

# Codes Correcteurs de Hamming

L3 informatique, Sécurité Appliquée

Jean-François COUCHOT

Université de Franche-Comté, UFR-ST













De la redondance avant tout







De la redondance avant tout Introduction Code de Hamming (7,4)



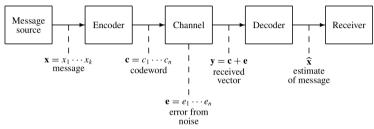


De la redondance avant tout Introduction Code de Hamming (7,4)



### Communication bruitée

Schéma théorique de communication avec correction <sup>1</sup>



- 1. A la source : un vecteur x doit être envoyé
  - ► Si x était transmis tel quel : n'importe quel bruit ajouté rendrait le message irrécupérable
  - ldée principale : ajouter de la redondance au message
- 2. Mot de code c envoyé sur le canal : vecteur contenant x et redondance
- 3. Erreur e : modélisée par un vecteur de bruit ajouté au mot de code
- 4. Estimation  $\hat{x}$  de x: construite à partir du vecteur reçu y



## Des erreurs partout



#### Taux d'erreur

Le taux d'erreur (BER pour Bit Error Ratio) est le rapport entre le nombre de bits erronés reçus par rapport au nombre total de bits transmis.

#### Ordres de grandeur 2 de BER

► Disques optique : 10<sup>-5</sup>

Ligne téléphonique :  $10^{-6}$ 

ightharpoonup Communication par fibres optiques :  $10^{-9}$ 



 $2. \ https://ljk.imag.fr/membres/Jean-Guillaume.Dumas/Publications/IntroductionAuxCodesCorrecteurs.pdf \\$ 



## Détecter, corriger : des exemples de codes

### Au temps des transmissions analogiques <sup>3</sup> de mauvaise qualité

Remplacer chaque lettre par un mot dont la première lettre coïncide avec la lettre à épeler. Sens de « Écho, Roméo, Roméo, Écho, Uniforme, Roméo » ?

#### Code détecteur : bit de parité

- Codage d'un caractère : nombre décimal entre 0 et  $\frac{\text{Car. Dec. Bin. Bin+Parité}}{\text{B}} = \frac{\text{Bin. Bin+Parité}}{\text{B}} = \frac{\text{Bin. Bin+Parité}}{\text{B}} = \frac{\text{Car. Dec. Bin. Bin+Parité}}{\text{B}} = \frac{\text{Bin. Bin+Parité}}{\text{$
- Ajout d'un 8ème bit de parité : la somme des 8 bits C 66 1000010 10000100 est paire :

Corriger 1 erreur? Détecter 2 erreurs?

a 97 1100001 1100001**1** 

#### Code correcteur : duplication

- ightharpoonup Répéter la totalité du message suffisamment ( $\geq 3 \times$ ) de fois (penser à « ââgmee »)
- ightharpoonup 011010 ightharpoonup 000.111.111.000.111.000
- ▶ Si au plus une erreur, message original associé à 100.111.110.001.110.000?

<sup>3.</sup> Rousseau, C., & Saint-Aubin, Y. (2009). Mathématiques et technologie. Springer Science & Business Media.



## Efficacité de la détection/correction?



#### Définition du rendement $\rho$

 $\rho: {\it rapport} \ \frac{{\it nombre de bits de message}}{{\it nombre de bits transmis}} \ toujours inférieur ou égal à 1.$ 

#### Rendements des codes précédents

- Parité :  $\rho_{\text{parite}} = 7/8 = 87.5\%$ , mais détection seulement
- ▶ Duplication :  $ho_{ ext{dupl}_3} = 1/3 \approx 33.3\%$ , avec correction



### **Autres exemples**

Formule de Luhn 4

- 1. Doubler un chiffre sur deux mod. 9, depuis l'avant dernier, de dr. à gche.
- 2. Somme de tous les chiffres multiple de  $10 \sim$  validité du nombre original.
- 3. Exemples : validité de "972-487-086" et "927-487-086"?
- 4. Rendement  $ho_{\rm Luhn_{CB}}=15/16\approx 93.8\%$ , mais détection uniquement

