INTRODUCTION À LA THÉORIE DE L'INFORMATION POUR LE CODAGE

Master Systèmes de Radiocommunications, année 2007.
Version : 2.1.

Samson LASAULCE
lasaulce@lss.supelec.fr

Table des matières

Chapitre 1

Introduction. Outils de base du cours

1.1 Objectifs du cours

Ce cours traitera assez peu de théorie de l'information au sens large. En effet, à part les définitions de base données dans ce chapitre, l'essentiel du cours portera sur l'application de la théorie de l'information aux communications. Dans ce cadre précis nous nous limiterons essentiellement à la notion de théorème de codage. Une des activités de la théorie de l'information pour les communications consiste à fournir, pour un canal donné, le théorème de codage correspondant. Un théorème de codage a au moins deux fonctions essentielles :

1. grâce à la démonstration de la partie dite "d'atteignabilité" fournir des idées et des concepts de codage dont on peut s'inspirer pour construire une stratégie pratique de codage canal.
2. grâce au résultat de la partie dite "réciproque", qui permet de valider l'optimalité de la stratégie de codage construite dans la partie atteignabilité, donner les limites ultimes en termes de débit d'information d'une communication fiable sur un canal donné.

Le cours comporte trois parties :

1. une partie qui contient les principaux outils de base de la théorie de l'information ;
2. une partie dédiée au canal mono-utilisateur, c'est-à-dire avec un seul émetteur et un seul récepteur (canal de Shannon, canal MIMO, canal multi-trajet, ...) ;
3. une partie dédiée à des canaux plus complexes, c'est-à-dire avec au moins trois terminaux (canal de diffusion, canal à relais, ...).

1.2 Outils de base

1.2.1 Information associée à un événement

On conçoit très bien qu'un événement certain ne contient pas de nouveauté ou d'information. Par exemple, lorsque la météo annonce une journée sans pluie dans le Sahara, on apprend peu de choses. Inversement l'annonce d'une journée de pluie est un renseignement précieux, car inattendu. On peut ainsi définir l'information associée à une réalisation x (événement) d'une variable aléatoire X discrète comme suit :

$$i(x) \triangleq -\log p(x). \tag{1.1}$$

Bien que cette quantité soit parfois utile dans certains calculs intermédiaires ou pédagogiquement, elle ne figure pas parmi les grandeurs d'usage de la théorie de l'information pour les communications.

1.2.2 Entropie

C'est Boltzmann qui a donné (1872) le premier une interprétation statistique de l'entropie thermodynamique ($S = k \ln W$) en utilisant l'hypothèse microcanonique qui consiste à dire que, pour un système isolé, tous les états microscopiques associés un état macroscopique donné sont équiprobables. Shannon, à partir de la définition de l'entropie thermodynamique de Boltzmann et de considérations axiomatiques sur l'information, a définie l'entropie statistique d'une source d'information "X" qui est plus générale que l'entropie de la thermodynamique statistique.

Soit X une variable aléatoire (VA) discrète d'alphabet $\mathcal{X} = \{x_1, ..., x_M\}$ avec $M = |\mathcal{X}|$, $|.|$ étant le cardinal de l'ensemble considéré. L'entropie ou l'incertitude de cette variable aléatoire est définie comme suit :

$$H(X) = -\sum_{x \in \mathcal{X}} \Pr[X = x] \log \Pr[X = x]. \tag{1.2}$$

En utilisant la notion d'information associée à un événement, l'entropie se traduit comme la moyenne de l'information associée à la variable aléatoire X.

Pour une variable aléatoire X continue de densité de probabilité $f_X(x)$ on parle d'entropie différentielle, celle-ci étant définie par :

$$H_d(X) = -\int_{x \in \mathbb{R}} f_X(x) \log f_X(x) dx. \tag{1.3}$$

Remarque 1. Le choix de la base du logarithme est arbitraire. Dans ce cours nous prendrons souvent la base 2, ce qui revient à mesurer l'information en bit. Précisons d'ailleurs qu'il faut distinguer le bit, qui est une unité de l'information, de l'élément binaire 0 ou 1. Un élément binaire d'une source binaire véhicule effectivement un bit d'information si les deux éléments de la source sont équiprobables. Sinon, l'élément binaire transporte moins d'information qu'un bit, le cas extrême étant la source déterministe qui véhicule zéro bit d'information.

Remarque 2. L'entropie d'une variable aléatoire discrète est, par définition, non-négative : $H(X) \geq 0$. Par contre l'entropie d'une variable aléatoire continue, appelée souvent entropie différentielle, peut être négative, on peut facilement le vérifier pour le cas gaussien (voir exemple 2). **Dans ce cours nous priviligierons, dans les définitions notamment, le cas discret** sachant que le cas continu peut s'obtenir en exploitant les contraintes "physiques" du problème et des arguments de discrétisation (discrétisation selon l'axe des temps : échantillonnage, discrétisation selon l'axe des amplitudes : quantification). On retiendra donc la propriété suivante pour une variable aléatoire discrète :

$$0 \leq H(X) \leq \log |\mathcal{X}|. \tag{1.4}$$

On remarquera que la borne supérieure de l'entropie d'une VA discrète est atteinte pour une distribution de probabilités uniforme. Pour le cas d'une VA continue on pourrait montrer que, pour une variance donnée, c'est la distribution gaussienne qui maximise l'entropie différentielle $H_d(X)$.

Exemple 1 : entropie d'une VA binaire. Soit $\mathcal{X} = \{0, 1\}$ avec $\Pr[X = 0] = p$ et $\Pr[X = 1] = \overline{p} = 1 - p$. Alors $H(X) = -p \log p - \overline{p} \log \overline{p} \triangleq H_2(p)$.

Exemple 2 : entropie d'une VA gaussienne scalaire et réelle. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$. On rappelle que la densité de probabilité d'une VA gaussienne et réelle est donnée par :

$$f_X(x) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{1.5}$$

avec $\mu = E(X)$ et $\sigma^2 = \text{Var}(X) = E(X-\mu)^2$. Nous avons donc

$$\begin{aligned} H(X) &= -\int_{\mathbb{R}} f_X(x)\log_2[f_X(x)]dx \\ &= -\int_{\mathbb{R}} f_X(x)\log_2\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right]dx \\ &= \frac{1}{2}\log_2(2\pi\sigma^2)\int_{\mathbb{R}} f_X(x)dx + \frac{1}{2\sigma^2\ln 2}\int_{\mathbb{R}}(x-\mu)^2 f_X(x)dx \\ &= \frac{1}{2}\log_2(2\pi\sigma^2) + \frac{\log_2(e)}{2} \\ &= \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) \\ &= \frac{1}{2}\log_2\left[2\pi e\text{Var}(X)\right]. \end{aligned} \tag{1.6}$$

Exemple 3 : entropie d'une VA gaussienne vectorielle et réelle. Soit $\underline{X} \sim \mathcal{N}(\underline{\mu}, \mathbf{R})$. On rappelle que la densité de probabilité d'une VA gaussienne réelle à n dimensions est donnée par :

$$f_{\underline{X}}(\underline{x}) = \frac{1}{|2\pi\mathbf{R}|^{\frac{n}{2}}}\exp\left[-\frac{1}{2}\left(\underline{x}-\underline{\mu}\right)^T\mathbf{R}^{-1}\left(\underline{x}-\underline{\mu}\right)\right] \tag{1.7}$$

avec $\underline{\mu} = E(\underline{X})$ et $\mathbf{R} = \text{Covar}(\underline{X}) = E[(\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})^T]$. En utilisant la même technique de calcul que pour le cas scalaire on peut montrer que :

$$H(\underline{X}) = \frac{1}{2}\log_2\left(|2\pi e\mathbf{R}|\right) \tag{1.8}$$

où $|\mathbf{M}|$ représente le déterminant de la matrice $\mathbf{M}$. On notera que si la VA est complexe, il faut remplacer les trois 2 de l'équation (1.7) par des 1, ce qui revient à remplacer les deux 2 de l'équation (1.8) par des 1 également.

1.2.3 Entropie conditionnelle

L'entropie ou l'incertitude d'une VA X sachant une autre VA Y est définie par :

$$H(X|Y) \triangleq -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \Pr[(X,Y)=(x,y)]\log\Pr[X=x|Y=y]. \tag{1.9}$$

On peut noter au moins quatre points intéressants quant à la définition de l'entropie conditionnelle :
- pour deux VA X et Y indépendantes $H(X|Y) = H(X)$;
- le conditionnement réduit l'incertitude : on a toujours $H(X|Y) \leq H(X)$;
- la remarque précédente est vraie en moyenne si bien que la relation suivante est fausse : $H(X|Y=y) \leq H(X)$. Justification intuitive : l'incertitude d'un juge sur une affaire peut augmenter avec l'apport d'un nouvel élément au dossier ;
- on utilise souvent la réécriture suivante de l'entropie conditionnelle :

$$H(X|Y) = \sum_{y\in\mathcal{Y}} \Pr[Y=y]H(X|Y=y). \tag{1.10}$$

1.2.4 Information mutuelle

L'information mutuelle entre deux VA discrètes, notées X et Y, est définie par :

$$I(X;Y) \triangleq \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \Pr[(\mathrm{X},\mathrm{Y})=(\mathrm{x},\mathrm{y})]\log\left[\frac{\Pr[(X,Y)=(x,y)]}{\Pr[X=x]\Pr[Y=y]}\right]. \tag{1.11}$$

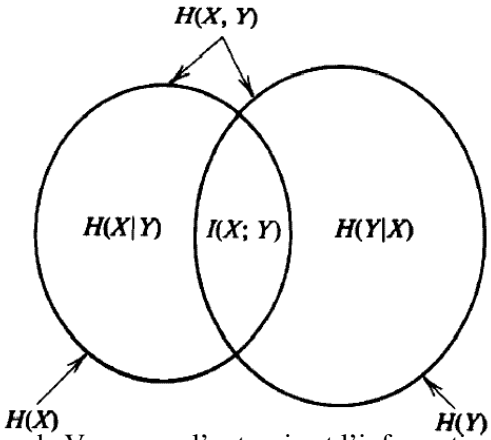

FIG. 1.1 – Diagramme de Venn pour l'entropie et l'information mutuelle

Commentaires sur la définition de l'information mutuelle.

– Le lecteur non averti notera l'importance du point-virgule au lieu d'une virgule dans la notation $I(X;Y)$. Le point-virgule indique comment il faut scinder au niveau du dénominateur la probabilité conjointe. Ainsi l'information mutuelle entre une VA X et un couple de VA (Y,Z) n'est, en général, pas égale à l'information mutuelle entre le couple de VA (X,Y) et la VA Z : $I(X;Y,Z) \neq I(X,Y;Z)$ d'où l'importance du " ;" pour préciser les deux entités entre lesquelles on veut évaluer l'information mutuelle.
– La définition de l'information mutuelle nous indique que cette quantité mesure le lien statistique entre deux VA. Ainsi pour deux VA indépendantes (i.e. $\Pr[(X,Y) = (x,y)] = \Pr[X = x]\Pr[Y = y]$) ce lien statistique est inexistant : $I(X;Y) = 0$. "Inversement" si on prend $Y \equiv X$ ce lien est maximal et $I(X;Y) = H(X) = H(Y)$.
– La définition pour le cas discret s'étend au cas continu en remplaçant la somme par une intégrale (Cf définitions de l'entropie). Les propriétés mentionnées ci-dessous valent à la fois pour les VA discrètes et continues.

Propriétés de l'information mutuelle.

– Lien avec l'entropie :

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && (1.12)\\ &= H(Y) - H(Y|X) && (1.13)\\ &= H(X) + H(Y) - H(X,Y). && (1.14) \end{aligned}$$

Ces relations peuvent être illustrées par un diagramme de Venn (voir figure 1.1).

– Non-négativité : $I(X;Y) \geq 0$.
– Symétrie : $I(X;Y) = I(Y;X)$.
– Concavité : pour une distribution de $Y|X$ fixée, disons $p(y|x)$, l'information mutuelle est concave en $p(x)$. Dans le cadre d'une communication sur le canal de Shannon cela signifie que pour un canal fixé il existe un codeur qui maximise l'information mutuelle.
– Convexité : pour une distribution de X fixée, disons $p(x)$, l'information mutuelle est convexe en $p(y|x)$. Dans le cadre d'une communication sur le canal de Shannon cela signifie que pour une stratégie de codage choisie il existe un canal qui minimise l'information mutuelle (scénario du "pire cas" pour l'ingénieur).

1.2.5 Information mutuelle conditionnelle

L'information mutuelle entre deux VA X et Y sachant la VA Z s'exprime simplement par :

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) && (1.15)\\ &= H(Y|Z) - H(Y|X,Z). && (1.16)\end{aligned}$$

1.2.6 Règles de chaîne

Pour l'entropie.
D'après les propriétés de l'information mutuelle (1.12)-(1.14) on voit que l'entropie d'une VA à deux dimensions s'écrit comme $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$. On peut généraliser cette formule, dite règle de chaîne pour l'entropie, au cas à n dimension. Soit $\underline{X} = (X_1, ..., X_n)$ un vecteur de VA. Notations : $X^{i-1} = (X_1, ..., X_{i-1})$, $X_i^j = (X_i, ..., X_j)$. Alors

$$H(\underline{X}) = \sum_{i=1}^{n} H(X_i|X^{i-1}) = \sum_{i=1}^{n} H(X_i|X_{i+1}^n) \tag{1.17}$$

avec les conventions : $X^0 \equiv \varnothing$, $X_{n+1}^n \equiv \varnothing$.

Pour l'information mutuelle.
Comme pour l'entropie on peut montrer que :

$$I(\underline{X};Y) = \sum_{i=1}^{n} I(X_i;Y|X^{i-1}) = \sum_{i=1}^{n} I(X_i;Y|X_{i+1}^n) \tag{1.18}$$

avec les mêmes conventions que pour la règle de chaîne pour l'entropie.

1.2.7 Chaînes de Markov

On dit que le triplet ordonné de variables aléatoires (X,Y,Z) est une chaîne de Markov si et seulement si :

$$p(z|x,y) = p(z|y). \tag{1.19}$$

Pour signifier que ce triplet ordonné de variables aléatoires est relié par une chaîne de Markov nous utiliserons la notation : $X - Y - Z$. Certains ouvrages utilisent les notations : $X \rightarrow Y \rightarrow Z$ ou $X \leftrightarrow Y \leftrightarrow Z$.

On peut signaler au moins cinq propriétés utiles des chaînes de Markov.
– En utilisant la règle de Bayes on peut remarquer que si $X - Y - Z$ alors $Z - Y - X$.
– Caractérisation en terme d'information mutuelle conditionnelle : $X - Y - Z \Rightarrow I(X;Z|Y) = 0$.
– Caractérisation en termes d'entropie conditionnelle : $X - Y - Z \Leftrightarrow H(Z|X,Y) = H(Z|Y)$ (équivalente à la caractérisation précédente).
– Chaîne de Markov extraite : $(X_1, ..., X_m) - Y - (Z_1, ..., Z_n) \Leftrightarrow \forall (i,j) \in \{1,..,m\} \times \{1,...,n\}, X_i - Y - Z_j$.
– Passage à l'intérieur : $(X_1, X_2) - Y - Z \Rightarrow X_1 - (X_2, Y) - Z$ et $(X_1, X_2) - Y - Z \Rightarrow X_2 - (X_1, Y) - Z$.
Considérons deux exemples où la notion de chaîne de Markov intervient très naturellement.

Exemple 1 : canal de diffusion.
Le canal de diffusion comprend un émetteur (signal émis : X) et deux récepteurs (signaux reçus : Y_1 et Y_2). La probabilité de transition de ce canal est notée $p(y_1, y_2|x)$. Si les différents signaux mis en jeu vérifient la relation de Markov $X - Y_1 - Y_2$ alors :

$$\begin{aligned} p(y_1, y_2|x) &= p(y_1|x)p(y_2|x, y_1) && (1.20)\\ &= p(y_1|x)p(y_2|y_1). && (1.21)\end{aligned}$$

On voit donc l'hypothèse de Markov "transforme" le canal de diffusion à une entrée et deux sorties en "parallèle" en un canal qui correspond à la mise en série de deux canaux mono-utilisateurs. Nous reviendrons sur ce canal dans le chapitre 3.

Exemple 2 : théorème du traitement de données.
Supposons que la VA Y soit le résultat d'un traitement déterministe de la VA X de type $Y = f(X)$. Si Y est une fonction de X alors on aura $S - X - Y$ où S représente n'importe quelle VA. Nous avons donc :

$$\begin{aligned} Y = f(X) &\Rightarrow S - X - Y \\ &\Rightarrow I(S;Y|X) = 0 \\ &\Rightarrow I(S;X) \geq I(S;Y). \end{aligned} \tag{1.22}$$

On voit donc qu'un traitement déterministe (ex : décodage) ne fait pas croître l'information mutuelle entre la source S et la grandeur traitée Y par rapport à l'information mutuelle entre la source S et X : c'est le théorème de traitement de données.

1.2.8 Séquences faiblement typiques (typicité entropique)

Lorsque, dans le langage courant, on parle de quelque chose ou de quelqu'un de typique, on suppose souvent implicitement que l'objet ou la personne considérée possède des caractéristiques proches des caractéristiques moyennes de la classe à laquelle il ou elle appartient. Exemple : on dira qu'un suédois est typique en taille si sa taille h est suffisamment proche de la moyenne nationale disons si $182 - 5$ cm $\leq h \leq 182 + 5$ cm. Pour traduire le degré de typicité désiré nous avons dû introduire un paramètre (valant 5 cm ici) qui définit la taille de l'intervalle qui définit si oui ou non un suédois est typique. La notion de séquence faiblement typique s'appuie sur la même idée mais la quantité moyenne en jeu est cette-fois l'entropie.

La définition d'une séquence typique au sens de l'entropie repose sur la loi faible des grands nombres qui est rappelée ci-dessous.

Loi faible des grands nombres
Soit $X_1, ..., X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) dont la distribution commune est notée $p(x)$. On a alors

$$\forall\, \varepsilon > 0, \lim_{n\to\infty} \Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X)\right| > \varepsilon\right] = 0 \tag{1.23}$$

avec $E(X) = \sum_{x\in\mathcal{X}} xp(x)$. Nous pouvons appliquer cette loi à une nouvelle suite de VA définie comme suit : $I_i = -\log p(X_i)$. Rappelons que p est une fonction déterministe. Les VA de cette nouvelle suite sont bien indépendantes et identiquement distribuées selon une même loi à savoir $p(x)$. En appliquant la loi des grands nombres à la suite de VA $I_1, ..., I_n$ nous obtenons :

$$\forall\, \varepsilon > 0, \lim_{n\to\infty} \Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} I_i - E\left[-\log p(X)\right]\right| > \varepsilon\right] = 0.$$

Nous voyons donc que $-\frac{1}{n}\log p(X_1, ..., X_n)$ converge en probabilité vers $H(X)$. C'est sur cette observation-clé que repose la définition de la typicité faible d'une séquence.

