Les codes correcteurs

Christophe Ritzenthaler

IRMAR (Rennes 1)

CIRM, Nov. 2016

Shannon : *A mathematical theory of communication* (1948)

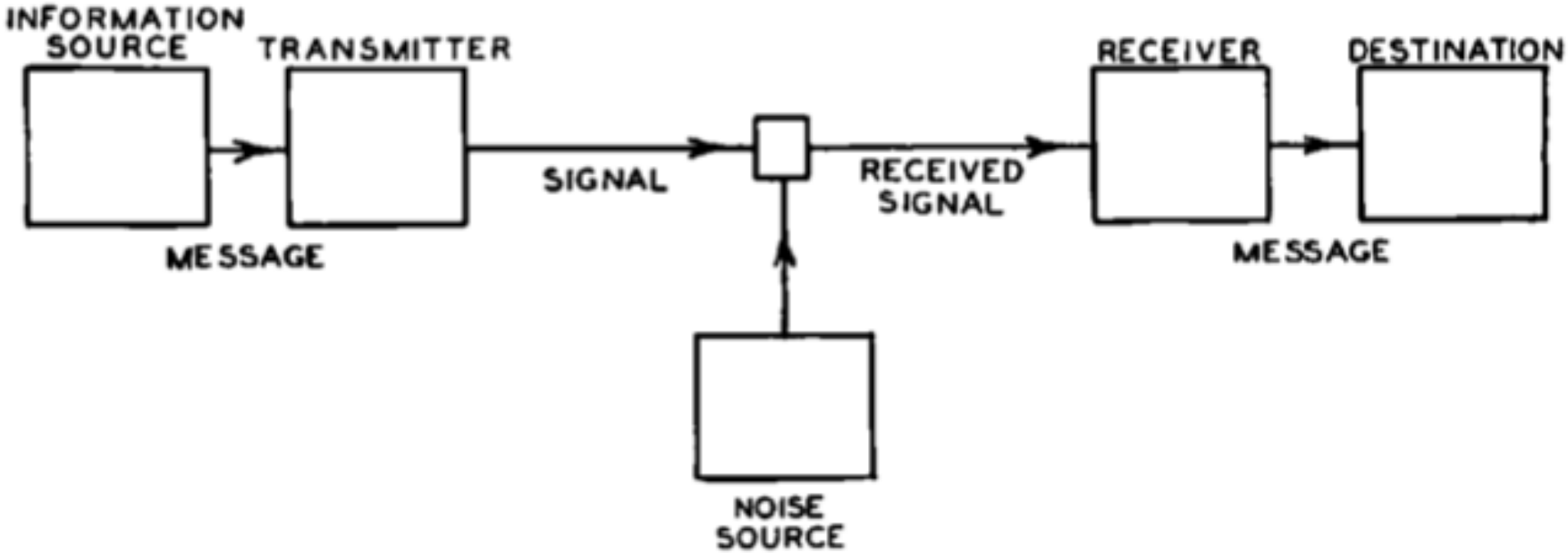


Fig. 1—Schematic diagram of a general communication system.

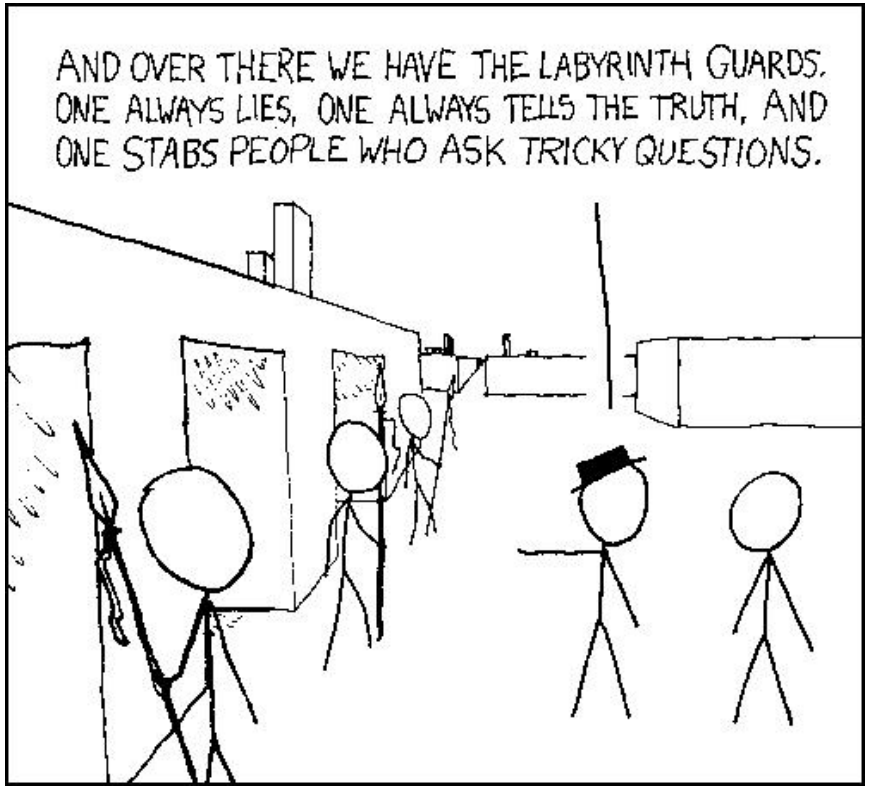
Altération d'un signal

Codage de l'information par des 0 et des 1 transmis sous forme de courants, ondes, etc.

