Métamorphoses et promesses du codage correcteur d'erreurs

Claude Berrou

30 avril 2008

L’information est vieille comme le monde ... et les erreurs aussi

Ce sont d’ailleurs les erreurs qui conduisent l’évolution de la vie

Mais l’homme n’aime pas les aléas et a inventé le principe de précaution

A la recherche du zéro défaut…

La théorie de l'information (1948) traite de :

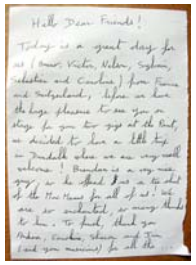
- l'écriture

- la sécurisation

- la protection

- le transport

- la restitution



de messages, indépendamment de leurs significations.

amour = haine = 35 bits en ASCII

La théorie de l'information (1948) fournit : les concepts

entropie	*entropie* (jourbon) > *entropie* (bonjour)
redondance	bonjour et bonne journée, blablabla
bruit	bonjoir
interférence	bonsoir bonjour
corrélation	bon???
distance	matin ≈ latin, amour ≠ haine
diversité	bonjour, good morning

Le codage correcteur : le principe

On veut transmettre "matin" dans une ambiance très bruyante

- Simple répétition :

"matin matin" est reçu "matin latin"

Aucune correction sûre possible

- Utilisation d'un synonyme :

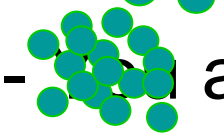
"matin aube" est reçu "matin auge"

Correction possible en se référant à un dictionnaire d'équivalence

Quid de "matin potron-minet" ?

Certaines langues sont mieux pensées que d’autres vis-à-vis de possibles erreurs :

