Théorie et codage de l'information

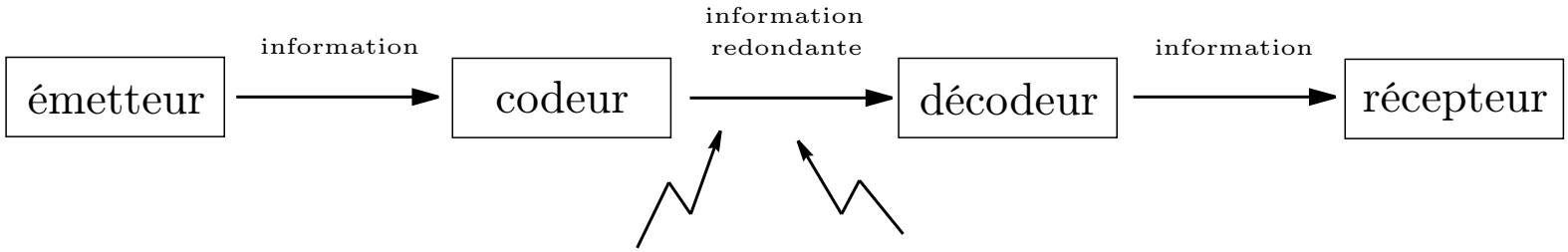
Canaux discrets

- Chapitre 4 -

Codage de canal

Motivations

Dans un système réel, le message reçu par le destinataire peut différer de celui qui a été émis par la source en raison de perturbations. On parle de *canal bruyant*.



Le codage de canal vise à introduire de la redondance dans le message

$\longrightarrow$ compenser l'érosion de l'information due au canal.

Modèles de canal discret

Modèle général

Un canal discret est un système stochastique acceptant en entrée des suites de symboles définies sur un alphabet $\mathcal{X}$, et émettant en sortie des suites de symboles définies sur un alphabet de sortie $\mathcal{Y}$.

Entrées et sorties sont liées par un modèle probabiliste :

$$P(Y_1 = y_1, \ldots, Y_m = y_m | X_1 = x_1, \ldots, X_n = x_n)$$

▷ **modèle trop général pour donner lieu à des développements simples**

Modèles de canal discret

Propriétés

Par souci de simplification, on fait des hypothèses sur le modèle de canal.

Propriété 1 (Canal causal). *Un canal est dit causal si*

$$P(Y_1 = y_1, \ldots, Y_m = y_m | X_1 = x_1, \ldots, X_n = x_n) = P(Y_1 = y_1, \ldots, Y_m = y_m | X_1 = x_1, \ldots, X_m = x_m)$$

quels que soient m et n tels que $m \leq n$.

Conséquence. En sommant les 2 membres de l'égalité sur $Y_1, \ldots, Y_{m-1}$, on vérifie

$$P(Y_m = y_m | X_1 = x_1, \ldots, X_n = x_n) = P(Y_m = y_m | X_1 = x_1, \ldots, X_m = x_m)$$

$\longrightarrow$ toute sortie est indépendante des entrées futures

Modèles de canal discret

Propriétés

On peut être amené à faire l'hypothèse suivante sur le comportement du canal.

Propriété 2 (Canal causal sans mémoire). *On dit qu'un canal causal est sans mémoire si, pour tout $k \geq 2$, on a :*

$$P(Y_k = y_k | X_1 = x_1, \ldots, X_k = x_k, Y_1 = y_1, \ldots, Y_{k-1} = y_{k-1}) \\ = P(Y_k = y_k | X_k = x_k).$$

Conséquence. La loi conditionnelle gouvernant le comportement du canal est entièrement déterminée par les lois conditionnelles instantanées :

$$P(Y_1 = y_1, \ldots, Y_m = y_m | X_1 = x_1, \ldots, X_n = x_n) = \prod_{k=1}^{m} P(Y_k = y_k | X_k = x_k).$$

$\longrightarrow P(Y_k = y_k | X_k = x_k)$ dépend éventuellement du temps

Modèles de canal discret

Propriétés

En remarquant que $P(Y_k = y_k | X_k = x_k)$ peut éventuellement dépendre du temps k, on est amené à introduire la propriété suivante.

