Floating Points

$x = (-1)^{s_x} \cdot m_x \cdot 2^{e_x}$

$y = (-1)^{s_y} \cdot m_y \cdot 2^{e_y}$

Annahme: $e_x \geq e_y$

Falls $e_x < e_y$: swap x & y

```
swap(x; y) {
    tmp := x;
    x := y;
    y := tmp;
}
```

Input: $s_x, m_x, e_x, s_y, m_y, e_y$

Ouptput: $x+y$ (s_z, m_z, e_z)

Annahme $e_x \geq e_y$

① Falls $s_x = s_y$ dann,

$s_z := s_x$;

$m_z := m_x + m_y \cdot 2^{e_y - e_x}$;

$e_z := e_x$;

② $s_x \neq s_y$

2a $e_x > e_y$

$s_z := s_x$

$m_z := m_x - m_y \cdot 2^{e_y - e_x}$;

$e_z = e_x$;

2b $e_x = e_y$

$m_z := m_x - m_y$

Falls $e_x > e_y \Rightarrow |x| > |y|$ $\quad e_x \geq e_y + 1$

$|x| = m_x \cdot 2^{e_x}$ $\quad 1 \leq m_x < 2$

$|y| = m_y \cdot 2^{e_y}$ $\quad 1 \leq m_y < 2$

$$|x| - |y| = m_x \cdot 2^{e_x} - m_y \cdot 2^{e_y}$$

$$\geq 2^{e_x} - 2 \cdot 2^{e_y}$$

$$= 2^{e_x} - 2^{e_y + 1}$$

$$\geq 0$$

$$x \cdot y = (-1)^{s_x} m_x \cdot 2^{e_x} \cdot (-1)^{s_y} \cdot m_y \cdot 2^{e_y}$$

$$= (-1)^{s_x \oplus s_y} m_x \cdot m_y \cdot 2^{e_x + e_y}$$

s_x	s_y	s_f
0	0	0
0	1	1
1	0	1
1	1	0

$s_f = s_x \oplus s_y$

$1 \le m_x < 2$

$1 \le m_y < 2$

$1 \le m_x \cdot m_y < 4$

$e_x + e_y$, 8 bits exponent.

$p = 10000101 \Rightarrow e_x = 6$

$q = 10000111 \Rightarrow e_y = 8$

$e_x + e_y = p - 128 + 1 + q$

$= (00000101)_2 + (00000001)_2 + q$

$= (00000110)_2 + (10000111)_2$

$= (10001101)_2$

m_x $1.1010 \cdot 2^{-4}$

$1.1101 \cdot 2^{-4}$

```
   1 1 0 1 0
 1 1 0 1 0
1 1 0 1 0
+) 1 1 0 1 0
-----------
1 0 1 1 1 1 0 0 1 0
```