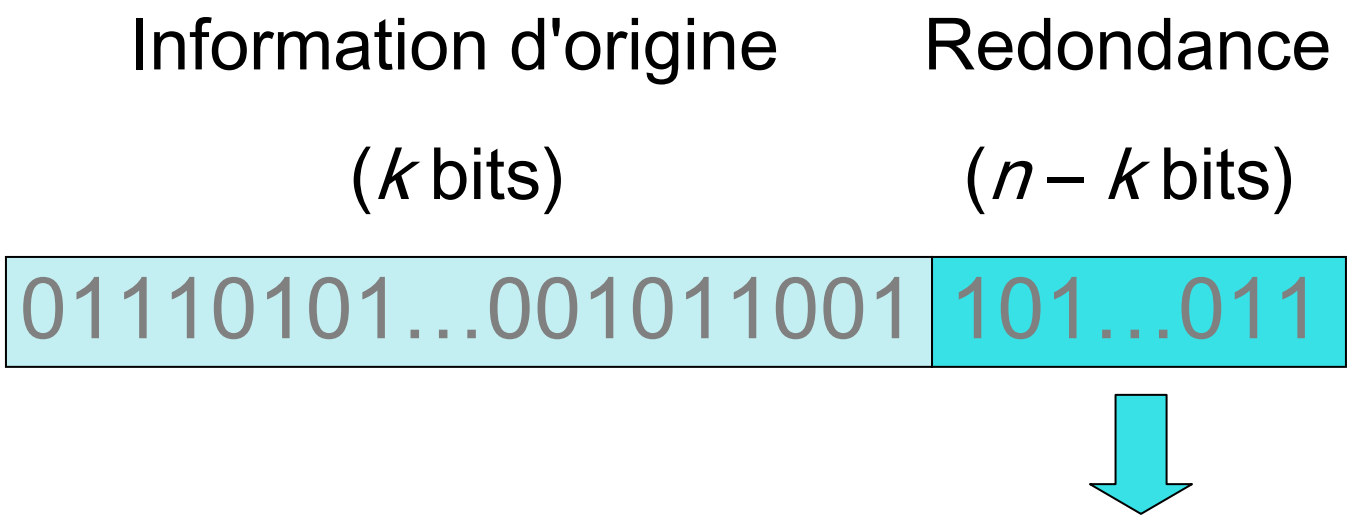
- Vous êtes d’accord



- You agree

Mais le texte est généralement plus long

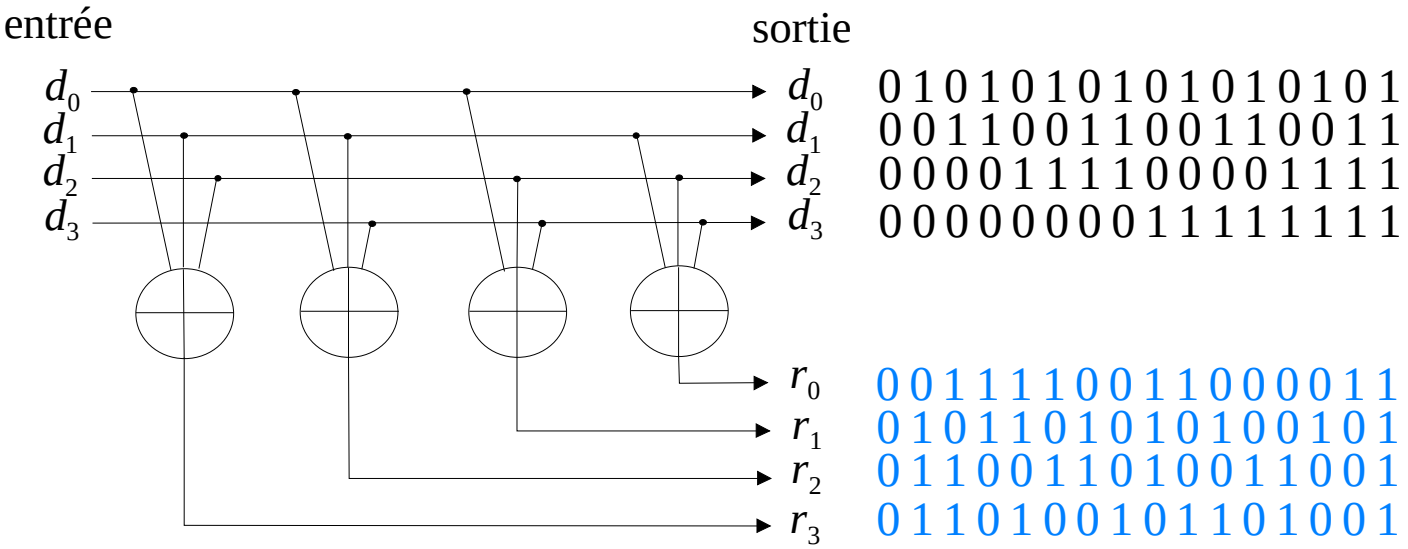
Le codage correcteur : de la redondance bien pensée



La redondance est construite selon une loi mathématique offrant une grande diversité :

Hamming, Golay, BCH, Reed-Solomon, convolutif, code produit, turbocode, LDPC…

Le premier code correcteur de l'histoire : le code de Hamming (étendu)