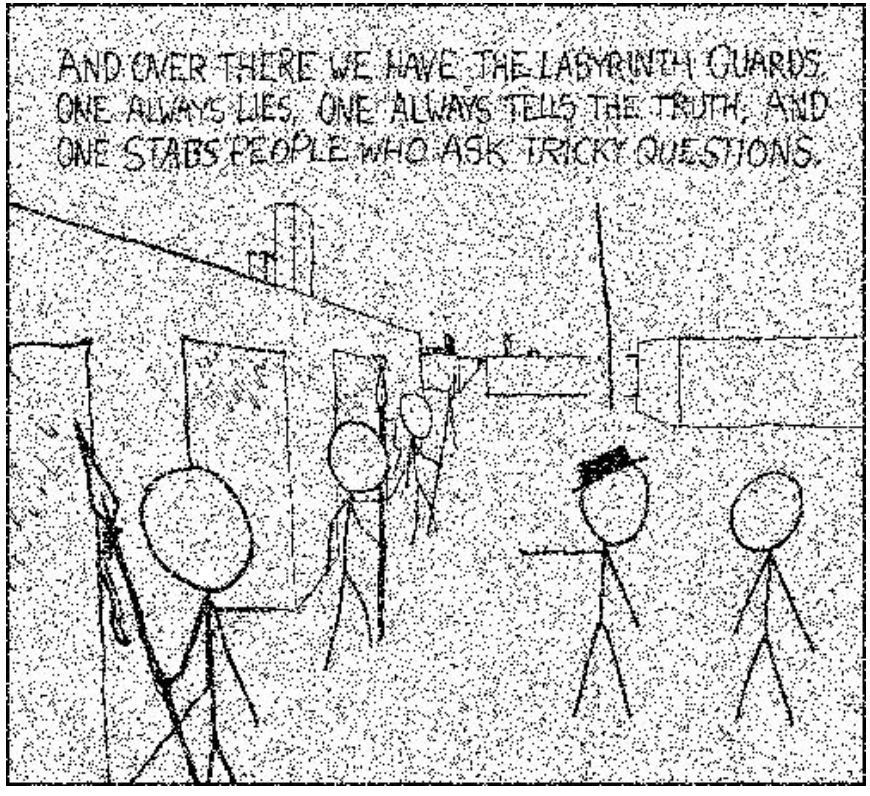
Erreurs de transmission :

- Remplacement de 0 par 1 et inversement ;
- ADSL : Probabilité d'erreur de 10^{-6} et connexion à $1\mathrm{Mo/s}$: en moyenne 8 bits erronés transmis chaque seconde !
- Un CD de bonne qualité $=$ 500000 erreurs (avant lecture).

Notre exemple



Original



Probabilité d'erreur $p = 1/10$

Détecter et corriger

Objectif de la théorie des codes :

coder les messages pour favoriser la détection et la correction des erreurs de transmission.

Détecter une erreur et après ?

- Renvoyer le message : pas toujours possible, manque de temps, nouvelles erreurs possibles ;
- Corriger l'erreur en tirant avantage des données reçues.

La redondance dans la vie courante

La redondance dans la vie courante

Au téléphone :

- Répétition du message ;
- Épeler le mot : C comme Claude, S comme Shannon (chat-nonne ?) ;
- Alphabet radio universel (Alpha, Bravo, Charlie, Delta,. . .).

La redondance dans la vie courante

Au téléphone :

- Répétition du message ;
- Épeler le mot : C comme Claude, S comme Shannon (chat-nonne ?) ;
- Alphabet radio universel (Alpha, Bravo, Charlie, Delta,. . .).

Sécurité sociale

Ici $80 + 2690549588157$ est divisible par 97 (Pourquoi 97 ?)

La redondance dans la vie courante

Au téléphone :

- Répétition du message ;
- Épeler le mot : C comme Claude, S comme Shannon (chat-nonne ?) ;
- Alphabet radio universel (Alpha, Bravo, Charlie, Delta,. . .).

Sécurité sociale

Ici $80 + 2690549588157$ est divisible par 97 (Pourquoi 97 ?)
Beaucoup d'autres exemples : codes barres, cartes bancaires, billets de banque, contraventions,. . .

Un peu plus loin de nous : les tambours parlants ou Tama

“Reviens à la maison” devient “fais que tes pieds reprennent le chemin du retour, fais que tes jambes prennent le chemin du retour, plante tes pieds et tes jambes dans le village qui nous appartient”.

Un peu plus loin de nous : les tambours parlants ou Tama

“Reviens à la maison” devient “fais que tes pieds reprennent le chemin du retour, fais que tes jambes prennent le chemin du retour, plante tes pieds et tes jambes dans le village qui nous appartient”.

Formalisons un peu : code à répétition R_3

Bit d'information	0	0	1	0	1	1	0
Mot transmis	000	000	111	000	111	111	000
Bruit		*	*		* *		
Message reçu	000	001	101	000	010	111	000
Erreur détectée		ici	ici		ici		
Erreur corrigée	0	0	1	0	0	1	0

On corrige ici selon le principe de maximum de vraisemblance :

- 0 s'il y a une majorité de 0 ;
- 1 s'il y a une majorité de 1.

