INF2 Principes des Systèmes Informatiques

le bit de mémoire

Olivier Ridoux

UNIVERSITÉ DE RENNES 1

Plan

- Le bit de mémoire
- Représentation binaire des données

INF2 Principes des Systèmes Informatiques

Le bit de mémoire

Le bit

- **Bit** ou ***binary digit***
  - étymologiquement : **chiffre binaire**
- Mémoire du choix d'**un** symbole parmi **deux**
  - 0 ou 1
  - ouvert ou fermé
  - Castor ou Pollux

A priori rien à voir avec un nombre !

N bits (1)

- **2 bits** = mémoire du choix parmi **2 × 2** (= 4) possibilités
  - – Athos, Porthos, Aramis ou d'Artagnan
- **3 bits** = mémoire du choix parmi **2 × 2 × 2** (= 8) possibilités
  - – Mercure, Venus, Terre, Mars, Jupiter, Saturne, Uranus ou Neptune

N bits (2)

- **n bits** = mémoire du choix parmi $\mathbf{2^n}$ possibilités
  - si N possibilités alors $n = \lceil \log_2 N \rceil$

M+P bits

- Mémoire du choix parmi $\mathbf{2^{M+P} = 2^M \times 2^P}$ possibilités

- Mémoire combinée du choix parmi $\mathbf{2^M}$ possibilités et du choix parmi $\mathbf{2^P}$

- Toutes les découpes sont possibles

Unités dérivés (1)

- **L'octet** = 8 bits

  $\longrightarrow 2^8 = 256$ possibilités

  – possible unité d'adressage de la mémoire

Unités dérivés (2)

- Le **mot** = 16, 32 ou 64 bits

  $\longrightarrow 2^{16} \approx$ 64 mille possibilités

  $\longrightarrow 2^{32} \approx$ 4 milliards de possibilités

  $\longrightarrow 2^{64} \approx$ 16 milliards de milliards de possibilités

  – possible unité d'adressage de la mémoire
  – IPv4 : adresse IP sur 32 bits
  – IPv6 : adresse IP sur 64 bits

Unités dérivés (3)

- Préfixes multiplicatifs
  - **kilo**
  - **méga**
  - **giga**
  - **téra**
  - **péta**

- Exemples : kilo-octets, gigabits

Unités dérivés (4)

- Unités complexes
  - bit par seconde
    - vitesse de transmission
  - bit par m^2
    - densité de stockage
  - € par bit
    - coût du stockage instantané
  - € par bit par s
    - coût du stockage entretenu

Deux normes (1)

- **Système international** (SI)
  - kilo, méga, giga, téra, péta, ...
  - 10^3, 10^6, 10^9, 10^{12}, 10^{15}, ..., $10^{3 \times n}$
  - **la norme scientifique !**
- **Système dichotomique**
  - kilo, méga, giga, téra, péta, ...
  - 2^{10}, 2^{20}, 2^{30}, 2^{40}, 2^{50}, ..., $2^{10 \times n}$
  - **la norme de fait !**

Deux normes (2)

- Écart faible, mais croissant

$$[\,(2^{10})^n - (10^3)^n\,] \;/\; (10^3)^n$$
$$= (2^{10} / 10^3)^n - 1$$

- Expression SI > expression dichotomique
  - exemple : clé USB

$$16 \text{ giga}_{SI} = 14{,}9 \text{ giga}_{DICHO}$$

Deux normes (3)

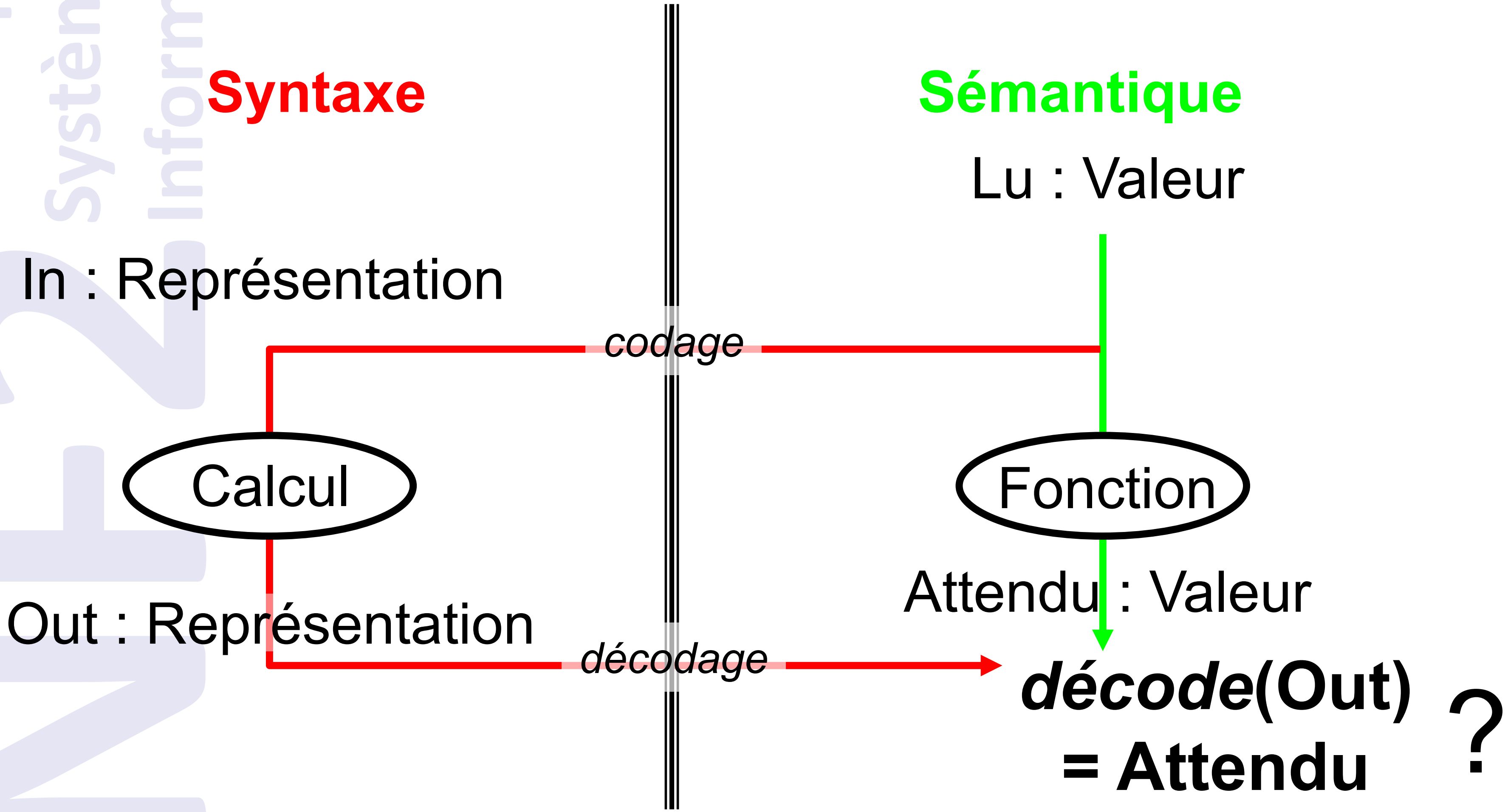
unité	dichotomique	SI	écart (%)
kilo	1024	1000	2,4
méga	1048576	1000000	4,85
giga	1073741824	1000000000	7,37
téra	1099511627776	1000000000000	9,95

