Université de Paris Cité
Master 1 MIC

Introduction aux codes polaires

Jonas Dos Santos, Nathan Fauvelle-Aymar

Mai 2025

Table des matières

1 Introduction

Dans son article fondateur de 1948 [12], Claude Shannon introduit les concepts clefs de la théorie de l'information tels que l'entropie, l'information mutuelle ou encore la capacité d'un canal. De son travail, on retient le plus souvent ses théorèmes fondamentaux [1]. Parmi eux, on retrouve notamment le théorème de codage de canal bruité, qui établit qu'on peut transmettre de l'information de manière fiable à travers un canal soumis à des perturbations, du moment que le débit d'information reste inférieur à la capacité du canal.

Cette démonstration est particulièrement remarquable par son originalité : au lieu de construire un code précis, Shannon adopte une approche entièrement nouvelle pour l'époque, fondée sur ce qu'on appelle aujourd'hui la méthode probabiliste. En sélectionnant un code aléatoirement dans un certain ensemble, Shannon montre que la probabilité qu'il soit tel qu'on le souhaite est strictement positive.

La méthode probabiliste sera plus tard popularisée par Paul Erdős, notamment célèbre pour avoir transformé cette technique en un outil central de la combinatoire moderne [2]. En 1998, Martin Aigner et Günter M. Ziegler évrivaient dans la préface de *Raisonnements Divins*[1] :

« *Paul Erdős aimait parler du Grand Livre, dans lequel Dieu inscrit les preuves parfaites des théorèmes mathématiques, suivant ainsi le dicton de G. H. Hardy : il n'y a pas de lieu permanent pour les mathématiques laides. Erdős disait aussi que l'on n'a pas besoin de croire en Dieu, mais qu'en tant que mathématicien, on doit croire au Grand Livre.* »

Ce livre, en hommage à Paul Erdős, regroupe, comme l'indique son sous-titre *Quelques démonstrations mathématiques particulièrement élégantes*, et on y trouve d'ailleurs bien la démonstration de Shannon – preuve, s'il en fallait, de son importance et élégance.

La démonstration de Shannon, bien que révolutionnaire [3], présente une limite importante : elle est non constructive. Autrement dit, si elle prouve l'existence de codes capables d'atteindre la capacité d'un canal, elle ne fournit aucun moyen explicite de les construire. Cette situation a longtemps laissé en suspens une question

1. Voir par exemple, Chambert-Loir dans [6]
2. Pour appuyer ce propos, voir notamment la préface d'Alon et Spencer dans [2]
3. Le mot peut paraître un peu fort mais le caractère "novateur" est certain : voir par exemple MacKay [10] p. 164 et Cover et Thomas[7] p.199

fondamentale : peut-on concevoir un code explicite et un algorithme de décodage efficaces qui atteignent cette fameuse capacité ? Il a fallu attendre plusieurs décennies pour voir apparaître des candidats sérieux, en voici une liste non exhaustive :

–C. Berrou, A. Glavieux et P. Thitimajshima présentent en 1993 les Turbo codes lors d'une conférence à Genève [5]. Ces nouveaux codes sont alors les premiers connus à approcher la limite de Shannon tout en ayant une complexité maximale raisonnable.

–En 1996, on "redécouvre" les codes LDPC (Low-density parity-check) par le travail de D. JC. MacKay et R. M. Neal [11] qui montrent que ces codes atteignent aussi bien la limite de Shannon que les Turbo Codes. Initialement proposés par R. Gallager dans sa thèse au MIT en 1960 [8], ces codes avaient été rarement étudiés ni utilisés car ils étaient trop exigeants en termes de calcul.

–En 2009, un tournant majeur s'opère : Erdal Arıkan introduit les codes polaires [3]. On est alors capable de construire explicitement des codes atteigant la capacité de certains canaux, tout en bénéficiant d'une complexité maximale de codage et de décodage plus que satisfaisante ($O(n \cdot \log n)$).
Initialement conçus pour être utilisé comme un "code interne" d'un schéma plus complexe, il s'est avéré que les codes polaires étaient suffisaments fiables pour être utilisés seuls [4].

Nous allons dans ce travail nous intéresser aux codes polaires en proposant une introduction à ce sujet. Comme nous allons le voir, ces codes reposent sur la polarisation de canaux, ce processus nous permettra de transformer un ensemble de canaux identiques en une combinaison de canaux soit parfaitement fiables soit complétement défaillants. À la lumière de cette construction, nous serons en mesure de démontrer que l'utilisation de ces canaux permet d'atteindre la borne de Shannon.

4. Sur ce sujet, voir l'article d'Arıkan, *On the Origin of Polar Coding*[4]

2 Préliminaires et notations

Définition 2.0.1.

Soit $y^N = (y_1, ..., y_N)$ un N-uplet d'un ensemble quelconque, on donne la notation y_i^j au vecteur $(y_i, y_{i+1}, ...y_{j-1}, y_j)$.

Définition 2.0.2.

*Un **canal binaire discret sans mémoire** est une application*

$$W : \mathcal{X} \to \mathcal{Y}, \text{ avec } \mathcal{X} = \{0, 1\}$$

qui, à chaque symbole $x \in \mathcal{X}$ en entrée associe un symbole $y \in \mathcal{Y}$ en sortie avec la probabilité $W(y \mid x)$ indépendamment des symboles précédents (absence de mémoire).

On notera $W^N : \mathcal{X}^N \to \mathcal{Y}^N$ avec $W^N(y^N \mid x^N) = \Pi_{i \leqslant N} W(y_i \mid x_i)$ le canal correspondant à l'utilisation en parralèle de N canaux W identiques.

Remarque 2.0.1.

*Nous adopterons la notation **B-DMC** (pour Binary Discrete Memoryless Channel) conformémant à l'usage courant dans la littérature anglophone.*

Définition 2.0.3.

*Soit W un B-DMC, il est dit **symétrique** s'il existe une permutation $\pi : \mathcal{Y} \to \mathcal{Y}$ telle que $W(y \mid 0) = W(\pi(y) \mid 1)$.*

Définition 2.0.4.

*Soit W un B-DMC, on définit alors la **capacité symétrique** de W :*

$$I(W) = \sum_{y \in \mathcal{Y}} \sum_{x \in \{0,1\}} \frac{1}{2} W(y \mid x) \log_2 \left(\frac{W(y \mid x)}{\frac{1}{2}W(y \mid 0) + \frac{1}{2}W(y \mid 1)} \right)$$

*On appelera souvent cette quantité le **taux** du canal W.*

Remarque 2.0.2. *Cette capacité représente l'information mutuelle entre $\mathcal{X}$ et $\mathcal{Y}$ lorsque $\mathcal{X} \sim \mathcal{U}(\frac{1}{2})$. Il y a égalité avec la capacité $I(W) = \max_{P_\mathcal{X}} I(\mathcal{X}, \mathcal{Y})$ de Shannon lorsque W est un canal symétrique, d'où le nom "capacité symétrique".*

Définition 2.0.5.

Soit W un B-DMC, on définit le paramètre de Bhattacharyya de W :

$$Z(W) = \sum_{y \in \mathcal{Y}} \sqrt{W(y \mid 0) W(y \mid 1)}$$

*On appelera cette quantité la **fiabilité** du canal W.*

Remarque 2.0.3. *La **fiabilité** d'un canal est une estimation (par le haut) de la probabilité de se tromper si l'on essaie de deviner le bit transmis après* 1 *seul usage du canal* W *par maximum de vraissemblance.*

Proposition 2.0.1.

SOit W *un B-DMC, on a les inégalités suivantes :*

$$I(W) \geqslant log_2(\frac{2}{1+Z(W)})$$

et

$$I(W) \leqslant \sqrt{1-Z(W)^2}$$

Démonstration. La démonstration de ces inégalités est admise. □

3 Idée générale et premières constructions

Comme expliqué dans l'introduction, la polarisation d'Arıkan est la transformation progressive d'un ensemble de canaux identiques vers un ensemble n'ayant que des canaux *synthétiques* aux 2 extrêmes : certains quasiments parfaits et d'autres quasiments inutilisables.
Dans la littérature sur le sujet, la polarisation de canaux est souvent énoncée comme un double processus :

- La combinaison de canaux : On combine plusieurs copies identiques et indépendantes d'un canal W pour former un canal vectoriel W_N
- La division des canaux : Le canal vectoriel W_N obtenu est divisé en N sous-canaux $W_N^{(i)}$ destinés transporter 1 bit d'information chacun

C'est lors de la seconde phase de ce processus que l'on distingue les canaux utiles (ceux qu'on a améliorés) et les canaux inutiles (ceux qu'on a dégradés).

