

Université du Québec
Institut National de la Recherche Scientifique
Énergie, Matériaux et Télécommunications

**Adaptation des taux et des puissances de transmission pour
les schémas Incremental Redundancy HARQ tronqués**

Par

Imene BEN SALEM

Mémoire présenté pour l'obtention du grade

Maître ès sciences (M.Sc.) en télécommunications

Jury d'évaluation

Président du jury
Examineur externe
Directeur de recherche
Codirecteur de recherche

Jean-Charles Grégoire *INRS-EMT*
Paul Fortier, *Université Laval*
Leszek SZCZECIŃSKI, *INRS-EMT*
Fabrice LABEAU, *Mc-Gill*

Remerciement

C'est en aveu du succès de ce mémoire que mes sincères mercis se vouent à mon directeur de recherche Prof. Leszek Szczecinski, professeur à l'INRS-EMT, pour sa serviabilité, ses compétences et ses directives fructueuses qu'il n'a cessé de me prodiguer tout au long de ce mémoire.

Je tiens également à présenter mes sincères remerciements à Prof. Fabrice Labeau, mon co-superviseur et professeur à McGill, pour sa collaboration, son soutien et ses conseils pertinents durant nos réunions.

J'adresse aussi ma profonde gratitude à tous mes amis, à ma famille et particulièrement à ma chère mère, mon cher père, mes deux frères Ahmed et Anes et à Hichem Garbouj qui ont été toujours présents pour moi contribuant ainsi à la réussite de ce mémoire.

Finalement, je remercie mes collègues à l'INRS, Mohamed Jabi, Rabî Meftehi et Achref Methenni, que j'ai côtoyés au quotidien pendant ces deux dernières années, pour leur aide, leurs remarques, et leurs conseils qui ont été pour moi d'un grand apport.

Imene Ben Salem

Leszek Szczecinski

Table des matières

Introduction	1
Positionnement du problème	2
Contributions	3
1 Un aperçu des schémas HARQ	5
1.1 Introduction	5
1.2 Du Automatic Repeat reQuest (ARQ) au Hybrid ARQ	5
1.2.1 ARQ	5
1.2.2 Hybrid ARQ	6
1.3 Métriques des performances des schémas Hybrid ARQ	8
1.3.1 Débit utile	8
1.3.2 Probabilité de coupure	9
1.3.3 Délai et capacité effective	10
1.4 État de l'art des expressions de la performance de HARQ	11
1.5 Hybrid ARQ avec allocation et adaptation des paramètres de l'émetteur	12
1.5.1 Allocation et adaptation	13
1.5.2 État de l'art des expressions de performance des systèmes de commu- nications avec allocation et adaptation des paramètres de l'émetteur	13
1.6 Conclusion	17
2 Adaptation des taux de transmission pour les schémas HARQ-IR avec CSIT	18
2.1 Introduction	18
2.2 Modélisation du système	19
2.3 Optimisation du débit utile	22
2.4 Adaptation des taux de transmission	25

2.5	Calcul du débit utile	30
2.6	Comparaison entre débit utile avec adaptation de taux, allocation de taux ou taux fixe	32
2.7	Conclusion	33
3	Adaptation des puissances de transmission pour les schémas HARQ-IR avec CSIT	34
3.1	Introduction	34
3.2	Adaptation des puissances de transmission pour la minimisation de la pro- babilité de coupure	35
3.2.1	Modélisation du système	35
3.2.2	Optimisation de la probabilité de coupure	36
3.2.3	Comparaison entre probabilité de coupure avec adaptation de puis- sance, allocation de puissance ou puissance fixe	40
3.3	Adaptation de la puissance des transmissions pour la maximisation de la capacité effective	42
3.3.1	Modélisation du système	42
3.3.2	Expression de la capacité effective pour les schémas HARQ	42
3.3.3	Maximisation de la capacité effective	45
3.3.4	Comparaison entre capacité effective avec adaptation de puissance, allocation de puissance ou puissance fixe	46
3.4	Conclusion	47
	Conclusion	47
	Annexe A	49
	Bibliographie	53

Table des figures

1	Modèle des systèmes de communications implémentant le schéma (H)ARQ	2
1.1	La capacité pour AWGN, la capacité pour les canaux avec évanouissement quand le CSI est disponible au niveau du récepteur (CSIR) et la capacité pour les canaux avec évanouissement quand le CSI est disponible au niveau du récepteur et de l'émetteur (full CSI) [1].	15
1.2	Les valeurs optimales de la puissance, quand le paramètre de l'évanouissement est 2 et le SNR moyen = 0 dB [2].	16
2.1	Trois trames regroupant des sous-mots de code de longueurs identiques ou différentes.	20
2.2	$R(\hat{\gamma}_1)$ pour $K = 3$; $\bar{\gamma} = 5dB$; $\delta = 0.5$; $\lambda = -1.6$	27
2.3	De gauche à droite, on représente $\rho_2(I_1, \hat{\gamma}_2, \hat{\gamma}_1 = 9 dB)$ et $\rho_3(I_2, \hat{\gamma}_3, \hat{\gamma}_1 = 9 dB)$, pour $\bar{\gamma} = 5 dB$; $K = 3$; $\bar{\gamma} = 9 dB$; $\delta = 0.5$; $R(\hat{\gamma}_1 = 9 dB) = 2.8 b/symb$; $\lambda = -1.6$	27
2.4	La superposition de ρ_2 et ρ_3 pour $\bar{\gamma} = 5 dB$; $K = 3$; $\bar{\gamma} = 9 dB$; $\delta = 0.5$; $R(\hat{\gamma}_1 = 9 dB) = 2.8 b/symb$; $\lambda = -1.6$	29
2.5	La densité de probabilité du décodage erroné après une, deux et trois transmissions conditionnées sur $\hat{\gamma}_1 = 9 dB$; Pour $\bar{\gamma} = 5 dB$; $K = 3$; $\bar{\gamma} = 9 dB$; $\delta = 0.5$; $R(\hat{\gamma}_1 = 9 dB) = 2.8 b/symb$; $\lambda = -1.6$	31
2.6	Comparaison entre débit utile avec adaptation de taux, allocation de taux ou taux fixe pour IR-HARQ avec $K = 2$	32
3.1	De gauche à droite, on représente $P_2(I_1, \hat{\gamma}_2)$ et $P_1(I_0 = 0, \hat{\gamma}_1)$ optimales, pour $\bar{\gamma} = 15 dB$; $K = 2$; $\delta = 0.5$; $R = 1 b/s$; $\lambda = -1.25 \cdot 10^{-4}$	38
3.2	$P_2(I_1, \hat{\gamma}_2)$ optimale, pour $\bar{\gamma} = -5 dB$; $K = 2$; $\delta = 0$; $R = 1 b/s$; $\lambda = -0.23$	39

3.3	La probabilité de coupure pour l'adaptation de puissance (f_K^{PA}), allocation de puissance (f_K^{PV}) et puissance fixe (f_K^{PF}). Pour $K = 2, K = 3, R = 1$ et $\delta = 0.5$	41
3.4	Diagramme à états finis pour IR-HARQ tronqué.	42
3.5	La capacité effective pour l'adaptation de puissance (E_c^{PA}), allocation de puissance (E_c^{PV}) et puissance fixe (E_c^{PF}). Pour $\bar{\gamma} = 5 \text{ dB}$, $K = 3$ et $R = 1$	46
3.6	P_1, P_2 et P_3 optimales pour les schémas IR-HARQ avec allocation de puissance, $\bar{\gamma} = 5 \text{ dB}$, $K = 3$ et $R = 1$	51