Deux normes (4)

- Une nouvelle norme pour le système dichotomique
  - kibi, mébi, gibi, tébi, ...
  - Ki, Mi, Gi, Ti, ...
  - 2^{10}, 2^{20}, 2^{30}, 2^{40}, ..., $2^{10 \times n}$
  - **CEI : Commission électrotechnique internationale**

Représentation binaire des données

Syntaxe et sémantique (1)



Syntaxe et sémantique (2)

- Une distinction fondamentale
- Inspirée de la linguistique
  - **le mot long est court**

    un mot ne peut pas être long et court à la fois
  - **le mot « long » est court**

    le mot « long » est effectivement court

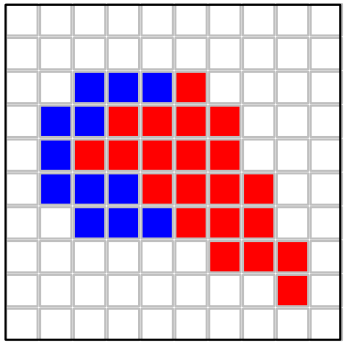
    les guillemets stoppent l'interprétation

Données complexes

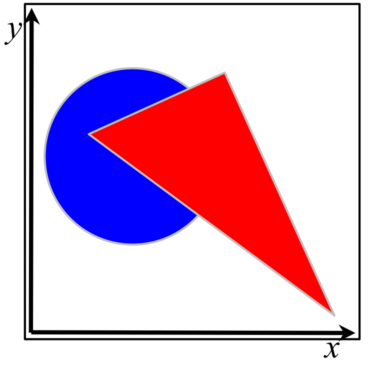
- Image
- Son

Images

- Représentation matricielle



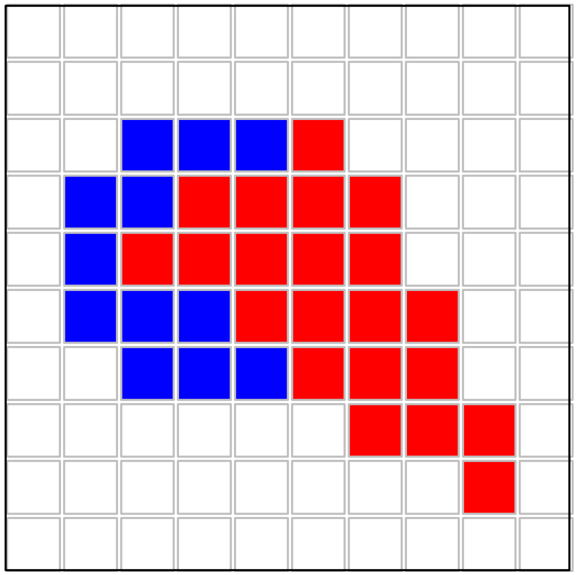
- Représentation vectorielle



Représentation matricielle (1)

- ***Bitmap***, ***raster***
- Une image est représentée par une grille de **pixels** (***picture elements***)
- Chaque pixel est décrit par des propriétés graphiques
  - noir ou blanc
  - niveau de gris
  - couleur
  - luminosité

  …



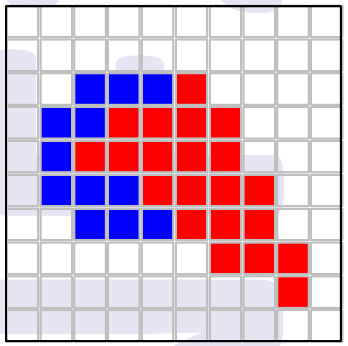
Représentation matricielle (2)

- **Définition** d'une image
  - nombre de pixels qui la représentent

    nb. lignes × nb. colonnes
  - question de représentation
- **Résolution** d'une image
  - nombre de pixels par unité de longueur

    ppp = point par pouce
  - question de rendu
  - à définition constante

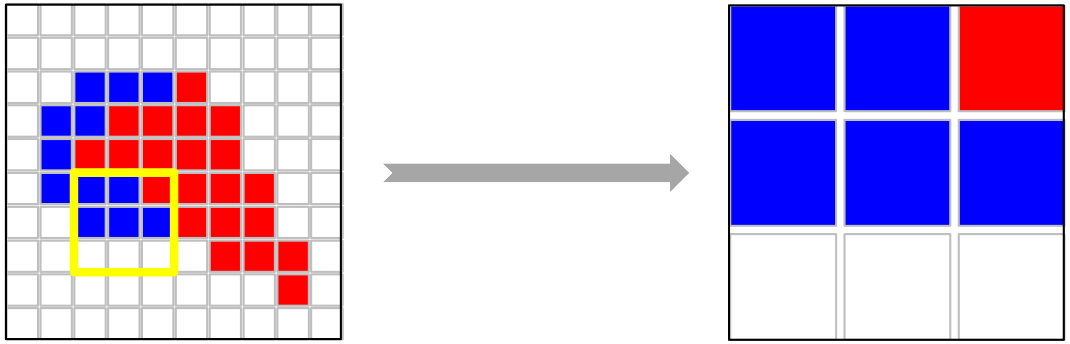
    résolution ↗ ⟷taille image ↘



d = 10 × 10
r = 5 ppp

Représentation matricielle (3)

- **Zoom** dans une image
  - moins de pixels par unité de longueur
  - **résolution ↘**



- Détermination des composants d'une image

  contours, objets, parties cachées
  - facile pour le spectateur
  - difficile ou impossible pour un programme

Représentation matricielle (4)

- Images photographiques
  - image satellite
  - image médicale

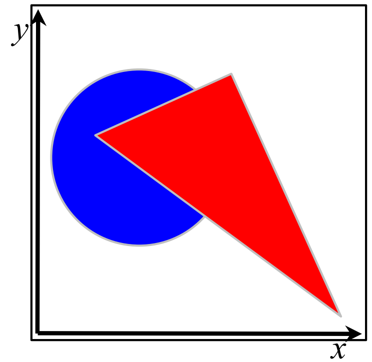
  …

- Outil de dessin matriciel

Représentation vectorielle (1)