Nous avons essayé de présenter ici le même double processus en nous inspirant de la manière dont Emre Telatar le présente [13].

3.1 Construction du canal vectoriel – Channel Combining

Soit W un B-DMC, on commence la récurrence au rang $n = 0$, on note alors $W_1 = W$. A la prochaine itération ($n = 1$), on duplique puis *polarise* le canal W_1, on obtient alors $W_2 : \mathcal{X}^2 \to \mathcal{Y}^2$ le canal défini par :

$$W_2(y_1^2 \mid x_1^2) = W^2(y_1^2 \mid u_1 \oplus u_2, u_2) = W(y_1 \mid u_1 \oplus u_2)W(y_2 \mid u_2)$$

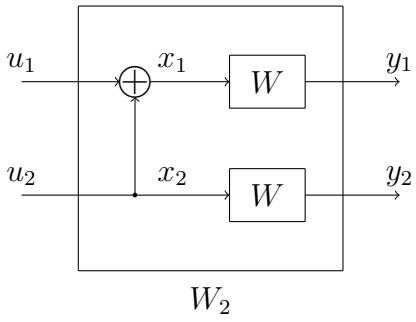

FIGURE 1 – Au rang $n = 1$ du processus de polarisation, le canal W_2

Au rang $n = 2$, en dupliquant et *polarisant* notre canal W_2, on obtient le canal $W_4 : \mathcal{X}^4 \to \mathcal{Y}^4$ défini par :

$$W_4(y_1^4 \mid u_1^4) = W_2(y_1^2 \mid u_1 \oplus u_2,\, u_3 \oplus u_4) \cdot W_2(y_3^4 \mid u_2,\, u_4)$$

Nous introduisons dans la Fig.2, qui représente notre nouveau canal vectoriel, une

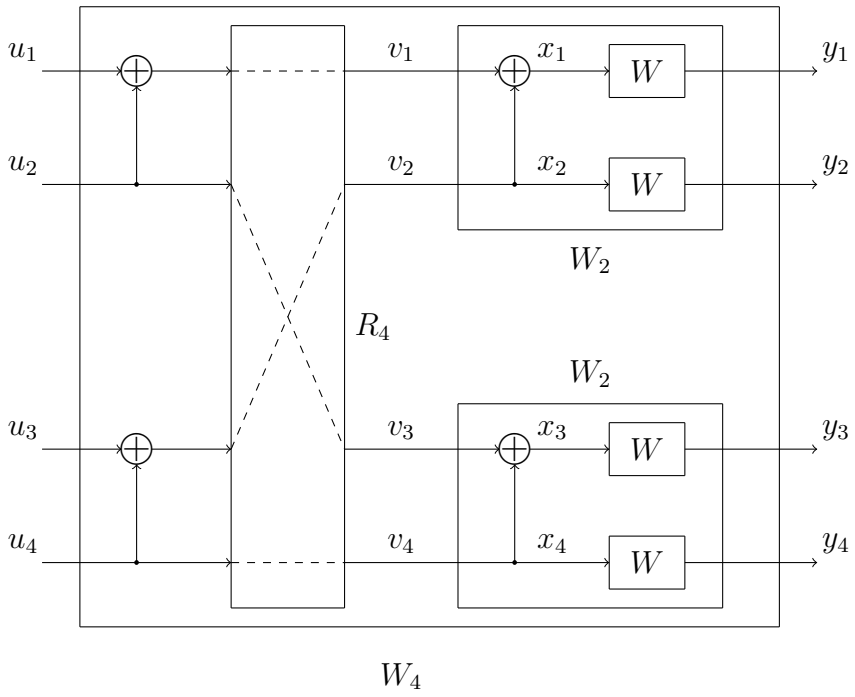

FIGURE 2 – Au rang $n = 2$ du processus de polarisation, le canal W_4

nouvelle notation : $R_4 : \mathcal{X}^4 \to \mathcal{X}^4$ qui, à une entrée $u_1^4 = (t_1, t_2, t_3, t_4)$ lui associe $v_1^4 = (t_1, t_3, t_2, t_4)$.

Posons :

$$\begin{aligned} \Phi_4 : \quad & W_4 \longmapsto W^4 \\ & u_1^4 \longrightarrow x_1^4 \end{aligned}$$

Cette application est simplement définie par :

$$\Phi_4(u_1^4) = R_4(u_1^4) \cdot G_4 = x_1^4 \qquad \text{où} \quad G_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

On a alors la relation :

$$W_4(y_1^4 \mid u_1^4) = W^4(y_1^4 \mid u_1^4 \cdot \Phi_4)$$

L'introduction et l'utilisation de R_4 et G_4 prennent sens lorsqu'on étudie le cas général de cette récurrence.

Dans le cas général (voir fig. 3), on construit le canal W_N en dupliquant et polarisant le canal $W_{N/2}$. On définit R_N comme étant la permutation par bits inversés (bit-reversal permutaiton).
Cette permutation est bien évidemment une application linéaire de $\mathcal{X}^2 \to \mathcal{X}^2$. On peut donc facilement affirmer que l'application Φ_N (forme générale de l'application Φ_4 définie plus haut) est linéaire. On a donc la relation :

$$W_N(y_1^N \mid u_1^N) = W^N(y_1^N \mid u_1^N \cdot \Phi_N) \tag{3.1.1}$$

3.2 Séparation en canaux *synthétiques* et études de leur légitimité – Channel splitting

Maintenant que nous avons vu la forme général de combinaison de canaux pour créer un canal vectoriel W_N, nous allons *séparer* ce dernier pour retrouver N canaux *synthétiques*.

Un premier problème se pose : il n'est pas immédiat que ces canaux *synthétiques* existent bien et soient licites, regardons de plus près.
La canal $W^- : U_1 \to Y_1 Y_2$ ne pose pas de problème, l'entrée U_1 est contrôlée par l'émetteur, la sortie (Y_1, Y_2) est bien observée par le récepteur.

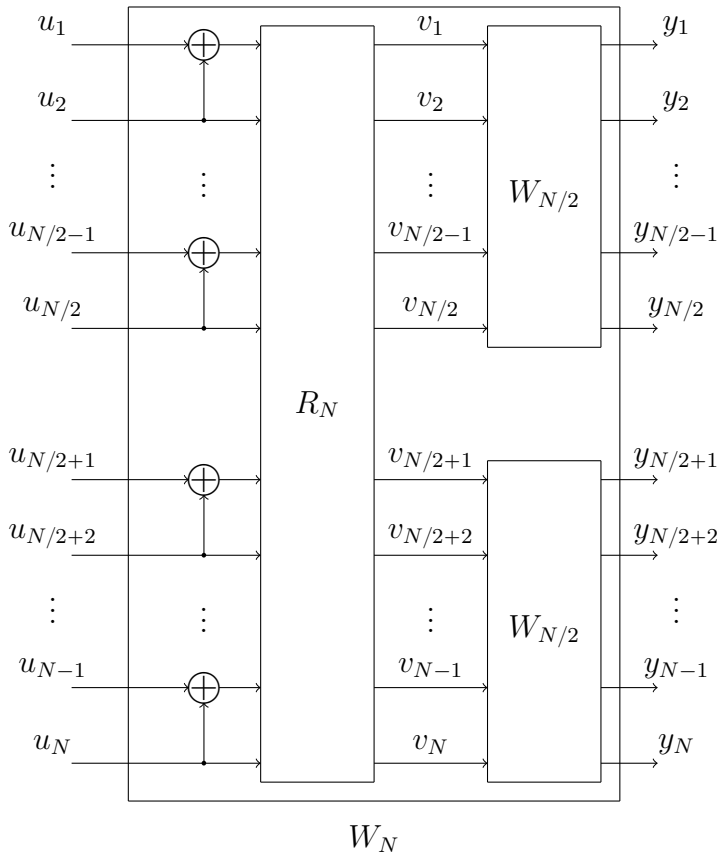

FIGURE 3 – Schéma de la construction du canal vectoriel W_N à partir de $W_{N/2}$

Quant à lui, le canal W^+ pose problème, l'entrée U_2 est bien contrôlée par l'émetteur mais la sortie (Y_1, Y_2, U_1) n'est pas observée par le récepteur (qui n'a pas accès à U_1). Mais, en pratique, le récepteur peut connaître U_1 (pas parfaitement) en essayant de l'estimer à partir de la sortie de W^-.