Formalisons un peu : code à répétition R_3

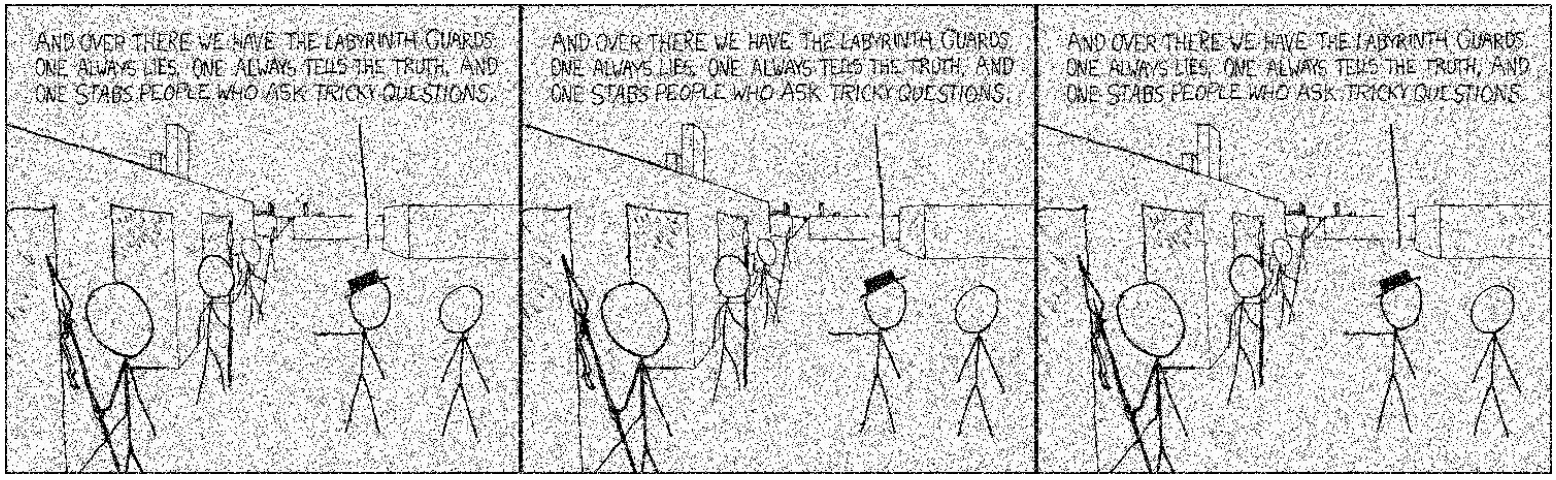
Bit d'information	0	0	1	0	1	1	0
Mot transmis	000	000	111	000	111	111	000
Bruit		*	*		* *		
Message reçu	000	001	101	000	010	111	000
Erreur détectée		ici	ici		ici		
Erreur corrigée	0	0	1	0	0	1	0

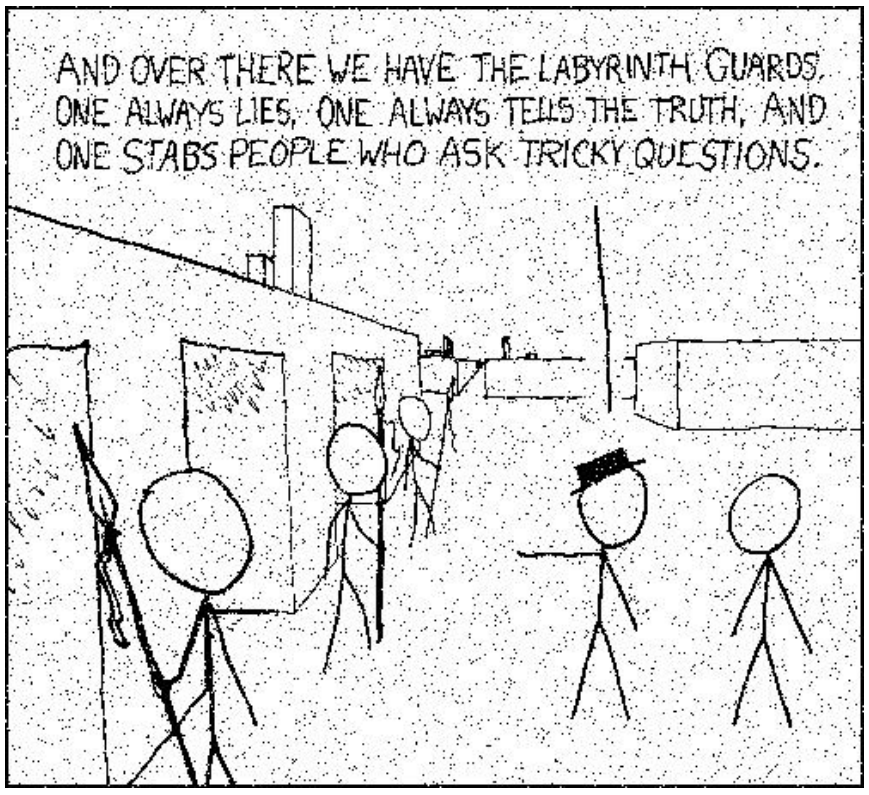
On corrige ici selon le principe de maximum de vraisemblance :

- 0 s'il y a une majorité de 0 ;
- 1 s'il y a une majorité de 1.

Correction si le nombre d'erreurs est inférieur à la moitié (ici 3/2).

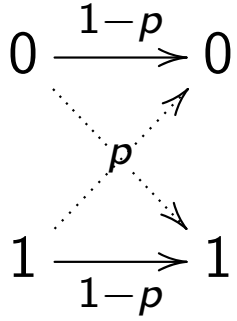
Exemple (la suite)





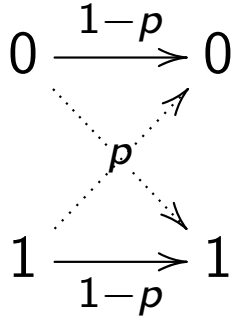
Un peu de probabilité

Soit $0 \leq p \leq 1$ la probabilité d'erreur



Un peu de probabilité

Soit $0 \leq p \leq 1$ la probabilité d'erreur

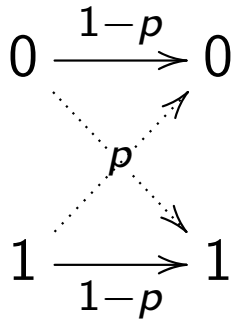


Dans le cas de R_3, les bruits non corrigibles sont :

- $**\ \ $, $*\ *$, $\ \ **$ avec probabilité $p^2(1-p)$;
- $***$ avec probabilité p^3.