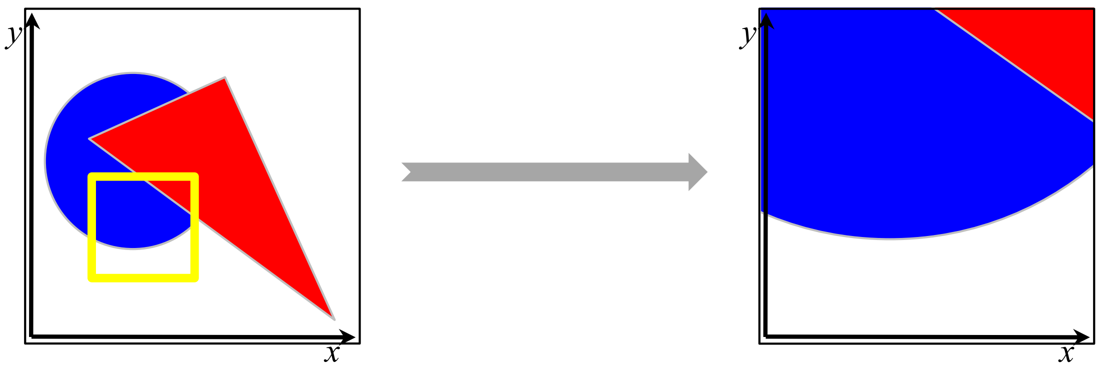
- Une image est représentée par des objets géométriques
- Chaque objet a des coordonnées et des dimensions
- Chaque objet a des propriétés graphiques
  - couleur
  - texture
  - niveau de gris

  …



Représentation vectorielle (2)

- Zoom facile sans perte de qualité



- Détermination des composants d'une image faite a priori

Représentation vectorielle (3)

- Outils de dessin vectoriel
- Cartographie numérique
- Images de synthèse

…

Son

- Son analogique ⟶ son numérique
- Son compressé avec perte
- Son symbolique

Son analog. ⟶ son numér. (1)

- Phénomène 2 fois continu
  - **énergie sonore** et **temps**
- **Double discrétisation**
  - de l'énergie sonore et du temps

Son analog. ⟶ son numér. (2)

- Le son est **échantillonné** tous les δt

  1/δt = **fréquence d'échantillonnage**

- **Théorème d'échantillonnage (Shannon-Nyquist)**
  - la fréquence d'échantillonnage doit être au moins le **double** de la plus haute fréquence du signal à échantillonner

    8 kHz pour le téléphone (voix 300-3400 Hz)

    44,1 kHz pour cédérom (oreille 20-20000 Hz)

  - sinon apparition de **fréquences fantômes**

Son analog. ⟶ son numér. (3)

- L'énergie sonore est **quantifiée** en 2^s **niveaux discrets**
- Chaque échantillon représente une énergie e par un niveau approchant
  - **bruit de quantification**

Son analog. ⟶ son numér. (4)

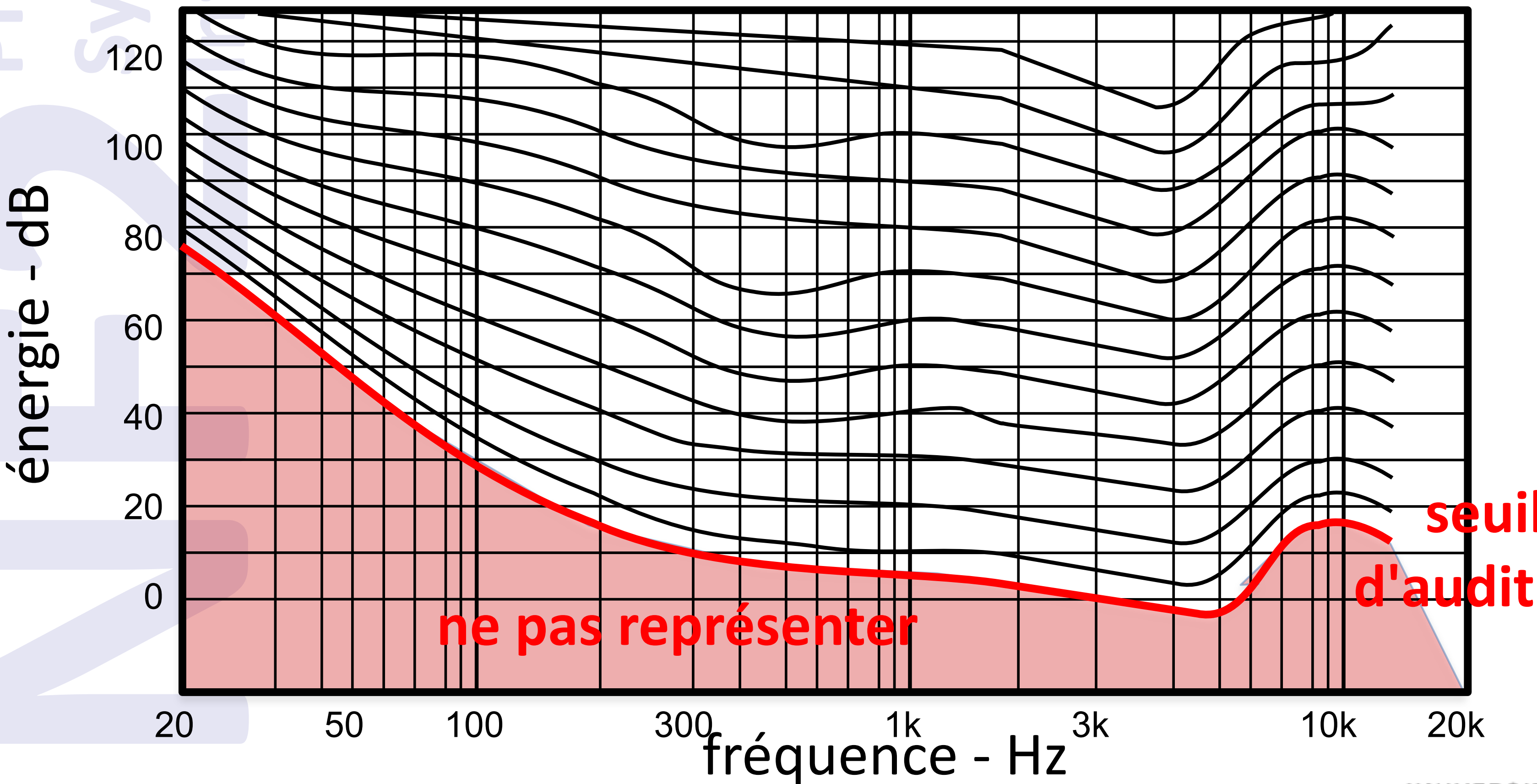
- Format **WAV**[eform audio format]
- À réserver au traitement du son

Son compressé avec perte (1)

- Tous les sons ne sont pas également audibles
- Ne pas représenter les sons quasiment inaudibles

Son compressé avec perte (2)

- Diagramme de Fletcher
  - intensités perçues égales en fonction de la fréquence



Son compressé avec perte (2)

- Format **WMA**, **MP3**
- Pour écoute seulement
  - impossible de traiter le son

Son symbolique (1)

- Format **MIDI**

  musical instrument digital interface

- Codage symbolique...
  - nom de note (hauteur), vélocité (intensité), contrôle du son

