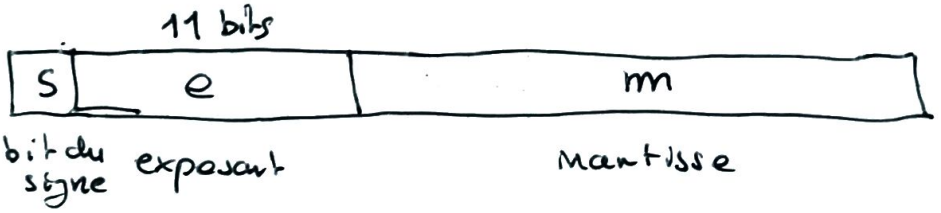
IEEE-754

Repr. EEE-754 d'un nombre dyadique $x \neq 0$:

$$x = \pm(1, b_1 \ldots b_k)_2 \cdot 2^e \quad \text{avec } \underbrace{e \in \mathbb{Z}}_{\text{exposant.}}$$



$$e' = e + 2^{10} - 1$$

$$e' \notin \{0, 2^{11} - 1\} \quad \text{(valeurs réservées)}$$

donc $e' \in [\![-1022, 1023]\!]$

---

en *float16* :

- repr de 1 : $+1,00000 \to 1$

$$(1, \underbrace{\qquad}_{10 \text{ bits}}) \cdot 2^{\underbrace{e'}_{5 \text{ bits}}}$$

$e' = e + 2^4 - 1$
$= e + 15$

$$(1, 0000\,000000) \cdot 2^0$$

$e' = 0 + 15 = 15$

$2^4\ 2^3\ 2^2\ 2^1\ 2^0$
$0\ \ 1\ \ 1\ \ 1\ \ 1$

$$(0.0111100000000000)_{2,\text{IEEE-754}}$$

repr de -2 : $-(1, \underline{\qquad})\ 2^1$

$e' = 1 + 15 = 16$

16 8 4 2 1
1 0 0 0 0

$$\to (1\,10000\,0000000000)_{2,\text{IEEE-754}}$$

- nombre qui suit 1 :

$$(0\,;01111\,;0000000001)_2$$

$$(1,0000000001)_2 \cdot 2^0$$

$$= 1 + 2^{-10}$$

$$\text{max} = (0\,;11110\,;1111111111)_2 =$$

$$\text{min} = (0\,;0001\,;0000000000)_2 =$$

nombre représenté par $(0\ \underbrace{10000}_{e'}\ 1001001000)_2$

$$e = e' - 15$$

$$= (10000)_2 - 15$$

$$= 16 - 15$$

$$= 1$$

$$+(1,1001001000)_2 \cdot 2^1$$

(• -1 -2 -3 -4 -5 -6 -7 -8 -9 -10)

$$= 2 + 2^0 + 2^{-3} + 2^{-6}$$

$$= 2 + 1 + \frac{1}{8} + \frac{1}{64}$$

Absorption:

$1 + 2^{53} = 2^{53}$ Donc $\mathbb{K}$ = n'est pas réflexif.

Cancellation:

diff entre deux qte très proche → grd perte de préc.