Nouvelle borne atteignable de la probabilité d'erreur pour des transmissions en paquets courts *

Dadja Toussaint ANADE AKPO[1], Jean-Marie GORCE[1], Philippe MARY[2]

[1]Univ. Lyon, INSA Lyon, CITI, INRIA
20 avenue Albert Einstein, 69621

[2]Univ. Rennes, INSA Rennes, IETR, CNRS
20 avenue des Buttes de Coësmes, 35708 Rennes

dadja-toussaint.anade-apko@insa-lyon.fr, jean-marie.gorce@insa-lyon.fr
philippe.mary@insa-rennes.fr

Résumé – La théorie de l'information en taille finie offre un cadre mathématique adapté à l'établissement de bornes fondamentales pour les systèmes de communication utilisant de petits paquets, comme les réseaux *internet of things* IoT. Dans ce papier, nous proposons une borne supérieure de la probabilité d'erreur atteignable pour une communication point-à-point (P2P) en bornant les probabilités d'outage et de confusion et qui est légèrement meilleure que celles proposées dans la littérature. L'originalité de l'approche réside dans la réécriture du test sur la densité d'information entre les distributions d'entrée et de sortie, couplé à un décodage itératif à seuil variable. Une application est faite pour le canal gaussien pour illustrer le gain.

Abstract – Finite blocklength information theory is a suitable mathematical framework to establish fundamental bounds for communication systems with short packets, like in internet of things (IoT) networks. In this paper, a new and better, w.r.t. existing literature, upper-bound on the achievable error probability for a point-to-point communication is proposed. The contribution lies in a reformulation of the test on information density between the channel input and output distributions, conjugated with an iterative decoding with variable decision threshold. An application is done in Gaussian channel to show the gain.

1 Introduction

La théorie de l'information en taille finie a pour objet l'étude des performances fondamentales des systèmes de communication, que ce soit au niveau du codage source ou canal, lorsque la taille des mots de codes n ne tend pas vers l'infini, comme cela est habituellement fait dans la théorie asymptotique de Shannon [1]. En ce qui concerne le problème du codage canal, Shannon a en particulier montré qu'il existait une séquence (n, M_n, ϵ_n) de codes atteignant un débit strictement positif avec une probabilité d'erreur, ϵ_n, tendant vers 0 :

$$R = \lim_{n\to\infty} \frac{\log_2(M_n)}{n} > 0, \tag{1}$$

avec $\lim_{n\to\infty} \epsilon_n = 0$ et M_n est la cardinalité de l'ensemble des messages à transmettre. Dans ce travail, nous nous intéressons à la problématique de la probabilité d'erreur atteignable dans un canal P2P lorsque $n < \infty$ utilisations de canal (u.c.) sont employés pour une taille d'ensemble de messages M fixée. Notons que Shannon et d'autres, comme Gallager [2], se sont intéressés à ce problème dès les années 50-60. Mais ce n'est que très récemment qu'une approche globale a été proposée pour dériver les bornes atteignables et les limites hautes, dites *converse*, des débits pour n'importe quel canal à une probabilité d'erreur fixée [3].

Pour la partie directe, i.e. atteignabilité, l'approche des auteurs de [3] consiste à considérer un test sur la densité d'information, définie en section 3, pour un mot de code donné. Les mots de code à tester sont ordonnés. Pour chaque mot de code testé, si la densité d'information excède un seuil (fixé) alors le mot de code testé est celui envoyé, sinon on passe au suivant et on s'arrête dès qu'un mot de code passe le test. La borne finale est obtenue par un argument de codage aléatoire [3, Th. 18]. La réciproque ou converse, appelée méta-converse, est basée sur l'établissement d'une relation entre la probabilité d'erreur moyenne observée sur un canal et la probabilité d'erreur d'un test d'hypothèse entre deux transformations, représentant le lien entre la sortie et l'entrée du canal.

MolanvianJazi s'est intéressé au cas du canal Gaussien et en étendant au cas multi-utilisateurs en liaison montante, MAC [4]. Son approche diffère de [3] dans le mesure où les auteurs bornent les probabilités d'outage, i.e. probabilité que le bon code ne passe pas le test, et de confusion, i.e. un mauvais code passe le test, avec des techniques différentes, i.e. théorème centrale limite pour les fonctions et changement de mesure pour l'outage, et en bornant la moyenne d'un produit d'exponentiel de variables aléatoires, pour la confusion.

