Сложение в формате с плавающей точкой

Содержание

Сложение чисел с фиксированной точкой

В формате с фиксированной точкой масштаб результата сложения такой же, что и масштаб операндов. Поэтому операция сложения особенностей не имеет и выполняется по правилам сложения дополнительных кодов.

Example

Выполнить сложение чисел $(-30.625 + -5.75)$ в формате с фиксированной точкой 16-разрядной сетке с масштабом $M = 2^{-3}$.

$-30.625 = (-11110.101)_2, -5.75 = (101.11)_2.$

Сложение чисел с фиксированной точкой

Решение.

$$-30.625 = \overset{15 \qquad\qquad\qquad\qquad 0}{\boxed{1111111100001011}}$$

$$-5.75 = \overset{15 \qquad\qquad\qquad\qquad 0}{\boxed{1111111111010010}}$$

Сложение в МДК:

$$\begin{array}{r} + \begin{array}{l} 1\ \ 11111111\ \ 00001011 \\ 1\ \ 11111111\ \ 11010010 \end{array} \\ \hline 1\ \ 11111110\ \ 11011101 \end{array}$$

ПРС отсуствует. Результат с масштабом 2^{-3}:

$$\overset{15 \qquad\qquad\qquad\qquad 0}{\boxed{1111111011011101}}$$

$1111111011011,101 = (-100100,011)_2 = -36.375$

Формат с плавающей точкой

$$X = m_X \cdot 2^{p_X},$$

где m_X — нормализованная мантисса числа X, p_X — порядок числа X, подобранный так, чтобы m_X была нормализованной.

Правила нормализации $X \neq 0$

Мантисса m_X получается из двоичного представления X переносом точки в такую позицию, чтобы целая часть была равна нулю, а в старшем разряде дробной части была единица:

$$(0.\underbrace{1xxx\cdots xxx}_{\text{разряды мантиссы}})_2$$

порядок p_X определяет на сколько разрядов нужно передвинуть запятую в мантиссе, чтобы получить исходное число.

Формат для примеров

15	14 6	5	4 0
X	XXXXXXXXX	X	XXXXX

- Разряды нормализованной мантиссы в прямом коде хранятся в разрядах [15 : 6].
- Порядок в прямом коде хранится в разрядах [5 : 0].

Целое, без потерь

$$-9 = (-1001)_2 = (-.1001) \cdot 2^4$$

15	14 … 6	5	4 … 0
1	100100000	0	00100

$$(-.100100000)_2 \cdot 2^4 = -9$$

Целое, с потерями

$$999 = (1111100111)_2 = (.1111100111) \cdot 2^{10}$$

15	14 … 6	5	4 … 0
0	111110011	0	01010

$$(.111110011)_2 \cdot 2^{10} = \boxed{998}$$

Дробное, без потерь

$$0.171875 = (0.001011)_2 = (.1011) \cdot 2^{-2}$$

15	14 — 6	5	4 — 0
0	101100000	1	00010

$$(.101100000)_2 \cdot 2^{-2} = 0,171875$$

Дробное, с потерями

$$-0.2 = (-0.[0011])_2 = (-.[1100]) \cdot 2^{-2}$$

15	14 … 6	5	4 … 0
1	110011001	1	00010

$$(-.110011001)_2 \cdot 2^{-2} = \boxed{\text{-0.19970703125}}$$

Ноль

15	14 6	5	4 0
0	000000000	0	00000

Правила сложения

$$X + Y = m_X \cdot 2^{p_X} + m_Y \cdot 2^{p_Y}$$

1. Порядки чисел выравниваются до большего, мантисса числа с меньшим порядком сдвигается вправо на модуль разности порядков.

$$\begin{cases} (p_X - p_Y) \geq 0, & m'_Y \leftarrow (m_Y \gg |p_X - p_Y|), m'_X \leftarrow m_X, \\ (p_X - p_Y) < 0, & m'_X \leftarrow (m_X \gg |p_X - p_Y|), m'_Y \leftarrow m_Y. \end{cases}$$

2. Получившиеся мантиссы складываются $m_R = m'_X + m'_Y$. При этом порядок результата: $p_R = \max(p_X, p_Y)$.
3. Выполняется нормализация результата, если он получился не нормализованым.