...très bas niveau

- attaque, maintien, et chute explicites
- δt entre chaque événement

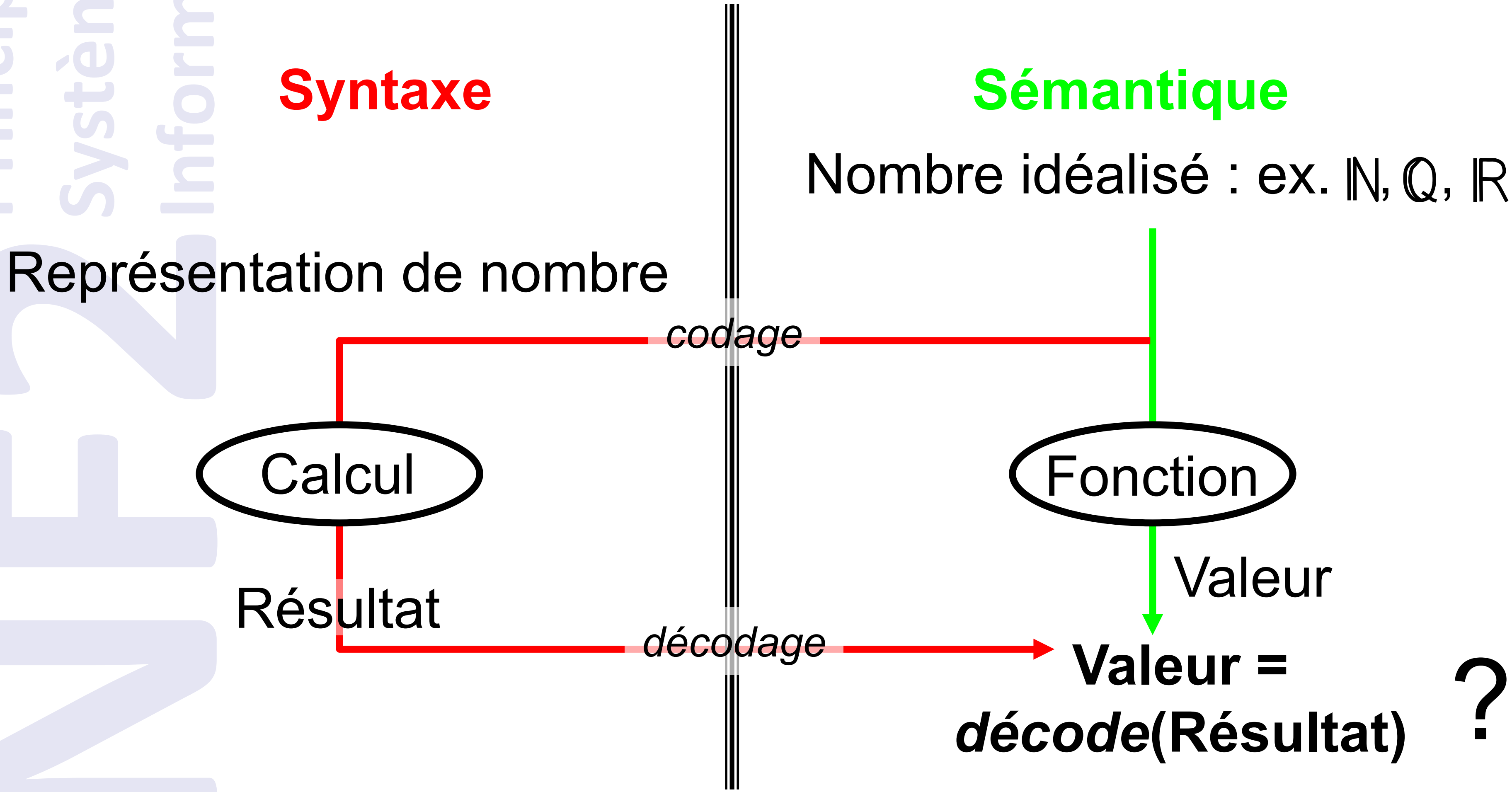
Son symbolique (2)