*This work has been (partly) funded by the French National Agency for Research (ANR) under grant ANR-16-CE25-0001 - ARBURST.

D'autres travaux se sont intéressés à l'obtention de bornes pour d'autres canaux, e.g. broadcast (BC) [5, 6]. Notons que dans les deux cas précédents, [3, 4], et autres travaux, les auteurs considèrent un seuil fixe par rapport auquel le test est effectué, malgré le fait qu'à mesure que des mots de code sont éliminés lors du décodage itératif, la probabilité de trouver le bon mot de code change et le seuil du test d'hypothèse considéré devrait varier aussi.

Cet article a pour objectif d'étudier l'influence d'un seuil variable sur la borne de la probabilité d'erreur atteignable dans le cas d'un canal Gaussien, point-à-point. Notre contribution consiste en la réécriture du test de décision, à chaque étape du décodage au niveau du récepteur, comme un test d'hypothèse binaire. Cette formulation permet au passage de justifier le choix d'un test sur la densité d'information découlant d'une adaptation d'un test de typicalité issu du régime asymptotique, i.e. $n \to \infty$. Le papier comprend les sections suivantes : le problème et les notations sont définis en section 2. La section 3 contient notre contribution sur le calcul de la borne de l'erreur. La section 4 compare notre borne avec celle de la littérature et la section 5 conclue l'article.

2 Définition du problème

2.1 Notation

Les variables aléatoires sont notées en majuscules, i.e. X, et leur réalisation en minuscule, i.e. x. Les ensembles sont caligraphiés avec leur dimension en indice supérieur. Ainsi $\mathcal{X}^n$ représente l'ensemble $\mathcal{X}$ de dimension n. Les lettres en gras désignent un vecteur aléatoire ou une réalisation du vecteur aléatoire selon qu'une majuscule ou minuscule est utilisée respectivement. Enfin, $\mathbb{E}_{\boldsymbol{X}}[.]$ désigne l'espérance suivant la distribution $P_{\boldsymbol{X}}$

2.2 Modèle du système étudié

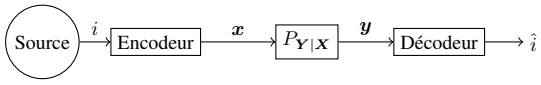


FIGURE 1 – Modèle du canal point-à-point

On considère un système de communication point à point tel que représenté par la figure 1. Dans ce schéma, une source cherche à transmettre un message $i \in \mathcal{W}$, tiré aléatoirement, et où $\mathcal{W} = \{1, 2, \ldots, M\}$, avec M la cardinalité des messages de la source, à un destinataire. Une fonction d'encodage associe un code $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathcal{X}^n$ au message i où n est la longueur du code en nombre d'utilisations de canal. Le canal est modélisé par une distribution de probabilité conditionnelle, $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y} \mid \boldsymbol{x})$, donnant l'observation $\boldsymbol{y} = (y_1, \cdots, y_n) \in \mathcal{Y}^n$ lorsque $\boldsymbol{x}$ est envoyé. Enfin, une fonction de décodage associe un message $\hat{i}$ à l'observation $\boldsymbol{y} \in \mathcal{Y}^n$. Le système est modélisé par le triplet $(\mathcal{X}^n, \mathcal{Y}^n, P_{\boldsymbol{Y}|\boldsymbol{X}})$.

Définition 1 ((n, M)-code)**.** *Soit un couple $(n, M) \in \mathbb{N}^2$, un (n, M)-code pour le système $(\mathcal{X}^n, \mathcal{Y}^n, P_{\boldsymbol{Y}|\boldsymbol{X}})$ est l'ensemble*

$$\{(\boldsymbol{x}(1), \mathcal{D}(1)), (\boldsymbol{x}(2), \mathcal{D}(2)), \ldots, (\boldsymbol{x}(M), \mathcal{D}(M))\}, \quad (2)$$

où $\boldsymbol{x}(i) \in \mathcal{X}^n \ \forall i \in \mathcal{W}$. De plus $\forall (i, j) \in \mathcal{W}^2$, $i \neq j$:

$$\mathcal{D}(i) \cap \mathcal{D}(j) = \emptyset, \quad \textit{et} \quad (3)$$

$$\bigcup_{i \in \mathcal{W}} \mathcal{D}(i) \subseteq \mathcal{Y}^n. \quad (4)$$

Le vecteur $\boldsymbol{x}(i)$ est le code associé au message i et les ensembles $\mathcal{D}(1), \cdots, \mathcal{D}(M)$, sont les régions de décision associées aux messages 1 à M respectivment.