#### Numéro de SS<sup>5</sup>

- 1. NIR : numéro de 13 chiffres (Genre, deux derniers chiffres de l'année de naissance, mois de naiss....)
- 2. Clef sur 2 chiffres : NIR mod. 97 (après gestion ev. des corses)
- 3. Rendement  $\rho_{\rm NSS}=13/15\approx 86.7\%$ , mais détection uniquement

#### En vrac

### Dans les QR codes 6, dans la DRAM 7, dans les OS,...

- 4. https://fr.wikipedia.org/wiki/Formule\_de\_Luhn pour les CB, les SIRET
- 5. https://fr.wikipedia.org/wiki/Num%C3%A9ro\_de\_s%C3%A9curit%C3%A9\_sociale\_en\_France#ancrage\_C
- $6. \ \mathtt{https://fr.wikipedia.org/wiki/Code\_QR}$
- 7. https://en.wikipedia.org/wiki/ECC\_memory







De la redondance avant tout

Introduction

Code de Hamming (7,4)



## Rappels d'algèbre de Bool



#### Somme et produit

- lackbox La somme + et le produit . sont définis dans les booléens  $\mathbb{B}=\{0.1\}$
- ► + : "OU exclusif"
- ▶ .: "ET"

+	0	1		0	1
0	0	1	0	0	0
1	1	0	1	0	1



### **Formalisation**

#### Contraintes pour corriger au plus 1 erreur

- Mot à transmettre de 4 bits :  $u = (u_1, u_2, u_3, u_4) \in \mathbb{B}^4$
- Mot de code de 7 bits transmis :  $v = (v_1, v_2, v_3, v_4, v_5, v_6, v_7) \in \mathbb{B}^7$

### Construction du mot de code $v = (u_1, u_2, u_3, u_4, v_5, v_6, v_7)$

$$v_5 = u_1 + u_2 + u_4$$
,  $v_6 = u_1 + u_3 + u_4$ ,  $v_7 = u_2 + u_3 + u_4$ .

▶ Bits redondants :  $v_5$ ,  $v_6$  et  $v_7 \sim \rho_{H_{(7,4)}} = 4/7 \approx 57.1\%$ 

### A réception de $w = (w_1, w_2, w_3, w_4, w_5, w_6, w_7) \in \mathbb{B}^7$

$$V_5 = w_1 + w_2 + w_4$$

$$V_6 = w_1 + w_3 + w_4$$

$$V_7 = w_2 + w_3 + w_4$$

lacksquare Utilisation de la table ightarrow

·/ ·/ -			
$w_5 = W_5$	$w_6 = W_6$	$w_7 = W_7$	postion de l'erreur
Т	Т	Т	Tout est OK
Т	Т	F	inverser w <sub>7</sub>
Т	F	Т	inverser w <sub>6</sub>
Т	F	F	inverser w <sub>3</sub>
F	Т	Т	inverser w <sub>5</sub>
F	Т	F	inverser w <sub>2</sub>
F	F	Т	inverser w <sub>1</sub>
F	F	F	inverser w <sub>4</sub>



## Exemple transmission de u = (1, 0, 1, 1)

o ,

Construction du mot de code

▶ Mot de code transmis : 
$$(1,0,1,1,v_5,v_6,v_7)$$
 avec  $v_5=u_1+u_2+u_4=1+0+1=0$  ,  $v_6=u_1+u_3+u_4=1+1+1=1$ ,  $v_7=u_2+u_3+u_4=0+1+1=0$ 

A réception de  $w = (w_1, w_2, w_3, w_4, w_5, w_6, w_7) = (1, 1, 1, 1, 0, 1, 0)$ 

$$W_5 = w_1 + w_2 + w_4 = 1 + 1 + 1 = 1 \neq w_5,$$
  
 $W_6 = w_1 + w_3 + w_4 = 1 + 1 + 1 = 1 = w_6,$   
 $W_7 = w_2 + w_3 + w_4 = 1 + 1 + 1 = 1 \neq w_7.$ 