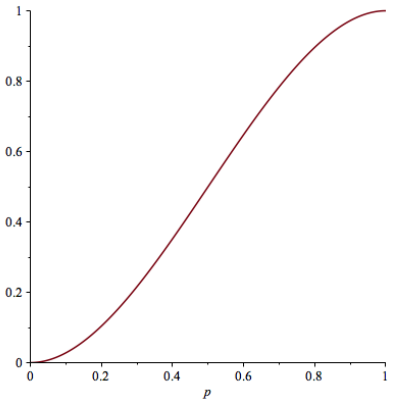
Un peu de probabilité

Soit $0 \leq p \leq 1$ la probabilité d'erreur



Dans le cas de R_3, les bruits non corrigibles sont :

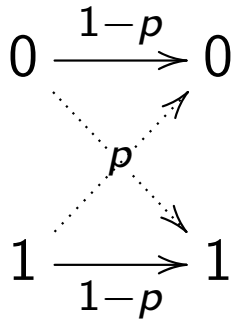
- $**$, $* *$, $**$ avec probabilité $p^2(1-p)$;
- $***$ avec probabilité p^3.



$$\mathcal{P}_3 : p \mapsto p^2(3-2p)$$

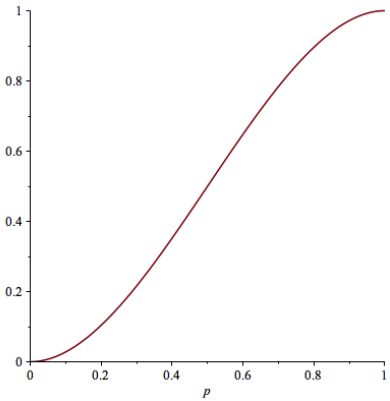
Un peu de probabilité

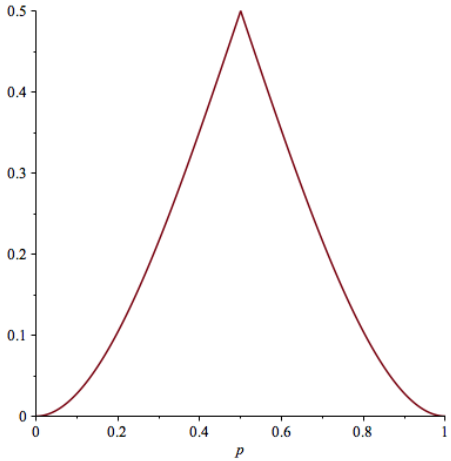
Soit $0 \leq p \leq 1$ la probabilité d'erreur



Dans le cas de R_3, les bruits non corrigibles sont :

- $**$, $* *$, $**$ avec probabilité $p^2(1-p)$;
- $***$ avec probabilité p^3.





$\mathcal{P}_3 : p \mapsto p^2(3-2p)$

$p \mapsto p^2(3-2p),\ 0 \leq p \leq 0,5$

$p \mapsto 1-p^2(3-2p),\ 0,5 \leq p \leq 1$

Généralisons à n répétitions

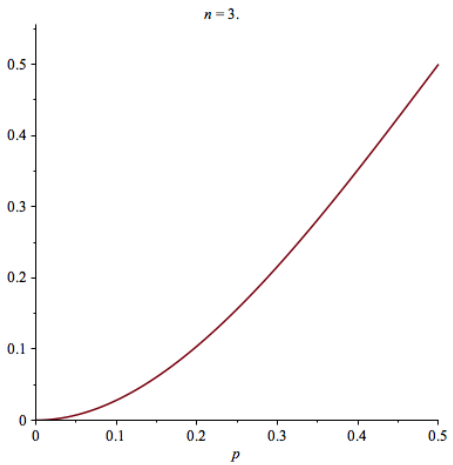
Dans le cas du code R_n avec une probabilité d'erreur $p < 1/2$, la probabilité que le bit ne soit pas corrigible est

$$\mathcal{P}_n(p) = \sum_{k \geq n/2} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Généralisons à n répétitions

Dans le cas du code R_n avec une probabilité d'erreur $p < 1/2$, la probabilité que le bit ne soit pas corrigible est

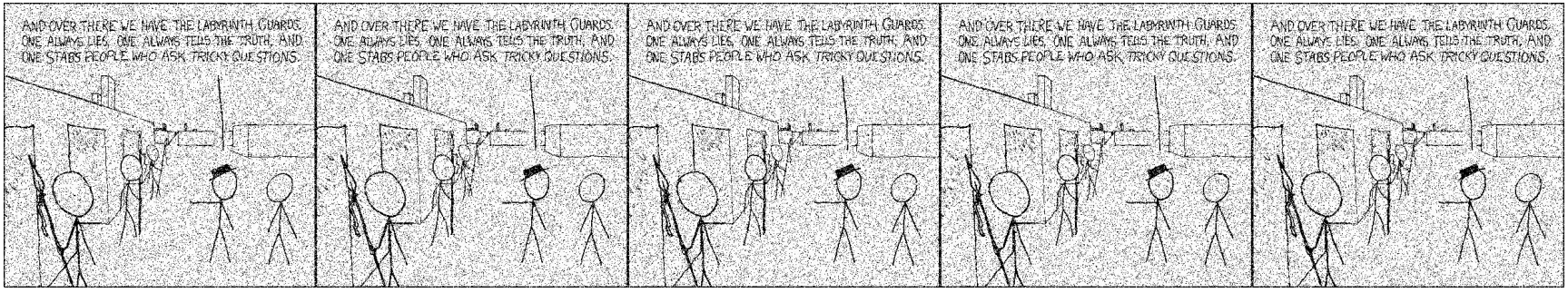
$$\mathcal{P}_n(p) = \sum_{k \geq n/2} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