Liste des acronymes

- **3GPP** 3rd Generation Partnership Project
- **ACK** Acknowledgement
- **AMC** Adaptative Modulation and Coding
- **ARQ** Automatic Repeat reQuest
- **AWGN** Additive White Gaussian Noise
- **CC** Chase Combining
- **CRC** Cyclique Redundancy Check
- **CSI** Channel State Information
- **CSIR** Channel State Information at the Receiver
- **CSIT** Channel State Information at the Transmitter
- **dB** décibel
- **DP** Dynamic Programming
- **EDGE** Enhanced Data Rates for GSM Evolution
- **FEC** Forward Error Correction
- **FSMC** Finite-State Markov Channel
- **GPRS** General Packet Radio Service
- **HARQ** Hybrid Automatic Repeat reQuest
- **IEEE** Institute of Electrical and Electronics Engineers
- **IR** Incremental Redundancy
- **LTE** Long Term Evolution
- **MAC** Medium Access Control
- **NACK** Negative Acknowledgement
- **OSI** Open Systems Interconnection
- **QoS** Quality of Service
- **RCPC** Rate-Compatible Punctured Convolutional codes
- **SNR** Signal-to-Noise Ratio

- **TCP** Transmission Control Protocol
- **TDMA** Time Division Multiple Access
- **WiMAX** Worldwide Interoperability for Microwave Access

Introduction générale

Depuis le début des années quatre-vingt-dix, nous assistons à un véritable engouement pour le développement des réseaux mobiles. Cet engouement est dû, d'une part, à un besoin de mobilité et, d'autre part, à la panoplie de nouveaux services offerts, comme l'accès Internet à haute vitesse, le téléchargement de fichiers audio et vidéo, la réception de programmes télévisés et la visiophonie. De tels services ne peuvent être disponibles sans une augmentation significative du débit du lien radio.

D'autre part, les utilisateurs exigent de plus en plus la même qualité de service sur leurs systèmes cellulaires que celle offerte dans un environnement câblé. Ainsi, le développement technologique des réseaux mobiles a connu un grand essor. Pour satisfaire de telles exigences en termes de débit et de QoS, de nouveaux protocoles et technologies ont été développés tels que l'adaptation du codage et de la modulation (AMC : Adaptive Modulation and Coding), les retransmissions des paquets erronés (HARQ : Hybrid Automatic Repeat Request), etc.

Ce mémoire s'intéresse au Hybrid Automatic Repeat Request qui assure et améliore la fiabilité des transmissions. Ce schéma a été adopté dès le début de la deuxième génération dans GPRS (General Packet Radio Service) et EDGE (Enhanced Data Rates for GSM Evolution). Il fait également partie des normes cellulaires de la troisième génération, et il est toujours utilisé dans les plus récentes normes de la quatrième génération, comme IEEE 802.16m (WiMAX) et LTE. Le succès récent de la norme 3GPP LTE [3] a présenté la technique HARQ comme une solution prometteuse pour améliorer les systèmes mobiles à débit de transmission élevé.

HARQ combine deux grandes classes de mécanismes gérant la fiabilité des transmissions sur les canaux bruités. Il s'agit des mécanismes réactifs ARQ (Automatic Repeat Request) et des mécanismes proactifs FEC (Forward Error Correction). Ce protocole complexe et sophistiqué a connu plusieurs améliorations passant de ARQ à HARQ type III.

Notre travail s'inscrit dans le cadre d'une optimisation des performances des systèmes

de communications implémentant ce mécanisme de retransmission.

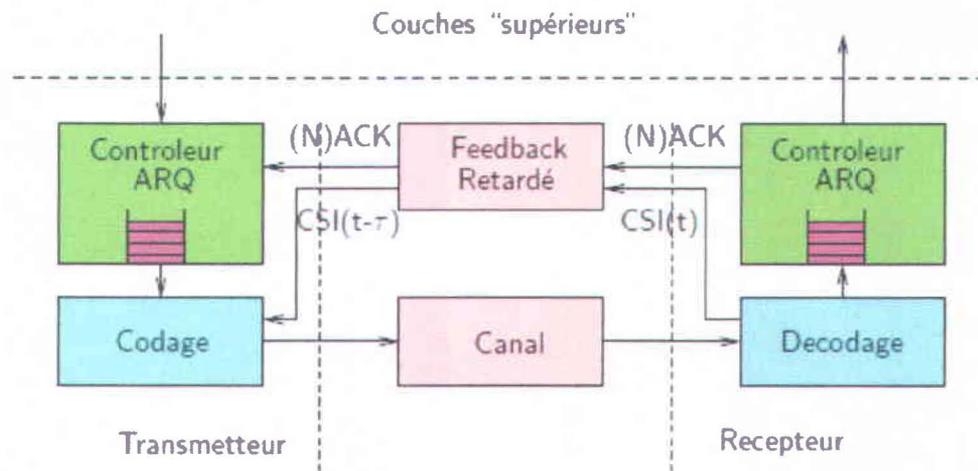


FIGURE 1 – Modèle des systèmes de communications implémentant le schéma (H)ARQ

Les schémas HARQ sont généralement implémentés dans la couche deux du modèle OSI. Comme le montre la figure (1), le canal d'acquiescement est utilisé pour transmettre une information sur l'état du décodage du paquet (ACK/NACK) au niveau du récepteur. D'autres informations nommées CSI (Channel State Information) peuvent être transmises avec ces messages d'acquiescement. Notre objectif est d'explorer comment utiliser des CSI retardées pour améliorer les performances des systèmes de communications.

Positionnement du problème

Généralement, les schémas HARQ sont étudiés sans CSI et avec des puissances et des taux de transmissions constants. Cependant, certains travaux se sont basés sur la variation de ces paramètres dépendamment du numéro de la transmission. Cette variation, qui ne dépend d'aucun acquiescements, est appelée allocation, par opposition à l'adaptation. L'allocation a fait ses preuves quant à l'amélioration des performances des systèmes de communications. Par exemple, [4] a présenté une étude des performances de Incremental Redundancy HARQ (définie dans la section 1.2.2) en terme de probabilité de coupure sur la base d'une allocation de la puissance des transmissions. Cependant, l'étude a été faite pour un HARQ tronqué avec un nombre maximal de retransmission égal à deux. L'allocation des

taux a été analysée dans [5] pour un nombre infini de retransmissions et dans [6] pour un schéma HARQ tronqué.

Concernant l'adaptation, elle a été étudiée principalement pour HARQ avec un schéma de codage et de modulation prédéfini comme c'est le cas pour [7], [8] et [9]. Pour une analyse plus générale (c.à.d. du point de vue de la théorie de l'information), on retrouve [10] qui a étudié l'adaptation de puissance avec un acquittement qui n'est pas retardé, mais plutôt discrétisé. [4] traite aussi l'adaptation de puissance dans un cas similaire à celui de HARQ. On cite également [11] qui a présenté une adaptation de puissance et de taux avec une contrainte sur le délai et le débordement de la file d'attente. Cependant, cette étude a été faite pour les schémas HARQ de type I.

Motivée par ces travaux, on s'intéresse également à l'adaptation des taux et de la puissance des retransmissions avec les schémas HARQ tronqués de type Incremental Redundancy. Notre analyse sera faite de point de vue de la théorie de l'information et donc sans contrainte sur le type de codage ni de la modulation utilisés. De plus, on exploitera différentes métriques pour l'analyse des performances de notre système d'étude.

Contributions

Les principales contributions de ce mémoire sont :

- Présenter une méthode d'optimisation de K taux des transmissions variables dans le but de maximiser le débit utile des systèmes de communications avec évanouissement de Rayleigh, CSIT retardé et HARQ de type Incremental Redundancy. Cette maximisation se base sur les méthodes d'optimisation avec programmation dynamique (DP) [12]. Une telle méthode n'a pas encore été utilisée pour l'adaptation des taux avec HARQ tronqué. DP a été uniquement utilisée dans l'adaptation des puissances dans un contexte similaire à celui de HARQ [13].
- Utiliser la méthode d'optimisation avec programmation dynamique pour l'adaptation de K puissances variables. Cette analyse est faite pour les systèmes de communications avec évanouissement de Rayleigh, CSIT retardé et Incremental Redundancy HARQ tronqué. L'optimisation des puissances est réalisée avec une contrainte sur la puissance moyenne de transmission. L'objectif est de minimiser la probabilité de coupure du système.
- Étudier la capacité effective pour les schémas IR-HARQ et présenter une optimisation de cette métrique basée sur l'adaptation des puissances de K transmissions quand CSI retardé est disponible au niveau de l'émetteur. La capacité effective traduit une

mesure du débit utile avec une contrainte sur le délai. Ainsi, elle présente un meilleur critère de performance comparé à celui du débit utile ou de la probabilité de coupure quand il s'agit d'étudier des services en temps réel. La capacité effective a été peu étudiée avec le protocole HARQ.