Imaginons maintenant 2 scénarios :

— Dans le premier, un **génie** nous souffle la vraie valeur $\hat{U}_1$ de U_1, on peut alors l'utiliser pour *décoder* U_2 en $\hat{U}_2 = \phi_2(Y_1 Y_2 \hat{U}_1)$ (ϕ est une fonction d'estimation quelconque). Nous avons donc une estimation de U_2 grâce au génie.

— Dans le second scénario, dit **implémentable**, on utilise une fonction d'estimation arbitraire pour trouver $\widetilde{U}_1 = \xi(Y_1, Y_2)$. On utilise maintenant notre première estimation pour estimer U_2 en $\widetilde{U}_2 = \phi_2(Y_1, Y_2, \widetilde{U}_1)$.

Intéressons nous maintenant aux évènements d'erreurs :

$$\mathcal{E}_{genie} = \{\hat{U}_2 \neq U_2\}, \qquad \mathcal{E}_{impl} = \{\widetilde{U}_2 \neq U_2\}$$

Si le récepteur aidé par le génie se trompe on est certain que l'erreur vient de la fonction ϕ.

Si maintenant le récepteur implémentable se trompe c'est que l'une des 2 fonction d'estimation (ξ ou ϕ) s'est trompée (ou les deux).
Or, si l'on construit ξ pour que $\hat{U}_1 = \widetilde{U}_1$ (c'est à dire que l'estimation soit la même en pratique que celle du génie), on s'assure que les erreurs lors de l'estimation du récepteur implémentable se produisent à cause de la fonction ϕ et uniquement celle-ci. Au final, on a :

$$\mathcal{E}_{genie} = \mathcal{E}_{impl.}$$

En conclusion, bien que U_1 ne soit pas directement observable par le récepteur, celui-ci peut l'estimer à partir des sorties du canal W^-, ce qui justifie la structure du canal W^+. Les performances du décodeur implémentable et du décodeur aidé par le génie ne sont pas exactement identiques mais les performances du second nous donne une borne pour les performances du premier, ce qui nous permet bien de considérer W^+ comme un canal authentique [5]

Dans le cas général, l'idée est de créer N canaux *synthétiques* $W_N^{(i)} : \mathcal{X} \to \mathcal{Y}^N \times \mathcal{X}^{i-1}$ de probabilité de transition :

$$W_N^{(i)}(y_1^N, u_1^{i-1} \mid u_i) = \sum_{u_{i+1}^N \in \mathcal{X}^{N-i}} \frac{1}{2^{N-i}} \cdot W_N(y_1^N \mid u_1^N) \tag{3.2.1}$$

Puisque nous avons supposé les bits en entrée *iid*, il est licite de voir apparaître une moyenne ici.

4 Propriétés du Channel splitting

4.1 Récursivité de l'opération

L'objectif de cette section est de démontrer que l'opération de *channel splitting* définie en 3.2.1 peut être effectuée *pas à pas*, c'est à dire récursivement.

Définition 4.1.1. *On dit que 2 canaux $W' : \mathcal{X} \to \widetilde{\mathcal{Y}}$ et $W'' : \mathcal{X} \to \widetilde{\mathcal{Y}} \times \mathcal{X}$ sont obtenus par une opération élémentaire à partir d'un canal $W : \mathcal{X} \to \mathcal{Y}$ s'il existe une fonction $f : \mathcal{Y}^2 \to \widetilde{\mathcal{Y}}$ telle que :*

$$W'(f(y_1^2) \mid u_1) = \sum_{u_2'} \frac{1}{2} W(y_1 \mid u_1 \oplus u_2') W(y_2 \mid u_2')$$

5. Le « genie-aided receiver » semble avoir été introduit par Jacobs et Berlekamp dans leur article *A Lower Bound to the Distribution of Computation for Sequential Decoding* [9]. Nous n'avons cependant pas eu accès à ce document.

$$W''(f(y_1^2), u_1 \mid u_2) = \frac{1}{2} W(y_1 \mid u_1 \oplus u_2) W(y_2 \mid u_2)$$

pour tout $u_1^2 \in \mathcal{X}$ *et* $y_1^2 \in \mathcal{Y}$.

Proposition 4.1.1.

L'opération qui transforme (W, W) *en* $(W_2^{(1)}, W_2^{(2)})$ *est une opération élémentaire*

Démonstration. On prend la fonction $f = id$. □

On va maintenant voir que cette opération est bien *élémentaire* à chaque étape de la récursivité.

Proposition 4.1.2.

L'opération qui transforme $(W_N^{(i)}, W_N^{(i)})$ *en* $(W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ *est une opération élémentaire. C'est à dire qu'on a :*

$$W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-2} \mid u_{2i-1}) =$$

$$\sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} \mid u_{2i})$$

et

$$W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1} \mid u_{2i}) =$$

$$\frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} \mid u_{2i})$$

où les notations $u_{m,o}^n$ *et* $u_{m,e}^n$ *désignent les bits d'indices impairs (odd) et pairs (even) de* u_m^n.

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], *p.*21.
Première égalité :
On a :

$$W_{2N}^{(2i-1)}(y_1^{2N}, u_1^{2i-2} \mid u_{2i-1}) = \sum_{u_{2i}^{2N}} \frac{1}{2^{2N-1}} W_{2N}(y_1^{2N} \mid u_1^{2N})$$

$$= \sum_{u_{2i,o}^{2N}, u_{2i,e}^{2N}} \frac{1}{2^{2N-1}} W_N(y_1^N \mid u_{2N}^{1,o} \oplus u_{1,e}^{2N}) W_N(y_{N+1}^{2N} \mid u_{1,e}^{2N})$$

$$= \sum_{u_{2i}} \frac{1}{2} \sum_{u_{2i+1,e}^{2N}} \frac{1}{2^{N-1}} W_N(y_{N+1}^{2N} \mid u_{1,e}^{2N}) \cdot \sum_{u_{2i+1,o}^{2N}} \frac{1}{2^{N-1}} W_N(y_1^N \mid u_{2N}^{1,o} \oplus u_{1,e}^{2N})$$

Or, en utilisant 3.2.1, la partie de droite se simplifie en :

$$W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i})$$

on réécrit alors :

$$= \sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}) \cdot \sum_{u_{2i+1,e}^{2N}} \frac{1}{2^{N-1}} W_N(y_{N+1}^{2N} \mid u_{1,e}^{2N})$$

et enfin, en utilisant encore une fois 3.2.1 :

$$= \sum_{u_{2i}} \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} \mid u_{2i})$$

Seconde égalité :
On a aisément :

$$W_{2N}^{(2i)}(y_1^{2N}, u_1^{2i-1} \mid u_{2i}) \quad = \quad \sum_{u_{2i+1}^{2N}} \frac{1}{2^{2N-1}} W_{2N}(y_1^{2N} \mid u_1^{2N})$$

$$= \frac{1}{2} \sum_{u_{2i+1,e}^{2N}} \frac{1}{2^{N-1}} W_N(y_{N+1}^{2N} \mid u_{1,e}^{2N}) \cdot \sum_{u_{2i+1,o}^{2N}} \frac{1}{2^{N-1}} W_N(y_1^N \mid u_{1,o}^{2N} \oplus u_{1,e}^{2N})$$

et, en utilisant 3.2.1 comme au dessus, on obtient :

$$= \frac{1}{2} W_N^{(i)}(y_1^N, u_{1,o}^{2i-2} \oplus u_{1,e}^{2i-2} \mid u_{2i-1} \oplus u_{2i}) \cdot W_N^{(i)}(y_{N+1}^{2N}, u_{1,e}^{2i-2} \mid u_{2i})$$

□

4.2 Evolution du taux et de la fiabilité des canaux

L'objectif de cette section est d'étudier comment évoluent le **taux** et la **fiabilité** des canaux lors d'une transformation par opération élémentaire.