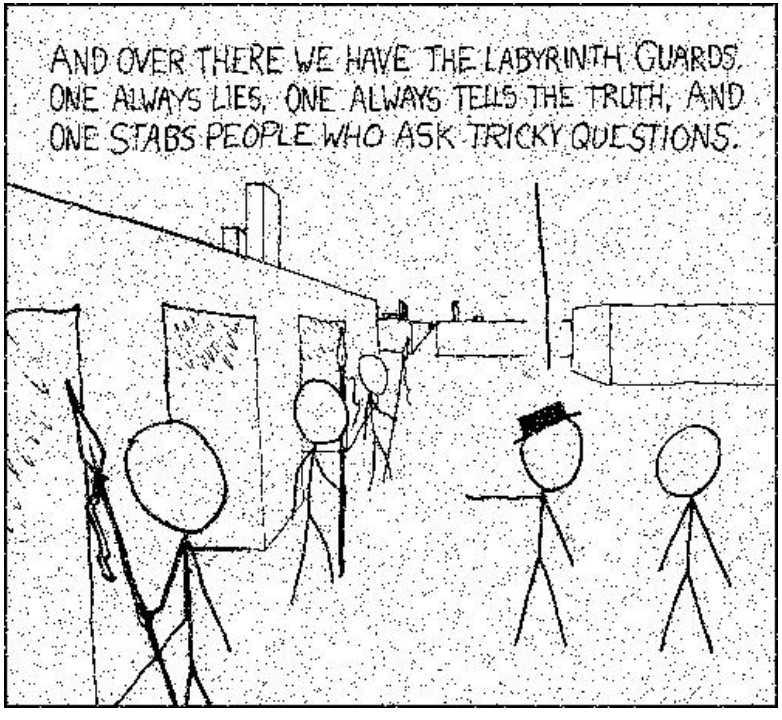


Théorème

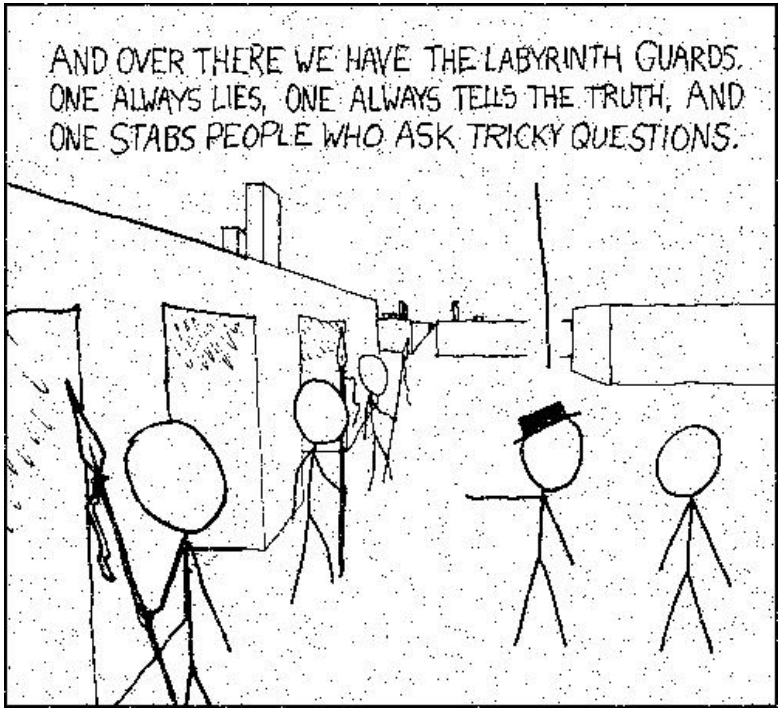
Soit $p < 1/2$ *alors* $\lim_{n \to \infty} \mathcal{P}_n(p) = 0$.

Exemple (la fin)



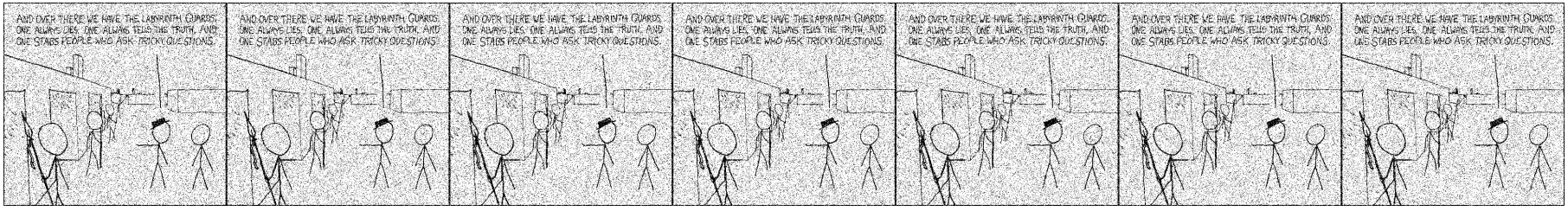


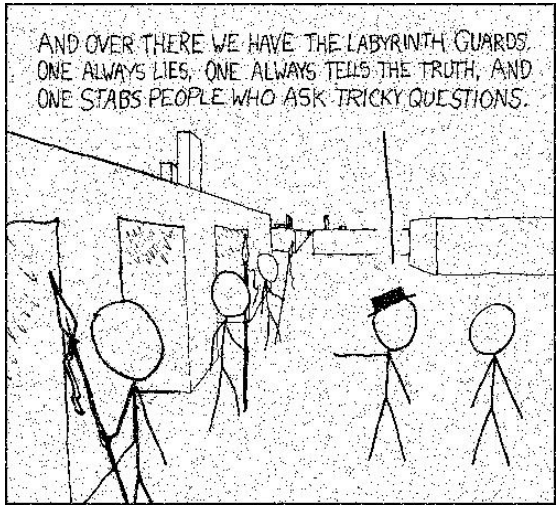
$n = 3$



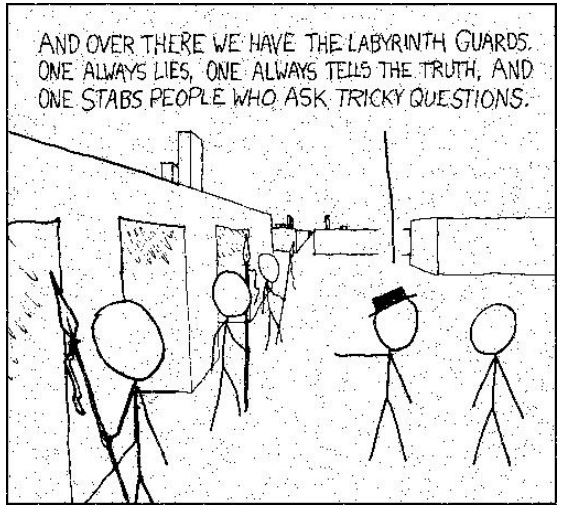
$n = 5$

Exemple (la fin)