Le mémoire est organisé comme suit : Le premier chapitre est consacré à la définition des schémas HARQ et des différentes métriques et notions utilisées à travers ce mémoire. On présente, dans le deuxième chapitre, une maximisation du débit utile par une adaptation des taux de transmission pour les systèmes implémentant le IR-HARQ tronqué. Le troisième chapitre est réservé à l'adaptation des puissances de transmission pour les schémas IR-HARQ tronqués afin de minimiser la probabilité de coupure dans un premier temps, et de maximiser la capacité effective dans un deuxième temps. De plus, on dérive l'expression exacte de la capacité effective pour les schémas HARQ tronqués. On termine par une conclusion générale dans le dernier chapitre.

Chapitre 1

Un aperçu des schémas HARQ

1.1 Introduction

Même si un aperçu complet de toutes les contributions faites pour HARQ est hors de portée, les notions introduites dans ce chapitre sont nécessaires pour comprendre le travail qui sera présenté tout au long de ce mémoire.

Ce chapitre est organisé comme suit : dans la section 1.2, on définit le schéma ARQ de base ainsi que ses différentes évolutions. La section 1.3 présente les différentes métriques utilisées pour mesurer la performance des schémas HARQ. La section 1.4 récapitule les principaux résultats obtenus dans la littérature en termes de performance HARQ. Enfin, l'application de HARQ dans les systèmes avec allocation ou adaptation des paramètres de l'émetteur ainsi que les travaux réalisés dans ce domaine sont introduits dans la section 1.5.

1.2 Du Automatic Repeat reQuest (ARQ) au Hybrid ARQ

1.2.1 ARQ

Le système initial ARQ a été inventé par Van Duuren [14] en 1943 afin d'assurer une transmission fiable des caractères. Le premier modèle commercial de ARQ, construit au Pays-Bas, a été mis en service en 1947 [15]. Ce système est basé sur un mécanisme qui informe l'émetteur sur l'état du décodage du paquet transmis. Un accusé de réception (ACK) ou un accusé de réception négatif (NACK) est envoyé à l'émetteur en conséquence.

Quand ARQ est implémenté, un code de détection d'erreur (généralement Cyclique Redundancy Check (CRC)) est utilisé. L'émetteur encode les données avec CRC, puis transmet

le paquet résultant sur le canal bruité. Le récepteur vérifie si le paquet reçu a été corrompu ou pas par le bruit du canal, et envoie à l'émetteur ACK ou NACK en conséquence. Si le paquet contient des erreurs, le récepteur rejette le paquet reçu. Dans ce cas, l'émetteur reçoit un NACK et retransmet le même paquet. Dans le cas contraire, l'émetteur reçoit un ACK. Il libère alors le paquet de sa file d'attente et le protocole recommence à nouveau avec le paquet de données suivant.

ARQ permet de transmettre sans erreurs, mais au détriment d'un retard supplémentaire. Un compromis entre la probabilité d'erreur du paquet et un retard de la transmission peut être obtenu avec la version dite ARQ tronquée [16], qui permet à chaque paquet d'être transmis au plus K fois. Le paramètre K est appelé crédit de transmission (également connu sous le nom de persistance).

1.2.2 Hybrid ARQ

Avec le schéma ARQ, les paquets transmis sont constitués de bits d'information et de bits de détection d'erreurs. Il n'existe pas de protection contre le bruit introduit par le canal lors de la transmission. Comme la fiabilité offerte par ARQ repose sur la répétition des données d'information, il est très efficace quand les conditions du canal sont bonnes, c'est-à-dire quand un faible nombre de retransmissions est requis.

D'une autre part, l'utilisation du codage FEC permet de corriger et ainsi récupérer le paquet même si les conditions du canal de transmission sont mauvaises, et ceci, en diminuant le taux de codage. En contrepartie, l'overhead dû aux taux de codage FEC pénalise l'efficacité de la transmission quand les conditions du canal sont bonnes.

Plusieurs solutions ont été proposées pour remédier au problème du ARQ en combinant des techniques de FEC avec le mécanisme de la répétition de paquets, ce qui a donné naissance à Hybrid ARQ. Une intéressante étude des articles clés analysant HARQ est présentée dans [17].

HARQ est classé en trois catégories selon la capacité de correction d'erreurs dans chaque transmission. Ces catégories sont : HARQ type I, type II et type III.

Hybrid ARQ type I

HARQ type I décrit un mécanisme ARQ combiné à un codeur FEC. En effet, un Cyclic Redundancy Check (CRC) est d'abord généré pour chaque message à transmettre puis ajouté aux données. Le paquet résultant est ensuite codé par un FEC de taux de codage R_0 .

L'intérêt principal de HARQ type I est d'utiliser la capacité de correction du code FEC afin de récupérer les bits d'information qui peuvent être bruités au cours de leurs transmissions, et diminuer ainsi la probabilité de retransmission du mécanisme ARQ.

Du côté du récepteur, si le paquet reçu contient toujours des erreurs, il est éliminé et un NACK est envoyé à l'émetteur. Par conséquent, HARQ type I définit les régimes HARQ avec une capacité de correction d'erreur constante à travers les retransmissions.

Toutefois, lorsque le canal est bon, le taux de codage R_0 diminue la quantité de bit d'informations reçus au niveau de chaque paquet accepté.

Hybrid ARQ type II

Contrairement au HARQ type I, HARQ type II combine les multiples sous-mots de code reçus à travers les retransmissions. Les deux principales méthodes de combinaisons sont : Chase Combining [18] et Incremental Redundancy [19], [20].

Chase combinaison (CC) : Pour HARQ type II avec CC ou CC-HARQ, les mêmes opérations que celles pour HARQ type I sont effectuées au niveau de l'émetteur, à savoir l'ajout du CRC et le codage des données avec FEC. Toutefois, si un accusé de réception négatif est transmis, le paquet est conservé au niveau de la file d'attente du récepteur. Dans ce cas, l'émetteur retransmet le même mot de code, qui sera combiné avec les retransmissions précédentes du même paquet. CC n'augmente pas le taux du codage à travers la combinaison des retransmissions, mais engendre une augmentation du SNR accumulé.

Incremental Redundancy (IR) : Introduit par Davida et Reddy [21], le HARQ de type II avec IR ou IR-HARQ est basé sur le principe suivant : les bits de correction d'erreurs sont envoyés au récepteur uniquement au besoin. En effet, la première transmission est constituée par les bits d'information et les bits de détection d'erreurs. Si le récepteur détecte des erreurs, la transmission est conservée au niveau de sa file d'attente et un NACK est envoyé à l'émetteur. Les retransmissions sont généralement différentes et contiennent les bits de parité permettant la correction des erreurs. Ainsi, après chaque retransmission, la redondance additionnelle combinée avec les retransmissions précédentes, engendre un mot de code avec une capacité de correction plus grande. Comme la redondance est envoyée par morceaux lors de la détection d'erreur, IR-HARQ est le système le plus polyvalent, et donne le meilleur compromis entre ARQ et FEC.

Hybrid ARQ type III

HARQ type III pourrait être considéré comme un cas particulier du HARQ type II où chaque transmission est décodable. En effet, chaque sous-mot de code contient suffisamment de bits d'information et de parité pour pouvoir le décoder. Chaque sous-mot de code pourrait ainsi être décodé seul ou combiné avec les autres transmissions du même paquet reçues précédemment.

1.3 Métriques des performances des schémas Hybrid ARQ

Divers indicateurs de performance ont été introduits dans [22] et [23] afin de donner une vue complète sur les performances d'un système avec les schémas HARQ. On cite :

- Le débit utile.
- La probabilité de coupure.
- Le délai et la capacité effective.

