TP : La norme IEEE-754

Partie A : Calculs préliminaires

1. Convertir les nombres binaires suivants en décimal :

$$(11,01)_2 \qquad (1110,10101)_2 \qquad (0,0010\,01)_2$$

2. Convertir les nombres réels suivants en binaire :

$$6,75 \qquad 11,5625 \qquad 0,40625$$

3. Tout nombre décimal peut s'écrire sous la forme $\pm M \times 10^E$, où $M \in [1;10[$ est appelé la **mantisse** et E l'**exposant**. Par exemple :

$$712,48 = 7,1248 \times 10^2 \qquad \Longrightarrow \qquad M = 7,1248 \quad E = 2$$

En binaire, c'est (presque) pareil : tout nombre s'écrit sous la forme $\pm 1,M \times 2^E$.
Ici, M est appelé **pseudo-mantisse**, car il ne représente que la partie décimale (la partie entière étant forcément 1, on parle de *bit implicite*). Exemple :

$$(101,001)_2 = (1,0100\,1)_2 \times 2^2 \qquad \Longrightarrow \qquad M = 01001 \quad E = 2$$

☞ *Pour M, on prend tous les bits après la virgule, même les 0 « inutiles »*

Calculer les valeurs de M et E pour chacun des exemples de l'exercice 1.

Partie B : La norme IEEE 754

La **norme IEEE 754** (pour *Institute of Electrical and Electronics Engineers*) est une norme de représentation des **nombres à virgule flottante** utilisée en informatique. Elle définit comment les nombres réels sont stockés et manipulés dans les ordinateurs, en particulier pour assurer la **portabilité**, la **précision** et la **cohérence des calculs** numériques. Elle définit le codage des nombres réels de la manière suivante :

Signe (S)	Exposant (E)	Pseudo-Mantisse (M)
1 bit	e bits	m bits

- Le bit de signe est égal à 0 pour les nombres positifs, 1 pour les nombres négatifs.
- En cas d'exposant négatif, on n'utilise pas le complément à 2 mais **on décale l'exposant** en lui ajoutant la valeur $d = 2^{e-1} - 1$ (où e est le nombre de bits utilisé pour le codage de l'exposant).
- Pour la pseudo-mantisse M, on complète si besoin avec des zéros à droite pour l'écrire sur m bits.
  Par exemple, si $m = 10$ bits et $M = 011$, alors on utilisera la valeur 011**0000000** pour M.

Un nombre représenté de cette manière a pour valeur (attention les yeux) :

$$(-1)^S \times (1 + \frac{M}{2^m}) \times 2^{E-d}$$

où M est donné en décimal, E est l'**exposant décalé** et $d = 2^{e-1} - 1$ est le décalage de l'exposant.

1. Dans cette question, on code l'exposant sur $e = 5$ bits, et la mantisse sur $m = 10$ bits.

   (a) Combien faut-il d'octets pour représenter un nombre de cette manière ?

   (b) Quelle est la valeur du décalage d'exposant d dans ce cas ?

   (c) On considère le nombre suivant, codé avec la norme IEEE-754 : `0 10001 0010000000`

      i. Donner les valeurs décimales de M et E.

      ii. Déterminer la valeur décimale de ce nombre grâce à la formule ci-dessus (ou par la méthode de votre choix).

   (d) On souhaite coder le nombre décimal 6,75 (déjà représenté en binaire dans l'exercice 2).

      i. Déterminer la valeur binaire de la pseudo-mantisse M, sans enlever d'éventuels 0 « inutiles ».

      ii. Déterminer la valeur binaire de l'exposant E (sans oublier le décalage).

      iii. En déduire la représentation binaire de 6,75 avec la norme IEEE-754.
      Au besoin, rajouter des 0 « à droite » pour écrire M sur 10 bits.

   (e) Donner de même les représentations des nombres -11,5625 et 0,40625.

   (f) On admet que le nombre $(0{,}1)_{10}$ admet pour écriture binaire $0,000110011001100\ldots$, où la séquence 1100 se reproduit à l'infini.

      i. Donner la représentation de $(0{,}1)_{10}$ avec la norme IEEE-754.
      (pour la mantisse, on arrondira le résultat...)

      ii. Donner la valeur décimale de cette représentation. Que remarque-t-on ?

2. Les formats les plus couramment utilisés avec la norme IEEE-754 sont les suivants :

   Nombre de bits utilisés

Type	Signe	Exposant	Pseudo-mantisse	**Total**
Simple précision (*float*)	1	8	23	**32**
Double précision (*double*)	1	11	52	**64**

   Donner la représentation de $(0{,}1)_{10}$ en simple précision, puis la valeur décimale de cette représentation.