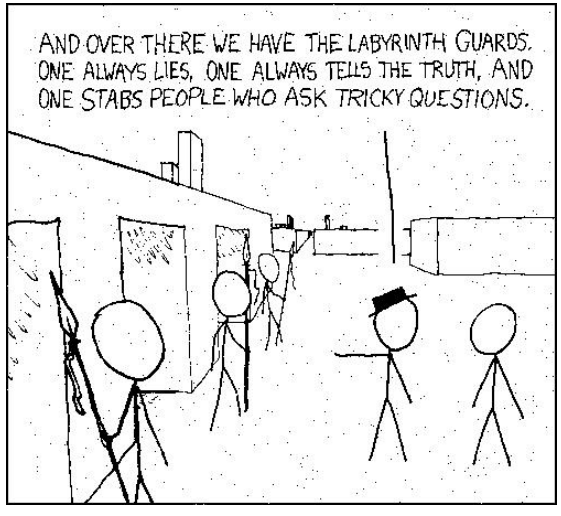




$n = 3$



$n = 5$



$n = 7$

Le code parfait ?

On peut donc rendre la probabilité d'avoir un message erroné aussi petite que l'on veut !

Le code parfait ?

On peut donc rendre la probabilité d'avoir un message erroné aussi petite que l'on veut !

Mais déjà pour $n = 3$

3 fois moins de chansons !

3 fois moins de débit !

Peut-on tout corriger avec un taux de transmission > 0 ?

Le deuxième théorème de Shannon

These results are the main justification for the definition of C and will now be proved.

Theorem 11. Let a discrete channel have the capacity C and a discrete source the entropy per second H. If $H \leq C$ there exists a coding system such that the output of the source can be transmitted over the channel with an arbitrarily small frequency of errors (or an arbitrarily small equivocation). If $H > C$ it is possible to encode the source so that the equivocation is less than $H - C + \epsilon$ where ϵ is arbitrarily small. There is no method of encoding which gives an equivocation less than $H - C$.

The method of proving the first part of this theorem is not by exhibiting a coding method having the desired properties, but by showing that such a code must exist in a certain group of codes. In fact we will average the frequency of errors over this group and show that this average can be made less than ϵ. If the average of a set of numbers is less than ϵ there must exist at least one in the set which is less than ϵ. This will establish the desired result.

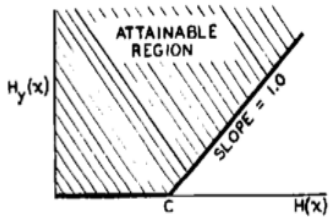
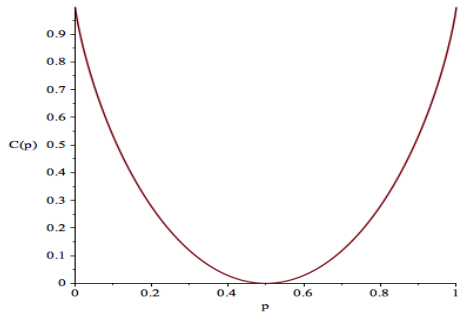


Fig. 9—The equivocation possible for a given input entropy to a channel.

The capacity C of a noisy channel has been defined as

$$C = \text{Max}\,(H(x) - H_y(x))$$

Il existe un code de longueur $n \to \infty$ avec une probabilité d'erreur aussi faible que l'on veut et un taux de transmission aussi proche que l'on veut de $C(p)$ (inférieurement).
Par contre tout code avec un taux de transmission $> C(p)$ n'est pas meilleur qu'un "décodage aléatoire".



AMERICAN MATHEMATICAL SOCIETY
MathSciNet
Mathematical Reviews

Previous | Up | Next

Citations
From References: 960
From Reviews: 99

MR0026286 (10,133e) 60.0X
Shannon, C. E.
A mathematical theory of communication.
Bell System Tech. J. **27,** (1948). 379–423, 623–656

"The discussion is suggestive throughout, rather than mathematical, and it is not always clear that the author's mathematical intentions are honorable."

Conséquences et interrogations

- Exemple : Probabilité d'erreur du canal $p = 10^{-3}$ alors $C(p) \approx 0,998$.

  Shannon : on peut donc corriger aussi sûrement que l'on veut des messages en rajoutant $1000/0,998 - 1000 \approx 12$ bits pour 1000 bits transmis.

- Avec R_3 il faut rajouter 2000 bits pour un taux d'erreur de $2 \cdot 10^{-6}$.