Typicité faible (ou entropique)
Soit $(x_1, ..., x_n) = \underline{x}$ une suite i.i.d de réalisations de la VA X de loi $p(x)$. Soit $\varepsilon > 0$. La séquence $\underline{x}$ est dite ε-typique si sa probabilité de réalisation peut être encadrée comme suit :

$$2^{-n[H(X)+\varepsilon]} \leq p(\underline{x}) \leq 2^{-n[H(X)-\varepsilon]} \tag{1.24}$$

où l'entropie $H(X)$ est calculée ici avec un log en base 2.

De par la loi des grands nombres on montre facilement que, pour n grand, une séquence a de forte chance d'être typique et que la probabilité de réalisation d'une telle séquence tend vers $2^{-nH(X)}$. Autrement dit les séquences typiques ont des probabilités de réalisations très proches : c'est la propriété d'équirépartion asymptotique (Cf analogie avec l'hypothèse micro-canonique en thermodynamique statistique). Nous noterons $A_\varepsilon^{(n)}(X)$ l'ensemble des séquences ε-typiques de taille n associé à la VA X de distribution $p(x)$:

$$A_\varepsilon^{(n)}(X) = \left\{ \underline{x} \in \mathcal{X}^n,\ 2^{-n[H(X)+\varepsilon]} \leq p(\underline{x}) \leq 2^{-n[H(X)-\varepsilon]} \right\}. \tag{1.25}$$

On peut montrer que :
– $\forall n \in \mathbb{N}, |A_\varepsilon^{(n)}(X)| \leq 2^{n[H(X)+\varepsilon]}$;
– $\exists n_0,\ \forall n \geq n_0, |A_\varepsilon^{(n)}(X)| \geq (1-\varepsilon)2^{n[H(X)-\varepsilon]}$.
Pour n suffisamment grand on peut donc évaluer la taille de l'ensemble des séquences ε-typiques : $|A_\varepsilon^{(n)}(X)| \simeq 2^{nH(X)}$.

La notion de typicité d'une séquence s'étend facilement à celle de typicité d'un couple de séquences. On dit que deux séquences $\underline{x}$ et $\underline{y}$ sont ε-typiques conjointement si et seulement si :

$$\begin{cases} 2^{-n[H(X)+\varepsilon]} & \leq & p(\underline{x}) & \leq & 2^{-n[H(X)-\varepsilon]} \\ 2^{-n[H(Y)+\varepsilon]} & \leq & p(\underline{y}) & \leq & 2^{-n[H(Y)-\varepsilon]} \\ 2^{-n[H(X,Y)+\varepsilon]} & \leq & p(\underline{x},\underline{y}) & \leq & 2^{-n[H(X,Y)-\varepsilon]} \end{cases}. \tag{1.26}$$

De même que pour la typicité individuelle on peut montrer que, pour n assez grand, le nombre de couples typiques $|A_\varepsilon^{(n)}(X,Y)|$ est environ égal à $2^{nH(X,Y)}$. Nous utiliserons les notions de typicité individuelles et conjointes dans la démonstration de la partie atteignabilité du second théorème de Shannon. La typicité est, en effet, un des outils qui permet de construire des codeurs et décodeurs asymptotiquement optimaux.

Pour bien illustrer la notion de typicité faible considérons le cas d'une VA binaire. Soit une source de Bernoulli X qui produit des symboles indépendants avec les probabilités suivantes $\Pr[X=0] = p = 0,9$ et $\Pr[X=1] = 1-p = 0,1$. Supposons que l'on considère des séquences issues de cette source et de longueur 5. L'entropie de cette source vaut exactement $0,469$ bit/symbole. Ce qui signifie que le nombre de séquences 0,01-typiques devrait être compris entre 4,9075 et 5,2598. Noter que nous exploitons les propriétés de l'ensemble typique alors que rien ne justifie a priori que nous sommes dans le régime asymptotique puisque n=5. Dans le cas de la source de Bernouilli on peut montrer que la condition de typicité entropique équivaut à une condition sur la moyenne empirique des séquences (on rappelle que $\mathbb{E}(X) = p = 0,9$), ce qui revient à dire qu'une séquence typique contient environ $np = 4,5$ uns. On peut vérifier qu'en prenant toutes les séquences de poids 4 et celle de poids 5 l'ensemble typique correspondant contient 91,86 % de la probabilité avec seulement 6 séquences parmi 32. Cela montre bien, que même sur des séquences assez petites, l'ensemble typique s'apparente à un ensemble de forte probabilité. On peut, de même, analyser avec profit la signification de la typicité sur le cas gaussien. On s'apercevrait alors que la condition de typicité entropique se ramène à une condition sur la variance empirique des séquences produites par la source considérée : une séquence de taille donnée serait typique si sa variance empirique est suffisamment proche de sa variance exacte (asymptotique).

1.2.9 Séquences fortement typiques

La notion de typicité faible est suffisante pour démontrer l'atteignabilité des théorèmes de codage pour des canaux tels que le canal mono-utilisateur et le canal à accès multiple. Cependant, pour certains canaux, tels que les canaux avec information adjacente [1] une autre notion de typicité est nécessaire : la typicité forte. On dit qu'une séquence $\underline{x}$ est $\varepsilon-$ fortement typique si :

$$\left| \frac{1}{n} N(a|\underline{x}) - p_X(a) \right| < \varepsilon \tag{1.27}$$

où la notation $N(a|\underline{x})$ indique le nombre d'éléments de la séquence $\underline{x}$ dont la valeur est égale à a.

De même on peut définir la typicité forte d'un couple de séquences. Soit $\underline{x}$ une suite de réalisations de la variable aléatoire de loi p_X. Soit $\underline{y}$ une suite de réalisations de la variable aléatoire Y de loi p_Y. Enfin, notons p_{XY} la loi conjointe du couple de VA (X,Y). On dit qu'un couple de séquences est (conjointement) $\varepsilon-$ fortement typique si :

$$\left|\frac{1}{n}N\left[(a,b)|(\underline{x},\underline{y})\right]-p_{XY}(a,b)\right|<\varepsilon \tag{1.28}$$

où la notation $N\left[(a,b)|(\underline{x},\underline{y})\right]$ indique le nombre de couples (x_i,y_i) coincidant avec le couple (a,b). On notera que, contrairement à la typicité faible, la typicité conjointe forte est spécifiée par une condition et non par trois. On peut, en effet, montrer que les typicités individuelles forte sont une implication de la typicité conjointe forte.

Chapitre 2

Transmission point-à-point

2.1 Notion de capacité

2.1.1 Cas du codage à répétition

Pami les mécanismes de correction d'erreur les plus intuitifs figure le codage à répétion. Par exemple, le professeur qui fait son cours va répéter certaines choses plusieurs fois pour donner plusieurs chances aux destinataires de l'information, les élèves, de comprendre. L'inconvénient est que la répétition a un coût en terme de débit d'information utile transmise. Illustrons ces propos sur un exemple, celui du canal binaire symétrique.

Supposons que l'alphabet des messages soit binaire : $W \in \{0,1\}$. De même, supposons que l'alphabet du signal émis à l'entrée du canal soit binaire : $X \in \{0,1\}$. Le codeur à répétition va associer à chaque message W le mot $\underline{X}$ formé des multiples répliques du message informatif. Par exemple pour $n = 5$:

$$\begin{aligned} 0 &\mapsto 00000 \\ 1 &\mapsto 11111. \end{aligned}$$

Le message "0" sera donc codé par le mot de code "00000". En sortie de canal, le récepteur va avoir une version bruité de ce mot, par exemple 00110. Le décodeur, en appliquant un décodage fondé sur la majorité de "0" ou de "1" va donc pouvoir corriger le mot reçu tant que le nombre d'erreurs est inférieur ou égal à $2 = \lfloor \frac{n-1}{2} \rfloor$. Ainsi, lorsque le nombre de répétition $n-1$ va tendre vers l'infini on est certain de retrouver le message informatif sans erreur après décodage majoritaire, ce que nous notons :

$$\lim_{n\to\infty} P_e^{(n)} = 0.$$

Le taux de codage de ce codeur, qui est mesuré en bit utile par symbole envoyé sur le canal vaut : $R = \frac{\log_2 |\mathcal{W}|}{n} = \frac{1}{n}$. Par conséquent,

$$\lim_{n\to\infty} R = 0.$$

Le codeur à répétition est donc asymptotiquement "fiable" mais a un débit informatif asymptotiquement nul ! Avant les travaux des chercheurs de Bell Labs notamment [2], [3] on pensait que pour avoir une transmission fiable il fallait décroître le débit informatif, le codage à répétition illustre ce mythe du début du 20$^{\text{e}}$ siècle. On peut montrer qu'il existe des codeurs qui permettent de transporter de l'information d'une manière aussi fiable que l'on veut pourvu que la quantité d'information utile par symbole n'excède pas une quantité limite qui est fixée par le canal de transmission. Heureusement cette quantité limite n'est pas nulle, comme le laisserait penser le codage à répétition. Cette quantité limite, qu'on appelle **capacité du canal**, vaut $1 - H_2(p)$ pour le canal binaire symétrique, p étant la probabilité d'erreur du canal.

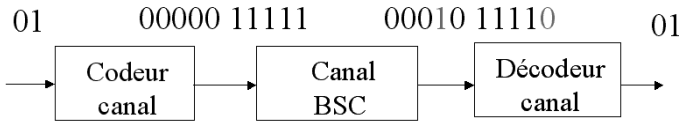

FIG. 2.1

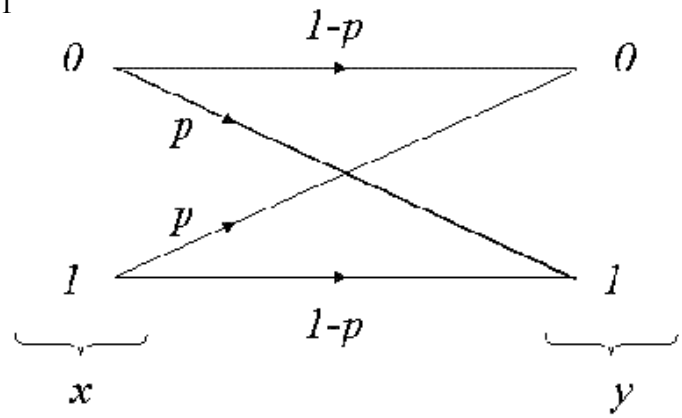

FIG. 2.2 – Canal binaire symétrique (BSC, binary symmetric channel)

2.1.2 Empilement de sphères sur le canal binaire symétrique

A partir d'un raisonnement, simplifié pour des raisons didactiques, nous allons établir la valeur du degré de redondance minimal assurant un décodage fiable pour le canal binaire symétrique (BSC, binary symmetric channel, voir figure 2.2. En effet, l'idée centrale est qu'il faut mettre de la redondance pour supprimer l'ambiguité au décodage. Avant de faire ce raisonnement nous allons ré-interpréter la procédure de codage-décodage pour le codeur à répétition. Pour le codeur à répétition sur le canal binaire symétrique, le rendement de codage vaut $\frac{1}{n}$ et la longueur des mots de code vaut n. Plus précisément $\underline{x} \in \{00...00, 11...11\}$. Ainsi on a seulement deux mots de code possibles alors que dans l'espace $\{0,1\}^n$ on a 2^n séquences possibles. Compte tenu du fait que le décodeur utilise une règle de décision majoritaire tout se passe donc au codage comme si on avait partitionné l'espace $\{0,1\}^n$ en deux parties : l'ensemble des mots de code située à une distance de Hamming[1] inférieure ou égale à $\frac{n-1}{2}$ du mot 00...00 et l'ensemble des mots de code située à une distance[2] (de Hamming) inférieure ou égale à $\frac{n-1}{2}$ du mot 11...11. Géométriquement le codage consiste à regrouper tous les mots de code situés dans la boule de rayon $\frac{n-1}{2}$ et de centre 00...00 ou 11...11, selon le message à transmettre. On voit donc que la contrainte de décodage "fiable" impose une certaine taille de boule, c'est le coût en terme de redondance de la suppression de l'ambiguité de décodage, et aussi que le nombre de boules correspond au nombre de messages que l'on peut transmettre sans ambiguité. Nous allons utiliser cette interprétation pour construire un codage plus performant dont la taille des boules n'est pas de $\frac{n-1}{2}$ mais est optimalement choisie en fonction du canal.

Nous supposons n grand. La sortie d'un canal BSC s'écrit $\underline{y} = \underline{x} \oplus \underline{e}$ où la notation $\oplus$ indique l'opération ou exclusif (XOR). Pour n assez grand, le mot d'erreur $\underline{e}$ contiendra un nombre de uns sensisblement égal à np, p étant la probabilité de transition du canal BSC. On est donc presque certain que le mot de code reçu $\underline{y}$ se situera dans la boule de centre $\underline{x}$ (i.e. le mot de code émis) et de rayon np (voir figure 2.3). Un codage possible permettant un décodage non ambigu est le suivant :

[1] La distance de Hamming entre deux vecteurs $\underline{x}$ et $\underline{y}$ est égale au poids du vecteur somme modulo 2 de ces deux vecteurs : $d_H(\underline{x},\underline{y}) = w(\underline{x} \oplus \underline{y})$

[2] On suppose ici implicitement que n est impair sachant qu'en pratique le cas n pair ne permet pas faire une décision majoritaire lorsque le nombre de uns et de zéros sont égaux.

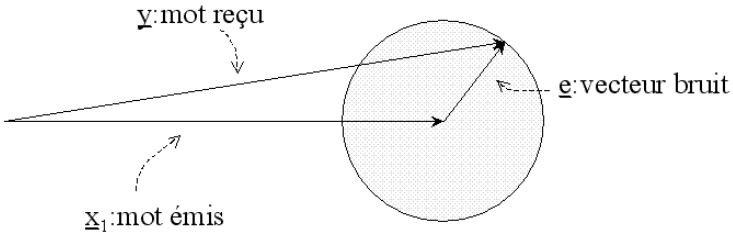

FIG. 2.3 – Codage optimal sur le canal BSC

– pour un message donné ($W \in \{1,...,M\}$), disons $W = 1$, choisir un mot de code $\underline{x}_1$;
– associer au message $W = 1$ tous les mots de code contenus dans la boule de centre $\underline{x}_1$ et de rayon np ;
– effectuer cette procédure pour tous les messages de $\{1,...,M\}$ où M est le nombre maximal de boules de rayon np ayant une intersection nulle asymptotiquement que l'on peut avoir dans $\{0,1\}^n$. Le nombre de bits d'information par symbole envoyé sur le canal atteint par cette stratégie vaudra alors exactement $\frac{\log_2 M}{n}$;
– le décodeur a juste à savoir à quelle boule appartient $\underline{y}$ pour remonter au message émis.

Supposons que pour $n \to \infty$ les différentes boules de rayon np sont choisies tel qu'elles ne s'intersectent pas et évaluons le nombre de boules que l'on peut "stocker" dans l'espace $\mathcal{Y}^n = \{0,1\}^n$. Pour cela il nous faut connaître le volume d'une boule de Hamming de rayon np. On peut facilement vérifier que le nombre de vecteurs contenus dans une boule de rayon k est donné par :

$$\mathrm{Vol}(\mathcal{B}(\underline{0};k)) = |\mathcal{B}(\underline{0};k)| = \sum_{i=0}^{k} \left(\begin{array}{c} n \\ i \end{array} \right) = \sum_{i=0}^{k} \frac{n!}{i!(n-i)!}.$$

Nous allons donner une approximation de ce volume pour n grand. Posons $\lambda_i = \frac{i}{n}$ et $\overline{\lambda}_i = 1 - \lambda_i$ et approximons tout d'abord $\dfrac{n!}{i!(n-i)!}$ en utilisant la formule de Stirling. Rappel :

$$\ln(n!) \underbrace{\sim}_{n\to\infty} n\ln n - n.$$

Par conséquent, si nous supposons à la fois n et k grands, nous avons

$$\begin{aligned}
\ln\left[\frac{n!}{k!(n-k)!}\right] &= \ln(n!) - \ln[(\lambda_k n)!] - \ln[(\overline{\lambda_k} n)!] \\
&\sim n\ln n - n - [\lambda_k n \ln(\lambda_k n) - \lambda_k n] - \left[\overline{\lambda_k} n \ln(\overline{\lambda_k} n) - \overline{\lambda_k} n\right] \\
&= n\ln n - n - \lambda_k n \ln\lambda_k - \lambda_i n \ln n + \lambda_k n - \overline{\lambda_k} n \ln\overline{\lambda_k} - \overline{\lambda_k} n \ln n + \overline{\lambda_k} n \\
&= n\ln n - n - (\lambda_k + \overline{\lambda_k}) n \ln n + (\lambda_k + \overline{\lambda_k}) n - n(\lambda_k \ln\lambda_k + \overline{\lambda_k} \ln\overline{\lambda_k}) \\
&= -n(\lambda_k \ln\lambda_k + \overline{\lambda_k} \ln\overline{\lambda_k}),
\end{aligned}$$

ce qui se réécrit

$$\left(\begin{array}{c} n \\ k \end{array} \right) \sim e^{[nH(\lambda_k)]}.$$

Comme on suppose $n \to \infty$ et $k \sim np$ on peut montrer que le terme dominant de la somme qui intervient dans l'expression du volume d'une boule de Hamming est le dernier terme i.e. le terme d'indice $i = k$: ceci est dû au comportement exponentiel des termes. Finalement nous avons :

$$|\mathcal{B}(\underline{0};k)| \sim e^{nH(\lambda_k)} = 2^{nH_2(\lambda_k)}.$$

Il est intéressant de voir que l'entropie est une grandeur qui intervient naturellement dans le volume des grandes boules de Hamming. Dans le cas du canal BSC le rayon de la boule est np et donc $\lambda_k = p$. Nous