- ► En supposant qu'il n'y ait qu'une erreur et par analyse du tableau précédent : il faut inverser le bit w<sub>2</sub>.
- Message corrigé : (1,0,1,1)





De la redondance avant tout

### Les codes systématiques de Hamming

Présentation générale

Matrices de contrôle et de génération

Erreur : détection et correction





De la redondance avant tout

### Les codes systématiques de Hamming

Présentation générale

Matrices de contrôle et de génération

Erreur : détection et correction



## Hamming $(2^k - 1, 2^k - k - 1)$

### Remarques

- $\triangleright$  2<sup>k</sup> 1 : taille du mot de code
- $\triangleright$  2<sup>k</sup> k 1 : taille du mot u à transmettre
- Pour k = 3, on retrouve Hamming(7,4)
- ▶ Ne corrige au plus qu'une seule erreur par mot de  $2^k 1$  bits
- Code systématique : « le mot u est complété avec des bits de redondance »
- ▶ Dans tout ce qui suit dans le chapitre : que des codes systématiques

#### Points clés

- ▶ Deux matrices  $G_k$  et  $H_k$  à valeur dans  $\mathbb{B}$  :
  - $ightharpoonup G_k$ : pour générer le mot de code v à partir du mot à transmettre u
  - $ightharpoonup H_k$ : pour contrôler w et le corriger éventuellement
- ► Le tout avec des produits matriciels







De la redondance avant tout

### Les codes systématiques de Hamming

Présentation générale

Matrices de contrôle et de génération

Erreur : détection et correction



## Matrices de contrôle $H_{\nu}$

#### Définition

$$H_k = \left(\frac{P_k}{I_k}\right) \qquad (1) \qquad I_k$$

- $I_{\nu}$ : matrice identité  $k \times k$
- $\triangleright$   $P_k$ : en ligne, tous les vecteurs non nuls de  $\mathbb{B}^k$  qui ne sont pas dans
- $ightharpoonup \sim H_k$ : contient tous les vecteurs non nuls de  $\mathbb{B}^k$ .
- $ightharpoonup \sim H_{\nu}: 2^k 1$  lignes et k colonnes

Exemple avec  $H_3$ 

$$H_3 = egin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

ou bien 
$$H_3' = egin{pmatrix} 1 & 1 & 1 \ 1 & 1 & 0 \ 1 & 0 & 1 \ \hline 0 & 1 & 1 \ \hline 1 & 0 & 0 \ 0 & 1 & 0 \ 0 & 0 & 1 \ \end{pmatrix}$$

$$H_3 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ \frac{1}{1} & \frac{1}{0} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
 ou bien  $H_3' = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \frac{0}{1} & 1 & 1 \\ \frac{1}{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  ou encore  $H_3'' = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ \frac{1}{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \dots$ 

## Matrices de génération $G_k$



#### Définition

$$G_k = \left(I_{2^k - k - 1} \mid P_k\right) \qquad (2)$$

 $\triangleright$   $P_k$ : même matrice que dans  $H_k$ 

$$G_k = (I_{2^k-k-1} \mid P_k)$$
 (2)  $I_{2^k-k-1}$ : matrice identité  $(2^k - k - 1) \times (2^k - k - 1)$ 

 $ightharpoonup 
ightharpoonup G_k: 2^k - k - 1 \text{ lignes et } 2^k - 1 \text{ colonnes}$ 

Exemple avec  $G_3$ 

$$\mathsf{si}\; H_3 = egin{pmatrix} 1 & 1 & 0 \ 1 & 0 & 1 \ 0 & 1 & 1 \ \frac{1}{1} & \frac{1}{1} & \frac{1}{0} & 0 \ 0 & 1 & 0 \ 0 & 0 & 1 \ \end{pmatrix} \;\; \mathsf{alors}\; G_3 = egin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \ 0 & 1 & 0 & 0 & 1 & 1 & 1 \ 0 & 0 & 0 & 1 & 1 & 1 & 1 \ \end{pmatrix}$$