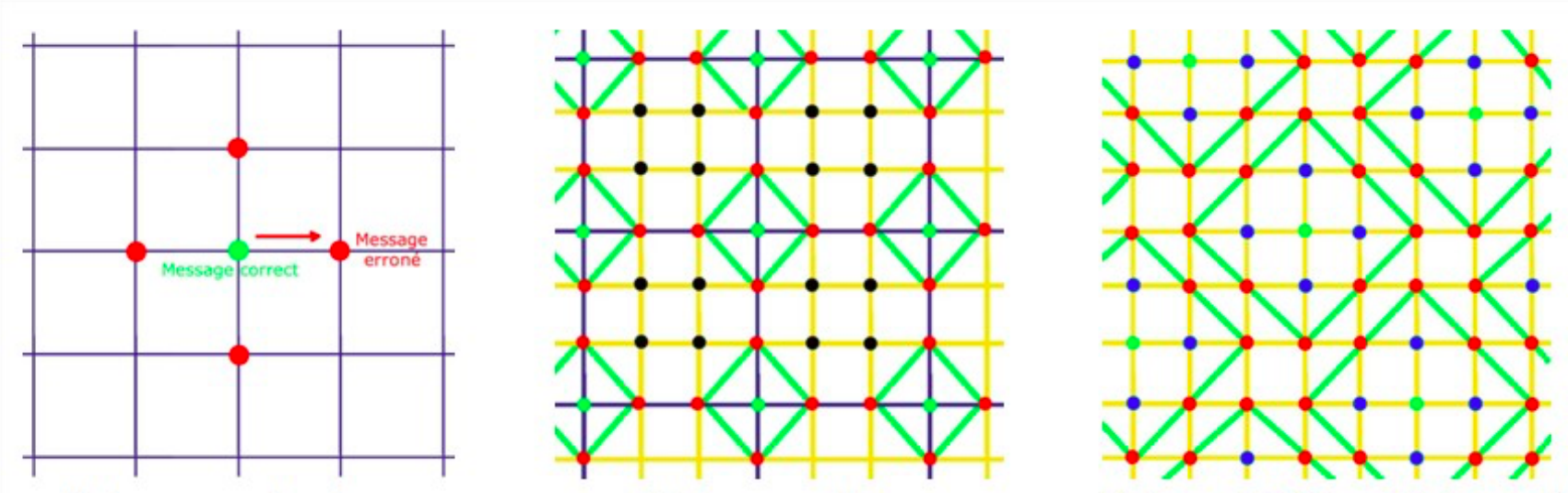
Conséquences et interrogations

- Exemple : Probabilité d'erreur du canal $p = 10^{-3}$ alors $C(p) \approx 0,998$.

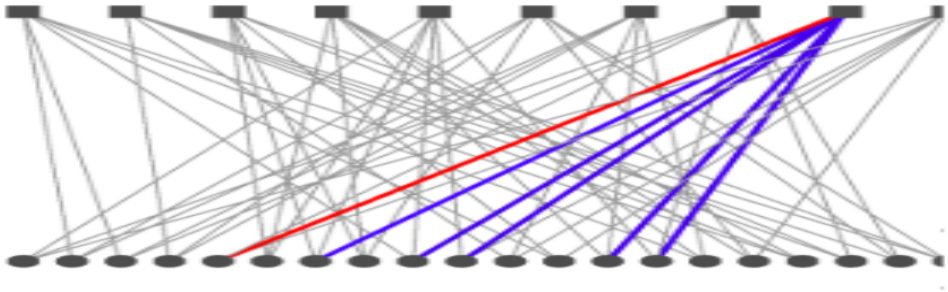
  Shannon : on peut donc corriger aussi sûrement que l'on veut des messages en rajoutant $1000/0,998 - 1000 \approx 12$ bits pour 1000 bits transmis.
- Avec R_3 il faut rajouter 2000 bits pour un taux d'erreur de $2 \cdot 10^{-6}$.
- Shannon ne dit pas comment construire ces codes !
- Dans la pratique, ils ne seraient de toute façon pas utilisables car leur longueur tend vers l'infini et surtout on ne sait pas les décoder de manière efficace.

Où sont les bons codes ?

Les codes linéaires $[n, k, d]$ de taux de transmission k/n et pouvant corriger $\lfloor \frac{d-1}{2} \rfloor$ erreurs par bloc.



Les turbo-codes et codes LDPC



Dans la vie de tous les jours

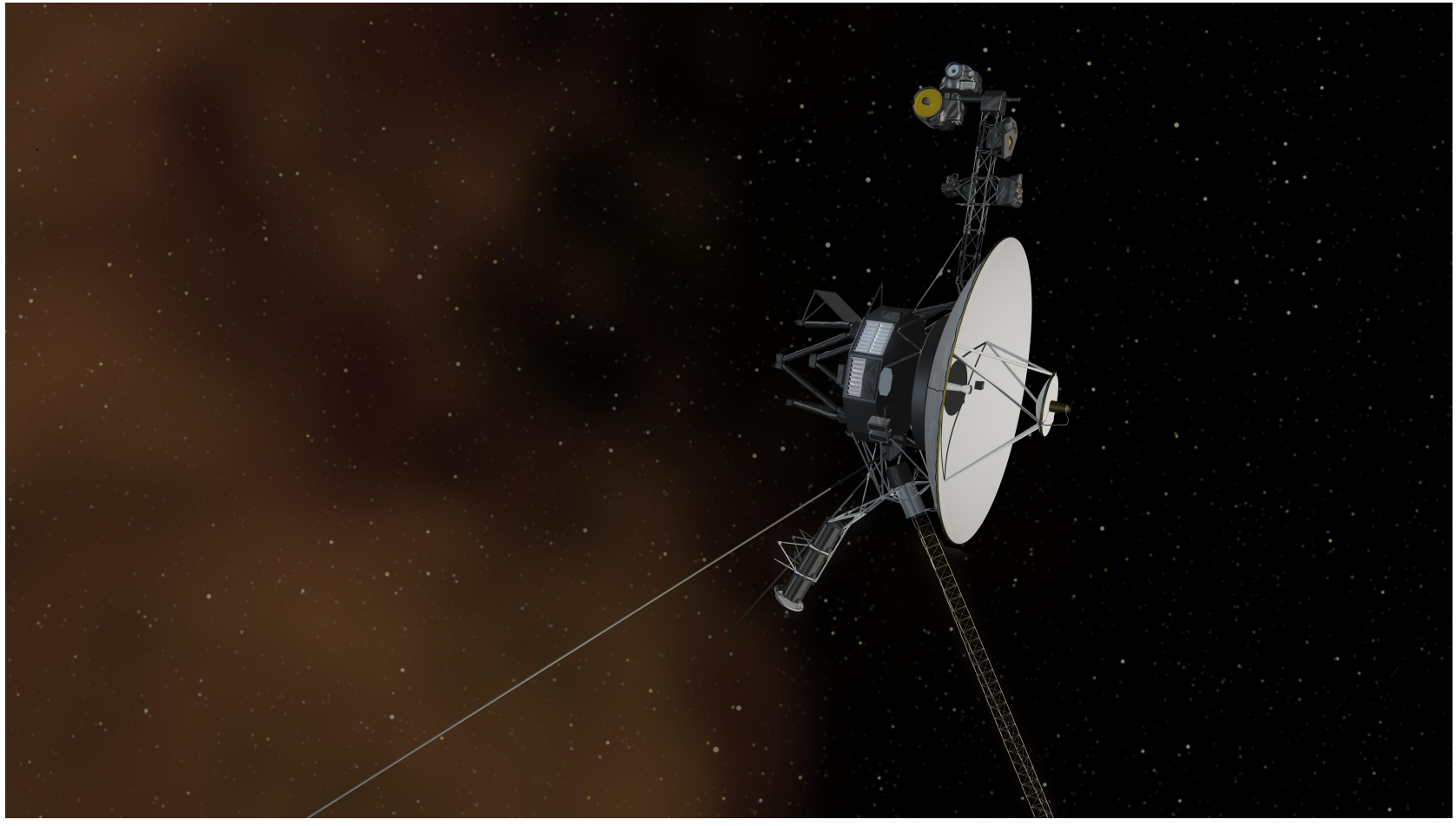
- Dans les CD : 2 codes de Reed-Solomon $[28, 24, 5]$ et $[32, 28, 5]$ entrelacés.