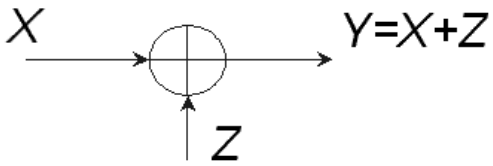

FIG. 2.4 – Canal AWGN

connaissons le volume de l'espace des mots de code reçus $|\mathcal{Y}^n| = 2^n$, nous connaissons le volume d'une "boule d'erreur"[3] $|\mathcal{B}(\underline{x}; np)| \sim 2^{nH_2(\lambda_k)}$. Par conséquent le nombre maximal de boules qui ne s'intersectent pas est $M = 2^{n[1-H_2(p)]}$ ou encore

$$\frac{\log_2 M}{n} = 1 - H_2(p).$$

On peut montrer que ce codage fiable est optimal et que le taux de codage maximal pour le BSC est donné par la valeur ci-dessus qui représente donc la capacité du canal :

$$C = 1 - H_2(p). \tag{2.1}$$

2.1.3 Empilement de sphères sur le canal gaussien

On peut refaire le raisonnement précédent pour un autre canal très utilisé le canal à bruit additif, blanc (la densité spectrale de puissance est constante), et gaussien (la densité de probabilité du bruit est une gaussienne). L'acronyme courant est canal AWGN (additive white Gaussian noise), voir figure 2.4

La sortie d'un canal AWGN s'écrit sous la forme : $Y = X + Z$ soit encore $\underline{y} = \underline{x} + \underline{z}$ dans un contexte de codage en blocs avec : $(\underline{x}, \underline{y}, \underline{z}) \in (\mathbb{R}^n)^3$, $E(X^2) = P$ et $Z \sim \mathcal{N}(0, N)$. Le volume d'une boule de rayon ρ dans $\mathbb{R}^n$ est donné par :

$$\text{Vol}(\mathcal{B}(\underline{0}; \rho)) = |\mathcal{B}(\underline{0}; \rho)| = \frac{(\pi\rho^2)^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)},$$

où Γ est la fonction Gamma qu'il est inutile de définir ici. On suppose, comme pour le canal BSC, que n est grand. Le rayon de la "boule de bruit" associée au vecteur $\underline{z}$ est environ égal à $\sqrt{nN}$. Le rayon de la boule qui délimite l'espace dans lequel vivent les vecteurs de sortie du canal est environ égal à $\sqrt{n(P+N)}$. Le nombre maximal de boules de rayon $\sqrt{nN}$ et qui ne s'intersectent pas, que l'on peut stocker dans l'espace de sortie est donc environ égal à :

$$\frac{|\mathcal{B}(\underline{0}; \sqrt{n(P+N)})|}{|\mathcal{B}(\underline{0}; \sqrt{nN})|} = \frac{(\pi nN)^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)} \frac{\Gamma\left(\frac{n}{2}+1\right)}{[\pi n(P+N)]^{n/2}} = \left(\frac{P+N}{N}\right)^{n/2} = M.$$

Au final le taux de codage maximal s'exprime donc par :

$$\frac{\log_2 M}{n} = C = \frac{1}{2}\log_2\left(1 + \frac{P}{N}\right). \tag{2.2}$$

Dans le cas des signaux complexes ($(X, Y, Z) \in \mathbb{C}^3$) il n'y a pas le facteur $\frac{1}{2}$ devant le logarithme.

2.1.4 Préliminaire au second théorème de Shannon

Le théorème de Shannon va généraliser les deux exemples que nous avons traités ci-dessus. L'idée centrale reste cependant la même : mettre de la redondance en quantité suffisante au codage pour éviter

[3] Cette boule représente ce qu'on appelle l'ensemble de décodage associé au mot de code émis.

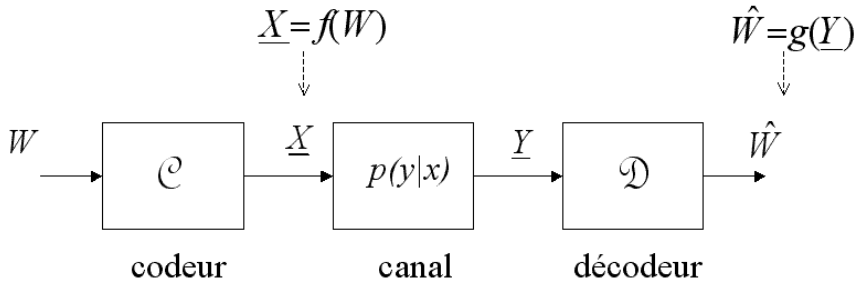

FIG. 2.5 – Canal de Shannon

l'ambiguïté au décodage. En raisonnant comme nous l'avons fait, nous montrerons que le nombre de séquences de sortie utiles du canal sera de l'ordre de $2^{nH(Y)}$ et que le coût en redondance de la suppression de l'ambiguïté de décodage est de l'ordre de $2^{nH(Y|X)}$. Au final on a bien un taux de codage de l'ordre de :

$$\frac{\log_2 M}{n} = \frac{1}{n}\log_2\left(\frac{2^{nH(Y)}}{2^{nH(Y|X)}}\right) = H(Y) - H(Y|X).$$

2.2 Définition du canal de Shannon

Le paradigme de Shannon [2], qui a été une avancée majeure en son temps, consiste à décomposer une transmission en 3 étapes auxquelles correspondent trois organes fondamentaux :
– la source ou l'émetteur ;
– le canal ou le médium de transmission ;
– et le destinataire ou le récepteur.
La source étant par nature aléatoire (puisqu'elle contient de l'information) est représentée par une variable aléatoire X. De même l'image de cette source à la sortie du canal est représentée par une variable aléatoire Y. Très naturellement le médium de transmission est représenté par sa probabilité de transition $p(y|x)$ qui caractérise la sortie pour une entré donnée. Bien souvent, par abus de langage, on dit que le canal de Shannon est simplement caractérisé par sa probabilité de transition $p(y|x)$. En fait, la définition de Shannon est plus précise et plus restrictive que cela puisqu'elle spécifie aussi l'émetteur et le récepteur. C'est l'objet de notre propos ici pour lequel la figure 2.5 nous sert de base.

1. La source d'information, représentée par la VA $W \in \{1,...,M\}$, est supposée *uniforme* : $\Pr[W = 1] = ... = \Pr[W = M] = \frac{1}{M}$. Étant donné que l'on veut exprimer la capacité du canal on se place dans la cadre d'un transfert optimal de l'information : l'entropie de la VA W est maximale sous l'hypothèse énoncée. Dans l'autre cas extrême de la source déterministe le canal ne véhiculerait aucune information.
2. Tout d'abord on suppose que $(X,Y) \in \mathcal{X} \times \mathcal{Y}$ avec $|\mathcal{X}| < \infty$ et $\mathcal{Y} < \infty$ c'est-à-dire que l'entrée et la sortie du canal appartiennent à des alphabets *discrets et finis*. Malgré cela, la formule de capacité du canal discret peut s'appliquer au cas des signaux continus grâce à des arguments classiques de discrétisation [4].
3. Le canal est supposé *sans mémoire et sans retour d'information*. Le caractère sans mémoire se traduit par
$$p(y_i|x_1,...,x_i,y_1,...,y_{i-1}) = p(y_i|x_i).$$
La sortie actuelle du canal ne dépend donc pas du passé, ce qu'on peut réécrire aussi sous forme de chaîne de Markov : $Y_i - X_i - (X_1,...,X_{i-1},Y_1,...,Y_{i-1})$. L'absence de retour d'information signifie que le codeur ne prend pas en compte les sorties du canal :
$$p(x_i|x_1,...,x_{i-1},y_1,...,y_{i-1}) = p(x_i|x_1,...,x_{i-1}). \tag{2.3}$$

De manière équivalente $X_i - (X_1,...,X_{i-1}) - (Y_1,...,Y_{i-1})$. En combinant ces deux propriétés on peut montrer que :

$$p(\underline{y}|\underline{x}) \triangleq p(y_1,...,y_n|x_1,..,x_n) = \prod_{i=1}^{n} p(y_i|x_i).$$

Pour montrer cette relation il suffit de développer $p(\underline{x},\underline{y})$:

$$\begin{aligned}
p(\underline{x},\underline{y}) &= p(x_1,y_1,...,x_n,y_n) \\
&= p(x_n,y_n|x_1,y_1,...,x_{n-1},y_{n-1})p(x_1,y_1,...,x_{n-1},y_{n-1}) \\
&= \prod_{i=1}^{n} p(x_i,y_i|x^{i-1},y^{i-1}) \\
&= \prod_{i=1}^{n} p(x_i|x^{i-1},y^{i-1})p(y_i|x_i x^{i-1},y^{i-1}) \\
&\overset{(a)}{=} \prod_{i=1}^{n} p(x_i|x^{i-1},y^{i-1})p(y_i|x_i) \\
&\overset{(b)}{=} \prod_{i=1}^{n} p(x_i|x^{i-1})p(y_i|x_i)
\end{aligned}$$

où pour (a) on utilise l'hypothèse "sans mémoire", pour (b) on utilise l'hypothèse "sans retour d'information" et enfin la notation v^i indique : $v^i = (v_1,...,v_i)$. En utilisant la dernière égalité on voit que :

$$\begin{aligned}
p(\underline{y}|\underline{x}) &= \frac{p(\underline{x},\underline{y})}{p(\underline{x})} \\
&= \frac{\prod_{i=1}^{n} p(x_i|x^{i-1})p(y_i|x_i)}{\prod_{i=1}^{n} p(x_i|x^{i-1})} \\
&= \prod_{i=1}^{n} p(y_i|x_i).
\end{aligned}$$

Par abus de langage on dit qu'un canal discret est sans mémoire si la condition (2.3) est vérifiée, sans repréciser l'absence de retour d'information de la part du récepteur à l'attention du codeur. Enfin on remarquera que nous avons supposé que le canal est implicitement *stationnaire* c'est-à-dire que les probabilités de transition ne dépendent pas du temps.

4. Le codeur est caractérisé par sa fonction de codage $\underline{X} = f(W)$ et son taux de codage

$$R \triangleq \frac{\log_2 M}{n}. \tag{2.4}$$

On notera qu'il s'agit d'un *codeur en bloc* alors qu'a priori on cherche la limite ultime du taux de codage pour n'importe quel type de codeur, convolutif par exemple. Nous verrons que cette restriction ne fait rien perdre en optimalité. De plus, il faut noter qu'en pratique les codeurs convolutifs peuvent être considérés commes des codeurs en blocs à cause de la fermeture du treillis associé. Le décodeur est aussi caractérisé par sa fonction de décodage : $\hat{W} = g(\underline{Y})$.

5. Enfin et ceci est fondamental, le canal est aussi défini par la *probabilité d'erreur au décodage* :

$$P_e^{(n)} \triangleq \Pr[\hat{W} \neq W]. \tag{2.5}$$

Il est important de noter que la valeur de la capacité est directement liée à la définition de la probabilité d'erreur. On notera que Shannon a défini la capacité par rapport à une probabilité d'erreur par bloc et non par symbole ou par bit.

Dans ce contexte bien précis nous pouvons énoncer le deuxième théorème de Shannon.

2.3 Deuxième théorème de Shannon

2.3.1 Énoncé du théorème

Théorème *(théorème de codage pour le canal de Shannon). Soit* $C = \max_{p(x)} I(X;Y)$. *Si* $R < C$ *alors il existe un codage qui permet d'avoir une probabilité d'erreur de décodage évanescente lorsque n tend vers l'infini. Inversement, s'il existe un code dont la probabilité d'erreur de décodage est évanescente alors nécessairement* $R \leq C$.

2.3.2 Commentaires

Comme nous le voyons, ce théorème comporte deux résultats : l'atteignabilité ($\Rightarrow$) et la réciproque ($\Leftarrow$). Tout comme pour le codage de source pour lequel le codage par séquences typiques montre (voir version 1.0 du polycopié)la possibilité d'atteindre la limite ultime du taux de compression sans perte et que l'inégalité de Kraft montre qu'intrinsèquement un codage sans préfixe (pour éviter toute ambiguité au décodage) ne peut pas avoir un taux inférieur à l'entropie, le codage de canal peut être analysé, au travers du théorème de codage de canal, de la même manière. Le premier résultat du théorème de codage est appelé *atteignabilité* car on dit qu'un taux de codage est atteignable lorsque la probabilité de décodage est aussi petite que l'on veut pour n suffisamment grand. Le second est appelé *réciproque* et affirme qu'il ne peut y avoir de codage ayant un taux supérieur à la capacité ayant une probabilité de décodage évanescente avec n. Avant de démontrer ce théorème nous donnerons les principales étapes et les raisonnements-clé. Avant cela commentons les résultats du théorème.

Tout d'abord ce théorème se veut rassurant lorsqu'on considère le cas particulier du codage à répétition. En effet nous avons vu qu'en faisant un décodage majoritaire, plus la taille des mots de code augmentait et plus la probabilité d'erreur de décodage s'approchait de zéro. Le codage à répétition permettait bien d'obtenir une probabilité d'erreur évanescente avec n mais d'un autre côté le rendement de codage tendait vers zéro également. Le second théorème de Shannon prédit l'existence de codes dont le taux peut valoir la capacité, a priori différente de zéro, et qui pourtant produisent une probabilité d'erreur de décodage évanescente. Le théorème affirme donc bien que pour un degré de redondance limité il est possible d'éviter toute ambiguïté de décodage en réception. En ce qui concerne le rôle de l'information mutuelle, nous reviendrons en détails sur celui-ci mais nous pouvons, avec les connaissances acquises jusqu'alors essayer de comprendre la présence de cette quantité de la manière suivante [4]. Il s'avère en effet que les performances limites d'un canal peuvent être atteintes, en théorie, en utilisant un codage basé sur la typicité. En effet l'idée est d'utiliser un codeur qui envoie des séquences typiques sur le canal. Le récepteur est alors censé recevoir des séquences typiques en sortie de canal. Le décodeur devine alors la séquence envoyée en cherchant l'unique séquence ou mot de code qui permet d'assurer la condition de typicité conjointe entre la séquence estimée et la séquence reçue. Or il y a environ $2^{nH(Y)}$ séquences typiques dans $\mathcal{Y}^n$ (Cf chapitre précédent). De plus, on peut montrer que pour une séquence typique d'entrée de canal donnée, il y a environ $2^{nH(Y|X)}$ séquences typiques possibles en sortie. Pour éviter toute ambiguïté de décodage on voit qu'on peut faire au plus $\frac{2^{nH(Y)}}{2^{nH(Y|X)}}$ sous-groupes distincts dans l'ensemble des séquences de sortie $\mathcal{Y}^n$. Puisqu'on a associé une séquence d'entrée à chaque sous-groupe on voit que l'on peut avoir un taux de codage pouvant aller jusqu'à $\frac{1}{n}\log_2\left(\frac{2^{nH(Y)}}{2^{H(Y|X)}}\right)$.

2.3.3 Démonstration de l'atteignabilité (résumé)

Il s'agit de montrer l'existence d'un code qui atteint le taux de codage annoncé avec une probabilité d'erreur aussi petite que l'on veut. Pour cela nous allons d'abord expliciter la stratégie de codage adoptée [2] et ensuite montrer que les probabilités d'erreur de codage et décodage sont évanescentes avec n.

Stratégie de codage
On génère des séquences selon une distribution de probabilité donnée $p(x)$ i.e. à partir d'une source X. On se restreint à un codage pour lesquels les composantes des mots de codes sont i.i.d. Cela va permettre d'exploiter la loi des grands nombres et il s'avère que cette hypothèse ne fait rien perdre en optimalité. On ne retient que les séquences qui sont typiques pour coder les messages de la source d'information W. Ce codage est dit aléatoire en ce sens que, pour un message donné, la réalisation de la séquence importe peu dans le codage, pourvu que la séquence retenue soit typique : dans le codage de Shannon, le dictionnaire des mots de code est tiré aléatoirement, on fait l'association message - mot de codes, le récepteur est informé du code utilisé par l'émetteur, on retire un nouveau dictionnaire, on réinforme le récepteur et ainsi de suite. Chaque message voit donc tous les mots de codes possibles et il n'y a pas un message en particulier qui génère plus d'erreur que les autres. On note que le codage reste déterministe en ce sens qu'à chaque message on associe une seule séquence et le récepteur connaît cette association à chaque tirage de dictionnaire, sans quoi le décodage serait ambigu. On peut montrer que les séquences en sortie de canal sont typiques. La règle de décodage est la suivante : on suppose qu'il n'y a qu'une seule séquence d'entrée du canal qui peut être conjointement typique avec celle reçue. On énumère toutes les séquences possibles d'entrée et on garde celle qui vérifie la condition de décodage énoncée.

Énumération des erreurs de codage et décodage
- La première erreur possible est au codage. En effet, on déclare une erreur lorsqu'en générant une séquence on obtient une séquence qui n'est pas typique. On peut montrer que la probabilité de cet événement est inférieure ou égal à ε , ce dernier pouvant être aussi petit que l'on veut lorsque n tend vers l'infini.
- La deuxième erreur possible est en réception, avant même décodage. En effet, on déclare une erreur lorsque la séquence de sortie considérée n'est pas typique. La probabilité de cet événement décroît également exponentiellement vers 0 lorsque n tend vers l'infini.
- En sortie de décodeur on déclare une erreur lorsqu'il existe une séquence d'entrée, autre que celle émise, qui est conjointement typique avec la séquence de sortie du canal. Sans perte de généralité supposons que ce soit le message numéro 1 qui ait été émis (parmi M messages possibles). On note E_i l'événement qui consiste à trouver la séquence $\underline{x}(i)$ conjointement typique avec $\underline{y} = \underline{y}(1)$. On déclare donc une erreur lorsque $\underline{x}(1)$ n'est pas conjointement typique avec $\underline{y}(1)$ et lorsque pour $i \geq 2$ la séquence $\underline{x}(i)$ est conjointement typique avec $\underline{y}$. La probabilité d'erreur de décodage s'exprime alors comme suit :

$$P_e^{(n)} = P(\overline{E}_1) + \sum_{i=2}^{M} P(E_i).$$

On peut majorer cette probabilité d'erreur simplement [4] :

$$\begin{aligned} P_e^{(n)} &= P(\overline{E}_1) + \sum_{i=2}^{M} P(E_i) \\ &\leq \varepsilon + P(\overline{E}_1) + \sum_{i=2}^{M} P(E_i) \\ &\leq \varepsilon + (M-1) \times 2^{-n[I(X;Y)-3\varepsilon]} \\ &= \varepsilon + (2^{nR}-1) \times 2^{-n[I(X;Y)-3\varepsilon]} \\ &\leq \varepsilon + 2^{nR} \times 2^{-n[I(X;Y)-3\varepsilon]} \\ &\leq \varepsilon + 2^{-n[I(X;Y)-R-3\varepsilon]}. \end{aligned}$$

Il est donc clair que si $R < I(X;Y) - 3\varepsilon$ alors la probabilité d'erreur de décodage est évanescente avec n. La démonstration reste vraie pour un choix de la distribution de probabilité d'entrée de canal qui maximise l'information mutuelle d'où la condition recherchée $R < C - 3\varepsilon$ avec $\lim_{n\to\infty} \varepsilon = 0$.