13+29

15	14 … 6	5	4 … 0
0	110100000	0	00100

\+

15	14 … 6	5	4 … 0
0	111010000	0	00101

m	p	прим.
0,110100000	0,00100	$X = 13$, ПК
0,111010000	0,00101	$Y = 29$, ПК
	00,00100 + 11,11011 11,11111	$p_X - p_Y < 0$, в МДК, денормализуется m_X
0,011010000	0,00101	X', денормализованное
00,011010000 + 00,111010000 01,010100000		$m_R = m'_X + m'_Y$, в МДК, ПРС!
01,010100000	0,00101	Нормализовать! $m_R \leftarrow m_R \gg 1$; $p_R \leftarrow p_R + 1$
0,101010000	0,00110	Рез-т!

15	14 … 6	5	4 … 0
0	101010000	0	00110

-17+14

15	14 … 6	5	4 … 0
1	100010000	0	00101

+

15	14 … 6	5	4 … 0
0	111000000	0	00100

m	p	прим.
1,100010000 0,111000000	0,00101 0,00100	$X = -17$, ПК $Y = 14$, ПК
	+ 00,00101 11,11100 00,00001	$p_X - p_Y \geq 0$, МДК, денормализуется m_Y
0,011100000	0,00101	Y', денормализованное
+ 11,011110000 00,011100000 11,111010000		$m_R = m'_X + m'_Y$, МДК
11,111010000	0,00101	Получить модуль мантиссы для представления в ПК!
00,000110000 00,001100000 00,011000000	0,00101 0,00100 0,00011	Нормализовать модуль! $m_R \leftarrow m_R \ll 1$; $p_R \leftarrow p_R - 1$ Нормализовать модуль! $m_R \leftarrow m_R \ll 1$; $p_R \leftarrow p_R - 1$ Нормализовать модуль! $m_R \leftarrow m_R \ll 1$; $p_R \leftarrow p_R - 1$
1,110000000	0,00010	Рез-т!

15	14 … 6	5	4 … 0
1	110000000	0	00010

-2+-2

15	14 … 6	5	4 … 0
1	100000000	0	00010

$+$

15	14 … 6	5	4 … 0
1	100000000	0	00010

m	p	прим.
1,100000000 1,100000000	0,00010 0,00010	$X = -2$, ПК $Y = -2$, ПК
	00,00010 + 11,11110 00,00000	$p_X - p_Y = 0$, порядки одинаковы
11,100000000 + 11,100000000 11,000000000		$m_R = m'_X + m'_Y$, МДК
$\underline{\underline{11}}$,000000000	0,00010	Получить модуль мантиссы для представления в ПК!
$\underline{\underline{01}}$,000000000	0,00010	Нормализовать модуль! $m_R \leftarrow m_R \gg 1$; $p_R \leftarrow p_R + 1$
1,100000000	0,00011	Рез-т!

15	14 … 6	5	4 … 0
1	100000000	0	00011

ПРС (переполнение разрядной сетки)

15	14 … 6	5	4 … 0
0	111100000	0	11111

+

15	14 … 6	5	4 … 0
0	110000000	0	11101

m	p	прим.
0,111100000 0,110000000	0,11111 0,11101	X, ПК Y, ПК
	+ 00,11111 11,00011 00,00010	$p_X - p_Y \geq 0$, МДК, денормализуется m_Y
0,001100000	0,11111	Y', денормализованное
+ 00,111100000 00,001100000 01,001000000		$m_R = m'_X + m'_Y$, МДК, ПРС мантиссы!
01,001000000	0,11111	Нормализовать! $m_R \leftarrow m_R \gg 1$; $p_R \leftarrow p_R + 1$
00,100100000	?,?????	ПРС порядка — настоящий ПРС в формате с ПЗ!

Генерация ошибки вычислений!

ПМР (потеря младших разрядов)

15	14 … 6	5	4 … 0
0	100001000	1	11110

$+$

15	14 … 6	5	4 … 0
1	111100000	1	11111

m	p	прим.
0,100001000	1,11110	X, ПК
1,111100000	1,11111	Y, ПК
	11,00010 + 00,11111 00,00001	$p_X - p_Y \geq 0$, МДК, денормализуется m_Y
1,011110000	1,11110	Y', денормализованное
00,100001000 + 11,100010000 00,000011000		$m_R = m'_X + m'_Y$, МДК
$\underline{\underline{00}}$,000011000	1,11110	Нормализовать! $m_R \leftarrow m_R \ll 1$; $p_R \leftarrow p_R - 1$
$\underline{\underline{00}}$,000110000	1,11111	Нормализовать! $m_R \leftarrow m_R \ll 1$; $p_R \leftarrow p_R - 1$
$\underline{\underline{00}}$,001100000	?,?????	p_R за пределом представления *отрицательных* чисел в ПК!