Dans cette section, nous allons définir les critères de performance que nous utiliserons tout au long de ce mémoire.

1.3.1 Débit utile

Le débit utile, noté par η , a été défini pour les systèmes segmentés ("slotted") dans [24]. Considérons N_b le nombre de bit d'information à transmettre et T la durée d'un segment ("slot"). Soient t le numéro du segment, $N_b(t)$ le nombre de bits décodé correctement jusqu'au segment t et $R(t)$ le taux correspondant en b/s . Le débit utile est alors donné par :

$$\eta = \lim_{t \rightarrow \infty} \frac{N_b(t)}{tT} \quad (1.1)$$

En se basant sur le "renewal-reward" théorème [24] et en considérant les hypothèses suivantes :

- L'évènement «l'arrêt de la transmission du paquet en cours» est un évènement récurrent.
- Une variable aléatoire \mathbb{R} représente la récompense obtenue suite à l'occurrence de l'évènement récurrent. En particulier, $\mathbb{R} = \frac{N_b}{T} = R$ b/s si la transmission s'arrête car un décodage correct a eu lieu au niveau du récepteur ; ou $\mathbb{R} = 0$ b/s si le paquet transmis n'a pas pu être décodé correctement après le nombre maximal de retransmissions disponibles.

- \mathbb{T} est une variable aléatoire représentant le temps entre deux occurrences consécutives de l'évènement récurrent.

Alors, le débit utile est donné par :

$$\eta = \frac{E[\mathbb{R}]}{E[\mathbb{T}]} \quad (1.2)$$

où $E[\mathbb{R}]$ est la récompense moyenne et $E[\mathbb{T}]$ est la durée moyenne d'inter-renouvellement.

Ainsi, le débit utile, mesuré en b/s ou encore en $b/symb$ ¹, est considéré comme le nombre moyen de bits reçus correctement divisé par le temps moyen nécessaire pour transmettre le paquet.

1.3.2 Probabilité de coupure

La probabilité de coupure est la probabilité qu'un paquet d'information ne soit pas transmis correctement par le processus HARQ. Cette probabilité dépend du schéma HARQ choisi.

Pour ARQ et HARQ type I tronqué avec un maximum de K transmissions, le décodage du paquet se base uniquement sur la dernière transmission reçue ainsi la probabilité de coupure, notée par f_K^I , est donnée par :

$$f_K^I = Pr\{I_K < R\} = Pr\{\log_2(1 + \gamma_K) < R\} \quad (1.3)$$

où $\gamma_k = |h_k|^2$ est le gain du canal à la $k^{\text{ième}}$ transmission.

Par contre, pour le schéma HARQ type II tronqué, la probabilité d'un décodage erroné dépend de la méthode de combinaison des paquets. On distingue alors :

$$f_K^{CC} = Pr\{I_K < R\} = Pr\left\{\log_2\left(1 + \sum_{i=1}^K \gamma_i\right) < R\right\} \quad (1.4)$$

$$f_K^{IR} = Pr\left\{\sum_{i=1}^K I_K < R\right\} = Pr\left\{\sum_{i=1}^K \log_2(1 + \gamma_i) < R\right\} \quad (1.5)$$

les probabilités de coupure pour HARQ-CC et HARQ-IR respectivement.

1. Le goodput est une autre métrique des performances des systèmes de communications. Il est défini comme étant le nombre de bits utiles reçus par symbole transmis [25]. Ainsi, le goodput est mesuré en $b/symb$. Cependant, par abus de langage, le goodput est souvent déclaré comme débit utile comme c'est le cas dans [24].

1.3.3 Délai et capacité effective

Le délai, noté D , est défini comme la durée moyenne nécessaire pour que le paquet d'information soit reçu sans erreurs.

L'expression du délai est donnée par :

$$D = T + \sum_{k=2}^K kT(f_{k-1} - f_k) \quad (1.6)$$

où f_k (la probabilité d'avoir un décodage incorrect après exactement k transmissions) est donnée par (1.3), (1.4) et (1.5) pour ARQ ou HARQ type I, CC et IR, respectivement (où $K = k$).

Il est à noter que la notion de délai qu'un utilisateur peut tolérer varie d'un service à un autre. Pour les services non temps réel comme la diffusion de données, le principal intérêt est d'augmenter le débit utile avec une contrainte laxiste sur le délai. Par contre pour les services en temps réel comme les vidéoconférences, vérifier une contrainte stricte sur le délai est plus important que d'assurer un débit utile très élevé. C'est pourquoi une métrique qui permet de pondérer le critère de délai et de le combiner à celui du débit utile serait plus pertinente que celle du délai. C'est dans ce contexte que la notion de capacité effective a été introduite dans [26].

La capacité effective est définie comme le taux d'arrivée constant maximal qu'un processus de service donné peut supporter afin de garantir une certaine qualité de service caractérisée par un paramètre θ . Ce paramètre est appelé exposant du QoS.

La capacité effective est donnée par :

$$E_c(\theta) = -\frac{1}{\theta} \log(\rho\{M\Phi(\theta)\}) \quad (1.7)$$

où :

- M est la matrice de transition du FSCM (Finite-State Markov Channel) qui modélise le service offert.
- $\rho\{\cdot\}$ désigne le rayon spectral d'une matrice.
- $\Phi(\theta) = \text{diag}(e^{v_1\theta}, e^{v_2\theta}, \dots, e^{v_K\theta})$ où v_k est le taux du service offert à la $k^{\text{ième}}$ transmission.

Pour comprendre la notion du délai introduite par le paramètre θ , on se réfère à [27] où Chang a montré que, pour un système de file d'attente dynamique avec des processus d'arrivée et de service stationnaires et ergodiques, la longueur de la file d'attente $Q(t)$ converge en distribution vers une variable aléatoire $Q(\infty)$ vérifiant :

$$\lim_{x \rightarrow \infty} \frac{\log(\Pr(Q(\infty) > x))}{x} = \theta \quad (1.8)$$

Le théorème ci-dessus affirme que la probabilité que la file d'attente dépasse un certain seuil x , décroît rapidement de manière exponentielle avec l'augmentation du seuil x . Ainsi, le paramètre θ indique le taux de décroissance exponentielle des probabilités de violation de QoS.

Une valeur faible de θ correspond à un taux faible de libération de la file d'attente, dans ce cas le système ne peut garantir qu'une condition souple sur le délai, tandis qu'une grande valeur de θ implique un taux important de libération de la file d'attente, ce qui signifie que des conditions plus strictes sur le délai peuvent être prises en charge.

1.4 État de l'art des expressions de la performance de HARQ

Les schémas HARQ ont été largement étudiés dans la littérature avec généralement deux objectifs : i) la dérivation des formules du débit utile et des différents critères de performance des régimes HARQ existants, ii) la conception de nouveaux schémas HARQ plus performants.

Cette section présente une brève revue non exhaustive des principales contributions faites dans le domaine i). Pour un bon aperçu des premiers principes HARQ, vous pouvez consulter [17] et [28].

Hagenauer est le pionnier dans l'étude des performances de HARQ-IR qu'il a présenté comme une application de son code RCPC [19], alors que [29] a probablement été le premier à analyser la performance de CC-HARQ, même si ce système a déjà été mentionné par Chase [18] comme une application de son code de combinaison. Cheng, [30], a tenté de répondre à la question intéressante suivante : quel est le meilleur régime entre CC-HARQ et IR-HARQ ? La réponse dépend de la qualité du canal ainsi que du taux de codage initial R_0 .

[24] est également une référence intéressante qui présente les expressions du débit utile et du retard pour ARQ, IR-HARQ et CC-HARQ. Dans le système de transmission traité dans [24], on considère des paquets de longueurs N_b bits qui sont codés en mots de code et divisés en K sous-mots de code constituant les K transmissions du schéma HARQ tronqué. On suppose que le temps est divisé en segments de durée T et que l'utilisateur occupe cette unité de temps pour transmettre un sous-mot de code. On suppose que le canal est

quasi-constant durant un segment et que le signal reçu est donné par :

$$Y_k = h_k \cdot X_k + z_k \quad (1.9)$$

où :

- $h_k > 0$ est l'évanouissement du canal,
- X_k est le signal émis d'énergie moyenne 1,
- z_k représente le bruit gaussien.