Proposition 4.2.1.

Soit (W', W'') 2 canaux obtenus par une opération élémentaire sur (W, W). On a alors :

$$I(W') + I(W'') = 2 \cdot I(W)$$

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], *p.*21.
On reprend les notations de 4.1.1 et on définit les variables aléatoires $U_1, U_2, X_1, X_2, Y_1, Y_2, \widetilde{Y}$ telles que ; (U_1, U_2) est uniformément distribuée sur $\mathcal{X}^2$; $(X_1, X_2) = (U_1 \oplus U_2, U_2)$, $P_{Y_1,Y_2|X_1,X_2}(y_1^2 \mid x_1^{2)=W(y_1|x_1)\cdot W(y_2|x_2)}$ et $\widetilde{Y} = f(Y_1, Y_2)$.
On a maintenant :

$$W'(\widetilde{y} \mid u_1) = P_{\widetilde{Y}|U_1}(\widetilde{y} \mid u_1)$$

et

$$W''(\widetilde{y}, u_1 \mid u_2) = P_{\widetilde{Y},U_1|U_2}(\widetilde{y}, u_1 \mid u_2)$$

Par ce qui précède, on obtient :

$$I(W') = I(U_1; \widetilde{Y}) = I(U_1; Y_1, Y_2)$$

et

$$I(W'') = I(U_2; \widetilde{Y}, U_1) = I(U_2; Y_1, Y_2, U_1)$$

et, puisque U_1 et U_2 sont $i.i.d$ on a :

$$I(U_2; Y_1, Y_2, U_1) = I(U_2; Y_1, Y_2 \mid U_1)$$

et enfin, par la *règle de chaînage* et par la relation entre (U_1, U_2) et (X_1, X_2), on a :

$$I(W') + I(W'') = I(U_1, U_2; Y_1, Y_2) = I(X_1, X_2; Y_1, Y_2)$$

Or, on a :

$$I(X_1, X_2; Y_1, Y_2) = I(X_1; Y_1) + I(X_2, Y_2) \quad et \quad I(X_1, Y_1) = I(X_2, Y_2) = I(W)$$

d'où l'égalité voulue. □

Proposition 4.2.2.

On a aussi :

$$I(W') \leqslant I(W'')$$

avec égalité ssi $I(W) = 0$ *ou* $I(W) = 1$.

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], $p.21$.
On garde les mêmes notations que dans la démonstration précédente :
On remarque que :

$$I(U_2; Y_1, Y_2, U_1) = I(U_2; Y_2) + I(U_2; Y_1, U_1 \mid Y_2) = I(W) + I(U_2; Y_1, U_1 \mid Y_2)$$

autrement, on a :

$$I(U_2; Y_1, Y_2, U_1) \geqslant I(W)$$

donc :

$$I(U_2; Y_1, Y_2, U_1) \geqslant I(U_1; Y_1, Y_2)$$

et ainsi on a bien l'inégalité voulue.

On a égalité si $I(U_2; Y_1, U_1 \mid Y_2) = 0$, c'est à dire si :

$$P_{U_1,U_2,Y_1|Y_2}(u_1^2, y_1 \mid y_2) = P_{U_1,Y_1|Y_2}(u_1, y_1 \mid y_2) \cdot P_{U_2|Y_2}(u_2 \mid y_2)$$

pour tout (u_1^2, y_1^2) et pour $P_{Y_2}(y_2) > 0$, c'est à dire :

$$P_{Y_1,Y_2|U_1,U_2}(y_1^2 \mid u_1^2) \cdot P_{Y_2}(y_2) = P_{Y_1,Y_2|U_1}(y_1^2 \mid u_1) \cdot P_{Y_2|U_2}(y_2 \mid u_2)$$

pour tout (u_1^2, y_1^2). On réécrit en :

$$W(y_2 \mid u_2)\left[W(y_1 \mid u_1 \oplus u_2)P_{Y_2}(y_2) - P_{Y_1,Y_2}(y_1^2 \mid u_1)\right] = 0$$

On remplace alors $P_{Y_2}(y_2)$ par $\frac{1}{2}W(y_2 \mid u_2) + \frac{1}{2}W(y_2 \mid u_2 \oplus 1)$ et $P_{Y_1,Y_2|U_1}(y_1^2 \mid u_1)$ par $\frac{1}{2}W(y_1 \mid u_1 \oplus u_2)W(y_2 \mid u_2) + \frac{1}{2}W(y_1 \mid u_1 \oplus u_2 \oplus 1)W(y_2 \mid u_2 \oplus 1)$ et on obtient :

$$W(y_2 \mid u_2)W(y_2 \mid u_2 \oplus 1)\left[W(y_1 \mid u_1 \oplus u_2) - W(y_1 \mid u_1 \oplus u_2 \oplus 1)\right] = 0,$$

qui est équivalent à :

$$W(y_2 \mid 0)W(y_2 \mid 1)\left[W(y_1 \mid 0) - W(y_1 \mid 1)\right] = 0.$$

Si la partie droite est égale à 0, on a alors $W(y_1 \mid 0) = W(y_1 \mid 1)$ et ainsi $I(W) = 0$. Sinon, si la partie gauche est nulle, on a $(W(y_2 \mid 0)W(y_2 \mid 1) > 0$ pour tout y_2 et ainsi $I(W) = 1$. □

Remarque 4.2.1.

L'inégalité dans la proposition précédente justifie l'utilisation d'exposants[6] $+$ *et* $-$ *pour désigner respectivement les canaux* W'' *et* W'*, ces exposants représentent le gain (resp. la perte) de capacité par rapport au canal initial* W*. Formellement, nous avons :*

$$I(W^-) \leqslant I(W) \leqslant I(W^+)$$

Les propositions précédentes nous indiquent que le **taux** est préservé lors d'une transformation par opération élémentaire. L'inégalité nous montre bien le processus de **polarisation** qui est en cours.

Proposition 4.2.3.

On a aussi :

$$Z(W'') = Z(W)^2$$

6. L'utilisation de ces exposants est due à Emre Telatar, qui la considère comme «[sa] plus importante contribution aux codes polaires»[13]

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], *p.*22.
Pour commencer, on a :

$$Z(W'') = \sum_{y_1^2,u_1} \sqrt{W''(f(y_1^2),u_1 \mid 0)} \cdot \sqrt{W''(f(y_1^2),u_1 \mid 1)}$$

$$= \sum_{y_1^2,u_1} \frac{1}{2}\sqrt{W(y_1 \mid u_1)W(y_2 \mid 0)} \cdot \sqrt{W(y_1 \mid u_1 \oplus 1)W(y_2 \mid 1)}$$

$$= \sum_{y_2} \sqrt{W(y_2 \mid 0)W(y_2 \mid 1)} \cdot \sum_{u_1} \frac{1}{2} \cdot \sum_{y_1} \sqrt{W(y_1 \mid u_1)W(y_1 \mid u_1 \oplus 1)}$$

$$= Z(W)^2$$

□

Proposition 4.2.4.

On a aussi :

$$Z(W') \leqslant 2 \cdot Z(W) - Z(W)^2$$

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], *p.*22.
On a :

$$Z(W') = \sum_{y_1^2} \sqrt{W'(f(y_1^2) \mid 0) \cdot W'(f(y_1^2) \mid 1)}$$

$$= \sum_{y_1^2} \frac{1}{2}\sqrt{W(y_1 \mid 0)W(y_2 \mid 0) + W(y_1 \mid 1)W(y_2 \mid 1)}$$

$$\cdot\sqrt{W(y_1 \mid 0)W(y_2 \mid 1) + W(y_1 \mid 1)W(y_2 \mid 0)}$$

$$\leqslant \sum_{y_1^2} \frac{1}{2}\left[\sqrt{W(y_1 \mid 0)W(y_2 \mid 0)} + \sqrt{W(y_1 \mid 1)W(y_2 \mid 1)}\right]$$

$$\cdot\left[\sqrt{W(y_1 \mid 0)W(y_2 \mid 1)} + \sqrt{W(y_1 \mid 1)W(y_2 \mid 0)}\right]$$

$$- \sum_{y_1^2} \sqrt{W(y_1 \mid 0)W(y_2 \mid 0)W(y_1 \mid 1)W(y_2 \mid 1)}$$

Or,

$$\sum_{y_1^2} \sqrt{W(y_1 \mid 0)W(y_2 \mid 0)W(y_1 \mid 1)W(y_2 \mid 1)} = Z(W)^2$$

d'où l'inégalité recherchée. Les détails supplémentaires et la justification de l'inégalité utilisée ne seront pas développés ici. □

Proposition 4.2.5. *On a aussi :*

$$Z(W') \geqslant Z(W) \geqslant Z(W'')$$

Démonstration. La démonstration est admise mais se trouve dans l'article d'Arıkan [3], *p.*22. □

Remarque 4.2.2. *Les propositions 4.2.3 et 4.2.4 impliquent que :*

$$Z(W') + Z(W'') \leqslant 2 \cdot Z(W)$$

On remarque alors que la ***fiabilité*** *est améliorée lors d'une transformation par opération élémentaire. De plus, la proposition 4.2.5 montre bien le processus de* ***polarisation*** *qui est en cours.*

Proposition 4.2.6.