Propriété 3 (Canal sans mémoire stationnaire). *On dit d'un canal sans mémoire qu'il est stationnaire si, quel que soit $k \geq 1$, on a :*

$$P(Y_k = y_k | X_k = x_k) = P(Y = y_k | X = x_k).$$

Notation. On note $(\mathcal{X}, \mathcal{Y}, \Pi)$ un canal discret sans mémoire, où Π est la matrice de transition définie par :

$$\Pi(i, j) = P(Y = y_j | X = x_i)$$

Modèles de canal discret

Notions de canal symétrique

Un canal est dit *symétrique* si les lignes de sa matrice de transition sont formées des mêmes éléments à l'ordre près, tout comme ses colonnes.

Exemples. Les matrices de transition suivantes sont celles de canaux symétriques.

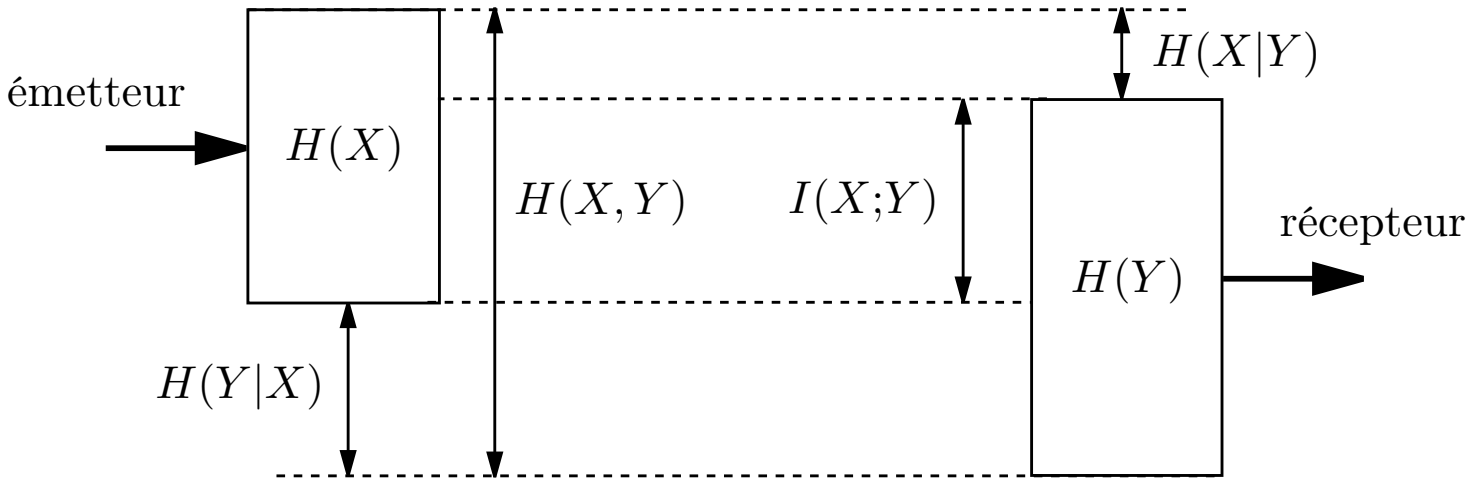
$$\Pi = \begin{pmatrix} p & q & 1-p-q \\ q & 1-p-q & p \\ 1-p-q & p & q \end{pmatrix},$$

$$\Pi = \begin{pmatrix} p & 1-p-q & q \\ q & 1-p-q & p \end{pmatrix},$$

où p et q sont des éléments de l'intervalle $[0, 1]$.

Capacité d'un canal sans mémoire

Présentation intuitive



$H(X)$ est la quantité d'information transmise par un canal sans bruit

$H(X|Y)$ est l'information requise pour supprimer l'ambiguïté sur l'entrée

$I(X;Y)$ est la quantité d'information transmise par le canal bruité

Capacité d'un canal sans mémoire

Définition

Définition 1. *On définit la capacité en information par symbole d'un canal par :*

$$C \triangleq \max_{P(X=x)} I(X;Y).$$

Précaution. On vérifie que $I(X,Y)$ est une fonction concave de la loi de X. En notant $f(x) = -x \log x$, on note qu'il s'agit d'une somme de fonctions concaves :

$$\begin{aligned} I(X;Y) &= \sum_i \sum_j p(i,j) \log \frac{p(i,j)}{p(i)\,p(j)} \\ &= \sum_i \sum_j p_i\, p_i(j) \log \frac{p_i(j)}{\sum_i p_i\, p_i(j)} \\ &= \sum_i p_i \left(\sum_j p_i(j) \log p_i(j) \right) + \sum_j f\left(\sum_i p_i p_i(j) \right). \end{aligned}$$

Capacité d'un canal sans mémoire

Calculs de capacités

Dans le cas général, le calcul direct de la capacité d'un canal s'avère compliqué. Toutefois, dans le cas d'un canal symétrique, le calcul s'effectue aisément.