En supposant que le canal d'acquittement est sans erreur ni délai, la formule du débit utile pour HARQ est donné par :

$$\eta = \frac{R(1 - f_K)}{1 + \sum_{k=1}^{K-1} f_k} \quad (1.10)$$

où :

- R représente le taux de transmission qui est donné par $R = \frac{N_k}{T}$,
- f_k représente la probabilité du décodage incorrecte après exactement k transmissions et varie selon le type du schéma HARQ adopté. Les expressions des f_k pour les différents schémas HARQ ont été définies dans la section 1.3.2.

Notons que pour les systèmes de communications sans schéma HARQ, une seule transmission est valable par paquet ($K = 1$). Ainsi dans ce cas le débit utile est donné par

$$\eta = R(1 - f_1) \quad (1.11)$$

1.5 Hybrid ARQ avec allocation et adaptation des paramètres de l'émetteur

En plus de la détermination des formules exactes des critères de performance des schémas HARQ, plusieurs travaux ont apporté des améliorations aux schémas HARQ de bases déjà existants. Parmi ces travaux, on peut citer [31] et [7] qui se sont basé sur une variation de la puissance ou du taux des retransmissions pour optimiser la performance du système. Il existe deux types de variations qu'on appelle allocation et adaptation. Dans cette section, on va commencer par définir ces deux termes, allocation et adaptation, et résumer les principaux résultats obtenus suite à l'ajustement de la puissance ou du taux de transmission du schéma HARQ.

1.5.1 Allocation et adaptation

Après la transmission d'un sous-mot de code, le récepteur utilise le canal d'acquiescement pour envoyer un seul bit indiquant un ACK ou un NACK. Le récepteur peut également transmettre d'autres informations sur ce canal d'acquiescement pour permettre à l'émetteur de s'adapter aux conditions du système. Ces informations sont appelées informations d'état du canal ou CSI et se réfèrent aux propriétés des canaux d'une liaison de communication. Le CSI décrit comment un signal se propage à partir de l'émetteur vers le récepteur et contient, par exemple, l'évanouissement, et l'affaiblissement de la puissance avec la distance. Le CSI permet à l'émetteur d'adapter ses transmissions aux conditions du canal. Cette adaptation peut être cruciale pour parvenir à une communication fiable avec des débits de données élevés comme dans le cas des systèmes à multiantennes.

Contrairement à l'adaptation, l'allocation ne fait pas intervenir le canal d'acquiescement qui servira uniquement à transmettre un seul bit indiquant un ACK ou un NACK. L'émetteur alloue différentes puissances ou taux aux retransmissions indépendamment de tout acquiescements.

1.5.2 État de l'art des expressions de performance des systèmes de communications avec allocation et adaptation des paramètres de l'émetteur

Les deux cas limites qui sont intéressants à traiter tout d'abord sont : le cas où aucune information CSI n'est disponible à l'émetteur et le cas où l'émetteur connaît parfaitement l'état du canal.

Considérons un système de communications avec évanouissement rapide. Le CSI est uniquement disponible au niveau du récepteur. L'émetteur reçoit un unique bit lui indiquant de retransmettre de nouveau ou de passer au paquet suivant. On suppose que les transmissions sont faites à puissance constante. Dans ce cas, la capacité est définie par [1] :

$$C_R = \mathbb{E} \left\{ \log_2 \left(1 + \gamma \frac{P}{N_0} \right) \right\} \quad \text{b/s/Hz} \quad (1.12)$$

où :

- l'opérateur $\mathbb{E}\{\cdot\}$ est l'espérance,
- P est la puissance du signal émis,
- N_0 est la puissance du bruit gaussien,
- γ est la variable aléatoire représentant le gain du canal.

Pour les schémas HARQ avec un nombre illimité de retransmissions, le système atteint la capacité ergodique définie par (1.12) mais le délai D est plus ou moins important selon l'état du canal et tend en moyenne vers ∞ . Par contre si un nombre maximal de K transmissions est imposé, η défini dans (1.10) est la valeur du débit utile que le système avec HARQ tronqué atteint.

D'autre part, si l'émetteur reçoit une information exacte sur l'état du canal transmis sur le canal d'acquiescement, peut-on dépasser cette capacité ergodique définie dans (1.12) ?

Le chapitre 5.4 du livre [1] explique que l'émetteur utilise le CSIT pour adapter sa puissance afin de s'ajuster à l'état de canal et permettre ainsi au récepteur de décoder correctement le paquet. Pour calculer la capacité obtenue dans ce cas, on doit résoudre le problème suivant :

$$C_T = \max_{A(\gamma)/E(A(\gamma)=1)} \int_0^\infty \log_2 \left(1 + A(\gamma) \gamma \frac{P}{N_0} \right) b/s/Hz \quad (1.13)$$

sous contrainte d'un SNR moyen $\bar{\gamma}$ égal à P/N (d'où la condition $E(A(\gamma)) = 1$).

$A(\cdot)$ est une fonction d'adaptation de puissance qui associe à chaque valeur du gain du canal γ la valeur de la puissance qui permet d'atteindre la capacité du système. La solution optimale est donnée par :

$$\frac{P}{N_0} A(\gamma) = \left[\frac{1}{\gamma_0} - \frac{1}{\gamma} \right]^+ \quad (1.14)$$

avec $x^+ = \max(0, x)$ et γ_0 une constante vérifiant $E(A(\gamma)) = 1$.

Ainsi la capacité est égale à :

$$C_T = \log_2 \left(1 + \frac{\gamma}{\gamma_0} \right) \text{ bits/s/Hz} \quad (1.15)$$

Cette stratégie d'adaptation de puissance est appelée remplissage ("waterfilling") et consiste à allouer plus de puissance quand γ est grande, profitant ainsi des conditions favorables du canal, ou bien allouer une puissance faible voir nulle quand le canal est mauvais.

Comme le montre la figure suivante (1.1), pour les faibles SNR, C_T est supérieure à C_R mais cette différence diminue avec l'augmentation du SNR jusqu'à tendre vers zéro.

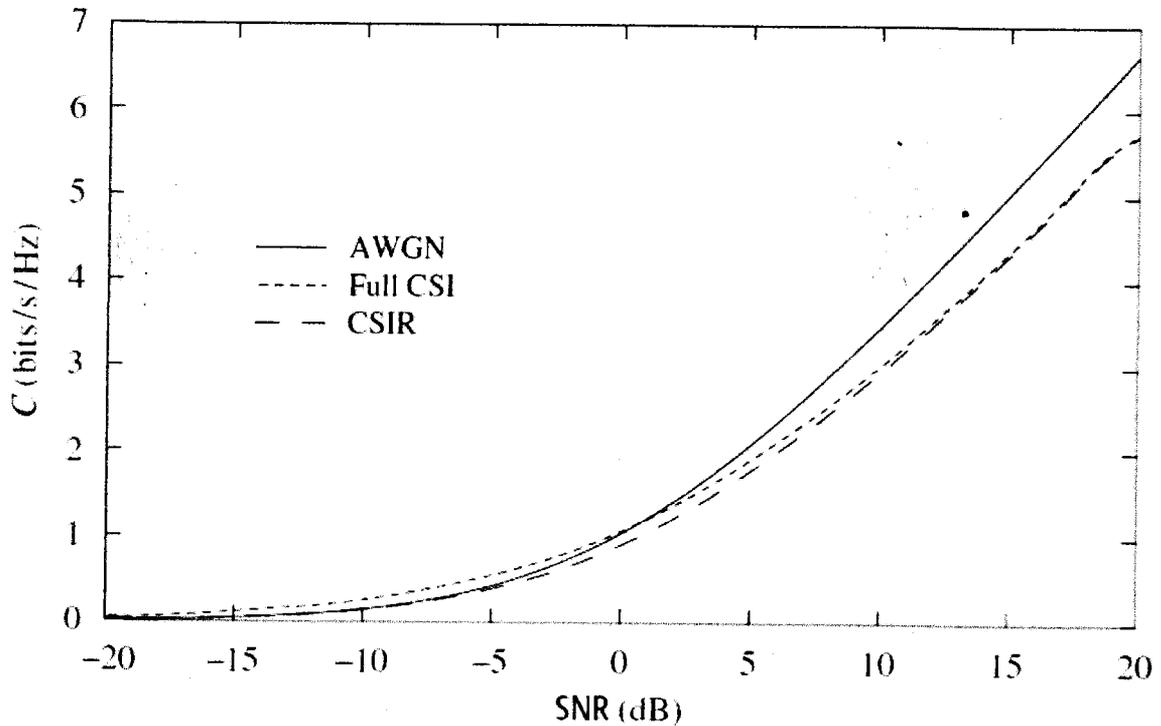


FIGURE 1.1 – La capacité pour AWGN, la capacité pour les canaux avec évanouissement quand le CSI est disponible au niveau du récepteur (CSIR) et la capacité pour les canaux avec évanouissement quand le CSI est disponible au niveau du récepteur et de l'émetteur (full CSI) [1].