Soit W un B-DMC, $N = 2^n$, pour tout $i \in \{1, \ldots, N\}$, la transformation par opération élémentaire $(W_N^{(i)}, W_N^{(i)}) \mapsto (W_{2N}^{(2i-1)}, W_{2N}^{(2i)})$ préserve le ***taux*** *et améliore la* ***fiabilité****, c'est à dire :*

$$I(W_{2N}^{(2i-1)}) + I(W_{2N}^{(2i)}) = 2 \cdot I(W_N^i)$$

et

$$Z(W_{2N}^{(2i-1)}) + Z(W_{2N}^{(2i)}) \leqslant 2 \cdot Z(W_N^i)$$

Démonstration. La preuve est directe en utilisant 4.1.2, 4.2.1, 4.2.3 et 4.2.4. □

5 Polarisation

Nous allons maintenant représenter notre travail par un arbre binaire (voir Fig.4), l'idée[7] va maintenant être d'étudier les processus aléatoires *se déplaçant* dans cet arbre. On appellera parfois ces processus des **trajectoires**.

Un indéxation naturelle des noeuds de cet arbre est la suivante :

- Au niveau 0 (la racine), le noeud est indéxé avec la séquence vide
- Au niveau 1, le noeud supérieur est indéxé avex la séquence $_0$ et le noeud inférieur avec la séquence $_1$

7. C'est l'idée originale d'Arıkan dans son article de 2009, plusieurs autres démonstrations ont été établis à ce jour mais nous avons préféré étudier cette approche.

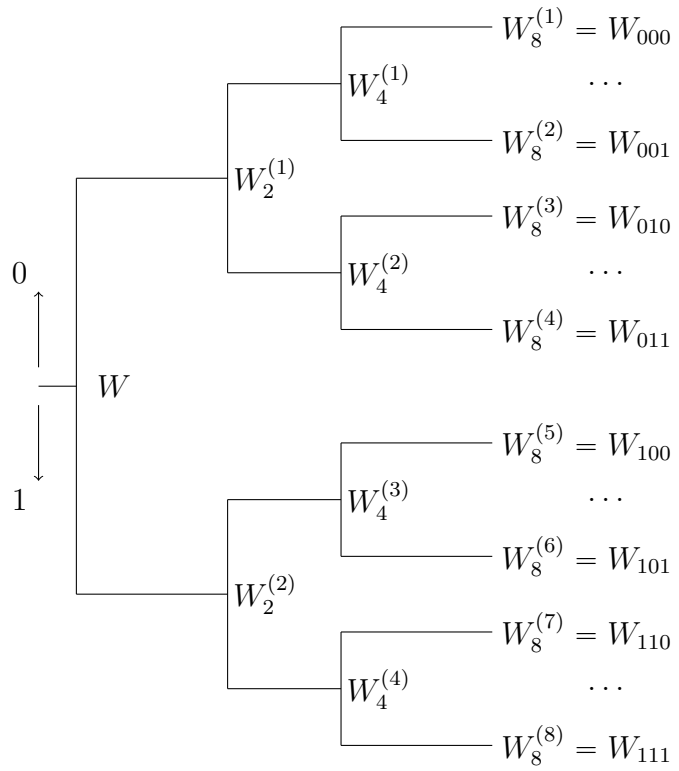

FIGURE 4 – Arbre binaire représentant le processus de construction des canaux

— Au niveau n, un noeud est indéxé avec la séquence $_{b_1 \ldots b_n}$ et ainsi, le noeud supérieur (resp. inférieur) en émanant est indéxé avec la séquence $_{b_1 \ldots b_n 0}$ (resp. $_{b_1 \ldots b_n 1}$)

Le canal $W_{2^n}^{(i)}$ est situé au noeud $_{b_1 \ldots b_n}$ avec $i = 1 + \sum_{1 \leqslant j \leqslant n} b_j 2^{n-j}$. Le canal $W_{2^n}^{(i)}$ pourra alors être noté $W_{b_1 \ldots b_n}$.

On définit maintenant un processus aléatoire $(K_n)_{n \in \mathbb{N}}$ sur cet arbre, on a donc $K_0 = W$ et, pour tout $n \in \mathbb{N}$, si $K_n = W_{b_1 \ldots b_n}$ alors $K_{n+1} = W_{b_1 \ldots b_n 1}$ ou $K_{n+1} = W_{b_1 \ldots b_n 0}$ avec une probabilité $\frac{1}{2}$ chacun. Le chemin emprunté par un processus K_n correspond donc à une séquence B_n de variables aléatoires de Bernouilli *i.i.d.*

Nous définissons aussi 2 autres processus aléatoires afin de suivre l'évolution du taux et de la fiabilité des canaux empruntés par un processus aléatoire K_n. On les note :

— $I_n = I(K_n)$

— $Z_n = Z(K_n)$

Rigouresement, les processus aléatoires que nous venons de définir peuvent être formalisés comme suit :

Soit l'espace probabilisé $(\Omega, \mathcal{F}, P)$ où $\Omega = (b_1, b_2, \ldots) \in \{0,1\}^\infty$, $\mathcal{F}$ est la tribu générée par les cylindres

$$S(b_1, \ldots, b_n) = \{\omega \in \Omega : \omega_1 = b_1, \ldots, \omega_n = b_n) \text{ pour tout } n \in \mathbb{N}$$

et la mesure de probabilité P est définie sur $\mathcal{F}$ telle que

$$P(S(b_1, \ldots, b_n)) = \frac{1}{2^n}$$

Pour tout $n \geqslant 1$, on définit $\mathcal{F}_n$ par la tribu générée par $S(b_1, \ldots, b_n)$ où $b_1, \ldots, b_n \in \{0,1\}^\infty$ et $\mathcal{F}_0 = \varnothing$. On a clairement $\mathcal{F}_0 \subset \ldots \subset \mathcal{F}$ et ainsi la suite $(\mathcal{F}_n)_{n\in\mathbb{N}}$ forme bien une filtration.
Pour $\omega = (\omega_1, \omega_2, \ldots) \in \Omega$ et $n \geqslant 1$, on définit :

$$B_n(\omega) = \omega_n, \quad K_n(\omega) = W_{\omega_1 \ldots \omega_n}, \quad I_n(\omega) = I(K_n(\omega)), \quad Z_n(\omega) = Z(K_n(\omega)).$$

et pour $n = 0$, on pose simplement :

$$K_0 = W, \quad I_0 = I(W), \quad Z_0 = Z(W).$$

De plus, la proposition suivante nous assure que ces processus aléatoires sont bien mesurables par rapport à la tribu $\mathcal{F}_n$.

Proposition 5.0.1.

La tribu $\mathcal{F}$ est la tribu borélienne de l'espace Ω.

Démonstration. La tribu borélienne est par définition la tribu engendrée par les ensembles ouverts de Ω. Or, dans notre topologie, les $S(b_1, \ldots, b_n) = \{\omega \in \Omega : \omega_1 = b_1, \ldots, \omega_n = b_n)$ forment une base des ouverts de Ω. □

5.1 Etude des processus aléatoires

On va maintenant étudier la convergence des processus aléatoire précédents :

Proposition 5.1.1.

Le processus $(I_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une martingale.