Au récepteur, le message d'indice i est décidé si $\boldsymbol{y} \in \mathcal{D}(i)$. En notant $\mathcal{D}^{\mathsf{c}}(i)$ l'ensemble complémentaire de $\mathcal{D}(i)$ dans $\mathcal{Y}^n$, la probabilité d'erreur de décodage associée au message d'indice $i \in \mathcal{W}$, notée $\lambda(i) \in [0, 1]$, est

$$\lambda(i) = \Pr\left[\boldsymbol{Y} \in \mathcal{D}^{\mathsf{c}}(i) \middle| \boldsymbol{X} = \boldsymbol{x}(i)\right]. \quad (5)$$

La probabilité d'erreur moyenne, notée $\bar{\lambda}$, est obtenue en moyennant sur les codes

$$\bar{\lambda} = \frac{1}{M} \sum_{i=1}^{M} \lambda(i). \quad (6)$$

Définition 2 ((n, M, ϵ)-code)**.** *$\forall \epsilon \in [0, 1]$, un $(n, M, \epsilon)-$code est un $(n, M)-$code atteignant une probabilité d'erreur moyenne qui satisfait $\bar{\lambda} \leq \epsilon$.*

Définition 3 (canal gaussien réduit sans mémoire)**.** *Un canal gaussien réduit sans mémoire est un triplet $(\mathcal{X}^n, \mathcal{Y}^n, P_{\boldsymbol{Y}|\boldsymbol{X}})$ tel que*

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_i - x_i)^2}{2}}. \quad (7a)$$

Les codes présentent une contrainte de puissance :

$$||\boldsymbol{x}(i)||^2 = nP, \ \forall i \in \mathcal{W}, \quad (7b)$$

où $||.||$ est la norme euclidienne.

3 Borne sur la probabilité d'erreur

3.1 Borne existante

Théorème 1 ([3], Th. 18)**.** *Il existe un (n, M, ϵ)-code avec une probabilité d'erreur moyenne telle que :*

$$\begin{aligned} \epsilon \leq\ & P_{\boldsymbol{X}\boldsymbol{Y}}\left[i(\boldsymbol{X}; \boldsymbol{Y}) \leq \log\left(\frac{M-1}{2}\right)\right] \\ & + \frac{M-1}{2} P_{\boldsymbol{X}} P_{\boldsymbol{Y}}\left[i(\boldsymbol{X}; \boldsymbol{Y}) > \log\left(\frac{M-1}{2}\right)\right], \end{aligned} \quad (8)$$

où $i(\boldsymbol{x}, \boldsymbol{y})$ est la densité d'information

$$i(\boldsymbol{x}; \boldsymbol{y}) = \log\left(\frac{P_{\boldsymbol{X}\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})}{P_{\boldsymbol{X}}(\boldsymbol{x}) P_{\boldsymbol{Y}}(\boldsymbol{y})}\right) = \log\left(\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y} \mid \boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}\right) \quad (9)$$

Le théorème 1 s'obtient en considèrant une série de test d'hypothèse binaire telle que

$$H_1 \quad : (\boldsymbol{X}, \boldsymbol{Y}) \sim P_{\boldsymbol{XY}}, \tag{10}$$
$$H_0 \quad : (\boldsymbol{X}, \boldsymbol{Y}) \sim P_{\boldsymbol{X}} P_{\boldsymbol{Y}}, \tag{11}$$

où H_1 signifie que l'entrée et la sortie sont dépendantes et tirées de la loi conjointe $P_{\boldsymbol{XY}}$ et H_0 signifie que l'entrée et la sortie sont indépendantes. Les probabilités *a priori* des hypothèses sont $\mathbb{P}[H_1] = 2/(M+1)$ et $\mathbb{P}[H_0] = (M-1)/(M+1)$. Cela implique que chaque test est effectué sous le même seuil de décision, ce qui est sous-optimal puisqu'à mesure que des codes sont écartés, le seuil de décision doit changer en conséquence pour les tests restants.