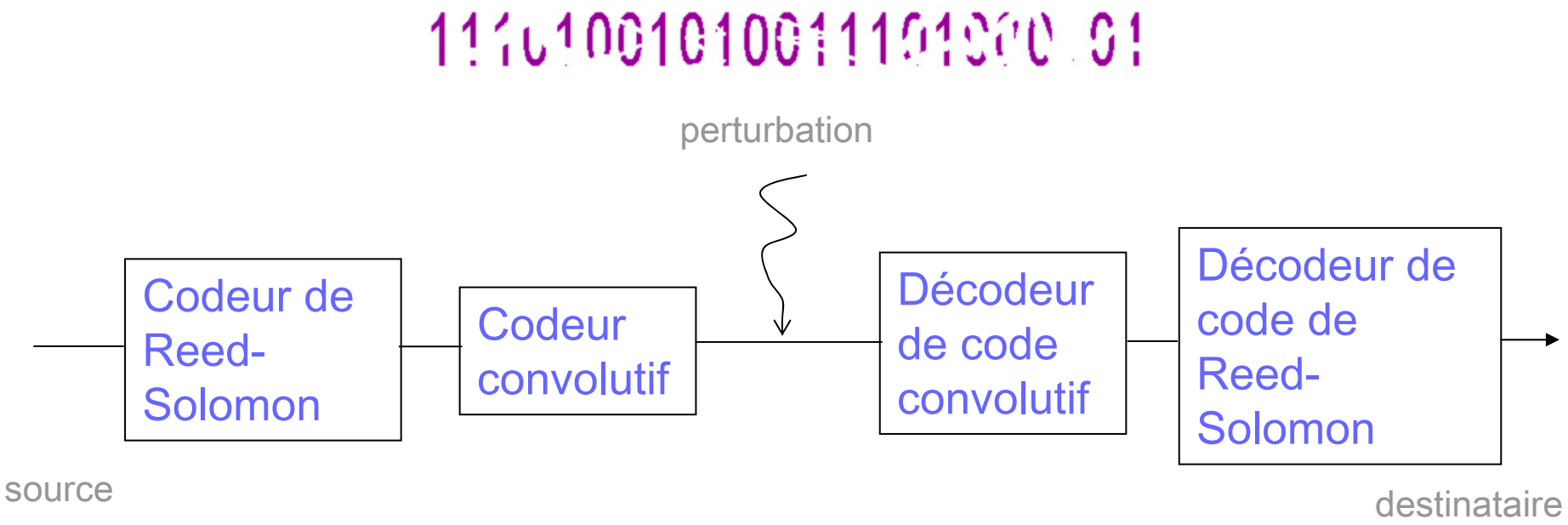


$$r_j = d_j + \sum_{p=0}^{3} d_p \quad \text{modulo } 2 \quad \text{pour } j = 0,...,3$$

La distance minimale est 4

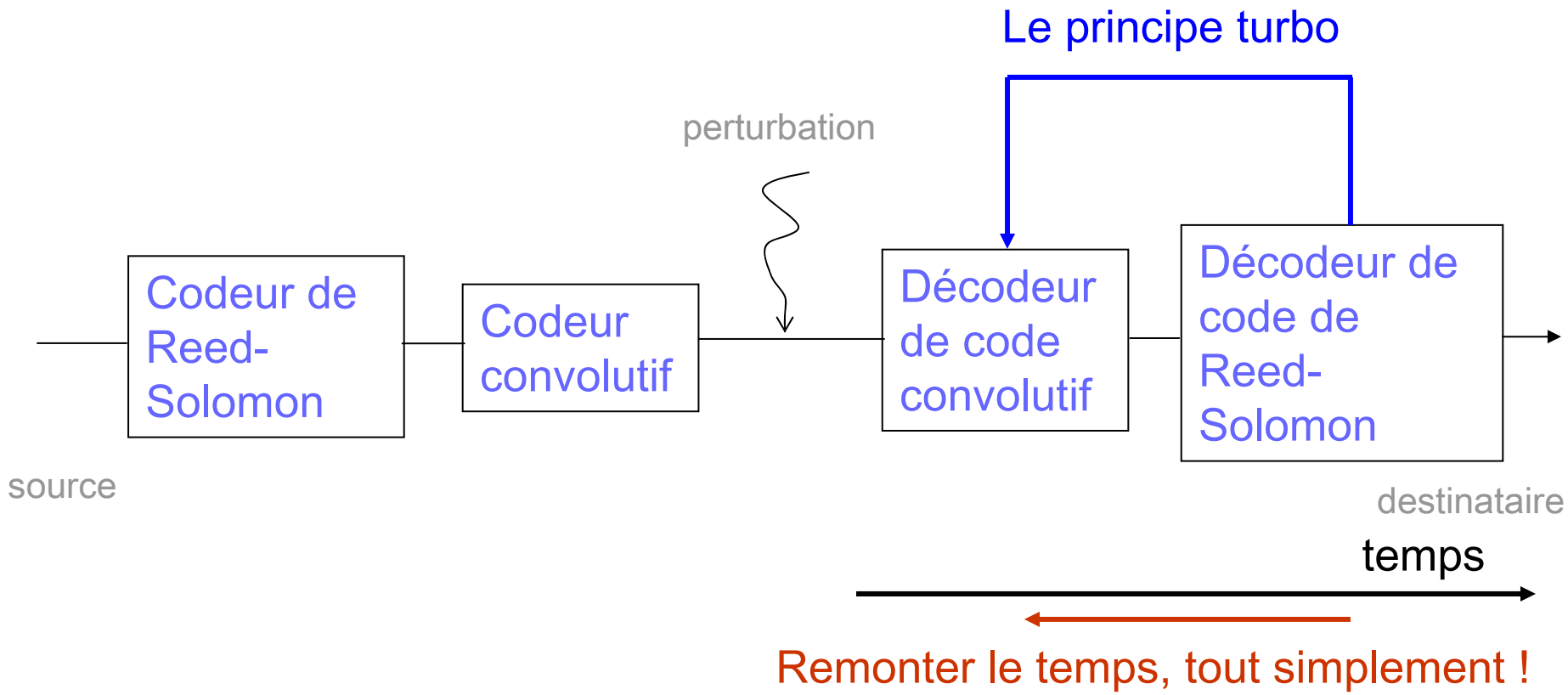
(On peut corriger une erreur)