- Interface avec instruments de musique électronique

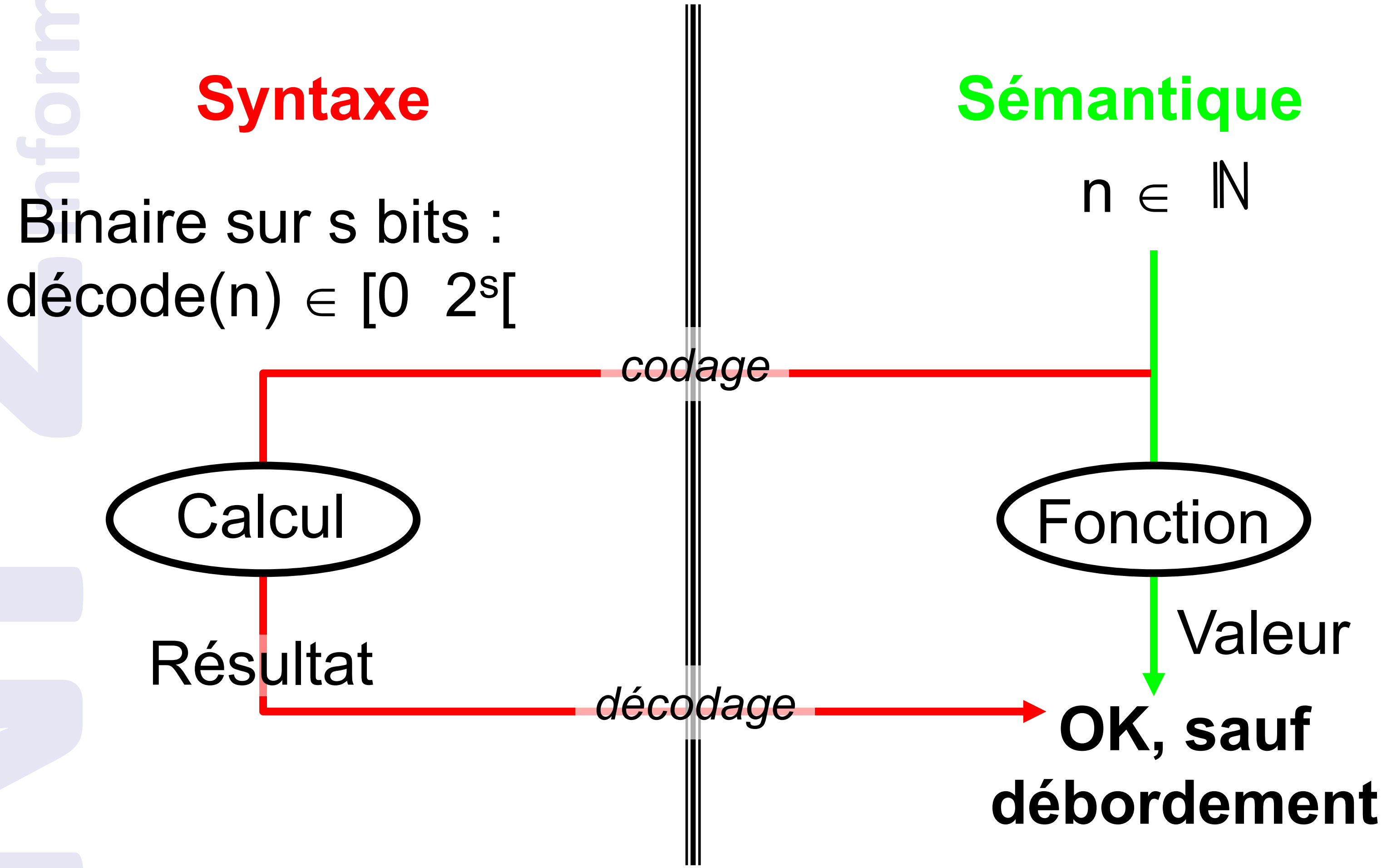
Données élémentaires

- Symboles
- Caractères
- Nombres

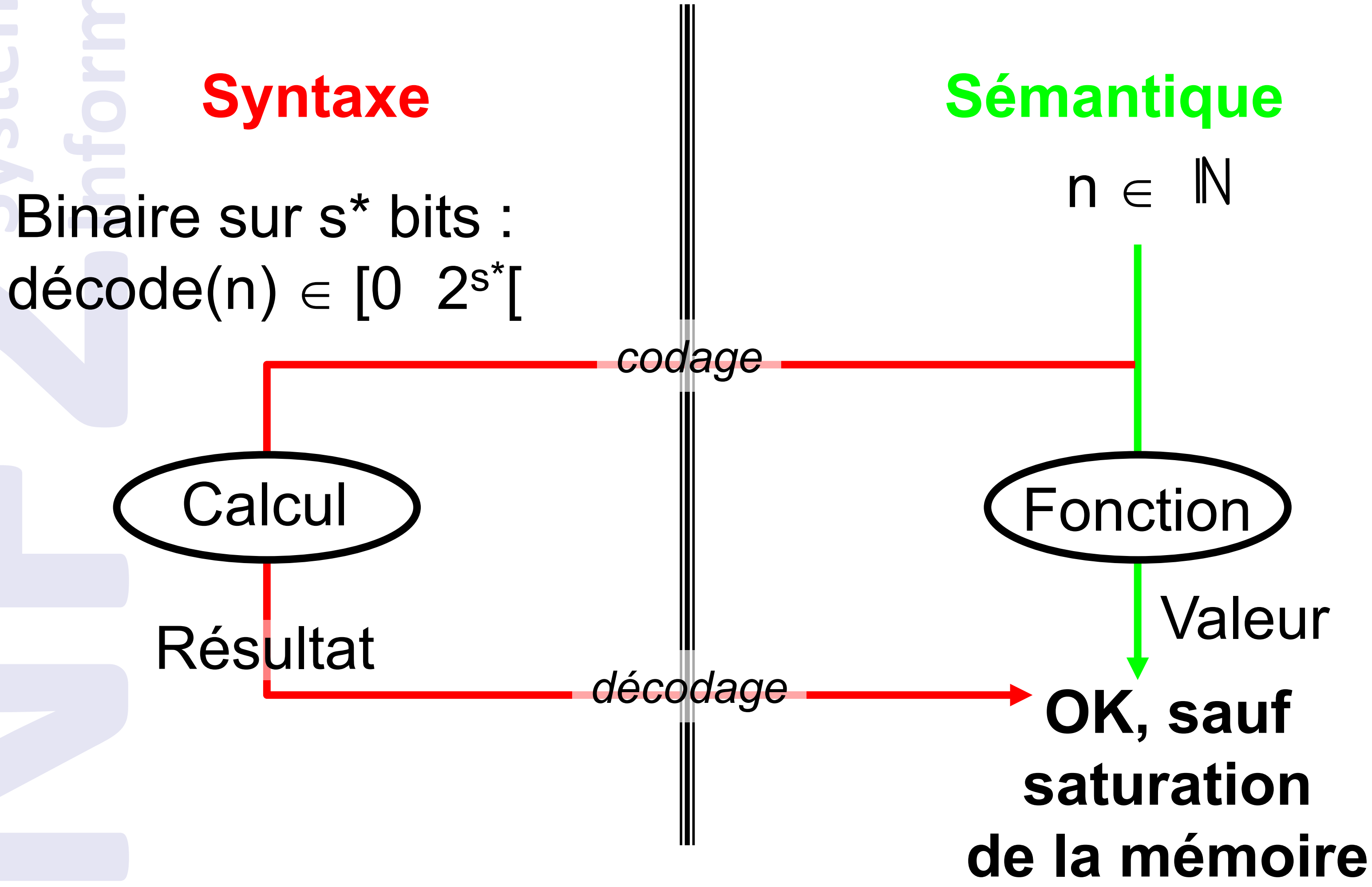
Syntaxe et sémantique des nombres



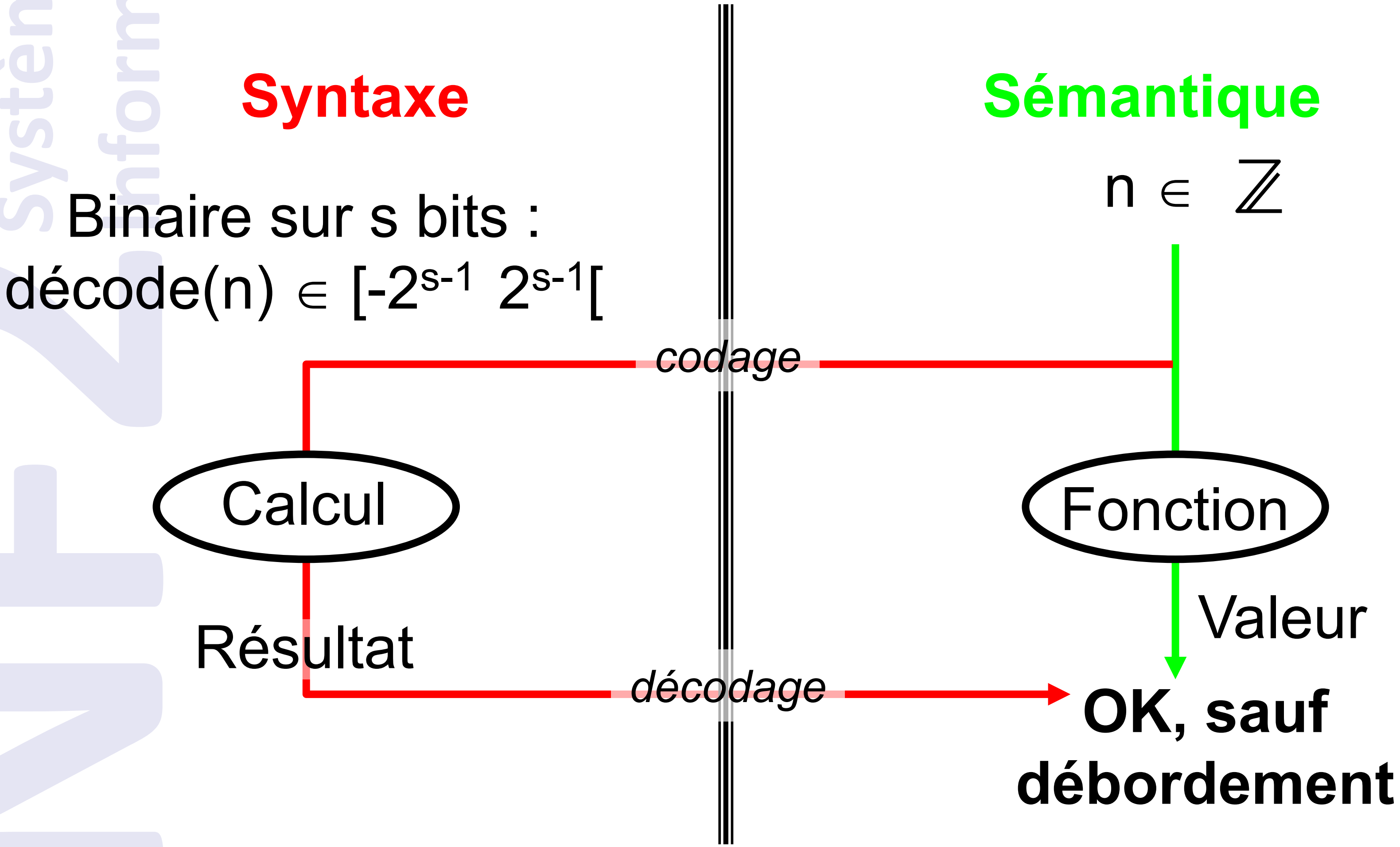
Syntaxe et sémantique des entiers : int, long, etc