Démonstration.

–Par construction, I_n est **adapté** à la filtration.
–On a $0 \leqslant I_n < 1$ et ainsi $E[|I_n|] < \infty$
– Soit $S(b_1, \ldots, b_n) \in \mathcal{F}_n$, on a :

$$E[I_{n+1} \mid \mathcal{F}_n] \quad = \quad E[I_{n+1} \mid S(b_1, \ldots, b_n)]$$

en utilisant maintenant la proposition 4.2.6, on trouve :

$$= \frac{1}{2} I(W_{b_1,\ldots,b_n,0}) + \frac{1}{2} I(W_{b_1,\ldots,b_n,1}) \quad = \quad I(W_{b_1,\ldots,b_n}) \quad = \quad I_n$$

□

Propriété 5.1.1.

Il existe un variable aléatoire I_∞ telle que I_n converge presque sûrement vers celle-ci. De plus, on a $E[I_\infty] = I_0$.

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], $p.8$.
On sait déjà que $E[\mid I_n \mid] < \infty$, ce qui implique que (I_n) est intégrable dans L^1.
De plus, $\exists M \in \mathbb{R}$ tel que $\mid I_n \mid \leqslant M$ presque sûrement pour tout n (c'est une suite bornée) donc (I_n) est équiintégrable. On utilise alors la propriété de convergence des martingales pour conclure que (I_n) converge presque sûrement dans L^1 vers une variable aléatoire I_∞.
Pour prouver la seconde affirmation, on utilise le fait que la convergence dans L^1 implique la convergence de l'espérance :

$$\lim_{n\to\infty} E[\mid I_n - I_\infty \mid] = 0 \quad \Longrightarrow \quad E[I_n] \underset{n\to\infty}{\longrightarrow} E[I_\infty]$$

Ensuite, on a :

$$E[I_{n+1} \mid \mathcal{F}_n] = I_n \quad presque \;\; srement$$

d'où :

$$E[E[I_{n+1} \mid \mathcal{F}_n]] = E[I_n]$$

et enfin, puisque $E[E[I_{n+1} \mid \mathcal{F}_n]] = I_{n+1}$ (théorème de l'espérance totale), on obtient :

$$E[I_{n+1}] = E[I_n]$$

La discussion précédente implique :

$$E[I_n] = E[I_0] = I_0 \quad car \;\; I_0 \;\; est \;\; une \;\; constante$$

On conclut finalement avec :

$$\lim_{n\longrightarrow\infty} E[I_n] = I_0 \quad et \quad E[I_n] \underset{n\longrightarrow\infty}{\longrightarrow} E[I_\infty] \quad \Longrightarrow \quad E[I_\infty] = I_0$$

□

Proposition 5.1.2.

Le processus $(Z_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ est une sur-martingale.

Démonstration. La démonstration est très similaire à celle de la proposition 5.1.1, nous traitons ici que le dernier point.

Soit $S(b_1, \ldots, b_n) \in \mathcal{F}_n$, on a :

$$E[Z_{n+1} \mid \mathcal{F}_n] \quad = \quad E[Z_{n+1} \mid S(b_1, \ldots, b_n)]$$

en utilisant maintenant la proposition 4.2.6, on trouve :

$$= \frac{1}{2} Z(W_{b_1,\ldots,b_n,0}) + \frac{1}{2} Z(W_{b_1,\ldots,b_n,1}) \quad \leqslant \quad Z(W_{b_1,\ldots,b_n}) \quad = \quad Z_n$$

□

Propriété 5.1.2.

Il existe un variable aléatoire Z_∞ telle que Z_n converge presque sûrement vers celle-ci. De plus, on a $Z_\infty \in \{0, 1\}$ presque sûrement.

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], *p.*9.
La preuve commence de la même manière que celle de la propriété 5.1.1, on a donc :

$$\lim_{n \to \infty} E[\mid Z_n - Z_\infty \mid] = 0$$

et aussi :

$$\lim_{n \to \infty} E[\mid Z_{n+1} - Z_n \mid] = 0$$

Or, en utilisant les propositions 4.2.3 et 4.1.2, on peut affirmer que $Z_{n+1} = Z_n^2$ avec une probabilité $\frac{1}{2}$ (en effet, si $K_n = W_{b_1 \ldots b_n}$, on a $K_n = W_{b_1 \ldots b_n 1}$ avec une probabilité de $\frac{1}{2}$ et alors $Z(W_{b_1 \ldots b_n 1}) = Z(W_{b_1 \ldots b_n})^2$). On a alors :

$$E[\mid Z_{n+1} - Z_n \mid] \geqslant \frac{1}{2} E[\mid Z_n^2 - Z_n \mid]$$

qu'on peut réecrire :

$$\lim_{n \to \infty} E[\mid Z_n(1 - Z_n) \mid] = 0$$

et vient alors :

$$E[\mid Z_\infty(1 - Z_\infty) \mid] = 0$$

On conclut alors facilement que $Z_\infty \in \{0, 1\}$ □

Nous avons maintenant tous les outils pour démontrer le théorème suivant, qui est un des résultats les plus importants du travail d'Arıkan.

Théorème 5.1.1.

Soit W un B-DMC, on note μ sa capicité symétrique $I(W)$. Soit $n \in \mathbb{N}$ et $N = 2^n$, lorsque n tend vers l'infini, les canaux $\{W_N^{(i)}\}$ se polarise. C'est à dire que, pour tout

$\delta \in]0,1[$, la fraction des indices $i \in \{1,\dots,N\}$ tels que $I(W_N^{(i)}) \in]1-\delta,1]$ tend vers μ et la fraction des indices tels que $I(W_N^{(i)}) \in [0,\delta[$ tend vers $1-\mu$.

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], *p.*9.
La propriété 5.1.2 affirme que $Z_\infty \in \{0,1\}$ presque sûrement, on utlise alors la proposition 2.0.1, et on en déduit que $I_\infty = 1 - Z_\infty$ presque sûrement, c'est à dire que $I_\infty \in \{0,1\}$. On a aussi :

$$E[I_\infty] = 1 \cdot \mathbb{P}(I_\infty = 1) + 0 \cdot \mathbb{P}(I_\infty = 0)$$

d'où :

$$\mathbb{P}(I_\infty = 1) = I_0 \quad et \quad \mathbb{P}(I_\infty = 0) = 1 - I_0$$

□

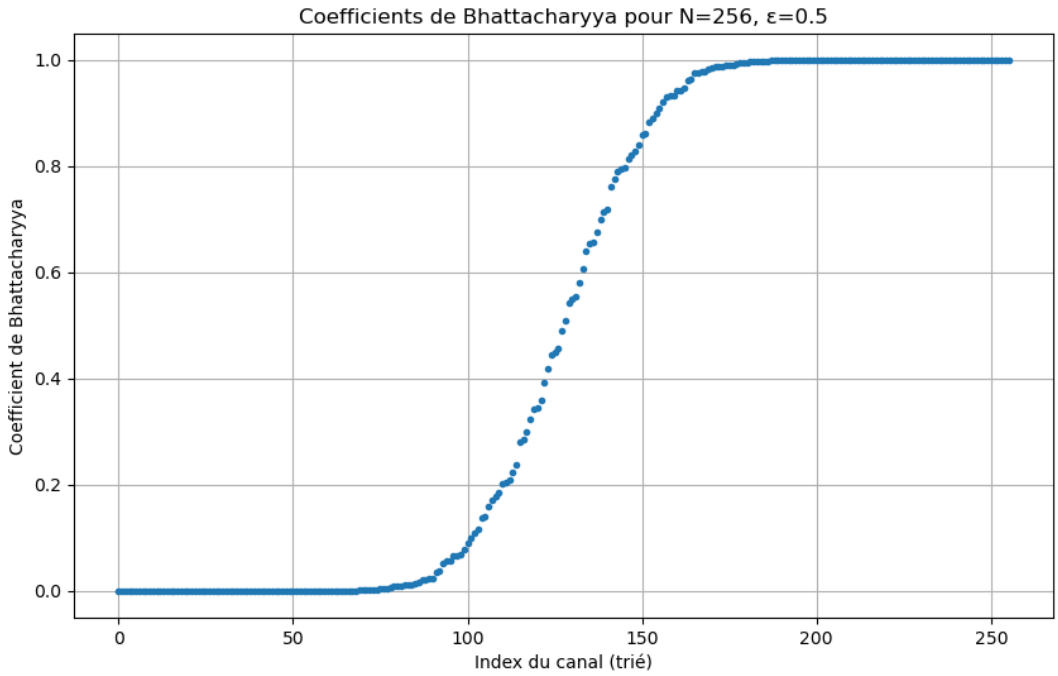

FIGURE 5 –

L'exemple est assez explicite : la proportion de canaux *médiocre* tends vers 0 lorsque $n \longrightarrow \infty$.