2.3.4 Démonstration de la réciproque (faible)

Nous voulons montrer que si un code a une probabilité évanescente lorsque n augmente alors le taux de ce code vérifie nécessairement : $R \leq C$. Pour cela l'outil clé est l'inégalité de Fano.

Tout d'abord faisons apparaître le taux de codage R. Comme nous l'avons fait pour l'atteignabilité, nous supposons que les messages de la source sont équiprobables. Donc $H(W) = \log_2 M = nR$. Puisqu'il faudra relier ce taux à $I(X;Y)$ nous écrivons que $I(W;\underline{Y}) = H(W) - H(W|\underline{Y})$. Nous utilisons tout le vecteur reçu car le décodage d'un message donné nécessite la connaissance de tout le vecteur reçu. Donc $nR = H(W|\underline{Y}) + I(W;\underline{Y})$. Il faut donc analyser les deux termes de cette somme. Pour cela nous introduisons la probabilité d'erreur de décodage. Nous introduisons la VA E avec $E = 0$ lorsque $\hat{W} = W$ et $E = 1$ sinon. Par définition de E : $H(E|W,\underline{Y}) = 0$. L'astuce est de développer la quantité $H(W,E|\underline{Y})$. Nous avons :

$$\begin{aligned} H(W,E|\underline{Y}) &= H(W|\underline{Y}) + H(E|\underline{Y},W) \\ &= H(E|\underline{Y}) + H(W|\underline{Y},E). \end{aligned}$$

Le premier terme de la somme s'exprime donc comme suit $H(W|\underline{Y}) = H(E|\underline{Y}) + H(W|\underline{Y},E)$, terme que nous pouvons majorer :

$$\begin{aligned} H(W|\underline{Y}) &= H(E|\underline{Y}) + H(W|\underline{Y},E) \\ &\leq H(E) + H(W|\underline{Y},E) \\ &= H_2(P_e^{(n)}) + \sum_e \Pr[E=e] H(W|E=e,\underline{Y}) \\ &= H_2(P_e^{(n)}) + (1-P_e^{(n)}) \underbrace{H(W|E=0,\underline{Y})}_{0} + P_e^{(n)} H(W|E=1,\underline{Y}) \\ &\leq H_2(P_e^{(n)}) + P_e^{(n)} \log_2(|\mathcal{W}|-1) \end{aligned}$$

Le second terme se majore comme suit :

$$\begin{aligned} I(W;\underline{Y}) &= H(\underline{Y}) - H(\underline{Y}|W) \\ &\leq \sum_{i=1}^{n} H(Y_i) - H(\underline{Y}|W) \\ &= \sum_{i=1}^{n} H(Y_i) - H(Y_i|W,Y^{i-1}) \\ &\leq \sum_{i=1}^{n} H(Y_i) - H(Y_i|W,Y^{i-1},X_i) \\ &= \sum_{i=1}^{n} H(Y_i) - H(Y_i|X_i) \\ &= \sum_{i=1}^{n} I(X_i;Y_i) \\ &\leq n \max_i I(X_i;Y_i) \\ &\leq nC \end{aligned} \tag{2.6}$$

où nous avons utilisé la chaîne de Markov suivante qui est vraie pour un canal sans mémoire et sans

retour de l'information : $(W, Y^{i-1}) - X_i - Y_i$. En effet,

$$\begin{aligned} p(\underline{y}|\underline{x}) &= \prod_{i=1}^{n} p(y_i|x_i) \\ \Rightarrow \quad p(y_1^{i-1}, y_{i+1}^{n}|\underline{x}) &= \sum_{y_i} \prod_{i=1}^{n} p(y_i|x_i) = \prod_{k\neq i} p(y_k|x_k) \\ \Rightarrow \quad p(y_i|\underline{x}, y^{i-1}, y_{i+1}^{n}) &= \frac{p(y_i^n|\underline{x})}{p(y^{i-1}, y_{i+1}^{n}|\underline{x})} = p(y_i|x_i) \end{aligned}$$

avec $\underline{x} = f(W)$ et $W = g(\underline{x})$. Au final nous voyons donc que s'il existe un codeur (dont la taille des mots de code n est a priori quelconque) tel que $P_e^{(n)} \to 0$ alors nécessairement $nR \leq nC$, ce qui est le résultat recherché.

2.4 Exploitation du second théorème de Shannon

Comme nous l'avons déjà souligné un théorème de codage pour un canal donné apporte au moins deux types d'informations : des idées et concepts de codage via la démonstration de l'atteignabilité et une caractérisation des performances limites via la réciproque qui assure qu'on ne peut pas faire mieux que le taux de codage trouvé dans l'atteignabilité.

L'atteignabilité nous donne au moins les renseignements suivants :

- les séquences ou mots de code utilisés par le codeur de Shannon ne sont pas structurés contrairement à un codage de Hamming par exemple. Comme le mot qu'on associe à un message d'information donné est choisi au hasard, on parle de codage *aléatoire*. L'activité de conception de bons codes qui a suivi les travaux de Shannon a longtemps été dominée par le codage algébrique qui lui est très structuré. Par contre, des codes plus récents tels que les turbo-codes [5] sont plus conformes à l'idée de codage aléatoire, la présence de l'entrelaceur au codeur va dans ce sens. En pratique l'avantage majeur de la structuration du codage est de diminuer la complexité de décodage. Les turbo-codes sont à la fois plus conformes à l'idée de codage aléatoire et possédent une structure qui permettent un décodage d'une complexité raisonnable.
- au codage et au décodage la notion de typicité est largement exploitée. Or, pour assurer la typicité individuelle ou conjointe on doit avoir des mots de code *longs*. La théorie semble donc indiquer que l'on peut obtenir de bons codes en prenant des mots de code longs ;
- pour s'affranchir de l'ambiguité au décodage il faut introduire de la *redondance*. Pour un médium de transmission parfait $Y \equiv X$ on pourrait transmettre $M_0 = |\mathcal{X}|^n = 2^{nH(X)}$ messages mais sur un canal quelconque on doit se restreindre à envoyer $M = \dfrac{2^{nH(X)}}{2^{nH(X|Y)}}$ à cause de l'ambiguïté sur $\underline{X}$ sachant $\underline{Y}$.

Nous allons maintenant exploiter l'expression de la capacité proprement dite pour le canal BSC puis pour le canal AWGN.

2.4.1 Capacité du canal BSC

Le canal BSC, bien que simple, existe toujours dans les systèmes de communications numériques puisqu'il y a toujours dans l'émetteur un point de mesure où la tension ne prend que deux valeurs (ex : 0V et 5V) et de même au récepteur. On note p le paramètre du canal BSC : $\Pr[Y = 1|X = 0] = \Pr[Y = 0|X = 1] = p$ et $\Pr[Y = 0|X = 0] = \Pr[Y = 1|X = 1] = 1 - p = \overline{p}$. On note également $\Pr[X = 0] = q$ et $\Pr[X = 1] = 1 - q = \overline{q}$. On veut calculer $I(X;Y)$. Comme le canal est donné il est préférable de calculer $H(Y|X)$ au lieu de $H(X|Y)$. On utilise donc l'expression suivante de l'information mutuelle : $I(X;Y) = H(Y) - H(Y|X)$. Pour calculer $p(y)$ on utilise la formule de Bayes marginalisée par rapport à

$x : p(y) = \sum_x p(y|x)p(x)$. L'expression de $H(Y)$ en fonction de p et q s'en déduit. Pour calculer $H(Y|X)$ on applique la définition : $H(Y|X) = -\sum_{(x,y)} p(y|x)p(x)\log_2 p(y|x)$. On trouve que $H(Y|X) = H_2(p)$. Pour obtenir la capacité du canal il suffit de maximiser $I(X;Y)$ par rappport à q. On trouve alors que pour $q = \frac{1}{2}$:

$$\max_q I(X;Y) = C = 1 - H_2(p).$$

Prenons un exemple d'utilisation de cette formule. Supposons qu'on puisse calculer ou estimer les performances en termes de taux d'erreurs binaire (TEB) brut d'un sytème sans codage. Cela nous donne une probabilité d'erreur qui correspond au paramètre p. On en déduit alors C qui va représenter le meilleur rendement possible qu'un bon codeur en bloc (assurant une transmission aussi fiable que l'on veut après décodage) peut atteindre. En choisissant un codeur en bloc donné on peut donc évaluer immédiatement la perte en débit d'information associée. On remarquera qu'on bon codeur sur un canal BSC doit être capable de corriger np erreurs tout en véhiculant nC bits d'information par mot de code. Le canal BSC donne donc une indication des performances limites pour les systèmes qui utilisent une décision ferme avant décodage, ce qui est en fait sous-optimal (il y a perte irréversible d'information). C'est pourquoi le canal AWGN est un meilleur modèle en termes de performances puisqu'il permet de prendre en compre le cas des décisions dites souples.

2.4.2 Capacité du canal BEC

Un autre cas particulier du canal discret est le canal à effacement binaire (BEC, binary erasure channel). Pour ce canal, l'alphabet d'entrée est binaire tout comme le canal BSC : $\mathcal{X} = \{0,1\}$. Cependant, les symboles d'entrées sont soit reçus sans erreur soit non reçus. L'alphabet de sortie du canal est ternaire : $\mathcal{Y} = \{0,1,e\}$. Lorsque $Y = e$ on dit que le symbole émis a été effacé. En appelant p la probabilité d'effacement nous avons $\Pr[Y = 1|X = 1] = \Pr[Y = 0|X = 0] = 1 - p$ et $\Pr[Y = e|X = 0] = \Pr[Y = e|X = 1] = p$. En apliquant la formule de la capacité donnée par le théorème de Shannon, tout comme nous l'avons fait pour le canal BSC on montre facilement que la capacité du canal à effacment s'exprime par :

$$C = 1 - p. \tag{2.7}$$

Comparativement aux modèles BSC et AWGN, ce modèle de canal a été peu utilisé pendant longtemps. Ce n'est que depuis quelques années que ce canal a subi un regain d'intérêt. En effet, le "quantum d'information" échangé étant le paquet de données dans de nombreux systèmes de transmission, le pourcentage de paquet reçus sans erreur apparaît comme un des critères naturels de performance dans ces systèmes. Ainsi, un paquet est vu du récepteur comme un symbole équivalent qui est soit juste soit faux, c'est-à-dire effacé. En pratique, on peut par exemple utiliser des codes de détection d'erreur (e.g. CRC, cyclic redundancy codes) pour savoir si un paquet est juste ou faux, cette connaissance n'étant pas parfaite. Dans ce contexte, on comprend le récent interêt pour ce type de canaux (voir e.g. [6, 7]) dont la capacité est plus facile à obtenir que le cas général (discret).

2.4.3 Capacité du canal AWGN

Nous avons déjà vu que la sortie d'un canal AWGN s'écrit : $Y = X + Z$. Le bruit Z est supposé gaussien $Z \sim \mathcal{N}(0,N)$ et indépendant du signal utile. On pourrait montrer qu'il s'agit, pour une distribution d'entrée et variance de bruit données, de la pire distribution de bruit [8]. Il s'avère que cette hypothèse de gaussiannité facile grandement les calculs de grandeurs telle que l'information mutuelle. Dans ces conditions, on peut montrer que la distribution optimale pour X est une gaussienne. Nous supposerons

donc que $X \equiv X^* \sim \mathcal{N}(0,P)$. On calcule donc l'information mutuelle en supposant que tout est gaussien, ce qui nous conduit directement à la capacité. On vérifie que :

$$\begin{aligned} H(Y^*) &= \frac{1}{2}\log_2\left[2\pi e(P+N)\right] \\ H(Y^*|X^*) &= \frac{1}{2}\log_2\left[2\pi eN\right] \end{aligned}$$

où nous avons utilisé

$$\begin{aligned} H(Y^*|X^*) &= \int_{x\in\mathbb{R}} f_{X^*}(x) H_d(Y^*|x)dx \\ &= -\int_{x\in\mathbb{R}} f_{X^*}(x) \int_{y\in\mathbb{R}} f_{Y^*|X^*}(y|x)\log_2 f_{Y^*|X^*}(y|x)dydx \\ &= -\int_{x\in\mathbb{R}} f_{X^*}(x) \int_{y\in\mathbb{R}} f_Z(y-x)\log_2 f_Z(y-x)dydx \\ &= -\int_{x\in\mathbb{R}} f_{X^*}(x) \int_{z\in\mathbb{R}} f_Z(z)\log_2 f_Z(z)dzdx \\ &= -\int_{x\in\mathbb{R}} f_{X^*}(x)dx \times \int_{z\in\mathbb{R}} f_Z(z)\log_2 f_Z(z)dz \\ &= H_d(Z). \end{aligned}$$

On retrouve donc que :

$$C = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right).$$

Nous allons maintenant répondre à deux questions auxquelles la formule de capacité ci-dessus donne des éléments quantitatifs.

Quelle est la puissance minimale d'émission pour avoir une transmission fiable ?
La puissance d'un symbole est $E(X^2) = P$. Étant donné qu'on veut connaître la valeur minimale d'énergie en émission on se place aux limites du canal : $R = C$. Chaque symbole véhicule donc C bits. Si on appelle P_b la puissance par bit d'information a bien : $P = CP_b$. Par conséquent on peut écrire :

$$C = \frac{1}{2}\log_2\left(1+\frac{CP_b}{N}\right).$$

Le rapport P_b/N minimal est donc donné par :

$$\left.\frac{P_b}{N}\right|_{\min} = \frac{2^{2C}-1}{C}.$$

Le plus petit rapport P_b/N est obtenu lorsque $C \to 0$. Or $\lim\limits_{C\to 0}\dfrac{P_b}{N} = 2\ln 2$ soit 1,4 dB. Dans le cas des signaux complexes cette limite vaut $\ln 2$ soit -1,6 dB. Pour une puissance de bruit donnée on peut donc calculer la puissance minimale de l'émetteur par bit d'information. En communications numériques, on utilise le plus souvent le rapport E_b/N_0 qui est le rapport de l'énergie par bit d'information à la densité spectrale de puissance de bruit. On peut relier ce rapport à P_b/N_0 en faisant intervenir la bande B.

Est-ce que les turbo-codes atteignent "presque" la limite de Shannon ?
Dans [5] les auteurs disent que les turbo-codes atteignent presque la limite de Shannon pour le canal gaussien. Les auteurs utilisent un turbo-codeur de rendement $\frac{1}{2}$ suivi d'une modulation à deux états (BPSK, binary phase shift keying). Après un nombre suffisant d'itérations au décodage, le TEB après décodage vaut 10^{-5} pour un E_b/N_0 égal à 0.7 dB. Les auteurs considèrent, arbitrairement, que la transmission est fiable pour cette valeur de TEB. L'efficacité ou le taux de codage de leur système, jugé fiable, est donc

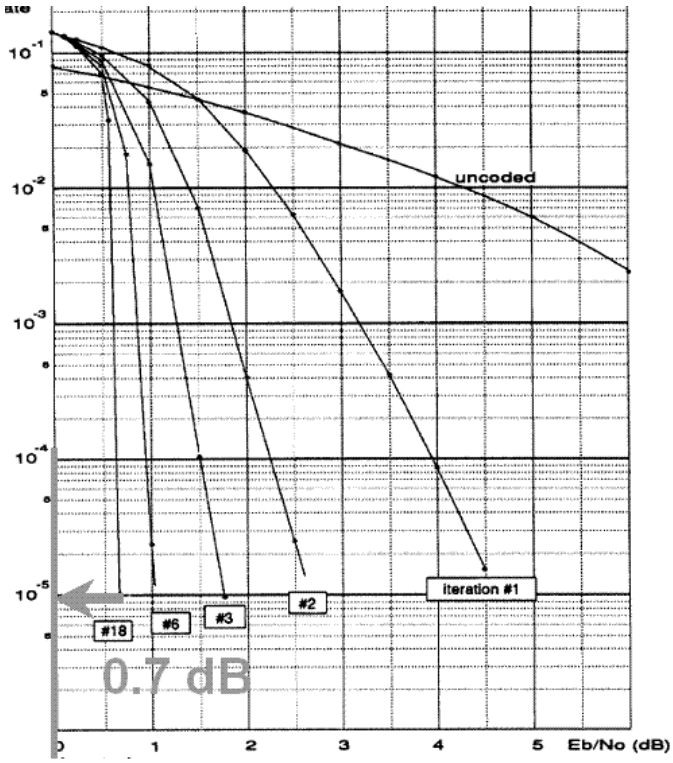

FIG. 2.6 – Performance du turbo-codeur de Berrou et al.

de $\frac{1}{2}$ bit par symbole pour un E_b/N_0 égal à 0.7 dB. Pour une même valeur de taux de codage, le rapport E_b/N_0 minimal nécessaire pour réaliser une transmission fiable sur un canal gaussien vaut 0 dB soit 1 en échelle linéaire. Le turbo-code de [5] est à "0.7 dB" de la limite de Shannon (figure 2.6), ce qui est proche en effet. Cependant, le côté arbitraire de la définition d'une transmission fiable (ici 10^{-5}) a son importance. En effet, pour des TEB plus petits, 10^{-9} par exemple on observerait un effet de pallier (error floor) qui agrandirait l'écart entre le E_b/N_0 nécessaire au turbo-codeur avec celui nécessaire en théorie. De plus la métrique de performance utilisée par [5] est le taux d'erreur par bit alors que Shannon définit la capacité par rapport à la probabilité d'erreur par bloc.

2.5 D'autres exemples pratiques de canaux mono-utilisateurs

Nous avons évoqué quelques conséquences du théorème de codage pour le canal de Shannon. Lorsqu'on a affaire à un canal qui ne vérifie pas les hypothèses du canal de Shannon on doit, a priori, trouver et démontrer un nouveau théorème de codage pour le cas considéré. De cette adaptation ou extension on pourra en général tirer des informations nouvelles propres à la nouvelle situation : des idées de codage grâce à l'atteignabilité et les performances limites grâce à l'expression de la capacité. C'est ce que nous allons faire pour quelques exemples d'intérêt pratique.