Malgré le fait que la stratégie du remplissage augmente le débit utile à long terme du système en transmettant lorsque le canal est bon, néanmoins, une augmentation du délai pour les faibles γ en découle. À cet égard, il est intéressant de comparer la stratégie de remplissage à celle d'inversion du canal.

La méthode d'inversion du canal consiste à transmettre avec beaucoup de puissance pour pouvoir contrecarrer les mauvaises conditions du canal et transmettre avec une faible puissance quand le canal est bon. Dans ce cas, la fonction d'adaptation de la puissance est donnée par :

$$A(\gamma) = \frac{1}{\gamma \mathbb{E}\{\frac{1}{\gamma}\}} \quad (1.16)$$

Avec cette stratégie le taux de transmission est maintenant le même pour toutes les valeurs de l'évanouissement ainsi le délai est indépendant des variations du canal. Cependant, il est à noter qu'une importante quantité de puissance est nécessaire pour inverser le canal quand il est mauvais. En bref, la stratégie d'inversion du canal est moins efficace en terme de puissance que celle du remplissage, mais plus efficace en terme de délai. Selon les applications (sensibles aux délais ou à la puissance), la meilleure stratégie est choisie pour avoir une meilleure performance du système.

Cependant, il existe des services qui sont sensibles aux délais, mais en même temps ne requièrent pas des contraintes très sévères sur la QoS. C'est dans ce contexte que s'inscrit l'article de Tang et Zhang [2] qui présente une analyse plus complète en optimisant la capacité effective pour les systèmes avec CSIT parfait. Comme on l'a énoncé précédemment, cette métrique tient compte du débit utile mais aussi du délai.

La figure suivante représente les valeurs optimales d'adaptation de puissance en fonction du SNR et de θ :

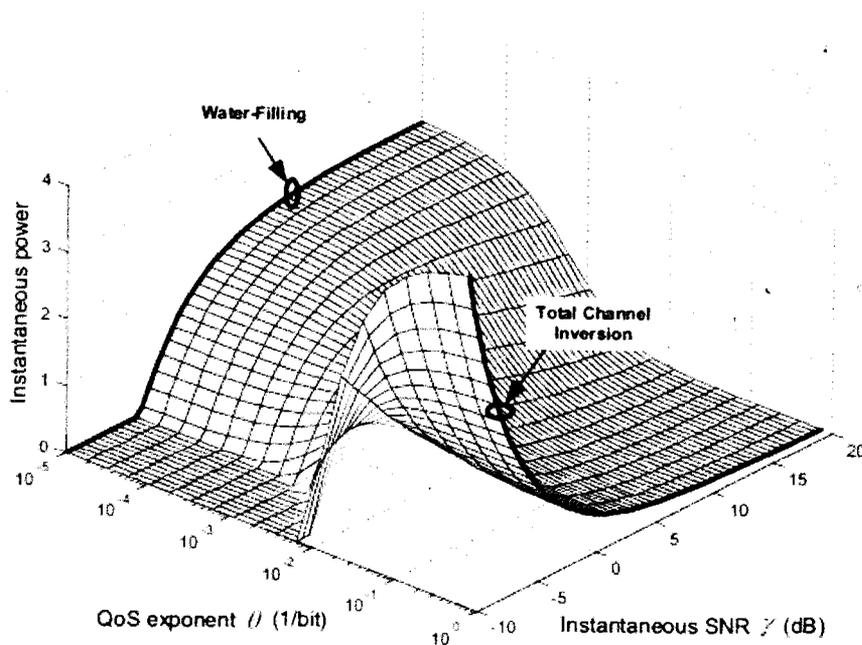


FIGURE 1.2 – Les valeurs optimales de la puissance, quand le paramètre de l'évanouissement est 2 et le SNR moyen = 0 dB [2].

Notons que ce schéma d'adaptation de puissance et de taux avec une contrainte sur le QoS se réduit au cas de remplissage quand le système peut tolérer un délai arbitrairement long ($\theta \rightarrow 0$). C'est le résultat attendu, car le remplissage est la stratégie optimale d'adaptation de puissance sans contrainte sur le délai. D'un autre côté, quand $\theta \rightarrow \infty$, le schéma d'adaptation de puissance et de taux converge vers l'algorithme d'inversion du canal qui caractérise les systèmes avec une contrainte stricte sur les délais.

Le schéma HARQ n'apporte aucun gain quand l'émetteur connaît parfaitement l'état du canal. Une seule transmission est suffisante pour avoir un décodage correct au niveau du récepteur. Cependant, dans les applications réelles, à cause des délais de transmissions et de traitements, l'émetteur ne peut pas avoir une information exacte sur l'état du canal, mais plutôt une information obsolète.

Une question se pose alors : Comment peut-on utiliser ce CSIT retardé pour pouvoir améliorer les performances d'un système avec HARQ tronqué en terme de débit utile, de probabilité de coupure et également en terme de capacité effective ?

1.6 Conclusion

Dans ce chapitre, nous avons présenté les principes fondamentaux des schémas HARQ qui seront utiles tout au long du mémoire. Sans être exhaustif, l'état de l'art exposé dans ce chapitre couvre la définitions des différents types HARQ ainsi que les métriques utiles pour l'évaluation de la performance des systèmes de transmission avec HARQ. Ensuite nous avons présenté les principaux résultats déjà obtenus pour les schémas HARQ avec CSIT où on a introduit la différence entre l'allocation et l'adaptation.

Le chapitre suivant sera consacré à l'optimisation du débit utile pour les systèmes de communications avec IR-HARQ en se basant sur l'adaptation des taux de transmission.

Chapitre 2

Adaptation des taux de transmission pour les schémas HARQ-IR avec CSIT

2.1 Introduction

Dans l'objectif d'améliorer les performances des systèmes de communications avec HARQ, beaucoup de travaux se sont penchés sur la variation du taux ou de la puissance des retransmissions. L'objectif est de s'approcher de plus en plus de la capacité ergodique tout en réalisant un nombre faible de retransmissions. L'exploitation du CSIT a été aussi étudiée dans les schémas HARQ avec AMC.

Dans ce contexte, on présente, dans ce chapitre, un problème d'optimisation qui permet d'utiliser le CSI retardé disponible au niveau de l'émetteur afin d'augmenter les performances d'un système de communications avec IR-HARQ tronqué en terme de débit utile.

Ce chapitre est organisé comme suit : dans la section 2.2, on définit le modèle du système étudié. On présente, dans la section 2.3, le problème d'optimisation du débit utile. La section 2.4 définit les taux optimaux maximisant le débit utile. Enfin, le calcul du débit utile optimal et sa comparaison avec celui à taux fixe et à allocation des taux sont présentés dans les sections 2.5 et 2.6 respectivement.