Théorème 1. *La capacité d'un canal symétrique* $(\mathcal{X}, \mathcal{Y}, \Pi)$ *est égale à* $I(X;Y)$ *dans le cas où l'entrée* X *suit une loi uniforme.*

Démonstration. L'entropie $H(Y|X = x_i) = -\sum_j p_i(j) \log p_i(j)$ est indépendante de i, les lignes i de Π étant formées des mêmes éléments : $H(Y|X)$ est donc indépendant de la loi de X.

On vérifie aisément que Y suit une loi uniforme si celle de X l'est. En effet :

$$p_j = \sum_i p_i \, p_i(j) = \frac{1}{q} \sum_i p_i(j)$$

est indépendant de j car les colonnes de Π sont constituées des mêmes termes. □

Calculs de capacités

Exemples

Canal binaire sans bruit. Ce canal reproduit en sortie le symbole d'entrée. En conséquence, on a $I(X;Y) = H(X)$ car $H(X|Y) = 0$.

$$C = 1 \text{ Sh/symb}$$

Canal binaire en dysfonctionnement. Ce canal reproduit en sortie toujours le même symbole, indépendamment de l'entrée. En conséquence, l'information mutuelle $I(X;Y)$ est nulle puisque $H(Y) = H(Y|X) = 0$.

$$C = 0 \text{ Sh/symb}$$

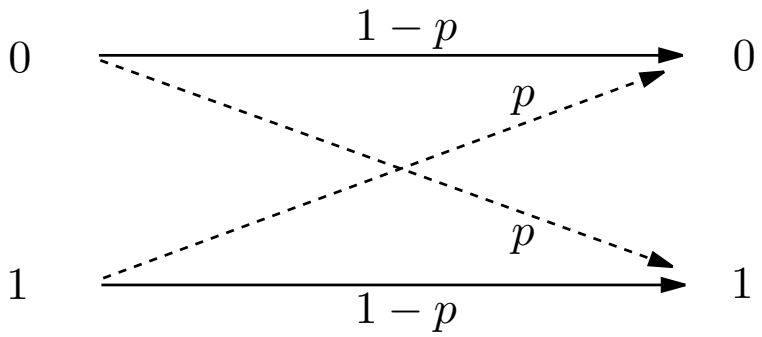
Calculs de capacités

Exemple du canal binaire symétrique

Le canal binaire symétrique est l'exemple le plus simple de canal bruyant. Sa matrice de transition est donnée par

$$\Pi = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

que l'on représente schématiquement ainsi



Calculs de capacités

Exemple du canal binaire symétrique

Afin d'évaluer la capacité en information de ce canal, calculons préalablement l'information mutuelle moyenne $I(X;Y)$:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - P(X=0)H(Y|X=0) - P(X=1)H(Y|X=1). \end{aligned}$$

Or, un calcul simple permet de montrer que $H(Y|X=x) = H_2(p)$, avec $x \in \{0,1\}$, ce qui entraîne que :