Le codage correcteur : l'état de l'art en 1990 (le code de la TNT)



Concaténation de codes

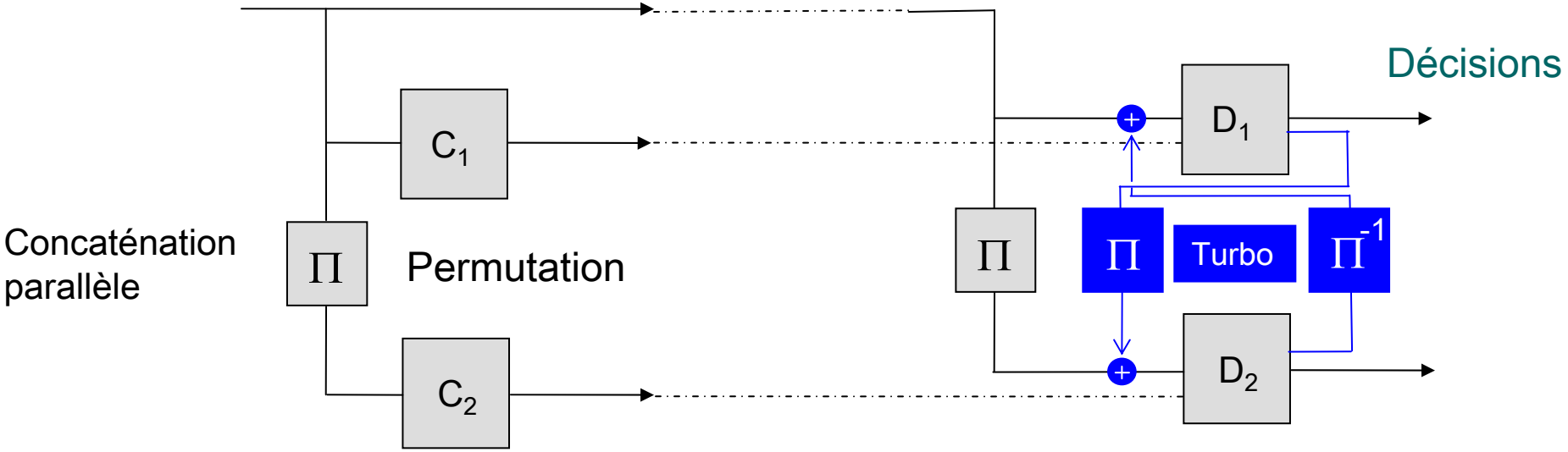
Le codage correcteur : l'état de l'art en 1990 (le code de la TNT)