2.5.1 Canaux MIMO gaussiens

Le modèle du canal gaussien scalaire est le suivant : $Y = X + Z$. Nous étudierons le cas des signaux complexes ici. A un instant donné, l'entrée du canal est donc scalaire et de même pour la sortie. Dans ce paragraphe nous étudions le cas où, à un instant donné, l'entrée et la sortie du canal sont vectorielles. En pratique cette situation apparaît, par exemple, lorsque l'émetteur et le récepteur utilisent plusieurs

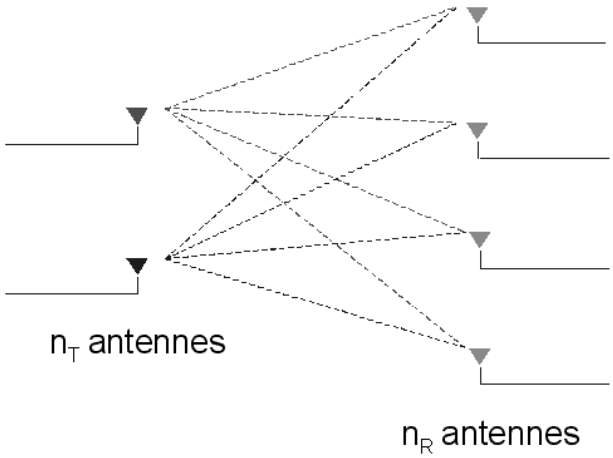

FIG. 2.7 – Canal MIMO

antennes ou des antennes multi-capteurs (figure 2.7). Ce sont les systèmes MIMO[4]. Le modèle du canal devient alors :

$$\underline{Y} = \mathbf{H}\underline{X} + \underline{Z} \tag{2.8}$$

où $\underline{X} = (X_1, ..., X_t)^T$, $\mathbf{H} \stackrel{d}{=} r \times t$ est la matrice de transfert du canal supposée constante, t étant le nombre d'antennes d'émission, r le nombre d'antennes de réception et $\underline{Z} \sim \mathcal{N}(\underline{0}, N\mathbf{I}_r)$. En supposant la matrice $\mathbf{H}$ connue du récepteur, on peut montrer que la capacité de ce canal est donnée par :

$$C = \max_{\mathbf{Q}} \log_2 \left| \mathbf{I}_r + \frac{1}{N}\mathbf{H}\mathbf{Q}\mathbf{H}^H \right| \tag{2.9}$$

avec $\mathbf{Q} = E(\underline{X}\underline{X}^H)$, $\mathrm{Tr}(\mathbf{Q}) \leq P$ et $|\mathbf{M}| = \mathrm{Det}(\mathbf{M})$.

Par rapport au cas scalaire on doit non seulement supposer $\underline{X}$ gaussien mais aussi trouver la meilleure structure spatiale pour le vecteur $\underline{X}$ soit $E(\underline{X}\underline{X}^H)$. Pour réaliser et interpréter l'optimisation correspondante nous allons utiliser les factorisations spectrales de matrices. On peut toujours décomposer $\mathbf{H}$ et $\mathbf{Q}$ de la manière suivante :

$$\begin{cases} \mathbf{H} &= \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H \\ \mathbf{Q} &= \mathbf{P}\mathbf{D}\mathbf{P}^H, \end{cases}$$

les matrices $\mathbf{U}$, $\mathbf{V}$, et $\mathbf{P}$ étant unitaires.

De plus nous allons choisir comme matrice de passage de $\mathbf{Q}$ la matrice $\mathbf{V}$: $\mathbf{P} = \mathbf{V}$. Noter que la matrice $\mathbf{\Lambda}$ est rectangulaire (en général) avec le choix de notre décomposition en valeurs singulières. On peut montrer qu'en restreignant l'ensemble de recherche de la matrice maximisant $I(\underline{X};\underline{Y})$ à l'ensemble des matrices ayant cette forme (i.e. $\mathbf{P} = \mathbf{V}$) on ne perd rien en optimalité (pour plus de détails voir [9]).

[4]MIMO : multiple input multiple output. SISO : single input single output.

Développons alors $I(\underline{X};\underline{Y})$:

$$\begin{aligned} I(\underline{X};\underline{Y}) &= \log_2 \left| \mathbf{I}_r + \frac{1}{N}\mathbf{H}\mathbf{Q}\mathbf{H}^H \right| \\ &= \log_2 \left| \mathbf{I}_r + \frac{1}{N}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^H\mathbf{P}\mathbf{D}\mathbf{P}^H\mathbf{V}\mathbf{\Lambda}^H\mathbf{U}^H \right| \\ &= \log_2 \left| \mathbf{I}_r + \frac{1}{N}\mathbf{U}\mathbf{\Lambda}\mathbf{V}^H\mathbf{V}\mathbf{D}\mathbf{V}^H\mathbf{V}\mathbf{\Lambda}^H\mathbf{U}^H \right| \\ &= \log_2 \left| \mathbf{I}_r + \frac{1}{N}\mathbf{U}\mathbf{\Lambda}\mathbf{D}\mathbf{\Lambda}^H\mathbf{U}^H \right| \\ &= \log_2 \left| \mathbf{I}_r + \frac{1}{N}\mathbf{\Lambda}\mathbf{D}\mathbf{\Lambda}^H \right| \\ &= \sum_{i=1}^{r} \log_2 \left(1 + \frac{d_i\lambda_i^2}{N} \right). \end{aligned} \tag{2.10}$$

Maximiser l'information mutuelle revient donc à maximiser

$$\sum_{i=1}^{r} \log_2 \left(1 + \frac{d_i\lambda_i^2}{N} \right)$$

par rapport aux d_i sous la contrainte de puissance totale $\sum_{i=1}^{r} d_i \leq P$. Cette réécriture de l'information mutuelle nous montre qu'optimiser la structure spatiale du codage revient à un changement de base et à la résolution d'un problème d'allocation de ressources entre r canaux gaussiens scalaires en parallèle de rapports signal-à-bruit $\frac{P_i}{N_i} = \frac{d_i\lambda_i^2}{N}$ avec $P_i = d_i$. Nous avons donc juste à maximiser le lagrangien suivant :

$$\mathcal{L}(P_1,..,P_t,\mu) = \sum_{i=1}^{r} \log_2 \left(1 + \frac{P_i}{N_i} \right) - \mu(\sum_{i=1}^{r} P_i - P),$$

d'où :

$$\begin{aligned} \frac{\partial \mathcal{L}(P_1,..,P_t,\mu)}{\partial P_i} &= \frac{1}{\ln 2} \frac{\frac{1}{N_i}}{1+\frac{P_i}{N_i}} - \mu \\ &= \frac{1}{P_i+N_i} - \mu. \end{aligned}$$

L'allocation optimale de puissance s'en déduit immédiatement :

$$P_i = \left| \begin{array}{cc} \omega - N_i & \text{si } N_i \leq \omega \\ 0 & \text{sinon} \end{array} \right.$$

où $\omega \triangleq \frac{1}{\mu \ln 2}$ est appelé le niveau d'eau. En efft, cette stratégie, suggerée par la théorie de l'information, s'appelle le "water-filling" (figure 2.8) : l'idée de la politique optimale d'allocation des ressources (finies) est de distribuer les ressources aux meilleurs canaux (N_i suffisamment petit) pour maximiser le débit global. Cette "politique" est utilisée dans d'autres contextes que les communications...

2.5.2 Canaux multi-trajets

Les canaux multi-trajets, que l'on rencontre par exemple en téléphonie cellulaire (échos de propagation), font partie des canaux qui se modélisent par un filtre :

$$Y_i = \sum_{j=0}^{\ell} h_j X_{i-j} + Z_i.$$

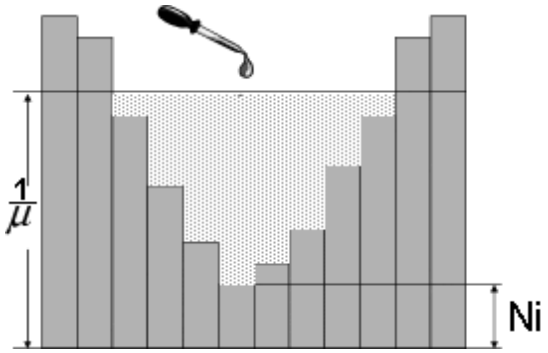

FIG. 2.8 – Principe du "water-filling"

Clairement ces canaux, dit à interférence entre symboles en communications numériques, sont à mémoire. On peut démontrer [10], [11] que la capacité de ces canaux est donnée par :

$$C = \int_{-\frac{1}{2}}^{\frac{1}{2}} \max\left\{\log_2\left(\frac{|H(f)|^2\xi}{N}\right), 0\right\} df \tag{2.11}$$

avec $H(f) = \sum_{i=0}^{\ell} h_i e^{-j2\pi fi}$ et ξ est déterminé par la contrainte

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \max\left\{\xi - \frac{N}{|H(f)|^2}, 0\right\} df = P$$

.

D'un point vue de l'atteignabilité du théorème de codage qui correspond à l'établissement de cette capacité, on peut montrer qu'une stratégie de codage optimale consiste à :
- coder en blocs avec des blocs longs (n grand) ;
- ajouter un préfixe dit cyclique en fin de chaque de bloc. Ce préfixe consiste en la répétition des ℓ premiers symboles de chaque bloc ;
- passer dans le domaine de Fourier en plaçant un pré-codeur en émission et un pré-décodeur en réception. Ces opérations consistant en de simples multiplications matricielles. Si on appelle $\underline{X}$ le bloc de données codées préfixé, on enverra sur le canal $\underline{\tilde{X}} = \mathbf{F}^H \underline{X}$ avec $(\mathbf{F})_{p,q} = \frac{1}{\sqrt{n}} e^{-j2\pi\frac{(p-1)(q-1)}{n}}$ et on appliquera au bloc reçu la matrice $\mathbf{F}$;
- appliquer une stratégie d'allocation de puissance entre les différents sous-canaux fréquentiels obtenus avec le changement de base (de Fourier) qui suit le principe du "water-filling".

Il existe des systèmes de communications qui suivent tout à fait cette stratégie de codage, ce sont les sytèmes OFDM (orthogonal frequency division modulation) à préfixe cyclique.

2.5.3 Canaux variables dans le temps

2.5.3.1 Canaux SISO à évanouissements rapides

Dans les exemples du canal MIMO et du canal multi-trajet le canal était supposé constant sur toute la transmission. En pratique, c'est le cas des canaux radio-mobiles, du canal inonospérique, etc. la fonction de transfert du canal peut varier dans le temps. En communications numériques un des modèles les plus utilisés est le canal de Rayleigh :

$$Y = HX + Z \tag{2.12}$$

où H est une VA telle que son module $R = |H|$ suit une loi de Rayleigh :

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} u(r) \tag{2.13}$$

où σ est le paramètre de la loi de Rayleigh et la fonction $u(r)$ est la fonction échelon d'Heaviside $u(r) = 1$ pour $r \geq 0$ et $u(r) = 0$ pour $r < 0$.

La capacité de ce canal, dit à évanouissements rapides, peut se calculer dans le cas où le récepteur connaît H à tout instant (on parle de communication cohérente) :

$$C(1,1) = E_H \left[\log_2 \left(1 + \frac{P|H|^2}{N} \right) \right]. \tag{2.14}$$

2.5.3.2 Canaux MIMO à évanouissements rapides

On peut facilement généraliser la formule de capacité ergodique du cas SISO (single-input single-output channel) au cas MIMO. On montre facilement que, lorsque le récepteur connaît la matrice de canal $\mathbf{H}$ à tout instant et que l'émetteur connaît ni les réalisations ni les statistiques de cette matrice, la capacité ergodique pour une communication MIMO s'exprime par :

$$C(t,r) = E_{\mathbf{H}} \left[\log_2 \left| \mathbf{I} + \frac{1}{N} \mathbf{H}\mathbf{H}^H \right| \right]. \tag{2.15}$$

Pour mieux évaluer physiquement l'apport des systèmes MIMO, concentrons-nous sur des cas particuliers. D'après [13], [14] on peut montrer que cette capacité s'approxime comme suit :

$$C(t,r) \sim \min\{t,r\} \times C(1,1)$$

lorsque $\frac{P}{N} \to \infty$ et

$$C(t,r) \sim r \times C(1,1)$$

lorsque $\frac{P}{N} \to 0$. On voit donc qu'il y a un gain en débit qui peut être substantiel quelque soit le RSB (rapport signal à bruit). Ceci est intéressant car cela veut dire qu'en pratique on peut espérer des gains en débits par rapport au cas SISO à faible RSB. En effet, pour un système SISO un moyen d'augmenter le débit est de densifier la constellation d'émission mais cela a un prix en termes de sensibilité au bruit. Le système HSPA (high speed packet access) utilise les deux techniques pour avoir de hauts débits.

2.5.3.3 Canaux SISO à évanouissements lents

Puisque les communications se font à temps discrets (au rythme symbole) nous avons implicitement supposé jusqu'à maintenant que le canal variait au rythme symbole. En pratique on rencontre fréquemment la situation où le canal varie au rythme bloc et est donc constant sur une durée, de quelques centaines de symboles typiquement. On parle de canaux à évanouissements lents. Sur chaque bloc tout se passe comme si on communiquait sur un canal gaussien mais comme le canal varie de bloc en bloc, l'information mutuelle est aussi une variable aléatoire. On pourrait associer un taux de codage maximal à chaque réalisation du canal mais comme en pratique le bloc a une taille limitée, on ne pourrait pas atteindre ni même approcher ce taux de codage théorique. Pour ce type de canal on utilise souvent un critère de performance plus adapté qui est la capacité de coupure, elle-même définie à partir de la probabilité de coupure [12]. Soit η un taux de codage cible. La probabilité de coupure du canal pour le débit η est définie par :

$$P_{out}(\eta) \triangleq P[I(X;Y|H) \leq \eta]. \tag{2.16}$$

Soit ε une probabilité cible. La capacité de coupure pour cette probabilité ε est alors définie comme suit :

$$C(\varepsilon) \triangleq \max_{R} \{R : P_{out}(R) \leq \varepsilon\}. \tag{2.17}$$

Pour le canal de Rayleigh la probabilité de coupure est facile à évaluer :

$$\begin{aligned} P_{out}(\eta) &= P[I(X;Y|H) \leq \eta] \\ &= P\left[\log_2\left(1+\frac{P|H|^2}{N}\right) \leq \eta\right] \\ &= P\left[|H|^2 \leq \frac{2^\eta - 1}{\rho}\right] \\ &= \int_0^{\frac{2^\eta-1}{\rho}} \frac{1}{\bar{\gamma}} e^{-\gamma/\bar{\gamma}} d\gamma \\ &= 1 - e^{\frac{1-e^\eta}{\rho}} \end{aligned} \tag{2.18}$$

où $\rho = \frac{P}{N}$, $\bar{\gamma} = E[|H|^2] = 1$ et $|H|^2 = \Gamma$ est distribuée selon une simple expotentielle : $f_\Gamma(\gamma) = \frac{1}{\bar{\gamma}} e^{-\gamma/\bar{\gamma}}$.

2.5.3.4 Canaux MIMO à évanouissements lents

Le calcul de la capacité de coupure pour les canaux MIMO à évanaouissements lents n'est pas simple en général. Par exemple, l'expression de la capacité de coupure est encore inconnue pour les canaux avec une composante de Rice ($E[\mathbf{H}] \neq \mathbf{0}$) et lorsqu'il y a de la corrélation spatiale entre les entrées et les sorties du canal.

Cependant, il existe un formalisme simple et élégant dans les régimes de forts RSB pour caractériser la capacité de coupure : il s'agit du compromis diversité - gain de multiplexage [15]. Soit le signal reçu :

$$\underline{y}(t) = \mathbf{H}\underline{x}(t) + \underline{z}(t). \tag{2.19}$$

La capacité ergodique s'exprime par

$$C = \mathbb{E}_H\left[\log\det\left(\mathbf{I} + \frac{\rho}{n_T}\mathbf{H}\mathbf{H}^H\right)\right]. \tag{2.20}$$

Cette quantité fait sens et correspond à la capacité de Shannon pour les canaux à évanouissements rapides, c'est-à-dire quand $\mathbf{H}$ varie au rythme symbole (que l'on noterait alors $\mathbf{H}(t)$). Cependant cette quantité constitue une borne supérieure du taux de transmission pour pour les canaux à évanouissements lents. Seulement une fraction de cette quantité peut-être atteinte. Cette fraction est appelée le gain de multiplexage et est défini par :

$$r \triangleq \frac{R}{\log\rho}. \tag{2.21}$$

En effet, lorsque $\rho \to \infty$,

$$C \sim \min(n_T, n_R)\log\frac{\rho}{n_T}, \tag{2.22}$$

expression qui montre la pertinence de la définition du gain de multiplexage.

On définit la diversité par :

$$\lim_{\rho\to\infty} \frac{\log P_{out}(R)}{\log\rho} \triangleq -d(r). \tag{2.23}$$

La théorie des communications [15] permet de montrer que le compromis optimal entre gain et diversité est donné par la fonction $d^*(r)$ qui est une fonction linéaire par morceaux dont les points intermédiaires dont donnés par :

$$d(k) = (n_T - k)(n_R - k)$$
$$k \in \{0, \min(n_T, n_R)\}$$

.

On peut ainsi grâce à ce compromis caratériser la sous-optimalité des codes spatio-temporels, tel que celui d'Alamouti pour lequel la matrice de précodage s'exprime par ([16]) :

$$\begin{matrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{matrix} .$$

Nous concluerons ce paragraphe en évoquant le cas particulier où le gain de multiplexage est nul où on veut utiliser le système MIMO en diversité totale pour privilégier la qualité de la communication representée par la vitesse de décroissance de la probabilité d'erreur en fonction du rapport signal-à-bruit. Lorsque $\rho \to \infty$

$$P_e \sim \frac{1}{\rho^{n_T \times n_R}}. \tag{2.24}$$

2.5.4 Canaux avec information adjacente

Voir version 1.0 du polycopié et les sujet et corrigé de l'examen de Théorie de l'Information de la session 2004 et aussi les références de base sur ce sujet [17, 18].

2.5.5 Canaux avec retour d'information

On peut montrer que pour un canal discret sans mémoire et point-à-point la connaissance strictement causale des sorties du canal par le codeur ne permet pas d'améliorer la capacité par rapport au cas sans retour d'information [4]. Cette conclusion ne se généralise pas nécessairement au cas des canaux multi-points. Par exemple, le retour d'information permet d'élargir la région de capacité des canaux à accès multiple [19].