2.2 Modélisation du système

Pour le système de transmission étudié, on considère que les bits d'information sont divisés en paquets de longueurs N_b . Chaque paquet est codé en un mot de code contenant N_s symboles x_1, x_2, \dots, x_{N_s} indépendants et aléatoires suivant la loi gaussienne complexe et centrée. Aucune contrainte sur N_s n'est imposée, ainsi un code avec un taux aléatoire peut être construit. Comme on étudie le cas de HARQ tronqué avec un maximum de K transmissions, ces symboles sont regroupés en K sous-mots de code X_1, X_2, \dots, X_K . Pour Incremental Redundancy HARQ, les sous-mots de code sont disjoints et composés de différents symboles x_i . On suppose que chaque sous-mot de code est de longueur $N_{s,k}$ symboles, $k = 1 \dots K$. Les longueurs des sous-mots de code peuvent être différentes. On suppose que la première transmission est de longueur $N_{s,1}$ non nulle et fixe.

Comme les longueurs des transmissions sont variables, une explication sur le mode de transmission choisi s'impose. Le système de transmission avec des segments fixes de durée T n'est pas optimal dans notre cas. En effet, les longueurs variables des sous-mots de code entraînent soit une perte des ressources si la longueur de la transmission est inférieure à la longueur du segment, soit une collision si au contraire la longueur de la transmission dépasse celle du segment. D'autre part, on peut supposer que le temps alloué à l'utilisateur varie selon les longueurs des transmissions. Ce scénario est envisageable pour les systèmes de communications à un seul usager, où émetteur et récepteur négocient le temps de transmission de chaque sous-mot de code. Cependant, en passant aux systèmes de communications multi-usagers, cette stratégie n'est plus réalisable, car chaque transmission implique une communication entre tous les usagers du système.

Afin d'éviter tous ces problèmes de transmission, on suppose que des sous-mots de code de différentes transmissions sont regroupés en une trame de longueur N_F symboles. Cette méthode de transmission, utilisée également dans [32], [33], consiste à attribuer une trame de durée fixe pour chaque utilisateur.

La figure (2.1) [34] représente un exemple de trames regroupant des sous-mots de code de longueurs variables. Les couleurs bleue, jaune et verte représentent respectivement la première, la seconde et la troisième tentative de transmission relative à un paquet Q_k , $k = 1 \dots 10$.

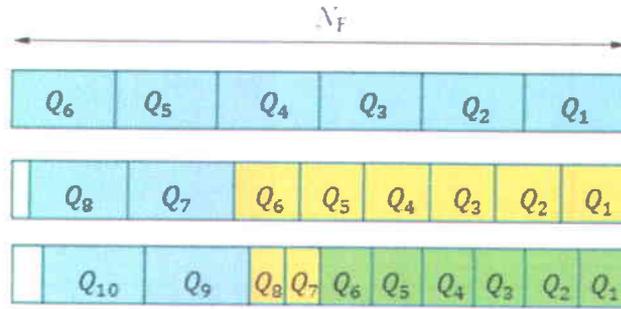


FIGURE 2.1 – Trois trames regroupant des sous-mots de code de longueurs identiques ou différentes.

On considère qu'un utilisateur envoie tout d'abord une trame contenant les premiers sous-mots de code (de couleurs bleues) de 6 paquets différents Q_i , $i = 1..6$. En supposant que la destination ne décode aucun des 6 paquets, l'émetteur reçoit alors un NACK et transmet les seconds sous-mots de code des paquets Q_i , $i = 1..6$ de longueurs différentes (couleurs jaunes). Comme la trame n'est pas totalement occupée, l'émetteur peut également transmettre les premiers sous-mots de code de deux nouveaux paquets Q_7 et Q_8 . De nouveau, on suppose que le récepteur ne décode aucun paquet, l'émetteur transmet alors le troisième sous-mot de code de Q_i , $i = 1..6$, le deuxième sous-mot de code de Q_i , $i = 7, 8$, et comme la trame n'est pas totalement occupée, il envoie également le premier sous-mot de code Q_9 et Q_{10} .

On note que l'espace vacant de la trame est relativement faible. Ainsi, l'utilisation des trames n'engendre qu'une perte minimale en terme de ressources (le temps). Cette méthode de transmission représente alors une solution pratique et efficace pour gérer des transmissions de longueurs variables dans les communications de type TDMA.

Pour notre modèle de système, on suppose que le canal reste constant durant la transmission du $k^{\text{ième}}$ sous-mot de code. Le signal reçu est donné par :

$$Y_k = h_k \cdot X_k + z_k \quad (2.1)$$

où :

- z_k représente le bruit gaussien centré de variance égale à 1.

– $h_k > 0$ est l'évanouissement du canal qu'on suppose constant durant la transmission d'un sous-mot de code mais varie indépendamment d'une transmission à une autre.

Pour simplifier les notations, soit :

$$R = \frac{N_b}{N_{s,1}} \quad (2.2)$$

représentant le taux de la première transmission.

Soit R_k le taux de la $k^{\text{ième}}$ transmission :

$$R_k = \frac{N_b}{N_{s,k}} = \frac{N_{s,1}}{N_{s,k}} \frac{N_b}{N_{s,1}} = \frac{R}{\rho_k} \quad b/symb \quad (2.3)$$

où ρ_k représente le rapport en terme de taux entre la première et la $k^{\text{ième}}$ transmission du même paquet.

Du point de vue de la théorie de l'information et pour les schémas IR-HARQ, un récepteur décode correctement un paquet transmis si l'information mutuelle cumulée après k transmissions est supérieure au taux global de transmission. Dans ce cas, il existe un code tel que la probabilité de coupure tend vers zéro. Cette condition peut être exprimée par :

$$\frac{1}{\sum_{l=1}^k N_{s,l}} \sum_{l=1}^k c_l N_{s,l} > \frac{1}{\sum_{l=1}^k N_{s,l}} N_b$$

$$I_k = \sum_{l=1}^k c_l \rho_l > R \quad (2.4)$$

où c_k représente l'information mutuelle entre X_k et Y_k de l'équation (3.1) conditionnée sur h_k :

$$c_k = I(Y_k; X_k | h_k) = \log_2(1 + \gamma_k) \quad (2.5)$$

$\gamma_k = |h_k|^2$ modélise le gain du canal à la $k^{\text{ième}}$ transmission.

On suppose que h_k suit une loi de Rayleigh, ainsi γ_k suit la loi exponentielle de moyenne $\bar{\gamma}$ donnée par :

$$p_{\gamma_k}(\gamma) = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma}{\bar{\gamma}}\right) \quad (2.6)$$

2.3 Optimisation du débit utile

Dans l'objectif d'adapter les taux de transmission R, R_2, \dots, R_K pour augmenter la probabilité d'un décodage correct, on se demande alors quelles sont les variables qui peuvent influencer le choix des taux? À partir de (2.4) et pour $k = 1$, la condition d'avoir un décodage correct s'écrit comme suit :

$$I_1 = \log_2(1 + \gamma_1) > R \quad (2.7)$$

Comme on l'a déjà mentionné dans le chapitre précédent, à cause des délais de transmission et de traitement, l'émetteur ne peut pas avoir une information exacte sur l'état du canal γ_1 mais plutôt une information obsolète. D'une manière générale, le récepteur estime l'évanouissement du canal \hat{h}_k et le transmet à l'émetteur à travers le canal d'acquiescement avec un retard d . d représente la différence en temps entre la valeur actuelle de l'état du canal et son estimation. Ainsi, la loi de probabilité de $\gamma_k = |h_k|^2$ sachant $\hat{\gamma}_k = |\hat{h}_k|^2$ est la loi chi carré non centrée, donnée par [35] :

$$p_{\gamma|\hat{\gamma}}(\gamma) = \frac{1}{(1-\delta)\sigma_e^2} \exp\left(-\frac{\gamma + \delta\hat{\gamma}}{(1-\delta)\sigma_e^2}\right) I_0\left(\frac{2\sqrt{\gamma\hat{\gamma}\delta}}{(1-\delta)\sigma_e^2}\right) \quad (2.8)$$

où :

- I_0 est la fonction Bessel modifiée d'ordre zéro et de premier type,
- δ représente la corrélation entre γ_k et $\hat{\gamma}_k$.