### Mot de code et syndrome

Mot de code pour  $(u_1, \ldots, u_{2k-k-1}) \in \mathbb{B}^{2^k-k-1}$ 

- Mot de code associé :  $v = uG_k \in \mathbb{B}^{2^k 1}$
- $ightharpoonup G_k = (I_{2^k-k-1} \mid P_k)$ : les  $2^k k 1$  premiers bits de v sont ceux de u
- ▶ Rendement  $\rho_{H_k} = (2^k k 1)/(2^k 1)$  : croissant pour k > 3

### Syndrome de $w \in \mathbb{B}^{2^k-1}$

- ▶ Soit  $w = (w_1, ..., w_{2^k-1})$  le mot à décoder.
- Le mot  $\sigma(w) = wH_k \in \mathbb{B}^k$ : appelé syndrome de w

Exemple avec u = (1, 0, 1, 1)

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & | & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & | & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & | & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & | & 1 & 1 & 1 \end{pmatrix}, H = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ \frac{1}{1} & \frac{1}{1} & \frac{1}{1} \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\triangleright$$
  $v = uG = (1, 0, 1, 1, 0, 1, 0)$ 

$$\sigma(v) = vH = (0,0,0)$$

$$ightharpoonup$$
 si  $w = (1, 0, 0, 1, 0, 1, 0)$ 

$$\sigma(w) = wH = (0,1,1)$$





De la redondance avant tout

### Les codes systématiques de Hamming

Présentation générale Matrices de contrôle et de génération

Erreur : détection et correction



21/26



Exemple avec  $\textit{G}_3 \times \textit{H}_3$ 

$$\begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & | & \mathbf{1} & 1 & 0 \\ 0 & 1 & 0 & 0 & | & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & | & 1 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} \mathbf{1} & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ \frac{1}{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{0} & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ \end{pmatrix}$$



7











$$\begin{pmatrix} 1 & 0 & 0 & 0 & | & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & | & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & | & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & | & 1 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ \frac{1}{1} & 1 & 1 \\ \frac{1}{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$





$$\begin{pmatrix} 1 & 0 & 0 & 0 & | & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & | & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & | & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & | & 1 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ \frac{1}{1} & 1 & 1 \\ \frac{1}{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$



$$\begin{pmatrix} 1 & 0 & 0 & 0 & | & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & | & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & | & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & | & 1 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ \frac{1}{1} & 1 & \frac{1}{1} \\ \frac{1}{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1+1}{1+1+1+1} \\ \frac{1+1}{1+1+1+1} \end{pmatrix}$$





7

Exemple avec 
$$G_3 \times H_3$$

Montrons que  $(G_k \times H_k)_{i,j} = 0$  pour pour toute ligne i et colonne j

$$ightharpoonup (G_k \times H_k)_{i,j} = P_{i,j} + P_{i,j} = 0$$



### Détection d'une erreur

#### Mot de code et Syndrome nul

Pour un code de Hamming $(2^k - 1, 2^k - k - 1)$ , un mot w est un mot de code si et seulement si son syndrome  $\sigma(w) = wH$  est nul

#### Preuve

- ightharpoonup  $\Rightarrow$  : v est un mot de code  $\sim v = uG \sim \sigma(v) = vH = uGH = (0)$
- $\blacktriangleright \Leftarrow : Soit \ w = (w_1, \dots, w_{2k-k-1}, w_{2k-k-1+1}, \dots, w_{2k-k-1+k})$

$$\sigma(w) = wH = (
\sum_{i=1}^{2^k - k - 1} w_i P_{i,1} + w_{2^k - k - 1 + 1}, \\
\sum_{i=1}^{2^k - k - 1} w_i P_{i,2} + w_{2^k - k - 1 + 2}, \\
\dots, \\
\sum_{i=1}^{2^k - k - 1} w_i P_{i,k} + w_{2^k - k - 1 + k})$$

$$\sigma(w) = 0 \rightsquigarrow w_{2^k - k - 1 + j} = \sum_{i=1}^{2^k - k - 1} w_i P_{i,j} \text{ pour tout } j, 1 \le j \le k$$
  
Ainsi  $w = (w_1, \dots, w_{2^k - k - 1})G \rightsquigarrow w$  est un mot de code.