Pour plus de détails, voir version 1.0 du polycopié.

Chapitre 3

Canaux à plus de deux terminaux

Dans le chapitre précédent notre propos s'est focalisé sur les transmissions faisant intervenir un seul émetteur et un seul récepteur. La liaison montante d'un système cellulaire est un exemple de systèmes où plusieurs émetteurs (les différents mobiles actifs) veulent communiquer avec un récepteur (la station de base). On peut imaginer une situation encore plus complexe : plusieurs émetteurs qui veulent communiquer avec plusieurs récepteurs via des noeuds pouvant jouant le rôle soit de simples relais ou de récepteurs intermédiaires. Comme le mentionne Cover dans [4] il n'existe pas, actuellement, une théorie de l'information pour les réseaux unifiée qui permettrait de donner la capacité (ou bien la région de capacité, nous y reviendrons) d'un tel système. Dans ce chapitre nous étudions quelques canaux qui font partie des "briques de bases" du canal général évoqué ci-dessus.

3.1 Canal à accès multiple

Si on isole une cellule d'un système radio-mobile dans laquelle plusieurs utilisateurs envoient leur signal à la station de base qui contrôle cette cellule, on se retrouve exactement dans la situation du canal à accès multiple (figure 3.1). On se restreint ici, par souci de simplicité, au cas de deux utilisateurs qui envoient leur message à un destinataire commun. Ce destinataire doit être capable de décoder les deux messages.

Étant donné qu'il y a deux flux d'information mis en jeu, et non plus un seul comme pour le canal mono-utilisateur, on ne parle plus de capacité mais de région de capacité. Ainsi, dans le cas 2 utilisateurs, on caractérise les performances limites du canal par une région à deux dimensions qui comprend tous les couples de taux de codage que l'on peut atteindre sur un tel canal avec une fiabilité aussi bonne que l'on veut sur le décodage des deux messages. On peut montrer que la région de capacité du canal à accès multiple discret à deux utilisateurs est donnée par l'ensemble des couples (R_1, R_2) vérifiant :

$$\left\{ \begin{array}{rcl} R_1 & \leq & I(X_1;Y|X_2) \\ R_2 & \leq & I(X_2;Y|X_1) \\ R_1+R_2 & \leq & I(X_1,X_2;Y) \end{array} \right. \tag{3.1}$$

pour toutes les distributions $p(x_1,x_2) = p(x_1)p(x_2)$ possibles. On voit que la description mathématique de la région du canal à accès multiple est assez intuitive. Le taux-somme, qui correspond au flux global d'information qui arrive au récepteur ne doit pas excéder celui d'un système MISO 2×1. Si les deux émetteurs étaient co-localisés, de manière à ce qu'ils puissent échanger leur connaissance parfaitement, la condition sur le taux-somme du canal MAC (multiple access channel) suffirait à décrire la région de capacité (région triangulaire dans ce cas). Comme la définition du canal MAC précise que les émetteurs ne coopérent pas, des conditions individuelles se rajoutent sur R_1 et R_2 qui traduisent le fait que les émetteurs agissent indépendamment. On peut étendre ce résultat au cas gaussien complexe et on trouve

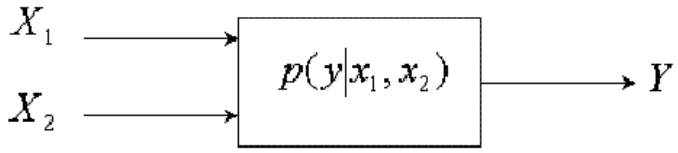

FIG. 3.1 – Canal à accès multiple à deux récepteurs

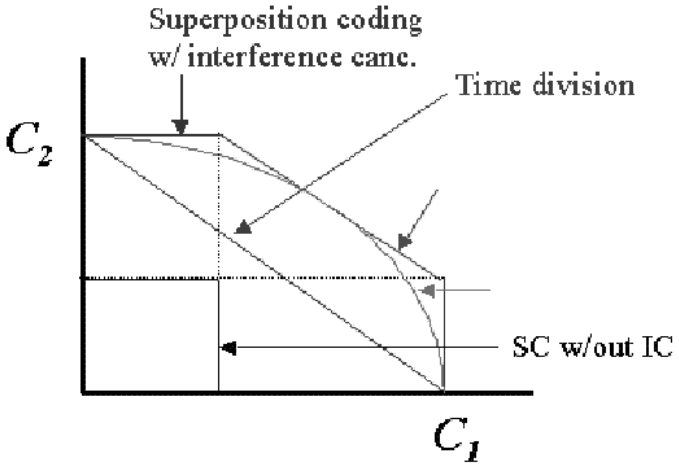

FIG. 3.2 – Région de capacité du canal à accès multiple à deux utilisateurs

que la région de capacité est donnée par l'ensemble des couples (R_1, R_2) vérifiant :

$$\begin{cases} R_1 & \leq & \log_2\left(1+\frac{P_1}{N}\right) \\ R_2 & \leq & \log_2\left(1+\frac{P_2}{N}\right) \\ R_1+R_2 & \leq & \log_2\left(1+\frac{P_1+P_2}{N}\right) \end{cases} \tag{3.2}$$

où P_1 est la puissance du premier émetteur, P_2 la puissance du second émetteur et N la puissance de bruit additif (blanc et gaussien) du récepteur. Pour démontrer l'atteignabilité du théorème de codage qui correspond à la région de capacité du MAC gaussien on peut montrer que la stratégie de décodage qui consiste à appliquer un décodage conjoint de (W_1, W_2) puis deux décodages successifs et individuels de W_1 et W_2 est optimale. En effet, comme $Y = X_1 + X_2 + Z$, le récepteur peut alors éliminer l'interférence pour chacun des deux flux. On notera que cette idée s'est traduite par des stratégies d'élimination d'interférence pratiques telle que le SIC (successive interference canceller) pour les systèmes CDMA. On notera, en ce qui concerne la stratégie d'accès multiple que la région de capacité est plus étendue que la région qu'atteindrait une stratégie TDMA (triangle formé par les points $(0,0)$, $(C_1,0)$ et $(0,C_2)$), ce qui indique l'existence de stratégies d'accès multiple plus performantes que le TDMA. Ainsi pour différentes stratégies d'accès multiple (TDMA, FDMA, CDMA, SDMA, ...) on peut comparer entre elles les régions de taux de codage atteignables par ces stratégies. Voici l' allure typique d'une région de capacité.

3.2 Canal de diffusion

Le canal de diffusion [20] correspond à la situation duale de celle du canal à accès multiple. Cette fois il s'agit d'un seul émetteur qui envoie un signal commun à plusieurs récepteurs (figure 3.3). Chaque récepteur doit extraire du signal qu'il reçoit le message qu'il intéresse. En effet, bien que pour ce qu'on appelle "diffusion" au sens courant (diffusion FM, diffusion TV,...) le canal de diffusion au sens de la théorie de l'information n'implique pas forcément que tous les récepteurs soient intéressés par le même message. Le cas du canal de diffusion couvre des situations plus générales telle que la liaison descendante

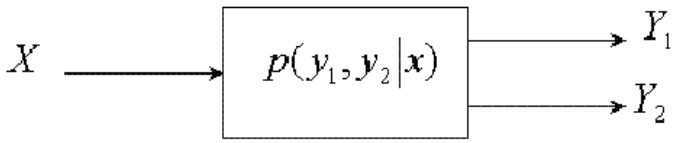

FIG. 3.3 – Canal de diffusion à deux récepteurs

d'un système cellulaire. Dans ce cas une seule base émettrice envoie un signal à plusieurs récepteurs qui décodent chacun ce qu'il leur est destiné.

A la différence du canal à accès multiple, la région de capacité du canal de diffusion discret caractérisée par sa probabilité de transition $p(y_1,y_2|x)$, n'est pas encore connue à ce jour dans le cas général. La région de capacité du canal de diffusion discret a cependant pu être déterminée dans des cas particuliers relativement utiles. Un de ces cas particuliers est le cas du canal de diffusion physiquement dégradé. C'est le cas où la probabilité de transition du canal s'écrit de la manière suivante :

$$p(y_1,y_2|x) = p(y_1|x)p(y_2|y_1).$$

L'hypothèse de dégradation physique revient à supposer que l'entrée et les deux sorties du canal de diffusion vérifie : $X - Y_1 - Y_2$. Cette hypothèse permet de ramener le canal de diffusion général, qui comporte deux canaux en parallèle, à un canal équivalent constitué de deux canaux à sortie scalaire en série. Autrement dit, une sortie a nécessairement des conditions de transmission moins favorables qu'une autre. Plus précisément, pour le récepteur ayant les moins bonnes conditions de transmissions, le signal émis X doit pouvoir s'écrire à partir du signal reçu de l'autre récepteur Y_1. Autant on peut considérer comme parfaitement réaliste le fait qu'un récepteur ait de meilleures conditions de transmission qu'un autre, le fait que le signal reçu par le plus mauvais récepteur se déduise du signal reçu par le meilleur récepteur peut être discutable[1]. Nous verrons cependant que l'hypothèse de dégradation physique est une hypothèse qui nous conduit à la région de capacité du cas gaussien. Pour ce qui est du cas discret, la région de capacité du canal de diffusion physiquement dégradé correspond à l'enveloppe convexe de tous les points (R_1,R_2) qui vérifient :

$$\begin{cases} R_1 & \leq & I(X;Y_1|U) \\ R_2 & \leq & I(U;Y_2) \end{cases} \tag{3.3}$$

pour toute distribution de la forme $p(u,x,y_1,y_2) = p(u,x)p(y_1|x)p(y_2|y_1)$. Nous pouvons faire quelques commentaires sur la région de capacité du canal de diffusion dégradé. Tout d'abord il n'y a pas de condition sur le taux somme, ceci est dû à l'hypothèse de dégradation physique. On voit aussi que la région de capacité dépend d'une variable auxiliaire U. Intuitivement, on peut voir cette variable auxiliaire comme un "paramètre d'action" qui permet de distribuer les ressources de l'émetteur entre les deux utilisateurs. Par exemple, lorsque $U \equiv \varnothing$, le taux de codage pour l'utilisateur 2 est nul alors qu'il est maximal pour $U \equiv X$. D'un point de vue du codage, elle indique qu'un codage par superposition permet d'atteindre les taux indiqués. Nous ne décrirons pas ici la technique de codage par superposition mais le principe peut être compris un considérant une situation de diffusion courante : le professeur qui parle à une classe d'élèves. Si son cours est bien conçu tous les élèves, y compris les moins bons, recevront et comprendrons les messages de base que le professeur veut que tout le monde ait compris. De plus les meilleurs éléments pourront bénéficier d'informations supplémentaires que le professeur aura superposé sur son cours de base. Ainsi dans le codage par superposition U représente le message qui est compréhensible par les deux récepteurs. Plus techniquement, cette variable auxiliaire joue le rôle de centre de nuage de points (un point représentant un mot de code) : les deux récepteurs sont capables de distinguer les centres

[1]Lorsque l'hypothèse de dégradation physique peut s'avérer trop restrictive, on peut utiliser les notions de "canal moins bruité" ou de canal "plus capable" définie par [22]. Bien que moins restrictives, ces hypothèses sont peu utilisées dans le domaine.

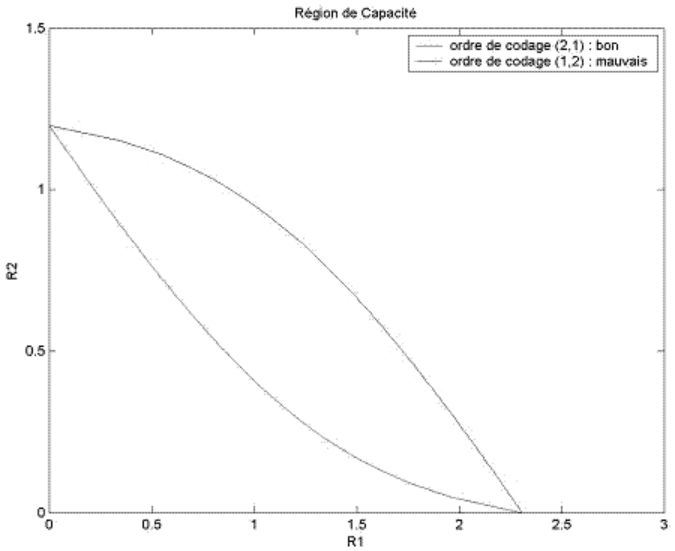

FIG. 3.4 – Région de capacité du canal de diffusion gaussien à deux récepteurs

de nuages alors que seul le meilleur peut distinguer les points à l'intérieur d'un nuage donné. On voit donc que la variable intermédiaire de codage U découle naturellement du caractère dégradé du canal. En quelques mots, la séparabilité du canal engendre une séparabilité au niveau de la procédure de codage. L'idée de codage par superposition peut se traduire facilement en un codage pratique.

Abordons maintenant le cas gaussien qui est donné par les équations des signaux reçus suivantes :

$$\left\{ \begin{array}{rcl} Y_1 & = & X+Z_1 \\ Y_2 & = & X+Z_2 \end{array} \right. \tag{3.4}$$

où $Z_1 \sim \mathcal{N}(0,N_1)$, $Z_2 \sim \mathcal{N}(0,N_2)$, $N_2 \geq N_1$ et $I(Z_1;Z_2)=0$. Le canal de diffusion gaussien ne vérifie pas la chaîne de Markov $X-Y_1-Y_2$ et n'est donc pas physiquement dégradé. Cependant, on peut montrer [21] que le canal défini par les équations $Y_1 = X+Z_1$, $Y_2 = Y_1+Z_2'$ avec $Z_2' \sim \mathcal{N}(0,N_2-N_1)$, canal qui est physiquement dégradé, a la même région de capacité que le canal de diffusion gaussien défini par (3.4). La région de capacité de ces deux canaux s'écrit :

$$\left\{ \begin{array}{rcl} R_1 & \leq & \log_2\left(1+\frac{\alpha P}{N_1}\right) \\ R_2 & \leq & \log_2\left(1+\frac{\overline{\alpha} P}{N_2+\alpha P}\right) \end{array} \right. \tag{3.5}$$

où αP et $\overline{\alpha} P = (1-\alpha)P$ sont les puissances attribuées à l'utilisateur 1 et 2 respectivement. Le paramètre réel permet donc de décrire toutes les répartitions de puissance au niveau de l'émetteur entre les deux flux à transmettre. On voit aussi dans le cas gaussien le codage par superposition, qui s'écrit simplement $X = X_1 + X_2$, permet de s'affranchir totalement de l'influence du flux dédié au moins bon récepteur sur le flux dédié au meilleur récepteur. En faisant varier α de 0 à 1 on obtient le graphe suivant pour la région de capacité :

Comme nous l'avons mentionné le codage par superposition se traduit par l'émission de $X = X_1 + X_2$ c'est-à-dire $U \equiv X_2$. Cette stratégie est optimale pourvu que l'on code dans le bon ordre. Il faut, en effet, commencer par coder le message de l'utilisateur 2 (car par hypothèse arbitraire $N_2 \geq N_1$) puis on code le message de l'utilisateur 1 en se servant de la connaissance de X_2 comme une interférence connue et que l'on peut donc compenser en émission [17].

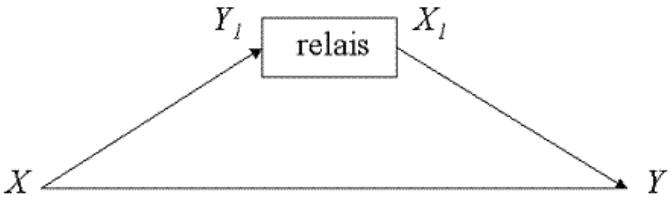

FIG. 3.5 – Canal à un relais

3.3 Canal à relais

Le canal à relais [23] est également un canal qui prend une importance croissante à mesure que des applications de type "réseaux ad hoc" ou bien de "diversité de coopération" [24] se développent. Considérons la situation la plus simple où un seul relais intervient (figure 3.5). Dans cette situation, le but est de se servir d'un relais pour améliorer la transmission d'un émetteur donné vers un récepteur donné. Le relais, dans sa définition originale [23], n'a pas nécessairement à décoder le message informatif de la source mais sa fonction est d'aider au mieux le récepteur.

Pour le canal à relais discret, là-aussi, la capacité n'est connue que dans des cas particuliers. Un cas particulier intéressant est le cas où le signal reçu par le récepteur peut être vue comme une version dégradée du signal reçu par le relais. On considère donc une situation où le relais est dans de meilleures conditions de réception que le récepteur, ce qui est souvent la situation attendue pour justifier l'intérêt du relais. Avec les notations adoptées sur la figure 3.5, on dit qu'un canal à relais est physiquement dégradé lorsque nous avons l'égalité suivante :

$$p(y_1,y_2|x,x_{12}) = p(y_1|x,x_{12}) \times p(y_2|x_{12},y_1).$$

La chaîne de Markov correspondant à cette dégradation physique est la suivante : $X-(X_{12},Y_1)-Y_2$. Dans ces conditions la capacité d'un canal à relais discret est donnée par :

$$C = \max_{p(x,x_{12})} \min\left\{I(X;Y_1|X_{12}), I(X,X_{12};Y_2)\right\}. \tag{3.6}$$

La présence du "min" indique que la capacité est limitée par le maillon le plus faible de la chaîne de transmission, qui se situe soit à l'émission qui s'apparente à une situation de diffusion soit à la réception qui s'apparente à une situation d'accès multiple. L'optimisation se fait alors sur la probabilité conjointe $p(x,x_{12})$ car il s'agit d'optimiser à la fois le codeur de l'émetteur et le codeur du relais.

Dans le cas général on ne connaît pas la capacité du canal à relais. Cependant, il existe des bornes inférieures et supérieures de la capacité de ce canal. On peut facilement montrer que la capacité est bornée supérieurement par la quantité suivante :

$$C \leq \max_{p(x,x_{12})} \min\left\{I(X;Y_1,Y_2|X_{12}), I(X,X_{12};Y_2)\right\}. \tag{3.7}$$

On peut interpréter physiquement cette borne car elle peut aussi être retrouvée en applicant le théorème du "max flow min cut" [25] qui s'utilise en théorie des flux. Le premier terme inidique le flux maximal que peut véhiculer les deux canaux issus de la source d'information alors que le deuxième représente une le maximum du flux total correspondants aux deux canaux qui arrivent au destinataire.