Syntaxe et sémantique des grands entiers : bigint, bignum, etc



Syntaxe et sémantique des entiers signés



Choix d'une représentation

- **Économie** de la représentation

  nb. valeurs codées / $2^{\text{nb. bits}}$

- **Unicité** de la représentation
  - comparaisons difficiles si représentations multiples

- **Faisabilité** des opérations
  - numération romaine et...

  ...numération positionnelle

Représentation en base 2 (1)

exemple
représentation de 71
sur 8 bits

- **$code_N(n) = <b_0,b_1,b_2,\ldots,b_{s-1}>$**
  b_i = parité de n div 2^i
- **$décode_N(<b_0,b_1,b_2,\ldots,b_{s-1}>)$**
  $= \sum_{i \in [0\ s-1]} (b_i \times 2^i)$

i	2^i	71 div 2^i	b_i	$b_i \times 2^i$
0	1	71	1	1
1	2	35	1	2
2	4	17	1	4
3	8	8	0	0
4	16	4	0	0
5	32	2	0	0
6	64	1	1	64
7	128	0	0	0

Représentation en base 2 (2)

- s bits codent 2^s nombres différents
  - efficacité maximale
- Les opérations se font comme en base 10
- Les tables sont plus simples qu'en base 10

+	0	1
0	0	1
1	1	10

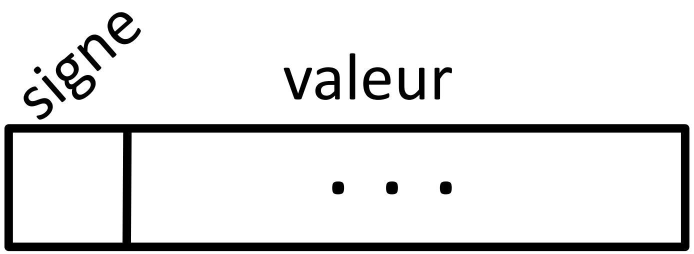
×	0	1
0	0	0
1	0	1

Représentation en base 2 (3)

0	**000**
1	**001**
2	**010**
3	**011**
4	**100**
5	**101**
6	**110**
7	**111**

Représentations des signes (1)

- **Signe + valeur**



1 bit de signe

s-1 bits de valeurs en base 2

- $\mathbf{code_{SV}(z) = <signe(z),code_N(|z|)>}$
- $\mathbf{décode_{SV}(<s,v>) = (-1)^s \times décode_N(v)}$

- calculs malcommodes...

...2 représentations de 0

Représentations des signes (1,5)

-3	111
-2	110
-1	101
-0	100
+0	000
+1	001
+2	010
+3	011

par représenté

+0	000
+1	001
+2	010
+3	011
-0	100
-1	101
-2	110
-3	111

par représentant

$code_{SV}(-2)+code_{SV}(+1)$

$= 110 + 001$

$= 111$

$\mathbf{= code_{SV}(-3)}$

Représentations des signes (2)

- **Valeur + biais**

$$\mathbf{code_{Zb}(n) = code_N(n+2^{s-1})}$$

$$\mathbf{décode_{Zb}(z) = décode_N(z) - 2^{s-1}}$$

  - représentation unique
  - calculs malcommodes

$$\mathbf{code_{Zb}(a)+code_{Zb}(b) = code_Z(a+b)+2^{s-1}}$$

Représentations des signes (2,5)

-4	000
-3	001
-2	010
-1	011
+0	100
+1	101
+2	110
+3	111

$\text{code}_{Zb}(-2)+\text{code}_{Zb}(+1)$

$= 010 + 101$

$= 111$

$\mathbf{= \text{code}_{Zb}(3)}$

UNIVERSITÉ DE RENNES 1

Représentations des signes (3)

- **Compléments à deux**

  $$\mathbf{code_{Z2}(n) = code_N(\ (n+2^s) \bmod 2^s\ )}$$

  – représentation unique

  – calculs commodes

  $$\mathbf{code_{Z2}(a) + code_{Z2}(b) = code_{Z2}(a+b)}$$

Représentations des signes (3,5)