$$I(X;Y) = H(Y) - H_2(p) \leq \log 2 - H_2(p).$$

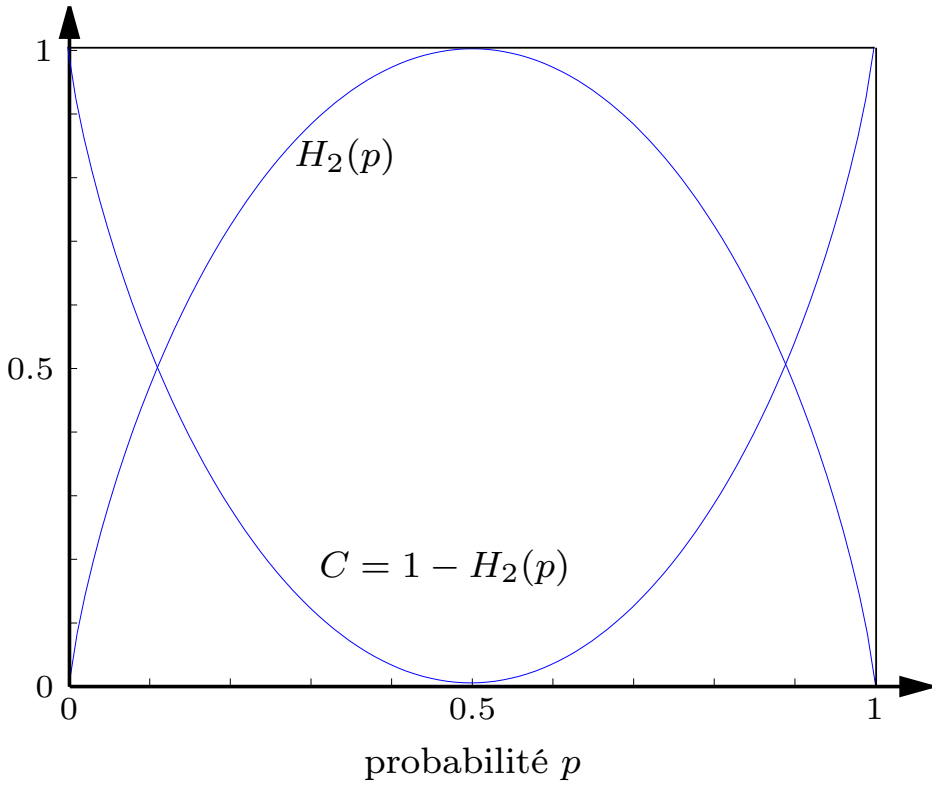
En conséquence, la capacité d'un canal binaire symétrique est donnée par :

$$C = 1 - H_2(p) \text{ Sh/symb}$$

Calculs de capacités

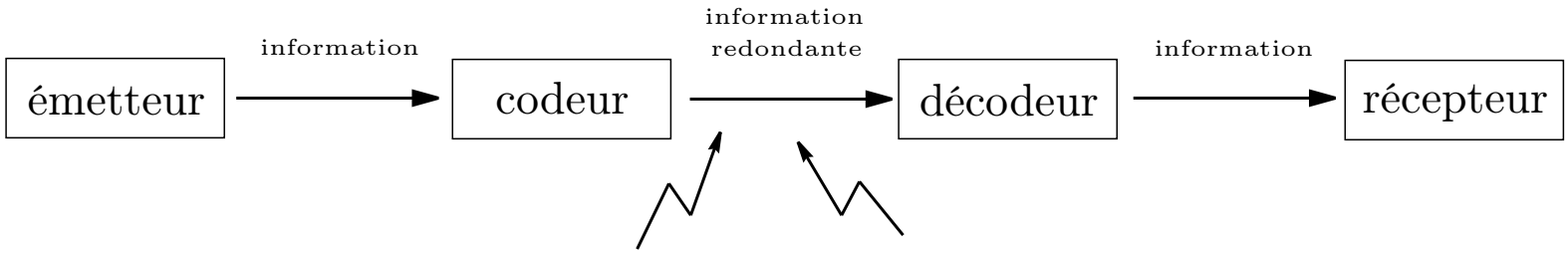
Exemple du canal binaire symétrique

La capacité en information d'un canal binaire symétrique a pour allure :



Codage de canal

Définitions préalables



Afin de détecter et/ou corriger les erreurs transmises, il est nécessaire d'ajouter des *symboles de contrôle* selon une règle $\mathcal{C}$, appelée *codage*.

▷ **le décodeur vérifie si la séquence reçue respecte $\mathcal{C}$**

Usage de redondance. On utilise des blocs de n symboles afin de transmettre k symboles d'information, avec $k < n$. Chaque bloc de longueur n est dit *mot du code*.