Pour les bornes inférieures, on peut en trouver autant qu'on peut immaginer de stratégies de codage. Pour le cas discret, l'article original sur le canal à relais [23] en donne deux, qui constituent encore aujourd'hui les principales stratégies à l'étude. Ces stratégies sont appelées "decode-and-forward" et "estimate-and-forward".

Decode-and-forward
Cette stratégie de codage consiste à imposer au relais de décoder le message émis par la source. Le relais et la source coopérent de manière à maximiser le débit de la source. Le taux de codage atteint par cette stratégie vaut :

$$R_{DF} = \max_{p(x,x_{12})} \min \{I(X;Y_1|X_{12}), I(X,X_{12};Y_2)\}.$$

On remarque que cette stratégie, sous-optimale dans le cas général, est optimale pour le canal à relais physiquement dégradé.

Estimate-and-forward
Ici, le relais n'essaye pas de décoder le message de la source mais cherche à envoyer de manière fiable une image du signal qu'il a reçu au récepteur. Cette image, que l'on note $\hat{Y}_1$, peut être vu dans le cas d'un canal continu comme une quantification du signal reçu par le relais. Le taux de codage atteint s'exprime par :

$$R_{EF} = \max_{p(x,x_{12},\hat{y}_1)} I(X;\hat{Y}_1,Y_2|X_{12}).$$

L'utilité de cette stratégie sur la stratégie DF peut se comprendre sur le cas où le relais est dans de mauvaises conditions de réception et donc il peut s'avérer plus profitable d'envoyer tel quel le signal reçu que de vouloir le décoder dans de mauvaises conditions.

Ces stratégies, élaborées pour le cas discret peuvent s'appliquer dans le cas continu. Par exemple, en appliquant DF dans le cas gaussien on obtient le taux de codage suivant :

$$R_{DF} = \max_{\rho} \min \left\{ \log_2\left(1+\frac{\rho P}{N_1}\right), \log_2\left(1+\frac{P+P_{12}+2\sqrt{\bar{\rho}PP_{12}}}{N_2}\right) \right\}$$

où $\rho \in [0,1]$.

Une stratégie de relayage propre au cas continu est la stratégie amplify-and-forward. Se reporter à l'examen 2006 pour avoir une notion de base sur cette stratégie consistant à renvoyer le signal reçu en le multipliant par un scalaire dans le cas du protocole scalaire et par une matrice dans le cas du protocole vectoriel.

3.4 Canal à accès multiple à émetteurs coopératifs

Pour une synthèse sur ce sujet voir la version 3.0 du polycopié (en construction) ou le chapitre de livre [26].

3.5 Canal de diffusion à récepteurs coopératifs

Pour une synthèse sur ce sujet voir la version 3.0 du polycopié (en construction) ou le chapitre de livre [26].

Annexe A

Sujet examen 2004

Examen de Théorie de l'Information
Mastère Recherche Radiocommunications
15/11/2004, Supélec

Note : épreuve sans aucun document mais calculatrices autorisées.
Durée de l'épreuve : 1 heure.

Problème à résoudre

Le but de ce problème est l'évaluation de la capacité d'un canal que l'on rencontre par exemple dans les applications de type tatouage d'image. Le canal considéré dans ce problème est représenté sur la figure 1. Sur cette figure, W représente le message informatif que l'on veut coder, X représente l'entrée du canal et a une puissance d'émission P, S représente un bruit gaussien de moyenne nulle et de variance Q, V représente également un bruit gaussien de moyenne nulle et de variance N, Y est la sortie du canal. Nous supposerons X, S et V indépendants. On voit, sur cette figure, que le codeur a accès à une certaine connaissance de la perturbation S. En fait nous supposerons que l'émetteur connaît parfaitement (et de manière non causale) toutes les réalisations de la variable aléatoire S. Cette dernière hypothèse ne remet nullement en cause les hypothèses précédentes. En particulier l'hypothèse l'indépendance entre X et S reste vraie (on suppose l'existence d'un mécanisme de codage qui permet de préserver l'indépendance de X et S).

1. On suppose, dans cette question uniquement, que l'émetteur (qui fabrique X) et le récepteur (qui utilise Y) ne disposent d'aucune information sur S. On admettra que la distribution optimale à imposer à X pour atteindre la capacité de ce canal est gaussienne. Donner, sans démonstration, la capacité de ce canal.

2. On veut maintenant trouver la capacité du canal de la figure 1. Il s'avère que le canal considéré appartient à une classe générale de canaux (appelée canaux de Gelfand et Pinsker). L'expression générale de la capacité de ces canaux est la suivante :

$$C = \max_{p(u,x|s)} \left[I(U;Y) - I(U;S) \right].$$

Dans notre problème la signification de U, qui est une variable auxiliaire de codage, n'interviendra pas. Nous retiendrons simplement que cette variable est supposée gaussienne dans notre cas (tout comme X). En effet nous supposerons que U est donnée par $U=X+\alpha S$ où α est une constante réelle.

2.1. Tout d'abord redémontrer que l'information mutuelle entre deux variables aléatoires A et B peut s'exprimer par:

(i) $I(A;B)=H(A)-H(A|B)$

(ii) $I(A;B)=H(A)+H(B)-H(A,B)$

où $H(.)$ désigne l'entropie.

2.2. Expliciter *I(U ;Y)* et *I(U ;S)* en fonction des données du problème.

2.3. Représenter l'allure de *I(U ;Y),* de *I(U ;S)* puis de *I(U ;Y)-I(U ;S)* en fonction du paramètre α pour $\alpha \in [-2,2]$ et $P=Q=N=1$. Graphiquement, quelle est la valeur de α qui permet d'atteindre la capacité dans ce cas ?

2.4. Pour *P, Q, N* quelconques retrouver analytiquement la valeur de α qui maximise la différence *I(U ;Y)-I(U ;S).* Vérifier la cohérence du résultat dans le cas particulier de la question 2.3.

2.5. En déduire la capacité du canal de la figure 1. Comparer à la capacité trouvée dans la question 1. Commenter.

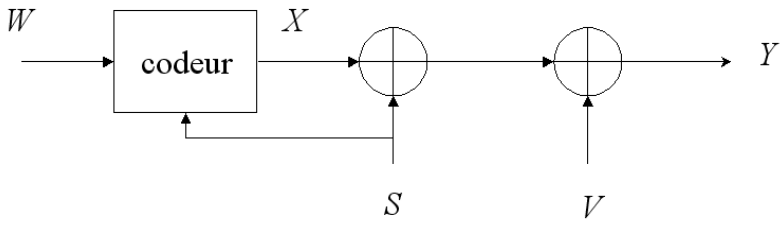

Figure 1: canal considéré dans ce problème

Fin de l'épreuve

Annexe B

Corrigé examen 2004

B.1 Remarque générale

L'exercice proposé ne faisait pas appel à d'autres notions que celles d'entropie et d'information mutuelle. En effet il fallait essentiellement savoir appliquer correctement la définition de l'entropie d'une variable aléatoire continue et la définition de l'entropie d'un couple de variable aléatoire. Rappelons que l'entropie d'une VA gaussienne X, disons centrée (ce qui était le cas dans cet exercice) et de puissance P, est donnée par :

$$\begin{aligned} H(X) &= -\int_{\mathbb{R}} f_X(x)\log_2[f_X(x)]dx \\ &= -\int_{\mathbb{R}} f_X(x)\log_2\Big[\frac{1}{\sqrt{2\pi P}}\exp\Big(-\frac{x^2}{2P}\Big)\Big]dx \\ &= \frac{1}{2}\log_2(2\pi P)\int_{\mathbb{R}} f_X(x)dx + \frac{1}{2P\ln 2}\int_{\mathbb{R}} x^2 f_X(x)dx \\ &= \frac{1}{2}\log_2(2\pi P) + \frac{\log_2(e)}{2} \\ &= \frac{1}{2}\log_2(2\pi eP). \end{aligned} \tag{B.1}$$

On peut facilement vérifier que pour obtenir l'entropie d'une somme de deux VA gaussiennes indépendantes et centrées il suffit de remplacer P par $E\big[X_1^2+X_2^2\big] = P_1+P_2$. On peut aussi vérifier facilement que pour obtenir l'entropie d'une VA à deux dimensions $\underline{X} = (X_1, X_2)$ il suffit de suivre la méthode de calcul donnée ci-dessus. On obtient alors :

$$\begin{aligned} H(X_1,X_2) &= -\int_{\mathbb{R}^2} f_{X_1,X_2}(x_1,x_2)\log_2[f_{X_1,X_2}(x_1,x_2)]dx_1dx_2 \\ &= -\int_{\mathbb{R}^2} f_{X_1,X_2}(x_1,x_2)\log_2\Big[\frac{1}{2\pi|R|}\exp\big(-[x_1,x_2]R^{-1}[x_1,x_2]^T\big)\Big]dx_1dx_2 \\ &= \frac{1}{2}\log_2\big[(2\pi e)^2|R_{X_1,X_2}|\big] \\ &= \frac{1}{2}\log_2\big[(2\pi e)^2(P_1P_2-P_{12}^2)\big] \end{aligned} \tag{B.2}$$

où

$$R = E\big\{[X_1,X_2][X_1,X_2]^T\big\} = \begin{pmatrix} E(X_1^2) & E(X_1X_2) \\ E(X_1X_2) & E(X_2^2) \end{pmatrix} \triangleq \begin{pmatrix} P_1 & P_{12} \\ P_{12} & P_2 \end{pmatrix}$$

et $|R| = \mathrm{Det}(R)$.

Enfin rappelons que la puissance d'une VA $Y = \alpha X$ est $E(Y^2) = \alpha^2 E(X^2)$ et non pas $\alpha E(X^2)$ comme c'était le cas dans la totalité des copies. La conséquence de ce détail est que la convexité par rapport à α des grandeurs mises en jeu est fortement modifiée ! Ainsi aucune copie n'a pu fournir une allure de courbe correcte d'une des informations mutuelles demandées.

B.2 Question 1

Le canal considéré est un simple canal gaussien de bruit $Z = S + V$. Les variables aléatoires étant indépendantes et centrées, la puissance de ce bruit est donc $Q + N$. Par conséquent, la capacité de ce canal gaussien est :

$$C = \frac{1}{2}\log_2\Big(1 + \frac{P}{Q+N}\Big).$$

B.3 Question 2

B.3.1 Expressions de l'information mutuelle

Pour montrer que $I(A;B) = H(A) - H(A|B) = H(A) + H(B) - H(A,B)$ il faut partir de la définition de l'information mutuelle

$$I(A;B) = \sum_{a,b} P(a,b)\log_2\Big[\frac{P(a,b)}{P(a)P(b)}\Big]$$

et appliquer la règle de Bayes : $P(a,b) = P(a|b)P(b) = P(b|a)P(a)$. J'ai utilisé ici les notations du cas discret, le cas continu s'en déduisant immédiatement par changement de notation.

B.3.2 Calcul de $I(U;S)$ et de $I(U;Y)$

Application directe de la première relation trouvée en (i).

$$\begin{aligned}
I(U;S) &= H(U) - H(U|S) \\
&\overset{(a)}{=} H(X + \alpha S) - H(X + \alpha S|S) \\
&\overset{(b)}{=} H(X + \alpha S) - H(X|S) \\
&\overset{(c)}{=} H(X + \alpha S) - H(X) \\
&\overset{(d)}{=} \frac{1}{2}\log_2\big[(2\pi e)(P + \alpha^2 Q)\big] - \frac{1}{2}\log_2\big[(2\pi e)P\big] \\
&= \frac{1}{2}\log_2\Big(\frac{P + \alpha^2 Q}{P}\Big)
\end{aligned}$$

(a) par définition de U, (b) par application du conditionnement, (c) car X est indépendante de S, (d) par l'entropie de la somme de deux VA gaussiennes indépendantes.

Application directe de la première relation trouvée en (ii).

$$\begin{aligned} I(U;Y) &= H(Y)+H(U)-H(Y,U) \\ &\overset{(e)}{=} H(X+S+V)+H(X+\alpha S)-H(Y,U) \\ &\overset{(f)}{=} \frac{1}{2}\log_2(P+Q+N)+\frac{1}{2}\log_2(P+\alpha^2 Q)-\frac{1}{2}\log_2\Big\{(2\pi e)^2 \mathrm{Det}\Big(E\big[(Y,U)(Y,U)^T\big]\Big)\Big\} \\ &\overset{(g)}{=} \frac{1}{2}\log_2(P+\alpha^2 Q)-\frac{1}{2}\log_2\Big\{(2\pi e)^2\big[(P+Q+N)(P+\alpha^2 Q)-(P+\alpha Q)^2\big]\Big\}. \end{aligned}$$

B.3.3 Allure des grandeurs considérées : $I(U;S), I(U;Y), R(\alpha) = I(U;Y) - I(U;S)$

B.3.4 Valeur de α qui maximise $R(\alpha) = I(U;Y) - I(U;S)$

Le taux $R(\alpha)$ se déduit sans calculs supplémentaires des questions précédentes :

$$R(\alpha) = \frac{1}{2}\log_2\Big[\frac{P(P+Q+N)}{PQ(1-\alpha)^2+N(P+\alpha^2 Q)}\Big].$$

En calculant la dérivée de $R(\alpha)$ on trouve facilement que la valeur optimale de α est $\alpha^* = \frac{P}{P+N}$.

Lorsque $P = Q = N = 1$, on a donc $\alpha = 0,5$ ce qui correspond bien aux courbes tracées dans la question précédente.

B.3.5 Capacité du canal

On voit qu'en $\alpha = \alpha^*$ la fonction $R(\alpha)$ vaut $\frac{1}{2}\log_2\Big(1+\frac{P}{N}\Big)$, ce qui signifie que la capacité de ce canal est la même que celle d'un canal gaussien sans la perturbation S. Ainsi, en utilisant un codage approprié il est possible de s'affranchir de la perte de performances que l'on avait observée dans la toute première question du problème. Ce résultat remarquable est dû à Max H. Costa qui a su exploiter la formule de capacité donnée par Gel'fand et Pinsker en 1981.

Annexe C

Sujet examen 2005

Note : épreuve sans aucun document mais calculatrices autorisées.
Durée de l'épreuve : 1 heure.

Problème à résoudre

On considère un canal à relais unique (figure E.1). Il y a donc un émetteur qui produit le signal $X \sim \mathcal{N}(0,P)$ (gaussien de moyenne nulle et de puissance P) et un relais qui produit le signal $X_{12} \sim \mathcal{N}(0,P_{12})$. Le relais produit son signal à partir du signal qu'il reçoit, c'est-à-dire $Y_1 = X + Z_1$ où $Z_1 \sim \mathcal{N}(0,N_1)$ est un bruit blanc gaussien et indépendant de X. Le récepteur, quant à lui, reçoit un signal qui comprend la contribution du relais et de l'émetteur : $Y_2 = X + X_{12} + Z_2$ où là encore $Z_2 \sim \mathcal{N}(0,N_2)$ est un bruit blanc gaussien et indépendant de X et X_{12}.

A partir de la théorie des flux on peut facilement montrer que la capacité de ce canal dans le cas général (signaux non nécessairement gaussiens), notée C, vérifie nécessairement l'inégalité suivante :

$$C \leq \min\{I(X;Y_1,Y_2|X_{12}), I(X,X_{12};Y_2)\}. \tag{C.1}$$

1. Fournir une explication intuitive à la borne supérieure de la capacité du canal à relais.
2. En supposant la chaîne de Markov $X-(X_{12},Y_1)-Y_2$, simplifier l'écriture de la borne supérieure.
3. On tient maintenant compte du fait que les signaux sont gaussiens. De plus on suppose que le signal émis a une structure particulière : $X = X_0 + \sqrt{\frac{\rho P}{P_{12}}} X_{12}$, ce qui crée une dépendance entre X et X_{12}. Le signal $X_0 \sim \mathcal{N}(0,\overline{\rho}P)$ est supposé indépendant de X_{12}. Le paramètre ρ est un paramètre de répartition de puissance entre les deux composantes du signal émis : $\rho \in [0,1]$ et $\overline{\rho} = 1-\rho$. Expliciter alors $I(X,X_{12};Y_2)$ en fonction des paramètres du problème (puissances).
4. De la même manière expliciter $I(X;Y_1|X_{12})$ en fonction des paramètres du problème.
5. En déduire l'expression de la borne supérieure dans le cas gaussien avec les hypothèses faites jusqu'alors. On notera cette borne $R_G(\rho)$.
6. On suppose maintenant que $N_2 > N_1$. Noter que le codeur doit maximiser $R_G(\rho)$ pour obtenir le meilleur débit de transmission possible avec le codage choisi. Montrer qu'il existe une valeur critique de la puissance P_{12} au-delà de laquelle R_G ne bénéficie plus d'une augmentation de la puissance de relayage.
7. Pour une puissance P_{12} fixée, quelle est la valeur de ρ qui maximise $R_G(\rho)$?