Et la physique est venue se mêler de codage …

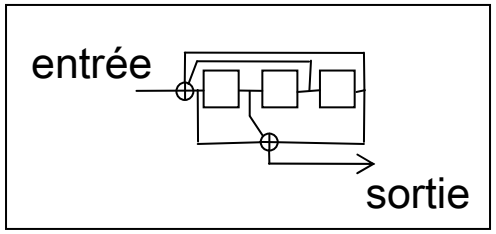
Les turbocodes

Données
binaires : …1101000110…



Traitement
probabiliste
itératif

Code élémentaire C

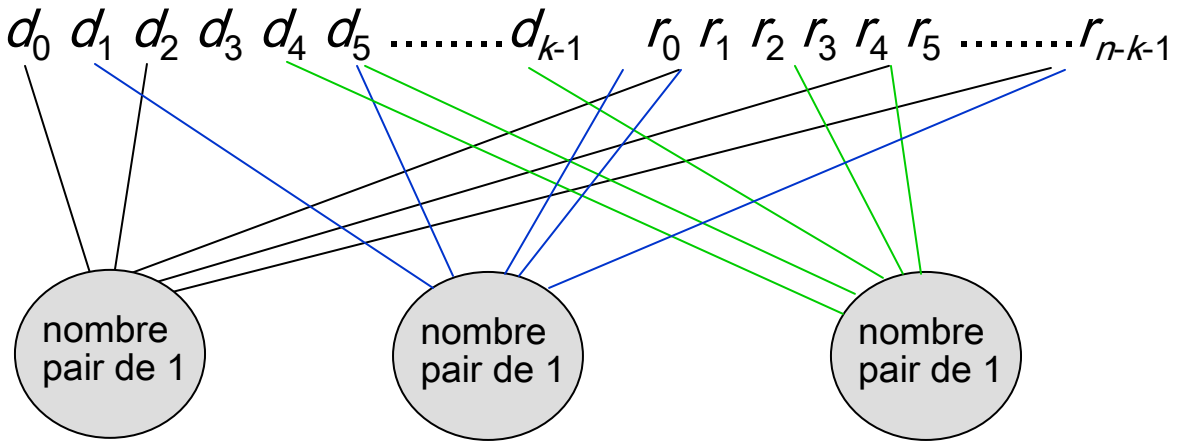


1997 : le retour de l'Amérique

avec les codes LDPC (*low density parity check*)

Bob Gallager
(1931-)

Le principe des codes LDPC



n-k contraintes

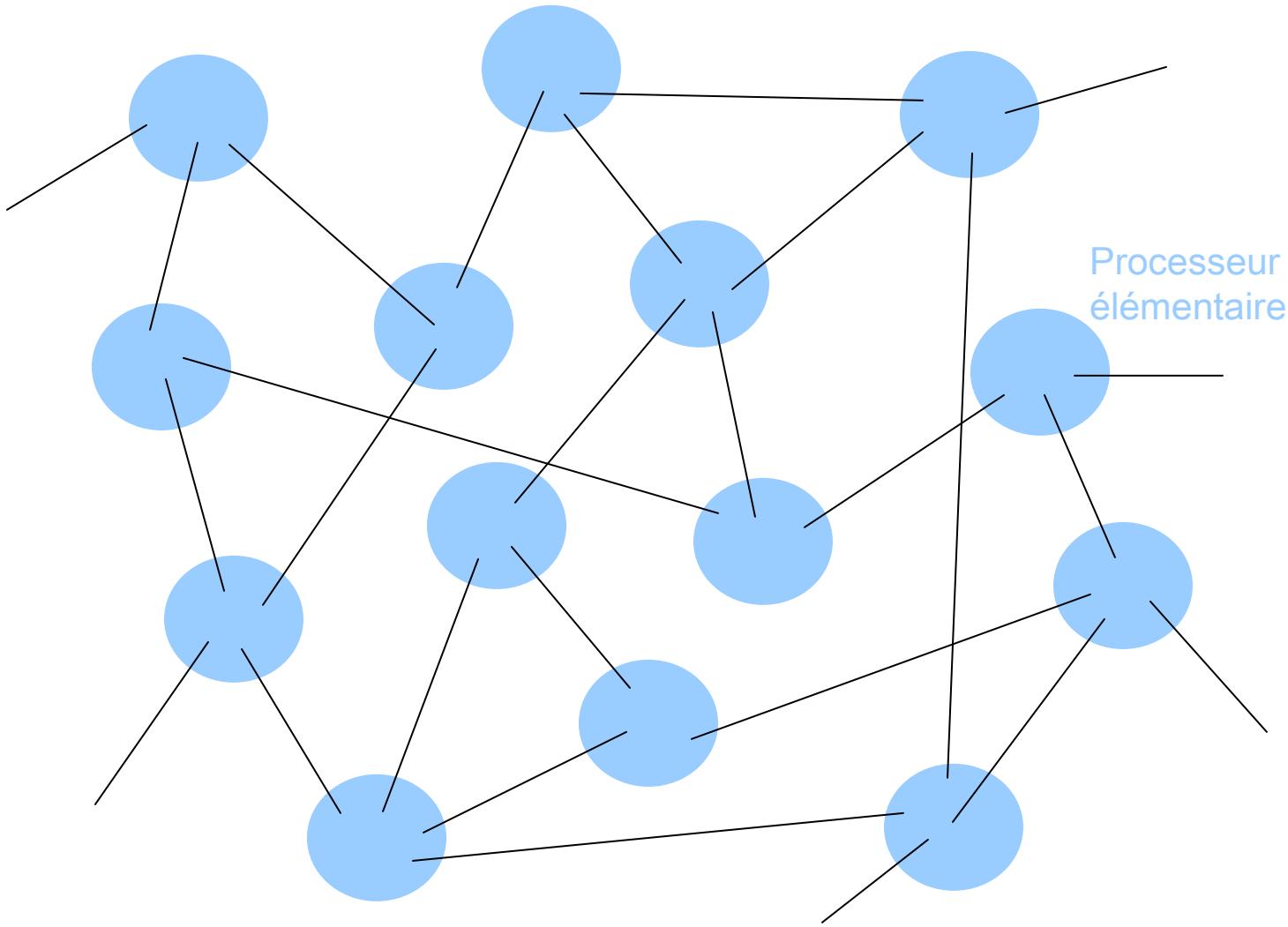
$$d_0 + d_2 + r_1 + r_5 + r_{n-k-1} = 0 \quad \text{modulo } 2$$
$$d_1 + d_5 + r_0 + r_1 + r_{n-k-1} = 0 \quad \text{modulo } 2$$
$$d_4 + d_5 + d_{k-1} + r_3 + r_5 = 0 \quad \text{modulo } 2$$