3.2 Nouvelle borne

Théorème 2. *Il existe un (n, M, ϵ)-code avec une probabilité d'erreur moyenne telle que*

$$\epsilon \leq \frac{1}{M} \sum_{k=1}^{M} \Bigg(P_{\boldsymbol{XY}} \Big[i(\boldsymbol{X}; \boldsymbol{Y}) \leq \log(M-k) \Big] + (M-k) P_{\boldsymbol{X}} P_{\boldsymbol{Y}} \Big[i(\boldsymbol{X}; \boldsymbol{Y}) > \log(M-k) \Big] \Bigg). \tag{12}$$

Démonstration. Pour une observation $\boldsymbol{y}$, une série de test d'hypothèse binaire est effectuée sur les M messages possibles, en commençant par le message d'indice 1 jusqu'à M. Le $k-$ème test s'énonce :

$$H_1 \quad : (\boldsymbol{X}(k), \boldsymbol{Y}) \sim P_{\boldsymbol{XY}} \tag{13}$$
$$H_0 \quad : (\boldsymbol{X}(k), \boldsymbol{Y}) \sim P_{\boldsymbol{X}} P_{\boldsymbol{Y}} \tag{14}$$

Si H_1 est décidé alors le message d'indice k est choisi au récepteur, sinon on passe au suivant et ce jusqu'à trouver un message dont l'indice satisfait H_1. Les probabilités *a priori* sont $\mathbb{P}[H_1] = 1/(M-k+1)$ et $\mathbb{P}[H_0] = (M-k)/(M-k+1)$. En effet, chaque message est choisi uniformément et au premier test, le message d'indice 1 a donc la probabilité $1/M$ d'avoir était envoyé ; tous les autres, $M - 1/M$. Si l'on arrive au $k-$ème test, cela implique que le message transmis ne fait pas parti des messages ayant les indices compris de 1 à $(k-1)$.

Pour le test d'hypothèse donné ci-dessus, la probabilité d'erreur minimal, ϵ_H, est donnée par

$$\epsilon_H = \mathbb{P}[H_1] \cdot P_{\boldsymbol{XY}} \Big[\log\Big(\frac{P_{\boldsymbol{XY}}[\boldsymbol{X}, \boldsymbol{Y}]}{P_{\boldsymbol{X}}[\boldsymbol{X}] P_{\boldsymbol{Y}}[\boldsymbol{Y}]} \Big) \leq \log\Big(\frac{\mathbb{P}[H_0]}{\mathbb{P}[H_1]} \Big) \Big] + \mathbb{P}[H_0] \cdot P_{\boldsymbol{X}} P_{\boldsymbol{Y}} \Big[\log\Big(\frac{P_{\boldsymbol{XY}}[\boldsymbol{X}, \boldsymbol{Y}]}{P_{\boldsymbol{X}}[\boldsymbol{X}] P_{\boldsymbol{Y}}[\boldsymbol{Y}]} \Big) > \log\Big(\frac{\mathbb{P}[H_0]}{\mathbb{P}[H_1]} \Big) \Big]. \tag{15}$$

Car si $\log\Big(\frac{P_{\boldsymbol{XY}}[\boldsymbol{x}, \boldsymbol{y}]}{P_{\boldsymbol{X}}[\boldsymbol{x}] P_{\boldsymbol{Y}}[\boldsymbol{y}]} \Big) > \log\Big(\frac{\mathbb{P}[H_0]}{\mathbb{P}[H_1]} \Big)$ pour un couple $(\boldsymbol{x}, \boldsymbol{y})$ alors l'hypothèse H_1 est choisie. Donc, la probabilité d'erreur du $k-$ème test s'écrit avec $i(\boldsymbol{x}; \boldsymbol{y})$ comme

$$\epsilon_{k,H} = \frac{1}{M-k+1} P_{\boldsymbol{XY}} \Big[i(\boldsymbol{X}; \boldsymbol{Y}) \leq \log(M-k) \Big] + \frac{M-k}{M-k+1} P_{\boldsymbol{X}} P_{\boldsymbol{Y}} \Big[i(\boldsymbol{X}; \boldsymbol{Y}) > \log(M-k) \Big]. \tag{16}$$