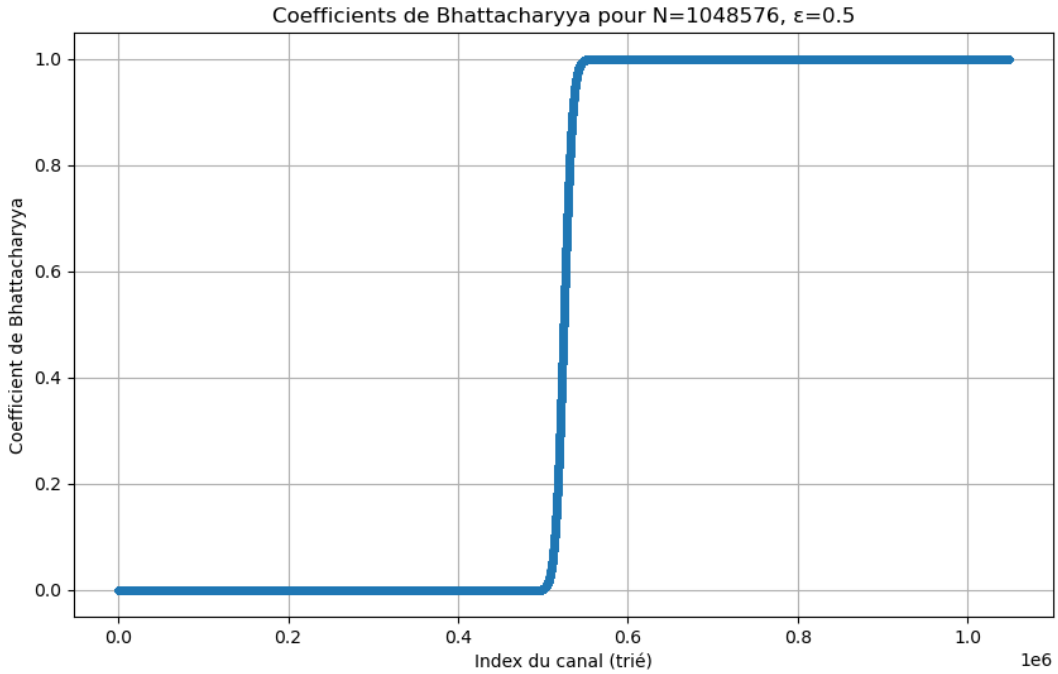

FIGURE 6 –

6 Des codes qui atteignent la borne de Shannon

Cette section est consacrée à l'établissement du résultat principal de notre travail : montrer qu'il existe un sous-ensemble assez grand (en proportion) de canaux de taux arbitrairement grand et suffisaments fiables.

Nous gardons les mêmes notations et nous nous plaçons dans le même espace probabilisé que dans la section précédente.

Définition 6.0.1.

Soit $m \geqslant 0$ et $\zeta \geqslant 0$, on définit l'ensemble des trajectoires (ou processus) ***ultimement polarisantes*** *:*

$$\Delta_m(\zeta) = \{\omega \in \Omega : Z_i(\omega) \leqslant \zeta, \quad \forall i \geqslant m\}$$

En d'autres termes, cet ensemble correspond à l'ensemble des trajectoires dont la ***fiabilité*** *devient arbitrairement petite ($\leqslant \zeta$) à partir d'un rang m.*

Le lemme suivant, nous permet d'isoler une proportion proche de $I_0 = I(W)$ de trajectoires **ultimement polarisantes** tout en garantissant qu'elles atteignent un seuil de fiabilité à partir un **rang fini** – on n'a pas besoin d'attendre une limite asymptotique.

Lemme 6.0.1.

Pour $\zeta > 0$ et $\delta > 0$ fixés, il existe un entier $m_0(\zeta, \delta)$ tel que :

$$\mathbb{P}(\Delta_{m_0(\zeta,\delta)}(\zeta)) \geqslant I_0 - \frac{\delta}{2}$$

où $I_0 = I(W)$.

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], *p.*23.
Fixons $\zeta > 0$, et posons :

$$\Omega_0 = \{\omega \in \Omega : \lim_{n\to\infty} Z_n(\omega) = 0\}.$$

qui correspond à l'ensemble des **trajectoires polarisantes**.
Soit $\omega \in \Omega_0$, puisque $Z_n(\omega) \underset{n\to\infty}{\longrightarrow} 0$, on peut affirmer qu'il existe $n_0(\omega, \zeta)$ tel que :

$$\forall n \geqslant n_0(\omega, \zeta), \quad Z_n(\omega) \leqslant \zeta$$

Il existe donc $k \geqslant 0$ tel que $\omega \in \Delta_k(\zeta)$ et ainsi :

$$\Omega_0 \subset \bigcup_{m=1}^{\infty} \Delta_m(\zeta)$$

d'où :

$$\mathbb{P}(\bigcup_{m=1}^{\infty} \Delta_m(\zeta)) \geqslant \mathbb{P}(\Omega_0) = I_0$$

Par le théorème de convergence monotone on affirme que :

$$\lim_{m\to\infty} \mathbb{P}(\Delta_m(\zeta)) = \quad \mathbb{P}(\bigcup_{m=1}^{\infty} \Delta_m(\zeta))$$

qui peut se réécrire :

$$\lim_{m\to\infty} \mathbb{P}(\Delta_m(\zeta)) \geqslant I_0$$

On conclut alors en notant que la suite $(\Delta_m(\zeta))_{m\geqslant 1}$ est croissante qu'il existe bien $m_0(\zeta, \delta)$ tel que $\mathbb{P}(\Delta_m(\zeta)) \geqslant I_0 - \delta/2$. □

Le prochain théorème nous permet de conclure notre travail car il permet de prouver que les codes polaires d'Arıkan atteigent la borne de Shannon.

Théorème 6.0.1.

Soit W un B-DMC tel que $I(W) > 0$ et soit $R < I(W)$. Pour $N \in \{1, 2, \ldots, 2^n, \ldots\}$ il existe une séquence d'ensembles $\mathcal{A}_N \subset \{1, \ldots, N\}$ telle que $\mid \mathcal{A}_N \mid \geqslant N \cdot R$ et $Z(W_N^{(i)}) \leqslant O(N^{-5/4})$ pout tout $i \in \mathcal{A}_N$.

Démonstration.

Nous nous appuyons sur la preuve d'Arıkan dans son article [3], *p.*9.
Soit $\omega \in \Delta_m(\zeta)$ et $i \geqslant m$, nous utilisions ici les inégalités de la proposition 4.2.6 pour écrire :

$$Z_{m+1}(\omega) \leqslant Z_m(\omega) \cdot \begin{cases} 2 & \text{si } b_{i+1}(\omega) = 0 \\ \zeta & \text{si } b_{i+1}(\omega) = 1 \end{cases}$$

d'où, pour $n > m$, on a :

$$Z_n(\omega) \leqslant Z_m(\omega) \cdot \prod_{i=m+1}^{n} \begin{cases} 2 & \text{si } b_{i+1}(\omega) = 0 \\ \zeta & \text{si } b_{i+1}(\omega) = 1 \end{cases}$$

On peut réécrire la partie de droite comme suit :

$$Z_n(\omega) \leqslant \zeta \cdot \prod_{i=m+1}^{n} 2^{1-b_i(\omega)} \cdot \zeta^{b_i(\omega)}$$

ou encore :

$$Z_n(\omega) \leqslant \zeta \cdot 2^{\sum_{i=m+1}^{n}(1-b_i(\omega))} \cdot \zeta^{\sum_{i=m+1}^{n} b_i(\omega))}$$