Codage de canal

Définitions préalables

Définition 2. *Soit $\mathcal{A} = \{a_1, \ldots, a_q\}$ un ensemble fini dit alphabet du code. Soit $\mathcal{A}^n$ l'ensemble de toutes les chaînes de longueur n sur $\mathcal{A}$. Tout sous-ensemble non vide $\mathcal{C}$ de $\mathcal{A}^n$ est dit code en bloc q-aire. Chaque chaîne dans $\mathcal{C}$ sera dite mot du code.*

Définition 3. *Si $\mathcal{C} \subset \mathcal{A}^n$ contient M mots du code, on dit alors que $\mathcal{C}$ est de longueur n et de taille M. On parle alors de (n, M)-code.*

Exemple. Le code $\mathcal{C}$ suivant est un (5,4)-code :

$$\mathcal{C} = \{11100, 01001, 10010, 00111\}$$

Codage de canal

Erreurs de détection

Définition 4. *On parle d'erreur de détection lorsque le mot $c \in \mathcal{C}$ a été émis et que l'on reçoit le mot d, avec $c \neq d$.*

Toute erreur de transmission ne peut être détectée que si le mot reçu n'est pas un autre mot du code. En conséquence, si $c \in \mathcal{C}$ est émis, on a :

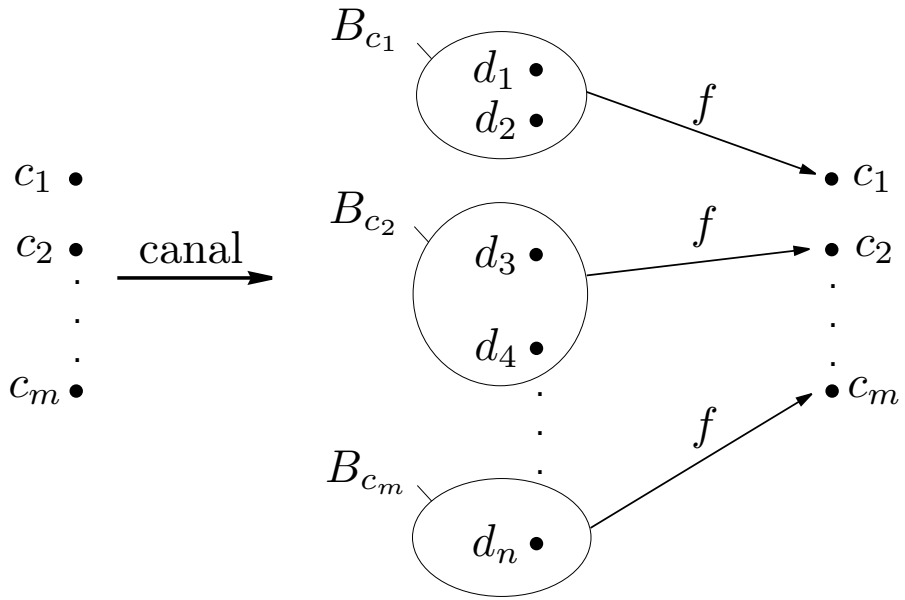
$$P(\text{erreur non détectée} \mid c \text{ est émis}) = \sum_{\substack{d \in \mathcal{C} \\ d \neq c}} P(d|c).$$

$$P(\text{erreur non détectée}) = \sum_{c \in \mathcal{C}} \sum_{\substack{d \in \mathcal{C} \\ d \neq c}} P(d|c)\ P(c).$$

Codage de canal

Erreurs de décision

Avant de parler d'erreur de décision, il faut introduire les *schémas de décision*.



Définition 5. *Un schéma de décision est une fonction partielle f de l'ensemble des chaînes reçues vers l'ensemble des mots du code.*

Codage de canal

Erreurs de décision

Définition 6. *On parle d'erreur de décision lorsque le mot $c \in \mathcal{C}$ a été émis, que d a été reçu et qu'il a été décodé par $f(d) \neq c$.*

La probabilité d'une erreur de décision sachant que c a été émis est définie par

$$P(\text{erreur de décodage} \mid c \text{ est émis}) = \sum_{\substack{d \in \mathcal{C} \\ d \notin f^{-1}(c)}} P(d|c),$$

et la probabilité d'une erreur de décodage est

$$P(\text{erreur de décodage}) = \sum_{c \in \mathcal{C}} P(\text{erreur de décodage} \mid c \text{ est émis})\, P(c).$$

Codage de canal

Second théorème de Shannon

En rappelant que la capacité d'un canal est la quantité maximale d'information qu'il peut transmettre, on peut énoncer le second théorème de Shannon.

Théorème 2. *Soit un canal discret et sans mémoire de capacité C. Pour tout nombre positif C' inférieur à C, il existe une suite $\mathcal{C}_k$ de codes r-aires associés aux schémas de décision f_k ayant les propriétés suivantes :*

- *$\mathcal{C}_k$ est un code de longueur k et de taux de transmission supérieur ou égal à C' ;*
- *la probabilité max. d'erreur de décodage tend vers 0 lorsque k tend vers l'infini :*

$$\lim_{k \to +\infty} P_{max}(k) = 0,$$

avec $P_{max}(k) = \max_{c \in \mathcal{C}_k} P(\text{erreur de décodage} \mid c \text{ est émis})$.