-4	100
-3	101
-2	110
-1	111
+0	000
+1	001
+2	010
+3	011

par représenté

+0	000
+1	001
+2	010
+3	011
-4	100
-3	101
-2	110
-1	111

par représentant

$\text{code}_{Z2}(-2)+\text{code}_{Z2}(+1)$

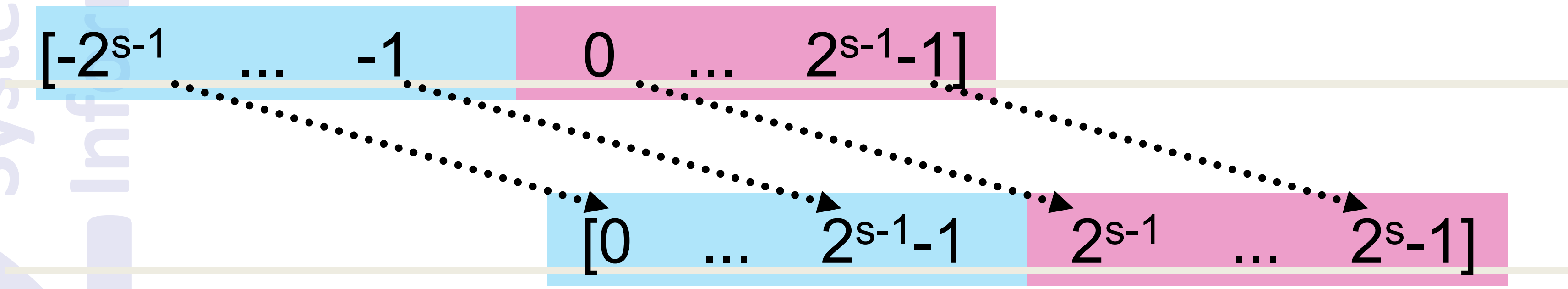
$= 110 + 001$

$= 111$

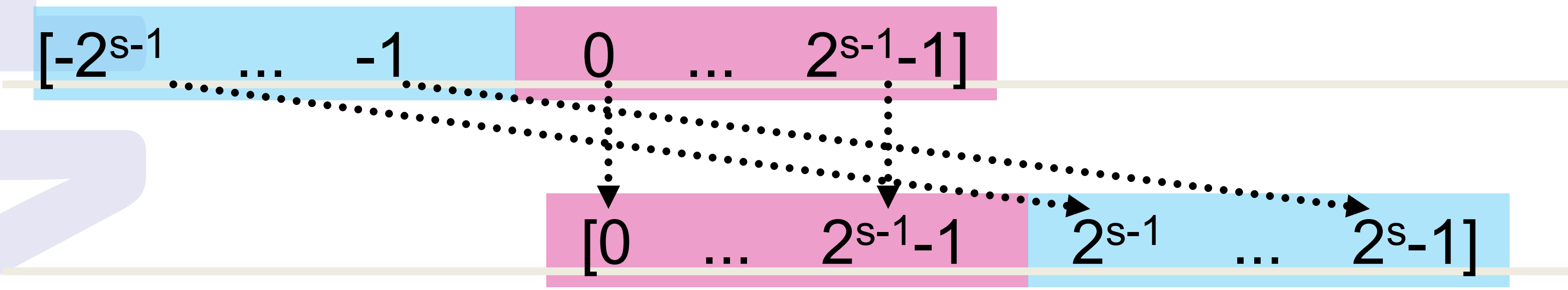
$\mathbf{= \text{code}_{SV}(-1)}$

Illustration - biais vs 2-complément

- Biais : + 2^{s-1}



- Complément à 2



Représentation des rationnels

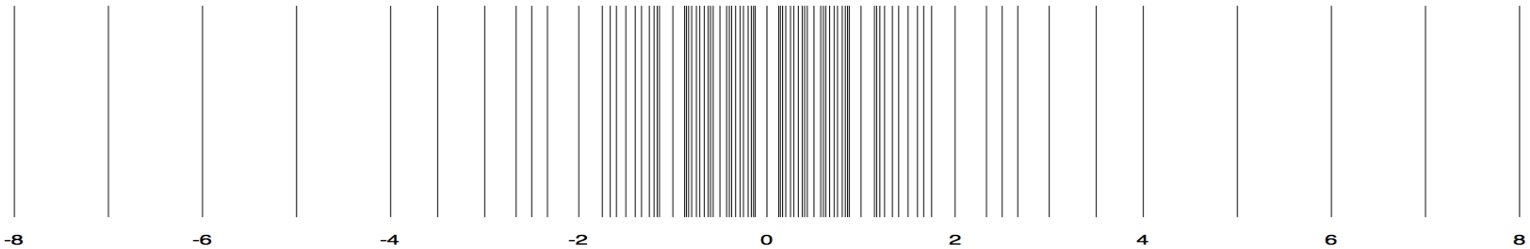
- Infinité de rationnels dans un intervalle fini
- Représentation décimale infinie
- Trois représentations finies
  - – numérateur / dénominateur
  - – virgule fixe
  - – virgule flottante

Numérateur / dénominateur (1)

$$\mathrm{code}_{Qnd}(q) = \langle \mathrm{code}_N((n)um), \mathrm{code}_N((d)én) \rangle$$

$$n/d = q$$

- Distribution irrégulière des rationnels représentables



4+4 bits → Représentables $\subsetneq$ [-8 +8[

17/19 $\in$ [-8 +8[, mais 17/19 $\notin$ Représentables

Numérateur / dénominateur (2)

- Représentations multiples

  1/2 = 2/4 = 3/6 = ...

  87 rationnels différents représentables avec 8 bits

- Calculs malcommodes

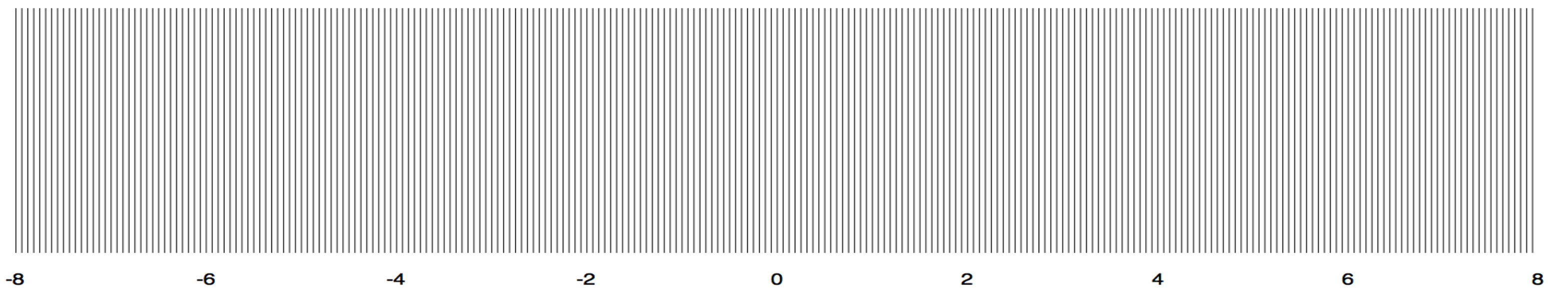
- Presque jamais utilisée
  - sauf rationnels en précision arbitraire sur *+* bits

Virgule fixe (1)

- Biais multiplicatif fixe

$$\mathbf{code_{Qvfx}(q) = code_N(arrondi(q \times D))}$$

- Distribution régulière des rationnels représentables



8 bits et D = 2^4

$\rightarrow$ Représentables $\subset$ [-8 +8[

Virgule fixe (2)

- Calculs commodes, représentations uniques
- **Mais, ordres de grandeur limités**
- Usages limités
  - matériels bas de gamme : **embarqué**
  - applications spécifiques : **comptabilité**

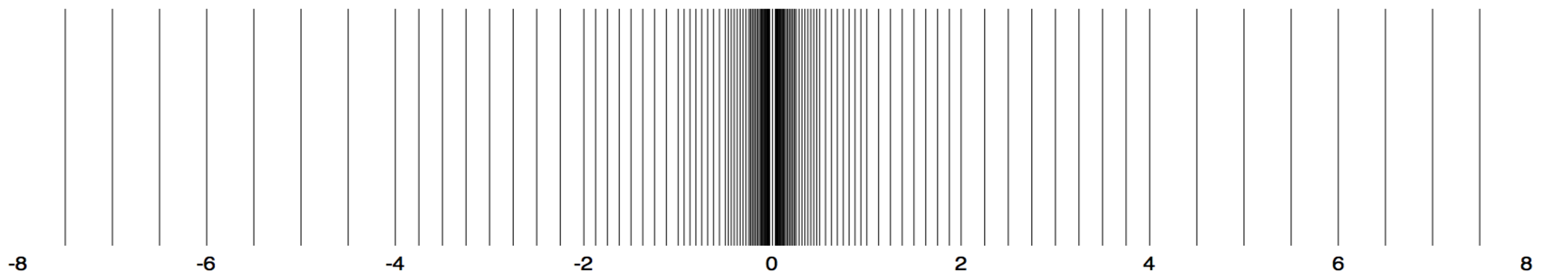
Virgule flottante (1)

