

**Objectif :** Comprendre la représentation des nombres réels en mémoire.

## 1 Problématique

Le 25 février 1991, à Dharan en Arabie Saoudite, un missile Patriot (figure 1) américain a raté l'interception d'un missile Scud irakien, ce dernier provoquant la mort de 28 personnes. La commission d'enquête a conclu à un défaut de l'horloge interne du missile. Cette dernière mesurait le temps en 1/10s.



FIGURE 1 – Missile Patriot

Pourquoi la représentation en mémoire du temps a engendré cette erreur ?

## 2 Représentation générale des nombres réels

### 2.1 Écriture scientifique

L'écriture scientifique des nombres réels répond à certaines règles :

- $1468 = +1,468 \times 10^3$
- $-891 = -8,91 \times 10^2$
- $0,00023 = 2,3 \times 10^{-4}$

La forme générale s'écrit :

$$\pm 1 \times \text{mantissee} \times 10^{\text{exposant}}$$

### 2.2 Représentation en mémoire

La représentation des nombres réels en mémoire s'appuie sur l'écriture scientifique mais :

- elle utilise la *base 2*,
- l'exposant est *biaisé* (décalé) d'une valeur  $d$  dépendante du format (32 ou 64 bits),
- la mantisse est comprise entre  $[1;2[$ .

La forme générale s'écrit :

$$(-1)^s \times m \times 2^{n-d}$$

## 3 La norme *IEEE 754*

### 3.1 Les choix effectués

C'est une norme mise au point par le *Institute of Electrical and Electronics Engineers*. Des choix techniques ont été pris :

- Cette représentation n'utilise pas le *complément à 2* pour stocker les exposants négatifs, mais un décalage d'une valeur  $d$ .
- La mantisse est un nombre de la forme 1,xxxxxx. Afin de gagner 1 bit en précision, on ne représente que les chiffres après la virgule.

### 3.2 Les formats

- *Simple précision* : Le nombre est représenté sur 32 bits.



L'exposant est représenté sur 8 bits donc des entiers entre 0 et 255. Il est décalé de  $d=127$  donc il est possible de représenter des exposants *signés* dans l'intervalle  $[-127;128]$ .

- *Double précision* : Le nombre est représenté sur 64 bits.



#### Activité 1 :

1. En s'appuyant sur le format 32 bits, donner la valeur du décalage  $d$  pour le format 64 bits.
2. En déduire les valeurs possibles pour l'exposant.

### 3.3 Un exemple

Considérons le mot de 32 bits :

$\overbrace{1}^{\text{signe}} \overbrace{100001110}^{\text{exposant}} \overbrace{101011011000000000000000}^{\text{mantisse}}$

- signe :  $(-1)^1 = -1$
- mantisse :  $1 + 2^{-1} + 2^{-3} + 2^{-5} + 2^{-6} + 2^{-8} + 2^{-9} = 1,677734375$
- exposant :  $(2^7 + 2^2 + 2^1) - 127 = 134 - 127 = 7$

Le nombre représenté est :

$$-1 \times 1,677734375 \times 2^7 = -214,75$$

### 3.4 Pour aller plus loin

La norme *IEEE 754* contient davantage de subtilités (représentation de 0, infini, dépassement de capacité, écart minimal...). Cette notion n'est pas au programme mais il peut être intéressant de lire la page Wikipédia correspondante :

[https://fr.wikipedia.org/wiki/IEEE\\_754](https://fr.wikipedia.org/wiki/IEEE_754)

## 4 Limites de la représentation

### 4.1 Convertir un nombre réel

**Activité 2 :**

1. En s'aidant de la page web <https://tinyurl.com/yyeymmln>, convertir 0,6875 en base 2.
2. Donner alors la représentation *en simple précision* de ce nombre.

**4.2 Erreur de calcul ?**

Le code ci-après renvoie un résultat surprenant.

```
1 >>> 0.1+0.2
```

Essayons d'expliquer ce résultat.

**Activité 3 :**

1. Convertir 0,2 en base 2.
2. Que peut-on en déduire sur la représentation de ce nombre en mémoire ?

**5 Imprécision du missile Patriot**

L'horloge interne du missile Patriot mesure le temps en  $1/10$ s soit 0,1s. Pour obtenir le temps en seconde, le système multipliait ce nombre par 10 en utilisant un registre de 24 bits en virgule fixe.

**Activité 4 :**

1. Convertir 0,1 en base 2. Que constate-t-on ?

Le registre de 24 bits contenait  $(0,00011001100110011001100)_2$  et induisait une erreur binaire de  $(0,0000000000000000000000011001100\dots)_2$ , soit approximativement 0,000000095s en notation décimale.

2. Le missile était allumé depuis 100 heures. Calculer le décalage *noté*  $\varepsilon$  entre l'horloge interne et le temps réel.
3. Un missile Scud volait à la vitesse de  $1676m.s^{-1}$ . Calculer la distance parcourue par le missile pendant la durée  $\varepsilon$ .