15	14 … 6	5	4 … 0
0	000000000	0	00000

$X + Y = X, Y \neq 0$?

15	14 … 6	5	4 … 0
0	111000000	0	11000

+

15	14 … 6	5	4 … 0
0	110000000	0	01110

m	p	прим.
0,111000000	0,11000	$X = 7 \cdot 2^{21}$, ПК
0,110000000	0,01110	$Y = 3 \cdot 2^{12}$, ПК
	+ 00,11000 11,10010 00,01010	$\|p_X - p_Y\| \geq 9$, МДК, m_Y денормализуется в 0
0,000000000	0,11000	$Y' = 0$?

15	14 … 6	5	4 … 0
0	111000000	0	11000

Характеристика

$$X = m_X \cdot 2^{p_X}.$$

Диапазон представления порядка p_X в n-разрядной сетке будет[1]:

$$p_X \in [-2^{n-1}, +(2^{n-1} - 1)]$$

Характеристика получается из порядка прибавлением фиксированной поправки Δ, такой, что левая граница представления обращается в ноль. Таким образом,

характеристика c_X — всегда положительное число.

$$c_X = p_X + \Delta, \tag{1}$$

где[a] $\Delta = +2^{n-1}$, а $c_X \in [0, 2^n - 1]$.

[a] Опять же только в случае использования дополнительного кода

[1] Если использовать дополнительный код

Свойства n-разрядной характеристики

- Характеристика — положительное число.
- Разность характеристик равна разности порядков.
- Если в процессе нормализации (или денормализации) порядок увеличивается (или уменьшается), то то же самое происходит и с характеристикой.
- Если для работы с характеристиками использовать ДК или МДК, о ПРС при нормализации легко судить по знаковому разряду: он не должен быть 1.
- Если использвется поправка $\Delta = 2^{n-1}$, то характеристика получается из дополнительного кода порядка инверсией знакового разряда.

Формат для примеров

15	14 … 6	5 … 0
X	XXXXXXXXX	XXXXXX

- Разряды нормализованной мантиссы в прямом коде хранятся в разрядах [15 : 6].
- Характеристика хранится в разрядах [5 : 0].
- $\Delta = 2^5 = 32 = (100000)_2$

Целое, без потерь

$\Delta = 2^5 = 32$

$$-9 = (-1001)_2 = (-.1001) \cdot 2^4$$

15	14 … 6	5 … 0
1	100100000	100100

$$(-.100100000)_2 \cdot 2^{36-\Delta} = -9$$

Целое, с потерями

$\Delta = 2^5 = 32$

$$999 = (1111100111)_2 = (.1111100111) \cdot 2^{10}$$

15	14 … 6	5 … 0
0	111110011	101010

$$(.111110011)_2 \cdot 2^{42-\Delta} = \boxed{998}$$

Дробное, без потерь

$\Delta = 2^5 = 32$

$$0.171875 = (0.001011)_2 = (.1011) \cdot 2^{-2}$$

15	14 … 6	5 … 0
0	101100000	011110

$$(.101100000)_2 \cdot 2^{30-\Delta} = 0,171875$$

Дробное, с потерями

$\Delta = 2^5 = 32$

$$-0.2 = (-0.[0011])_2 = (-.[1100]) \cdot 2^{-2}$$

15	14 … 6	5 … 0
1	110011001	011110

$$(-.110011001)_2 \cdot 2^{30-\Delta} = \boxed{\text{-0.19970703125}}$$

Ноль

15	14 6	5 0
0	000000000	000000

Правила

Исходя из формулы (1), почти не меняются

$$X + Y = m_X \cdot 2^{c_X - \Delta} + m_Y \cdot 2^{c_Y - \Delta}$$

1. Характеристики чисел выравниваются до большей, мантисса числа с меньшей характеристикой сдвигается вправо на модуль разности характеристик.

$$\begin{cases} (c_X - c_Y) \geq 0, & m'_Y \leftarrow (m_Y \gg |c_X - c_Y|), m'_X \leftarrow m_X, \\ (c_X - c_Y) < 0, & m'_X \leftarrow (m_X \gg |c_X - c_Y|), m'_Y \leftarrow m_Y. \end{cases}$$

2. Получившиеся мантиссы складываются $m_R = m'_X + m'_Y$. При этом характеристика результата: $c_R = \max(c_X, c_Y)$.
3. Выполняется нормализация результата, если он получился не нормализованым.