Or $\sum_{i=m+1}^{n}(1 - b_i(\omega)) = (n - m) - \sum_{i=m+1}^{n} b_i(\omega)$.
On peut donc finalement écrire :

$$Z_n(\omega) \leqslant \zeta \cdot 2^{n-m} \cdot \left(\frac{\zeta}{2}\right)^{\sum_{i=m+1}^{n} b_i(\omega)}$$

Soit $0 < \eta < \frac{1}{2}$, on définit :

$$\mathcal{U}_{m,n}(\eta) = \{\omega \in \Omega : \sum_{i=m+1}^{n} b_i(\omega) > (\frac{1}{2} - \eta) \cdot (n - m)\}$$

Cet ensemble contient les trajectoires qui ont une proportion de $b_i = 1$ supérieure à $(\frac{1}{2} - \eta)$, or on sait par les propositions de la section précédente que ceux-ci représentent les canaux avec la meilleur fiabilité.
Les trajectoires (c'est à dire canaux) qui nous intéressent dans le cadre de ce théorème sont donc ceux de l'ensemble :

$$\Delta_m(\zeta) \cap \mathcal{U}_{m,n}(\eta)$$

avec η et ζ raisonnablements petits.
En choisissant par exemple $\zeta_0 = 2^{-4}$ et $\eta_0 = \frac{1}{20}$, si $\omega \in \Delta_m(\zeta_0) \cap \mathcal{U}_{m,n}(\eta_0)$, on trouve

que [8] :

$$Z_n(\omega) \leqslant 2^{-4-5(n-m)/4} \qquad (*)$$

Il nous reste à démontrer que les trajectoires vérifiant (*) surviennent avec une fréquence significative.
On utilise alors le lemme 6.0.1 pour obtenir $m_0(\zeta, \delta)$ tel que

$$\mathbb{P}(\Delta_{m_0(\zeta,\delta)}(\zeta)) \geqslant I_0 - \frac{\delta}{2}$$

On utilise maintenant le théorème de Chernoff-Hoeffding pour estimer :

$$\mathbb{P}(\mathcal{U}_{m,n}(\eta)) \leqslant e^{-D(\frac{1}{2}-\eta\|\frac{1}{2})(n-m)}$$

où $D(p\|q)$ est la divergence de Kullback-Leibler. Mais puisque nous travaillons en binaire, on utilise le log_2 et on simplifie en :

$$\mathbb{P}(\mathcal{U}_{m,n}(\eta)) \leqslant 2^{-(1-h(\frac{1}{2}-\eta))\cdot(n-m)}$$

ou encore :

$$\mathbb{P}(\mathcal{U}_{m,n}(\eta)) \geqslant 1 - 2^{-(1-h(\frac{1}{2}-\eta))\cdot(n-m)}$$

où $h(p)$ est l'entropie de Shannon en base 2.
Puisque $1-2^{-(1-h(\frac{1}{2}-\eta))\cdot(n-m)}$ est une fonction (de n) croissante et de limite 1 lorsque n tend vers $+\infty$, il existe un entier $n_0(m, \eta, \delta)$ tel que $1-2^{-(1-h(\frac{1}{2}-\eta))\cdot(n_0(m,\eta,\delta)-m)} \geqslant 1-\frac{\delta}{2}$ et tel que, $\forall k < n_0$, $1 - 2^{-(1-h(\frac{1}{2}-\eta))\cdot(k-m)} < 1 - \frac{\delta}{2}$.
En prenant les valeurs η_0 et ζ_0 définie plus haut et les valeurs $m_1 = m_0(\zeta_0, \delta)$ et $n_1 = n_0(m_1, \eta_0, \delta)$ on obtient la borne suivante :

$$\mathbb{P}(\Delta_m(\zeta_0) \cap \mathcal{U}_{m_1,n}(\eta_0)) \geqslant I_0 - \delta, \quad n \geqslant n_1$$

Nous avons donc montré que les trajectoires ayant une forte proportion de 1 et étant ultimement polarisantes apparaissent régulièrement. Nous n'avons plus qu'un pas à faire pour montrer le résultat attendu.
Posons $c = 2^{-4+5m_1/4}$ et :

$$\mathcal{C}_n = \{\omega \in \Omega : Z_n(\omega) \leqslant c \cdot 2^{-5n/4}\}, \quad n \geqslant 0$$

Cet ensemble représente les trajectoires dont la **fiabilité** finale (au rang n) est très élevée. On notera que $\Delta_m(\zeta) \cap \mathcal{U}_{m,n}(\eta) \subset \mathcal{C}_n$ pour tout $n \geqslant n_1$.

8. Ces constantes sont celles proposées par Arıkan dans son article [3]

D'où, on déduit :

$$\mathbb{P}(\mathcal{C}_n) \geqslant I_0 - \delta, \quad \forall n \geqslant n_1$$

On peut aussi calculer la fréquence des $\omega \in \mathcal{C}_n$ comme suit :

$$\mathbb{P}(\mathcal{C}_n) = \sum_{\omega_1^n \in \mathcal{X}^n} \mathbb{P}(w_1^n) \cdot \mathbb{1}\{Z(W_{w_1^n}) \leqslant c \cdot 2^{-5n/4}\}$$

d'où, en posant $\mathcal{A}_N = \{i \in \{1, \ldots, N\} : Z(W_N^{(i)}) \leqslant c \cdot N^{-5/4}\}$ et en remarquant la bijection entre les ensembles $\mathcal{A}_n$ et $\{Z(W_{w_1^n}) \leqslant c \cdot 2^{-5n/4}\})$ on obtient :

$$\mathbb{P}(\mathcal{V}_n) = \frac{1}{N} \cdot \mid \mathcal{A}_N \mid$$

ce qui termine la preuve. □

Remarque 6.0.1.

Ce théorème suffit à démontrer que les codes polaires permettent d'atteindre la borne de Shannon (il faut utiliser la proposition 2.0.1).

7 Conclusion

Ce présent travail avait pour but de présenter les fondements des codes polaires en étudiant le mécanisme de polarisation sur lequel ceux-ci reposent. Nous avons montré comment il était possible de, en partant de canaux binaires identiques, obtenir une répartition *extrême* entre canaux fiables et non fiables.
Nous pensons avoir atteint notre objectif en mettant en evidence que cette technologie permet la construction de codes exploitant uniquement les canaux les plus performants, rendant rendant ainsi possible une transmission à un taux arbitrairement proche de la capacité du canal, conformément au théorème de Shannon.
L'étude de l'encodage et du décodage n'a pas été l'objet principal de ce travail. Cependant, nous avons développé plusieurs programmes pour expérimenter concrètement cette technologie. Si l'encodage se révèle relativement simple, la conception d'un décodeur efficace s'est montrée plus délicate. Cette question reste un enjeu important dans la recherche actuelle, même si des avancées significatives ont déjà été accomplies.

Références

[1] Martin Aigner and Günter M Ziegler. *Raisonnements divins : quelques démonstrations mathématiques particulièrement élégantes*. Springer Science & Business Media, 2006.

[2] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

[3] Erdal Arikan. Channel polarization : A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on information Theory*, 55(7) :3051–3073, 2009.

[4] Erdal Arıkan. On the origin of polar coding. *IEEE journal on Selected Areas in Communications*, 34(2) :209–223, 2015.

[5] Claude Berrou, Alain Glavieux, and Punya Thitimajshima. Near shannon limit error-correcting coding and decoding : Turbo-codes. 1. In *Proceedings of ICC'93-IEEE International Conference on Communications*, volume 2, pages 1064–1070. IEEE, 1993.

[6] Antoine Chambert-Loir. *Théorie de l'information - Trois théorèmes de Claude Shannon*. Calvage & Mounet, 2022.

[7] Thomas M Cover. Elements of information theory, 2006.

[8] R. Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*, 8(1) :21–28, 1962.

[9] I. Jacobs and E. Berlekamp. A lower bound to the distribution of computation for sequential decoding. *IEEE Transactions on Information Theory*, 13(2) :167–174, 1967.

[10] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[11] David JC MacKay and Radford M Neal. Near shannon limit performance of low density parity check codes. *Electronics letters*, 32(18) :1645–1646, 1996.

[12] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3) :379–423, 1948.

[13] Emre Telatar. Isit 2017 | emre telatar | the flesh of polar codes | 2017-06-29.