10010 +

2 = 00010 =

    _______
     1   (riporti)

    10100
```

Quindi E' = **10100**. Infine la rappresentazione del quadruplo sarà A' = **0|10100|1101000000**.