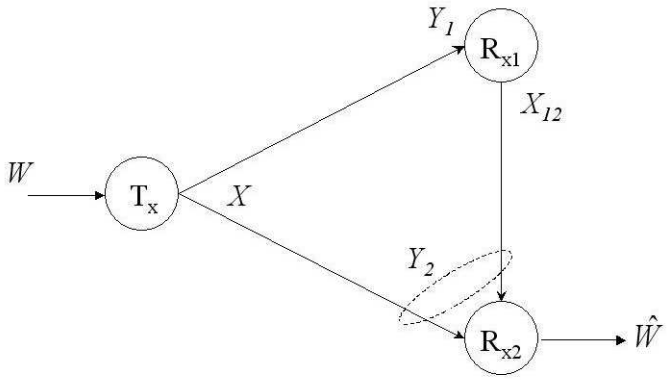

FIG. C.1 – Canal à relais

Annexe D

Corrigé examen 2005

1. *Question sur 3 pts*. On considère deux sections différentes à travers lesquelles passe tout le flux d'information. La première section est une section au niveau de l'émetteur, toute l'information de la source est contenue dans les deux sous-flux, soit la variable aléatoire X. Le signal X traverse cette section et se retrouve intégralement dans le couple de VA (Y_1, Y_2). L'information mutuelle ne peut donc pas excéder $I(X;Y_1,Y_2|X_{12})$ qui, elle, correspond au cas où les deux récepteurs seraient colocalisés ou bien avec un canal inter-récepteur parfait. On fait un raisonnement identique pour la section indiquée sur la figure, au niveau du récepteur. Toute l'information contenue dans Y_2 provient du couple (X, X_{12}), d'où le terme $I(X,X_{12};Y_2)$, qui correspond au cas où les deux émetteurs seraient colocalisés ou bien avec un canal inter-émetteur parfait.
2. *Question sur 3 pts*. Règle de chaîne : $I(X;Y_1,Y_2|X_{12}) = I(X;Y_2|X_{12}) + I(X;Y_1|X_{12},Y_1)$. Chaîne de Markov : $I(X;Y_2|X_{12},Y_1) = 0$ car $p(y_2|x_{12},y_1,x) = p(y_2|x_{12},y_1)$. Donc $I(X;Y_1,Y_2|X_{12}) = I(X;Y_1|X_{12})$.
3. *Question sur 5 pts*. Par définition $I(X,X_{12};Y_2) = H(Y_2) - H(Y_2|X,X_{12})$. On a $Y_2 = X_0 + \left(1+\sqrt{\frac{\rho P}{P_{12}}}\right) X_{12} + Z_2$. Puisque Y_2 est centrée et réelle, on sait que $H(Y_2) = \frac{1}{2}\log(2\pi e E(Y_2^2))$. Puisque les trois composantes de Y_2 sont indépendantes entre elles, on a $E(Y_2^2) = \overline{\rho}P + \left(1+\sqrt{\frac{\rho P}{P_{12}}}\right)^2 P_{12} + N_2 = P + P_{12} + 2\sqrt{\rho P P_{12}} + N_2$. Comme $H(Y_2|X,X_{12}) = H(Z_2|X,X_{12}) = H(Z_2) = \frac{1}{2}\log(2\pi e N_2)$. Au final $I(X,X_{12};Y_2) = C\left(\frac{P+P_{12}+2\sqrt{\rho P P_{12}}}{N_2}\right)$ où $C(x) = \frac{1}{2}\log(1+x)$.
4. *Question sur 5 pts*. Par définition $I(X;Y_1|X_{12}) = H(Y_1|X_{12}) - H(Y_1|X_{12},X)$. On a $Y_1 = X + Z_1 = X_0 + \sqrt{\frac{\rho P}{P_{12}}} X_{12} + Z_1$. Donc $H(Y_1|X_{12}) = H(X_0+Z_1|X_{12}) = H(X_0+Z_1) = \frac{1}{2}\log(2\pi e(\overline{\rho}P + N_1))$. Et $H(Y_1|X_{12},X) = H(Z_1|X_{12}) = H(Z_1) = \frac{1}{2}\log(2\pi e N_1)$. En conclusion, $I(X;Y_1|X_{12}) = C\left(\frac{\overline{\rho}P}{N_1}\right)$.
5. *Question sur 2 pts*. On a tout simplement $R_G(\rho) = \min\left\{C\left(\frac{P+P_{12}+2\sqrt{\rho P P_{12}}}{N_2}\right), C\left(\frac{\overline{\rho}P}{N_1}\right)\right\}$.
6. *Question sur 2 pts*. En choisissant P_{12} de manière à ce que les deux arguments du min soit égaux, on montre que la puissance de relayage critique correspondante est $P_{12}^{(crit)} = P\frac{N_2-N_1}{N_1}$. Si P_{12} excède cette valeur critique, c'est le terme $C\left(\frac{\overline{\rho}P}{N_1}\right)$ qui limite le débit de transmission.
7. *Question sur 2 pts*. Si $P_{12} \geq P_{12}^{(crit)}$, il faut choisir $\rho = 0$. Sinon, pour P_{12} donnée, il faut trouver les valeurs de ρ qui sont soultions de l'équation données par l'égalité des deux arguments du min. Le meilleur ρ est donné par : $\bar{\rho}^* = \frac{a_1^2+a_3^2 \pm 2\sqrt{a_1^2 a_3^2}}{a_2^2}$ où $\rho^* \in [0,1]$, $a_0 = P + P_{12}$, $a_1 = \sqrt{P P_{12}}$, $a_2 = P\frac{N_2}{N_1}$, $a_3^2 = a_1^2 + a_2^2 - a_0 a_2$.

Annexe E

Examen 2006

Examen écrit de Théorie de l'Information : Master SRC

Novembre 2006

Note : tous documents autorisés, calculatrice autorisée.
Durée de l'épreuve : 1 h 30.

Problème à résoudre

On s'intéresse au choix de la stratégie de relayage dans le cadre du canal à relais. Comme le montre la figure E.1 ce canal comporte un émetteur, un relais et un récepteur. Le but final du problème proposé est de savoir s'il vaut mieux que le relais amplifie et transfère sans modification au récepteur le signal reçu par le relais ou bien que le relais décode, recode et transfère le signal qu'il a reçu. Pour cela nous adopterons une modélisation simple pour laquelle toutes les variables aléatoires mises en jeu seront réelles, gaussiennes et de moyennes nulles.

Les équations des différents signaux reçus s'écrivent comme suit :

$$\begin{cases} Y_1 &= X + Z_1 \\ Y_2 &= X + Z_2 \\ Y_{12} &= X_{12} + Z_{12} \end{cases}$$

où $X \sim \mathcal{N}(0,P)$ est le signal émis par la source, $Z_1 \sim \mathcal{N}(0,N_1)$ est le bruit du canal source - relais, $Z_2 \sim \mathcal{N}(0,N_2)$ est le bruit du canal source - récepteur, X_{12} est le signal émis par le relais et $Z_{12} \sim \mathcal{N}(0,N_{12})$ est le bruit du canal relais - récepteur. Les variables aléatoires X, Z_1, Z_2, Z_{12} sont supposées indépendantes.

1. Étude de la stratégie "amplify-and-forward"
   Pour cette stratégie le signal de coopération, émis par le relais, s'exprime par : $X_{12} = a_{12}Y_1$.
   (a) Sachant que le puissance du relais est contrainte par $E(X_{12}^2) = P_{12}$, expliciter la constante d'amplification a_{12}.
   (b) Expliciter la quantité $H(Y_2,Y_{12}|X)$. On pourra pour cela considérer le couple (Y_2,Y_{12}) comme un vecteur à deux composantes et appliquer la formule de l'entropie conditionnelle pour une variable aléatoire vectorielle.
   (c) Expliciter la quantité $H(Y_2,Y_{12})$. En déduire l'expression de l'information mutuelle $I(X;Y_2,Y_{12})$.
2. Étude de la stratégie "decode-and-forward"

Maintenant on suppose que le signal émis par le relais s'écrit sous la forme suivante : $X_{12} = d_{12}\hat{X} = d_{12}(X + \delta X)$ où δX est une variable aléatoire traduisant les erreurs de décodage au niveau du relais et qui n'est pas décorrélée de X : $E(X\delta X) \neq 0$. On suppose que la quantité $E(X - \hat{X})^2 = E(\delta X)^2$ est connue : $E(\delta X)^2 = \varepsilon$.

(a) Sachant que l'on impose $E(X^2) = P$, $E(\hat{X}^2) = P$ et $E(X_{12}^2) = P_{12}$ expliciter le coefficient d'amplification d_{12}.

(b) On maintient tout au long du problème les contraintes de puissances données ci-dessus. Exprimer $E(X\delta X)$ en fonction des données du problème.

(c) Expliciter la quantité $H(Y_2, Y_{12}|X)$. On pourra pour cela considérer le couple (Y_2, Y_{12}) comme un vecteur à deux composantes et appliquer la formule de l'entropie conditionnelle pour une variable aléatoire vectorielle.

(d) Expliciter la quantité $H(Y_2, Y_{12})$. En déduire l'expression de l'information mutuelle $I(X; Y_2, Y_{12})$ en fonction des données du problèmes.

3. A quelle condition la stratégie "amplify-and-forward" est-elle meilleure que la stratégie "decode-and-forward" ?

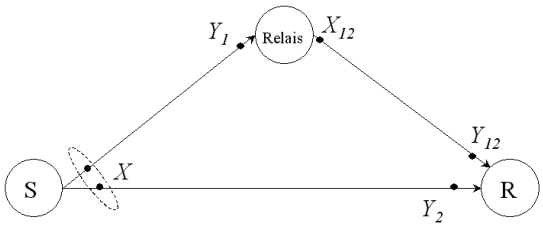

FIG. E.1 – Canal à relais

Annexe F

Examen 2006

Corrigé de l'examen écrit de Théorie de l'Information – Session 2006 Master SRC

Samson Lasaulce

Question 1.a.

Nous avons $E(X_{12}^2) = E[a_{12}Y_1^2] = a_{12}^2 E(Y_1^2) = a_{12}^2(P+N_1)$, puisque X et Z_1 sont indépendants. Donc

$$a_{12}^2 = \frac{P_{12}}{P+N_1}. \tag{F.1}$$

Question 1.b.

En appliquant la règle de chaîne pour l'entropie, nous avons $H(Y_2,Y_{12}|X) = H(Y_2,Y_{12},X) - H(X)$. Les variables aléatoires en jeu étant gaussiennes et réelles, le premier terme est donné par $H(Y_2,Y_{12},X) = \frac{1}{2}\log_2|2\pi e\mathbf{R}_{Y_2,Y_{12},X}|$ où

$$\begin{aligned}
|\mathbf{R}_{Y_2,Y_{12},X}| &= \begin{vmatrix} E(Y_2^2) & E(Y_2Y_{12}) & E(Y_2X) \\ E(Y_{12}Y_2) & E(Y_{12}^2) & E(Y_{12}X) \\ E(XY_2) & E(XY_{12}) & E(X^2) \end{vmatrix} && \text{(F.2)} \\
&= \begin{vmatrix} P+N_2 & a_{12}P & P \\ a_{12}P & P_{12}+N_{12} & a_{12}P \\ P & a_{12}P & P \end{vmatrix} && \text{(F.3)} \\
&\overset{(a)}{=} \begin{vmatrix} N_2 & a_{12}P & P \\ 0 & P_{12}+N_{12} & a_{12}P \\ 0 & a_{12}P & P \end{vmatrix} && \text{(F.4)} \\
&= N_2P[(P_{12}+N_{12}) - a_{12}^2P] && \text{(F.5)}
\end{aligned}$$

où l'égalité (a) découle du fait que le déterminant est invariant par le changement : colonne 1 $\rightarrow$ colonne 1 moins la colonne 3.

Le deuxième terme de l'entropie conditionnelle étant simplement donné par $H(X) = \frac{1}{2}\log_2(2\pi eP)$ nous trouvons que

$$H(Y_2,Y_{12}|X) = \frac{1}{2}\log_2\left\{(2\pi e)^2 N_2[(P_{12}+N_{12}) - a_{12}^2 P]\right\}. \tag{F.6}$$

Question 1.c.

Nous savons que $H(Y_2,Y_{12}) = \frac{1}{2}\log_2|2\pi e\mathbf{R}_{Y_2,Y_{12}}|$, calcul qui est un cas particulier de la question précédente :

$$\begin{aligned} |\mathbf{R}_{Y_2,Y_{12}}| &= \begin{vmatrix} E(Y_2^2) & E(Y_2Y_{12}) \\ E(Y_{12}Y_2) & E(Y_{12}^2) \end{vmatrix} && \text{(F.7)} \\ &= \begin{vmatrix} P+N_2 & a_{12}P \\ a_{12}P & P_{12}+N_{12}P \end{vmatrix} && \text{(F.8)} \\ &= (P+N_2)(P_{12}+N_{12}) - a_{12}^2P^2. && \text{(F.9)} \end{aligned}$$

L'expression de l'information mutuelle $I(X;Y_2,Y_{12}) = H(Y_2,Y_{12}) - H(Y_2,Y_{12}|X)$ s'en déduit alors. Après simplifications on trouve que :

$$I_{af}(X;Y_2,Y_{12}) = \frac{1}{2}\log_2\left(1 + \frac{P}{N_2} + \frac{a_{12}^2P}{P_{12}+N_{12}-a_{12}^2P}\right). \tag{F.10}$$

Question 2.a.

Même type de calcul que pour 1.a. On trouve :

$$d_{12}^2 = \frac{P_{12}}{P}. \tag{F.11}$$

Question 2.b.

En développant, $E(\hat{X}^2) = E(X^2) + 2E(X\delta X) + E(\delta X^2) = P$, d'où

$$E(X\delta X) = -\frac{\varepsilon}{2}. \tag{F.12}$$

Question 2.c.

Même type de calcul que pour 1.b. On trouve :

$$\begin{aligned} |\mathbf{R}_{Y_2,Y_{12},X}| &= \begin{vmatrix} P+N_2 & d_{12}\left(P-\frac{\varepsilon}{2}\right) & P \\ d_{12}\left(P-\frac{\varepsilon}{2}\right) & P_{12}+N_{12} & d_{12}\left(P-\frac{\varepsilon}{2}\right) \\ P & d_{12}\left(P-\frac{\varepsilon}{2}\right) & P \end{vmatrix} && \text{(F.13)} \\ &= N_2\left[(P_{12}+N_{12})P - d_{12}^2\left(P-\frac{\varepsilon}{2}\right)^2\right]. && \text{(F.14)} \end{aligned}$$

Question 2.d.

Nous savons que $H(Y_2,Y_{12}) = \frac{1}{2}\log_2|2\pi e\mathbf{R}_{Y_2,Y_{12}}|$, calcul qui est un cas particulier de la question précédente. L'expression de l'information mutuelle $I(X;Y_2,Y_{12}) = H(Y_2,Y_{12}) - H(Y_2,Y_{12}|X)$ s'en déduit alors. Après simplifications on trouve que :

$$I_{df}(X;Y_2,Y_{12}) = \frac{1}{2}\log_2\left[1+\frac{P}{N_2}+\frac{d_{12}^2P\left(1-\frac{\varepsilon}{2P}\right)^2}{P_{12}+N_{12}-d_{12}^2P\left(1-\frac{\varepsilon}{2P}\right)^2}\right]. \tag{F.15}$$

Question 3.

On peut alors montrer facilement que

$$I_{af}(X;Y_2,Y_{12}) > I_{df}(X;Y_2,Y_{12}) \Leftrightarrow \frac{\varepsilon}{2P} > 1-\sqrt{\frac{P}{P+N_1}}. \tag{F.16}$$

Nous avons donc à notre disposition un seuil explicite sur le niveau de bruit de décodage introduit par le protocole decode-and-forward au delà duquel il vaut mieux faire un simple amplify-and-forward.

Bibliographie

[1] S. I. Gelfand and M. S. Pinsker, "Coding for channels with random parameters", *Problems of Control and Information Theory*, Vol 9, No 1, pp. 19–31, 1980.

[2] C. E. Shannon, "A mathematical theory of communication", *Bell Labs Journal*, 1948.

[3] R. W. Hamming, "Error-detecting and error-correcting codes", *Bell Systems Technical Journal*, 1950.

[4] T. M. Cover and J. A. Thomas, "Elements of Information Theory", *Wileys Series in Telecommunications*, 2ème édition, 2006.

[5] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding : Turbo-codes", *International Conference on Communications*, Genève, 1993.

[6] K. Salamatian and R. Khalili, "An Information Theory for Erasure Channels", Conf.on Commun., Control and Computing, Allerton, Sep. 2005.

[7] R. Gowaikar, A.F. Dana, R. Palanki, B. Hassibi and M. Effros, "Capacity of Wireless Erasure Networks", IEEE Trans. on Inform. Theory, Vol. 52, Issue 3, pp. 789–804, March 2006.

[8] S. N. Diggavi and T. M. Cover, "The Worst Additive Noise Under a Covariance Constraint", *IEEE Transactions on Information Theory*, 47(7) : 3072–3081, November 2001.

[9] E. Telatar, "Capacity of Multi-antenna Gaussian Channels", *Europ. Trans. Telecom.*, Vol. 10, pp 585–595, Nov. 1999

[10] R. G. Gallager, "Information Theory and Reliable Communication", *Wiley*, 1968.

[11] W.Hirt and J. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference", *IEEE Transactions on Information Theory*, Vol. 34, pp 380–388, May 1988.

[12] Ozarow, L.H. ; Shamai, S. ; Wyner, A.D., "Information theoretic considerations for cellular mobile radio", *Vehicular Technology, IEEE Transactions on*, Volume 43, Issue 2, May 1994 Page(s) :359–378.

[13] L. Zheng and D. N. C. Tse, "Diversity and multiplexing : A fundamental tradeoff in multiple antenna channels", *IEEE Transactions on Information Theory*, Vol. 49, No 5, May 2003.

[14] D. Tse and P. Viswanath, "Fundamentals of Wireless Communication", *Cambridge University Press*, 2005.

[15] L. Zheng, D.N. Tse, "Optimal Diversity-Multiplexing Tradeoff in Multiple Antenna Channels", *Proc. Allerton Conf. Comm., Control, Computing*, Monticello, IL, pp. 835–844, Oct. 2001.

[16] S. M. Alamouti, "A simple transmit diversity technique for wireless communications", *IEEE Journal on Selected Areas in Communications*, Vol. 16, No. 8, pp. 1451-Ű1458.

[17] M. H. M. Costa, "Writting on Dirty Paper", *IEEE Trans. on Inform. Theory*, Vol. IT-29, Issue 3, pp. 439–441, May 1983.

[18] S. I. Gel'fand and M. S. Pinsker, "Coding for Channel with Random Parameters", *Problems of Control and Inform. Theory*, Vol. 9 (1), pp. 19–31, 1980.

[19] N. T. Gaarder and J. K. Wolf, "The Capacity Region of a Multiple-Access Discrete Memoryless Channel Can Increase with Feedback", *IEEE Transactions on Information Theory*, Vol. 21, No. 1, pp. 100–102, 1975.

[20] T.M. Cover, "Broadcast channels", *IEEE Trans. Inf. Theo.*, Vol. 18(1), pp. 2–14, 1972.

[21] P. Bergmans, "Random coding theorem for broadcast channels with degraded components", *IEEE Transactions on Information Theory*, IT-19(2), pp. 197–207, March 1973.

[22] A. A. El Gamal, "The capacity of a class of broadcast channels", *IEEE Transactions on Information Theory*, Vol. 25, No.2, pp 166-169, March 1979.

[23] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel", *IEEE Trans. Inform. Theory*, IT-25(5) : 572-584, 1979.

[24] A. Sendonaris, E. Erkip and B. Aazhang, "User cooperation diversity-Part II : Implementation aspects and performance analysis", *IEEE Trans. on Comm.*, COM-51(11) : 1939-1948, 2003.

[25] L. R. Ford and D. R. Fulkerson,, "Flows in networks", *Princeton university press*, NJ, 1962.

[26] E. V. Belmega, S. Lasaulce and M. Debbah, "Capacity of Cooperative Channels : Three Terminals Case Study", Cooperative Wireless Communication, *to be published by Auerbach Publications, CRC Press, Taylor and Francis Group*, 2007.