- - - - - - - - - - - - - - - - -

Chaque grandeur dispose de plusieurs estimations

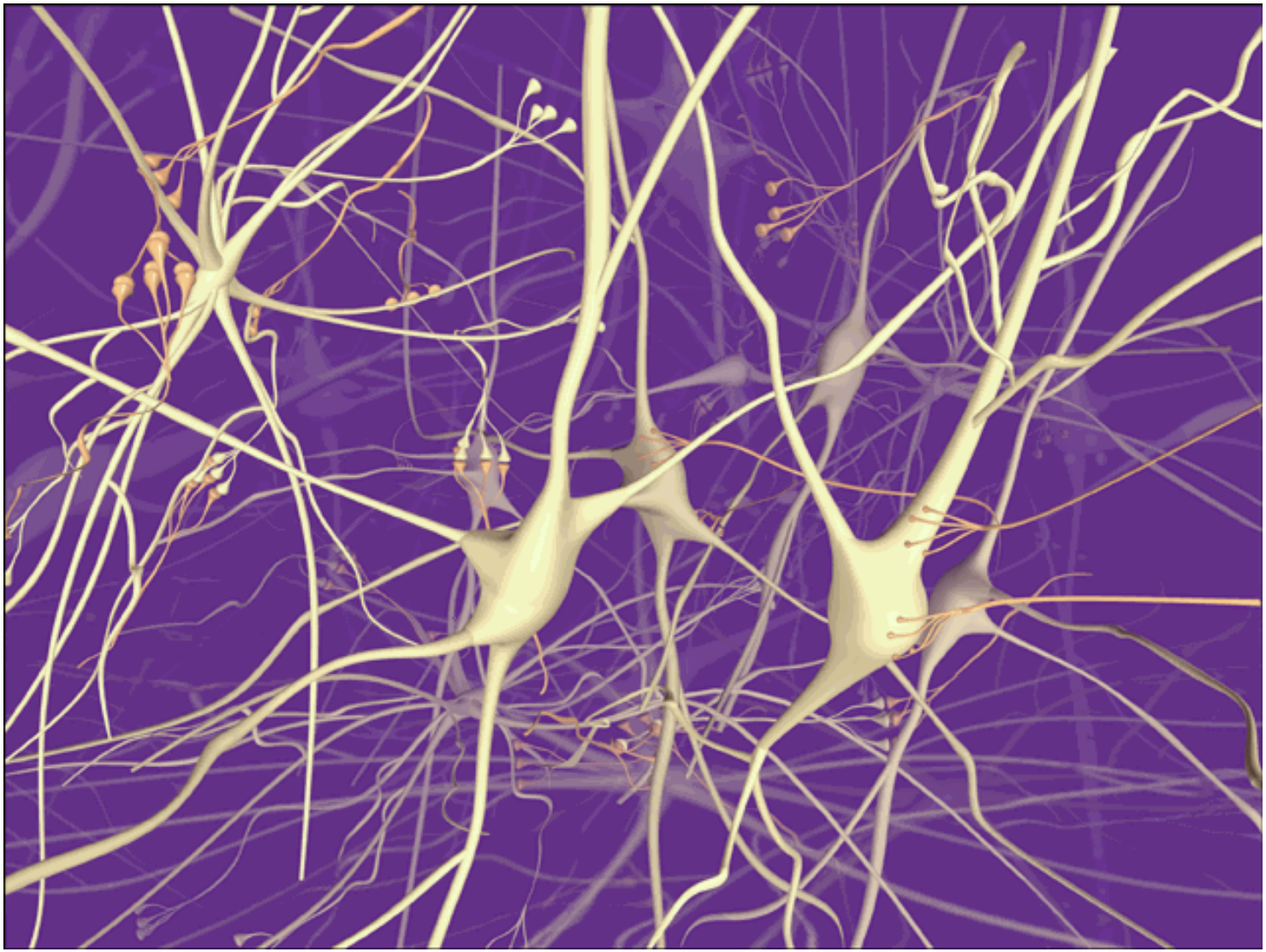
Par exemple, d_5 a été reçu en tant que tel et peut être aussi estimé en tant que $d_1 + r_0 + r_1 + r_{n-k-1}$ ou $d_4 + d_{k-1} + r_3 + r_5$