- Biais multiplicatif variable

  $\textbf{code}_{Qvfl}(q) = <$ **(s)igne, (m)antisse, (e)xposant** $>$

  $s \times m \times base^{e} \approx q$, **avec** $m \in [1/base \;\; 1[$

- **Distribution irrégulière**
  - très peu dense pour grandes valeurs absolues
  - se densifie vers petites valeurs absolues, mais vide autour de 0



Virgule flottante (2)

- Calculs commodes, mais précision délicate à gérer
- Représentation unique, sauf zéro
- **Beaucoup de configurations illégales**

  1+4+3 bits → Représentables ∈ [-7,5 7,5]

  mais seulement 129 rationnels différents
- Normalisée : IEEE 754
- Très utilisée, mais...

Virgule flottante

- Additions
  - soit F un « gros » flottant, f un « petit » flottant

$$\mathbf{F+f \rightarrow F}$$

même si $\mathbf{décode_{Qvfl}(F) + décode_{Qvfl}(f) \neq décode_{Qvfl}(F)}$

- Soustractions
  - soit f_1 et f_2 deux « petits » flottants

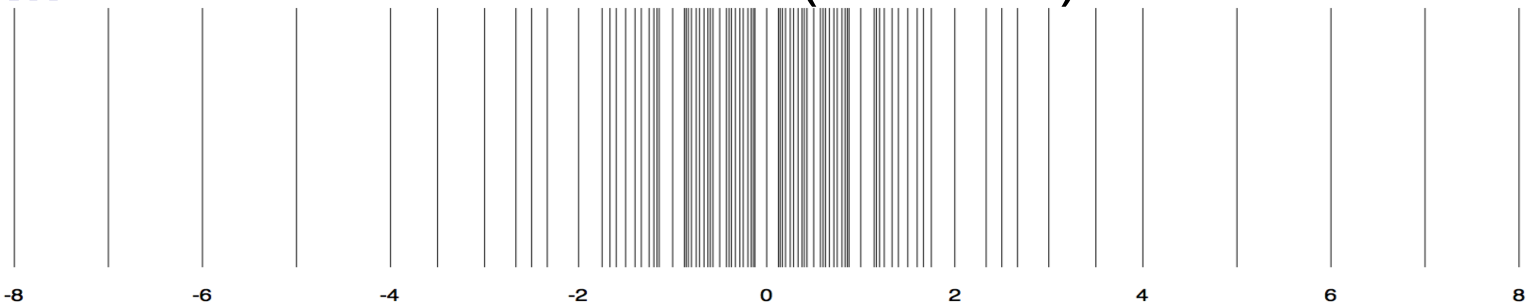
$$\mathbf{f_1\text{-}f_2 \rightarrow 0}$$

même si $\mathbf{décode_{Qvfl}(f_1) - décode_{Qvfl}(f_2) \neq décode_{Qvfl}(0)}$

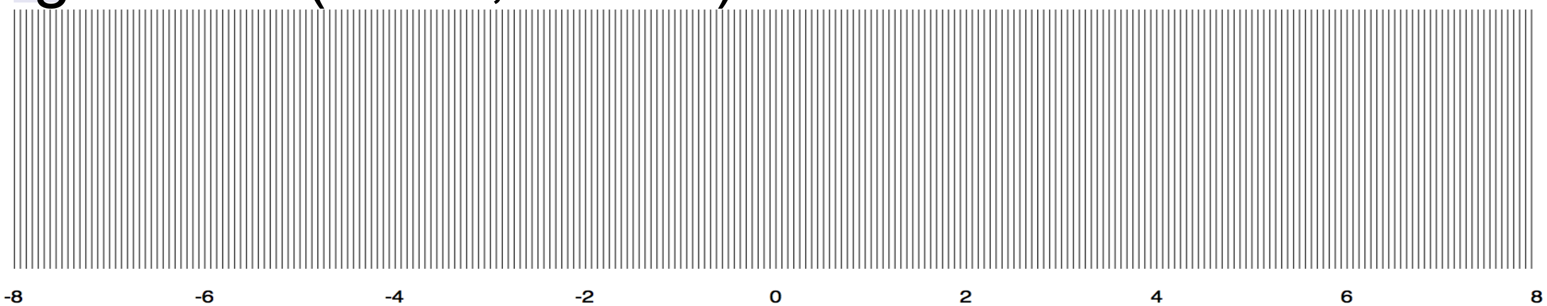
- Qualité numérique des algorithmes !

Distributions des représentations

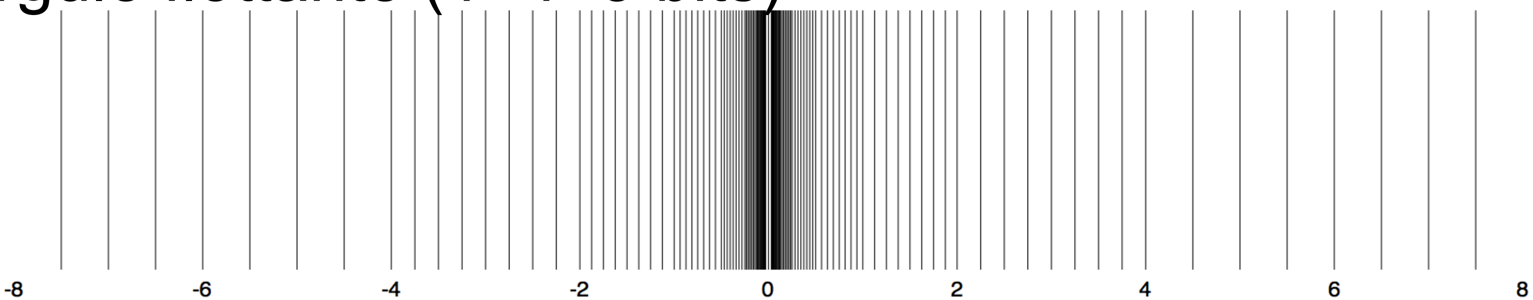
- Numérateur / dénominateur (4+4 bits)



- Virgule fixe (8 bits, D = 2^4)



- Virgule flottante (1+4+3 bits)



Conclusion (1)

- n bits $\leftrightarrow$ 2^n **symboles**
- Deux systèmes d'unités
  - SI : $(10^3)^k$, optimiste
  - dichotomique : $(2^{10})^k$, standard de fait
  - écart non négligeable

Conclusion (2)

- Relation syntaxe-sémantique
- Des approximations
- Des limitations
- Des risques d'erreur
- **Des compromis !**