En prenant en compte que le $k-$ème test a une probabilité de $\mathbb{P}[k] \leq \frac{M-k+1}{M}$ d'être effectué, la moyenne de la probabilité d'erreur moyenne de décodage, notée $\mathbb{E}[\bar{\lambda}]$, par l'argument du codage aléatoire est donnée par

$$\mathbb{E}[\bar{\lambda}] = \sum_{i=1}^{M} \frac{M-k+1}{M} \cdot \epsilon_{k,H} \tag{17}$$

$$\leq \frac{1}{M} \sum_{k=1}^{M} \Bigg(P_{\boldsymbol{XY}} \Big[i(\boldsymbol{X}; \boldsymbol{Y}) \leq \log(M-k) \Big] + (M-k) P_{\boldsymbol{X}} P_{\boldsymbol{Y}} \Big[i(\boldsymbol{X}; \boldsymbol{Y}) > \log(M-k) \Big] \Bigg). \tag{18}$$

□

La nouvelle borne (12) contient une somme sur M qui peut complexifier son évaluation. Le corollaire suivant donne une borne supérieure plus simple après quelques opérations algébriques.

Corollaire 1. *Il existe un (n, M, ϵ)-code avec une probabilité d'erreur moyenne telle que :*

$$\epsilon \leq \frac{1}{M} \mathbb{E}_{\boldsymbol{XY}} \Bigg[\frac{1}{2} \Big(e^{2 \min(|i(\boldsymbol{X};\boldsymbol{Y})|^+, \log(M)) - i(\boldsymbol{X};\boldsymbol{Y})} - e^{-i(\boldsymbol{X};\boldsymbol{Y})} \Big) + M - e^{\min(|i(\boldsymbol{X};\boldsymbol{Y})|^+, \log(M))} \Bigg] \tag{19}$$

où $|.|^+ = \max(0, .)$ et $\mathbb{E}_{\boldsymbol{XY}}[.]$ désigne l'espérance suivant la distribution de probabilité $P_{\boldsymbol{XY}}$.

3.3 Application au canal gaussien sans mémoire

Théorème 3. *Il existe un code (n, M, ϵ) pour le canal gaussien sans mémoire, défini en* (7)*, tel que :*

$$\epsilon \leq \mathbb{E}_{\boldsymbol{Y}}[\bar{\lambda}(\boldsymbol{Y})], \tag{20}$$

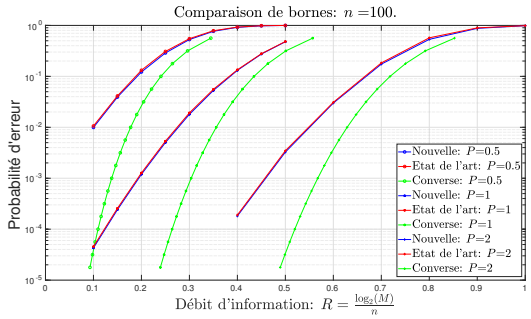
avec

$$\bar{\lambda}(\boldsymbol{y}) \leq \frac{1}{M} \Big[M \cdot F_{D|Y^n}(h_1|\boldsymbol{y}) - F_{D|\boldsymbol{Y}}(h_0|\boldsymbol{y}) - \frac{1}{2} e^{c_0} \int_{h_0}^{h_1} e^{w\alpha_0} f_{D|Y^n}(w|\boldsymbol{y}) dw - \frac{1}{2} e^{-c_0} \int_{h_0}^{h_1} e^{-w\alpha_0} f_{D|\boldsymbol{Y}}(w|\boldsymbol{y}) dw + \frac{M^2-1}{2} e^{-c_0} \int_{h_1}^{1} e^{-w\alpha_0} f_{D|\boldsymbol{Y}}(w|\boldsymbol{y}) dw \Big],$$