- Dans les QR-codes : (Reed-Solomon et BCH)

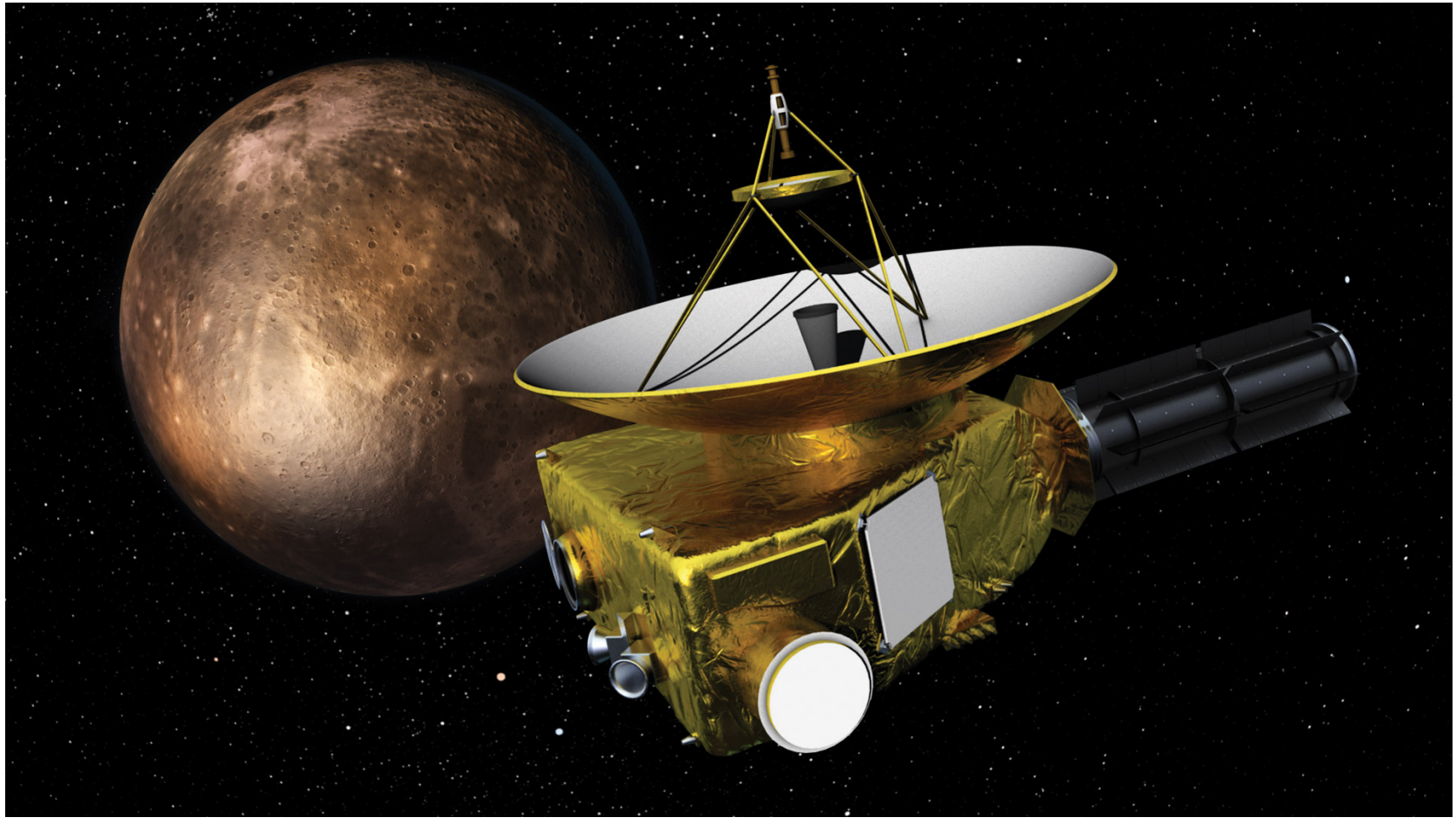
- Dans l'espace : Mariner (Reed-Müller [32, 6, 16]),

, .

- Dans l'espace : Mariner (Reed-Müller $[32, 6, 16]$),Voyager (Golay $[24, 12, 8]$), .



- Dans l'espace : Mariner (Reed-Müller [32, 6, 16]),Voyager (Golay [24, 12, 8]), New Horizon (turbo-codes, LDPC).



Merci de votre attention !

Pourquoi 97 ?

Soit $n = [K, C]$ un numéro de sécurité sociale valide : $K + C$ est divisible par 97.
Supposons que le i-ème chiffre de K soit possiblement faux, i.e. on transmet $K' = K + e10^i$ avec $0 \leq c_i + e \leq 9$. On a que 97 divise

$$K' + C = K + e10^i + C = K + C + e10^i$$

si et seulement si 97 divise $e10^i$.
Comme 97 est premier $e10^i$ est divisible par 97 si et seulement si $e = 0$.
On a donc détecter une erreur !