23/26

### Correction d'une erreur

Correction lorsque le syndrome  $\sigma(w)$  n'est pas nul

Soit un code de Hamming $(2^k - 1, 2^k - k - 1)$  et w tel que  $\sigma(w)$  n'est pas nul :

- ▶ il existe une ligne i de H qui lui est égale et
- le vecteur  $w'=(w_1,\ldots,w_{i-1},\overline{w_i},\ldots,w_{2^k-1})$  est un mot de code.

Preuve : montrons que  $\sigma(w')_c=0$  pour  $1\leq c\leq k$ 

$$\sigma(w')_c = \sum_{l=1, l\neq i}^{2^n-1} w_l \cdot H_{l,c} + \overline{w_i} \cdot H_{i,c}$$
(3)

$$\sigma(w)_{c} = \sum_{l=1, l\neq i}^{2^{n}-1} w_{l}.H_{l,c} + w_{i}.H_{l,c} = H_{i,c}$$
(4)

- ► Si  $w_i = 1$ , de (4), on déduit que  $\sum_{l=1, l \neq i}^{2^k 1} w_l . H_{l,c} = 0$  et donc  $\sigma(w')_c = 0$ .
- Si  $w_i = 0$ , de (4), on déduit que  $H_{i,c} = \sum_{l=1,l\neq i}^{2^k-1} w_l \cdot H_{l,c}$  et donc  $\sigma(w')_c = \sum_{l=1,l\neq i}^{2^k-1} w_l \cdot H_{l,c} + \left(\sum_{l=1,l\neq i} w_l \cdot H_{l,c}\right) = 0$



## Retour à l'exemple avec u = (1, 0, 1, 1)



#### Rappels

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & | & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & | & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & | & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & | & 1 & 1 & 1 \end{pmatrix}, H = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ \frac{1}{1} & \frac{1}{1} & \frac{1}{1} \\ \frac{1}{1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\triangleright$$
  $v = uG = (1, 0, 1, 1, 0, 1, 0)$ 

$$ightharpoonup \sigma(v) = vH = (0,0,0)$$

$$ightharpoonup$$
 si  $w = (1, 0, 0, 1, 0, 1, 0)$ 

Détection de l'erreur et correction

- v = (1, 0, 1, 1, 0, 1, 0) et  $\sigma(v) = (0, 0, 0)$  : v est un mot de code et le mot original était donc (1, 0, 1, 1)
- w = (1, 0, 0, 1, 0, 1, 0) et  $\sigma(w) = wH = (0, 1, 1) = H_3$ : le mot de code associé à w est w' = (1, 0, 1, 1, 0, 1, 0); le mot original était donc (1, 0, 1, 1)



### Pour aller plus loin

- les codes de redondance cyclique (CRC) : Hamming, BCH <sup>8</sup>, Reed-Solomon <sup>9</sup>
- les codes convolutifs 10
- les turbo-codes 11 :
  - standards 3G et 4G
  - ▶ la NASA lors de Mars reconnaissance Orbiter (1999 . . .2017)
  - ► la norme IEEE 802.16 (WiMAX)
- les codes de parité à faible densité (LDPC) 12 :
  - standard 5G
  - télévision numérique terrestre
  - Norme Wifi 802.11
  - disques SSD

<sup>10.</sup> https://en.wikipedia.org/wiki/Convolutional\_code



12. https://en.wikipedia.org/wiki/Low-density\_parity-check\_code





<sup>8.</sup> https://en.wikipedia.org/wiki/BCH\_code

<sup>9.</sup> https://en.wikipedia.org/wiki/Reed%E2%80%93Solomon\_error\_correction