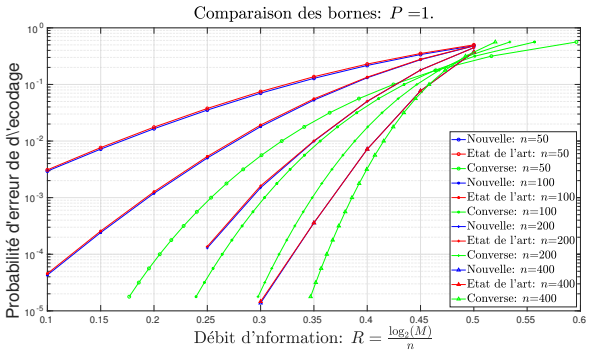
$$c_0 = -\frac{1}{2}(||\boldsymbol{y}||^2 + nP) - \frac{n}{2}\log(2\pi) - \log\left(P_{\boldsymbol{Y}}(\boldsymbol{y})\right),$$
$$\alpha_0 = ||\boldsymbol{y}|| \cdot \sqrt{nP},$$
$$h_0 = \min\left(1, \max(-1, -\frac{c_0}{\alpha_0})\right),$$
$$h_1 = \min\left(1, \max(-1, \frac{\log(M) - c_0}{\alpha_0})\right),$$
$$f_{D|\boldsymbol{Y}}(t|\boldsymbol{y}) = \frac{e^{\alpha_0 t}(1-t^2)^{\frac{n-3}{2}}}{r(\boldsymbol{y})},$$
$$r(\boldsymbol{y}) = \int_{-1}^{1} e^{\alpha_0 s}(1-s^2)^{\frac{n-3}{2}} ds,$$
$$F_{D|\boldsymbol{Y}}(d|\boldsymbol{y}) = \int_{-1}^{d} f_{D|\boldsymbol{Y}}(t|\boldsymbol{y}) dt,$$
$$P_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{\Gamma(n/2)}{\sqrt{\pi}\Gamma(n-1/2)}(2\pi)^{-n/2} e^{-\frac{||y^n||^2+nP}{2}} r(\boldsymbol{y}).$$

4 Résultats numériques

Dans cette section, nous illustrons la comparaison numérique entre la borne du théorème 3 et celle du théorème 1 pour le canal gaussien et la converse de Polyanskiy [3, Th. 41]. La mention *Nouvelle* designe la borne (20) et "Etat de l'art", la borne (8).



(a) $n = 100$



(b) $P = 1$

FIGURE 2 – Probabilités d'erreurs obtenues avec le théorème 1 en rouge, le théorème 3 en bleu et le converse en vert.

La figure 2 compare la nouvelle borne de la probabilité d'erreur obtenue avec le théorème 3 avec celle de l'état de l'art obtenue avec le théorème 1. Les figures 2a et 2b présentent la probabilité d'erreur atteignable pour différentes valeurs de P quand $n = 100$ et pour $n \in \{50, 100, 200, 400\}$ lorsque $P = 1$ respectivement. On remarque que le gain de la nouvelle borne est marginale pour le canal gaussien. La nouvelle borne a un gain relatif moyen de 6% par rapport à l'ancienne.

Une étude exhaustive nous a permis de constater que le gain relatif de la nouvelle borne (12) par rapport à (8) peut atteindre 25% dans le meilleur des cas. Ce gain est possible pour les canaux dont la densité de l'information mutuelle $i(\boldsymbol{x};\boldsymbol{y})$ se concentre autour de la valeur $\log\left(\frac{M-1}{2}\right)$.

5 Conclusion

Dans cet article, une nouvelle borne d'atteignabilité sur la probabilité d'erreur pour une communication point à point avec des paquets courts a été proprosée. L'approche proposée se base sur l'application d'un seuil variable sur le test d'hypothèse binaire permettant d'obtenir une probabilité d'erreur atteignable plus faible que celle proposée dans la littérature sur un décodeur à seuil. Le décodeur à seuil étant utilisé dans littérature pour pallier à la complexité du décodeur de maximum de vraisemblance. Cependant une expression analytique moyennée sur les observations reste encore à trouver. Enfin, cette approche pourrait être appliquée à l'étude des bornes atteignables pour les systèmes multi-utilisateurs en taille finie.

Références

[1] C. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 1948.

[2] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, no. 1, pp. 3–18, Jan 1965.

[3] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, pp. 2307–2359, December 2010.

[4] E. MolavianJazi and J. Laneman, "A Second-Order Achievable Rate Region for Gaussian Multi-Access Channels via a Central Limit Theorem for Functions," *IEEE Transactions on Information Theory*, vol. 61, pp. 6719–6733, December 2015.

[5] A. Unsal and J.-M. Gorce, "The Dispersion of Superposition Coding for Gaussian Broadcast Channels," in *IEEE Information Theory Workshop 2017*, Kaohsiung, Taiwan, Nov. 2017. [Online]. Available : https://hal.archives-ouvertes.fr/hal-01643260

[6] V. Y. F. Tan and O. Kosut, "On the Dispersion of Three Network Information Theory Problems," *IEEE Transaction on Information Theory*, vol. 60, no. 2, pp. 1–184, 2014.