13+57

15	14 … 6	5 … 0
0	110100000	100100

\+

15	14 … 6	5 … 0
0	111001000	100110

m	c	прим.
0,110100000 0,111001000	100100 100110	$X = 13$ $Y = 57$
	0,100100 + 1,011010 1,111110	$c_X - c_Y < 0$, ДК, денормализуется m_X
0,001101000	0,100110	X', денормализованное
00,001101000 + 00,111001000 01,000110000		$m_R = m'_X + m'_Y$, в МДК, ПРС!
$\underline{\underline{01}}$,000110000	0,100111	Нормализовать! $m_R \leftarrow m_R \gg 1$; $c_R \leftarrow c_R + 1$
0,100011000	100111	Рез-т!

15	14 … 6	5 … 0
0	100011000	100111

-17+14

15	14 … 6	5 … 0
1	100010000	100101

+

15	14 … 6	5 … 0
0	111000000	100100

m	c	прим.
1,100010000 0,111000000	100101 100100	$X = -17$ $Y = 14$
	+ 0,100101 1,011100 0,000001	$c_X - c_Y \geq 0$, ДК, денормализуется m_Y
0,011100000	0,100101	Y', денормализованное
+ 11,011110000 00,011100000 11,111010000		$m_R = m'_X + m'_Y$, МДК
11,111010000	0,100101	Получить модуль мантиссы для представления в ПК!
00,000110000 00,001100000 00,011000000	0,100101 0,100100 0,100011	Нормализовать модуль! $m_R \leftarrow m_R \ll 1$; $c_R \leftarrow c_R - 1$ Нормализовать модуль! $m_R \leftarrow m_R \ll 1$; $c_R \leftarrow c_R - 1$ Нормализовать модуль! $m_R \leftarrow m_R \ll 1$; $c_R \leftarrow c_R - 1$
1,110000000	100010	Рез-т!

15	14 … 6	5 … 0
1	110000000	100010

ПРС (переполнение разрядной сетки)

15	14 … 6	5 … 0
0	111100000	111111

+

15	14 … 6	5 … 0
0	110000000	111101

m	c	прим.
0,111100000 0,110000000	111111 111101	X Y
	+ 0,111111 1,000011 0,000010	$c_X - c_Y \geq 0$, ДК, денормализуется m_Y
0,001100000	0,111111	Y', денормализованное
+ 00,111100000 00,001100000 01,001000000		$m_R = m'_X + m'_Y$, МДК, ПРС мантиссы!
$\underline{\underline{01}}$,001000000	0,111111	Нормализовать! $m_R \leftarrow m_R \gg 1$; $c_R \leftarrow c_R + 1$
$\underline{\underline{00}}$,100100000	$\underline{\underline{1}}$,000000	$c_R < 0$, выход за правую границу представления — ПРС!

Генерация ошибки вычислений!

ПМР (потеря младших разрядов)

15	14 … 6	5 … 0
0	100001000	000010

$+$

15	14 … 6	5 … 0
1	111100000	000001

m	c	прим.
0,100001000	000010	X
1,111100000	000001	Y
	0,000010 + 1,111111 0,000001	$c_X - c_Y \geq 0$, МДК, денормализуется m_Y
1,011110000	0,000010	Y', денормализованное
00,100001000 + 11,100010000 00,000011000		$m_R = m'_X + m'_Y$, МДК
$\underline{\underline{00}}$,000011000	0,000010	Нормализовать! $m_R \leftarrow m_R \ll 1$; $c_R \leftarrow c_R - 1$
$\underline{\underline{00}}$,000110000	0,000001	Нормализовать! $m_R \leftarrow m_R \ll 1$; $c_R \leftarrow c_R - 1$
$\underline{\underline{00}}$,001100000	0,000000	Нормализовать! $m_R \leftarrow m_R \ll 1$; $c_R \leftarrow c_R - 1$
$\underline{\underline{00}}$,011000000	$\underline{\underline{1}}$,111111	$c_R < 0$, выход за левую границу представления — ПМР!

15	14 … 6	5 … 0
0	000000000	000000