Le traitement moderne et quasi-optimal de l'information : distribué et probabiliste



small + *communication* *is beautiful*

Forte similitude avec le réseau neuronal biologique



"Si vous voulez comprendre la vie, ne cherchez pas du côté des matières visqueuses vibrantes et palpitantes, pensez aux technologies de l'information"
(Richard Dawkins, *The blind watchmaker*, 1986, Norton, p. 112)

L'intelligence artificielle

Conférence de Dartmouth (1956)

Informatique
Cognitique

Réseau de neurones
Systèmes experts

Electronique

Parallélisme
Logique floue

Théorie de l'information

?

Principe turbo, propagation de croyance

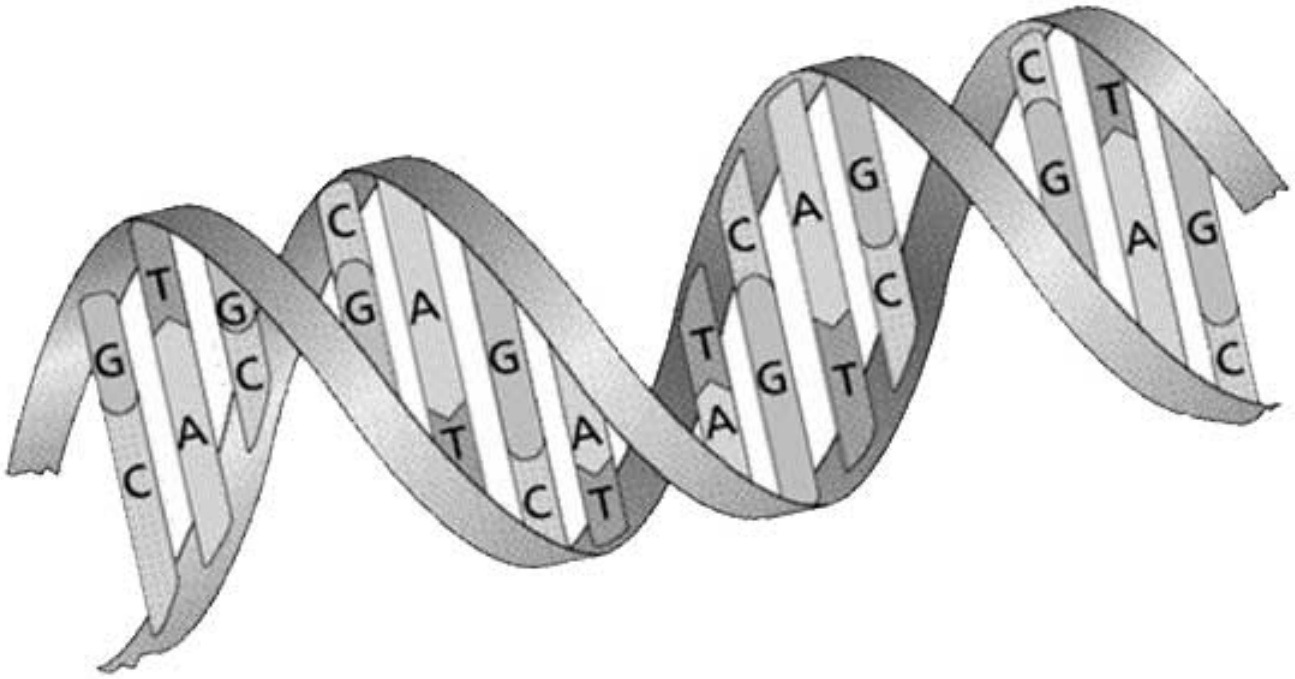
Traitement analogique de l'information

Croiser les connaissances et savoir-faire

$\Rightarrow$ pensée artificielle (2030 ?)

En 2030 une capacité de calcul d'un dollar aura la même performance que celle du cerveau humain (Ray Kurzweil)

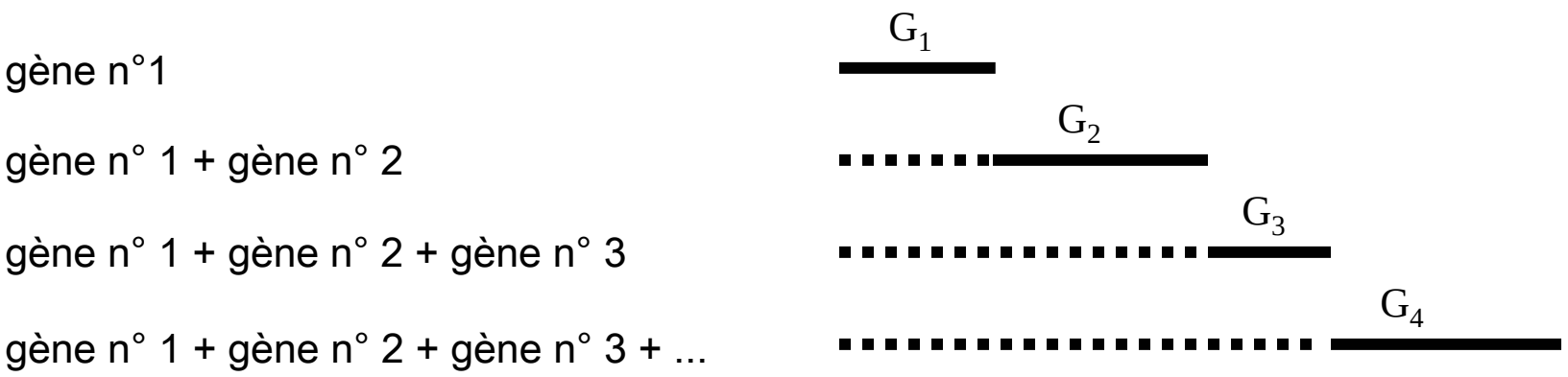
Codage et évolution



Codage et évolution

(d'après les travaux de Gérard Battail)

- La conservation des gènes depuis des temps très lointains est incompatible des taux d'erreurs de réplication (~10^{-4})
- Les mutations (non délétères) des espèces semblent d'autant plus nombreuses que les génomes sont courts



Enchaînement (concaténation) de contraintes syntaxiques comparables à celles des codes correcteurs

Plus il y a de gènes, plus il y a de redondance

Codage et évolution

Conséquences

- L'évolution (mutation/sélection) n'a pas été uniforme au cours du temps
- Plus l'espèce est ancienne (plus son génome est court), plus elle présente de variétés
- Plus l'espèce est récente, moins elle est susceptible de subir des mutations importantes (les fonctions vitales sont les mieux protégées)
- L'importance de la sélection/compétition dans l'évolution est moindre pour les espèces évoluées (puisqu'il y a moins de mutations non létales)

Conclusion

La redondance est au cœur des systèmes vivants et artificiels. Elle est nécessaire à leur bon fonctionnement.

Le codage correcteur, qui est la "science de la redondance", s'est libéré de la tutelle stricte des mathématiques et s'ouvre à d'autres applications que les télécommunications.