▷ **aucune preuve constructive de ce théorème n'est connue à ce jour**

Codage de canal

Définition d'une métrique sur $\mathcal{C}$

Les codes détecteur/correcteur reposent sur une structure algébrique/géométrique.

Définition 7. *Soient x et y des chaînes de même longueur sur le même alphabet. La distance de Hamming $d_{Ham}(x, y)$ entre x et y est par définition le nombre de positions pour lesquelles x et y diffèrent.*

Exemple : $d_{Ham}(10112, 20110) = 2$

Théorème 3. *L'espace $(\mathcal{A}^n, d_{Ham})$ est un espace métrique, autrement dit la distance de Hamming vérifie les propriétés suivantes pour tout x, y et z de $\mathcal{A}^n$:*

1. $d_{Ham}(x, y) = 0 \iff x = y$
2. $d_{Ham}(x, y) = d_{Ham}(y, x)$
3. $d_{Ham}(x, y) \leq d_{Ham}(x, z) + d_{Ham}(z, y)$.

Codage de canal

Décodage par maximum de vraisemblance

On considère un canal binaire symétrique caractérisé par :

$$P(1|0) = P(0|1) = p \quad P(0|0) = P(1|1) = 1 - p \quad \text{avec :} \quad p < 0.5.$$

En notant c le mot envoyé et d le mot reçu, $d_{Ham}(c, d)$ correspond au nombre d'erreurs de symboles dues au canal. En conséquence :

$$P(d|c) = p^{d_{Ham}(c,d)} \, (1-p)^{n - d_{Ham}(c,d)}.$$

▷ **$P(d|c)$ est maximale lorsque $d_{Ham}(c, d)$ est minimale**

Théorème 4. *Pour le canal binaire symétrique avec une probabilité d'erreur $p < 0.5$, la règle de décodage par maximum de vraisemblance est équivalente à la règle de décodage par minimum de distance.*

Codage de canal

Distance minimale d'un code

Dans l'idée de pouvoir utiliser le décodage par minimum de distance, on est amené à poser les définitions suivantes.

Définition 8. *La distance minimale du code $\mathcal{C}$ est définie par*

$$d(\mathcal{C}) = \min_{x,y \in \mathcal{C}} d_{Ham}(x,y).$$

Définition 9. *On parle de (n, M, d)-code pour évoquer un code de longueur n, de taille M et de distance minimale d.*

Exemple : le code binaire $\mathcal{C} = \{11100, 01001, 10010, 00111\}$ est un (5,4,3)-code.

Codage de canal

Codes t-détecteurs d'erreurs

On définit un code t-détecteur d'erreurs ainsi.

Définition 10. *Un code $\mathcal{C}$ est t-détecteur d'erreurs si, dès qu'au plus $t \geq 1$ erreurs se produisent dans un mot du code, le mot résultant n'est pas un mot du code.*

Le code $\mathcal{C}$ est dit exactement t-détecteur lorsqu'il est t-détecteur mais pas $(t+1)$-détecteur.

On démontre aisément le résultat suivant :

Théorème 5. *Un code $\mathcal{C}$ est exactement t-détecteurs d'erreurs si et seulement si*

$$d(\mathcal{C}) = t + 1.$$

Codage de canal

Codes t-correcteurs d'erreurs

On définit un code t-correcteur d'erreurs ainsi.

Définition 11. *Un code $\mathcal{C}$ est t-correcteur si le décodage par minimum de distance peut corriger les erreurs de taille inférieure ou égale à t dans tout mot du code.*

Un code est dit exactement t-correcteur s'il est t-correcteur mais pas $(t+1)$-correcteur. Ceci signifie que toute erreur de taille t est corrigées mais qu'il existe au moins une erreur de taille $t+1$ qui est décodée incorrectement.

On démontre le résultat suivant :

Théorème 6. *Un code $\mathcal{C}$ est exactement t-correcteur d'erreurs si et seulement si*

$$d(\mathcal{C}) = 2t+1 \quad ou \quad d(\mathcal{C}) = 2t+2$$