δ est donné par :

$$\delta = J_0^2(2\pi f_D \tau) \quad (2.9)$$

où J_0 est la fonction Bessel d'ordre zéro et de premier type et f_D représente la fréquence de Doppler.

Ainsi, pour $k = 1$, l'idée est d'adapter R en fonction de $\hat{\gamma}_1$ d'où $R(\hat{\gamma}_1)$.

Pour $k > 1$, la condition sur R_k , ou simplement ρ_k , pour avoir un décodage correct est donnée par :

$$I_k = I_{k-1} + \rho_k \log_2(1 + \gamma_k) > R$$

$$\rho_k > \frac{R - I_{k-1}}{\log_2(1 + \gamma_k)}$$

Ainsi, l'adaptation de ρ_k se basera sur trois paramètres R (ou $\hat{\gamma}_1$), $\hat{\gamma}_k$ et I_{k-1} d'où $\rho_k(\hat{\gamma}_1, \hat{\gamma}_k, I_{k-1})$ pour $k = 2..K$.

Notant que si $I_{k-1} > R$, $\rho_k = 0$ car la $k^{\text{ième}}$ transmission n'est nécessaire que si la transmission numéro $(k - 1)$ n'a pas été décodée correctement.

On suppose que le canal d'acquiescement qui véhicule les CSI, ($\hat{\gamma}_k$ et I_{k-1}), en plus des messages ACK et NACK est sans erreurs.

Remarquons que si R est fixe et que $\rho_k = 1$, $k = 2..K$ on retrouve le cas de IR-HARQ avec des taux fixes traité dans [24]. D'autre part si R, R_2, \dots, R_K varie selon l'indice k de transmission mais indépendamment de tout CSI, alors on retrouve le cas d'allocation des taux traité dans [36].

Notre objectif, à travers ce chapitre, est d'optimiser le débit utile par le biais de l'adaptation des taux de transmission et d'évaluer le potentiel du CSI retardé à augmenter le débit utile par comparaison à celui avec taux fixe et avec allocation des taux.

Le débit utile, donné par le rapport entre la moyenne des bits reçus correctement et le nombre de canaux utilisés, s'écrit comme :

$$\eta = \frac{\bar{N}_b}{\bar{N}_s} \quad (2.10)$$

où :

$$\bar{N}_b = \mathbb{E}_{\hat{\gamma}_1} \{ N_b(\hat{\gamma}_1)(1 - F_{I_K|\hat{\gamma}_1}(R(\hat{\gamma}_1))) \} \quad (2.11)$$

$F_{I_K|\hat{\gamma}_1}(R(\hat{\gamma}_1))$ représente la probabilité d'avoir un décodage incorrect après K transmissions sachant $\hat{\gamma}_1$. En d'autres mots, il s'agit de la probabilité de coupure conditionné par l'estimation de l'état du canal de la première transmission :

$$F_{I_K|\hat{\gamma}_1}(R(\hat{\gamma}_1)) = Pr\{I_K < R(\hat{\gamma}_1)|\hat{\gamma}_1\} = \int_0^{R(\hat{\gamma}_1)} p_{I_K|\hat{\gamma}_1}(x)dx \quad (2.12)$$

On note par $p_{I_K|\hat{\gamma}_1}$ la densité de probabilité de $I_K|\hat{\gamma}_1$.

La moyenne du nombre de canaux utilisés est donnée par :

$$\bar{N}_s = \sum_{k=1}^K \bar{N}_{s,k} \quad (2.13)$$

où $\bar{N}_{s,k}$ représente le nombre moyen de canaux utilisés durant la $k^{\text{ième}}$ transmission, ainsi :

$$\bar{N}_{s,k} = \mathbb{E}_{\gamma_1, \dots, \gamma_{k-1}; \hat{\gamma}_1, \dots, \hat{\gamma}_k} \{N_{s,k}(I_{k-1}, \hat{\gamma}_1, \hat{\gamma}_k)\} \quad k > 1 \quad (2.14)$$

Chaque $N_{s,k}$ dépend de I_{k-1} est donc des variables aléatoires $(\gamma_1, \dots, \gamma_{k-1}; \hat{\gamma}_1, \dots, \hat{\gamma}_{k-1})$, de $\hat{\gamma}_1$ et de $\hat{\gamma}_k$ ce qui explique toutes les variables aléatoires sur lesquelles la moyenne $\mathbb{E}\{\cdot\}$ est effectuée.

En divisant par $N_{s,1}$, l'expression du débit utile est donnée par :

$$\eta = \frac{\mathbb{E}_{\hat{\gamma}_1} \{R(\hat{\gamma}_1)(1 - F_{I_K|\hat{\gamma}_1}(R(\hat{\gamma}_1)))\}}{1 + \sum_{k=2}^K \mathbb{E}_{\gamma_1, \dots, \gamma_{k-1}; \hat{\gamma}_1, \dots, \hat{\gamma}_k} \{\rho_k(I_{k-1}, \hat{\gamma}_1, \hat{\gamma}_k)\}} \quad (2.15)$$

Pour simplifier les notations, l'expression du débit utile sera donnée par :

$$\eta = \frac{\mathbb{E}_{\hat{\gamma}_1} \{R(\hat{\gamma}_1)(1 - F_{I_K|\hat{\gamma}_1}(R(\hat{\gamma}_1)))\}}{1 + D(\rho_2, \dots, \rho_K)} \quad (2.16)$$

Afin d'optimiser le débit utile (2.16), on définit le problème d'optimisation auxiliaire suivant :

$$U(c) = \max_{R, \rho_2, \dots, \rho_K} \mathbb{E}_{\hat{\gamma}_1} \{R(\hat{\gamma}_1)(1 - F_{I_K|\hat{\gamma}_1}(R(\hat{\gamma}_1)))\} \quad s.t. \quad D(\rho_2, \dots, \rho_K) = c \quad (2.17)$$

Alors, le débit utile optimal est donné par :

$$\hat{\eta} = \max_c \eta(R^{*,c}, \rho_2^{*,c}, \dots, \rho_K^{*,c}) \quad (2.18)$$

avec $\{R^{*,c}, \rho_2^{*,c}, \dots, \rho_K^{*,c}\} = \arg \max(U(c))$

Le problème d'optimisation dans (2.17) présente une seule contrainte. Ceci implique que, même si le problème primaire est de dimension infinie le problème dual équivalent est unidimensionnel. Cependant, et d'une manière générale, résoudre le problème primaire et son problème dual ne donne pas la même solution sauf si la «duality gap» est nulle. La «duality gap» est définie comme la différence entre la solution primaire et la solution duale. Il a été démontré, dans [37], que la «duality gap» est nulle tant que les fonctions non-convexes à optimiser sont moyennées par des densités de probabilités strictement continues. Ainsi, dans notre cas, il est plus simple et équivalent d'optimiser le problème dual de (2.17).

En définissant le multiplicateur de Lagrange λ , la fonction duale est donnée par :

$$J(\lambda) = \max_{R, \rho_2, \dots, \rho_K} \mathbb{E}_{\hat{\gamma}_1} \{R(\hat{\gamma}_1)(1 - F_{I_K|\hat{\gamma}_1}(R(\hat{\gamma}_1)))\} + \lambda D(\rho_2, \dots, \rho_K) \quad (2.19)$$

Finalement, optimiser η est équivalent à maximiser $\eta(R^{*,c}, \rho_2^{*,c}, \dots, \rho_K^{*,c})$ par rapport c qui, à son tour, est équivalent à maximiser $\eta(R^{*,\lambda}, \rho_2^{*,\lambda}, \dots, \rho_K^{*,\lambda})$ par rapport à λ :

$$\hat{\eta} = \max_{\lambda} \eta(R^{*,\lambda}, \rho_2^{*,\lambda}, \dots, \rho_K^{*,\lambda}) \quad (2.20)$$

avec $\{R^{*,\lambda}, \rho_2^{*,\lambda}, \dots, \rho_K^{*,\lambda}\} = \arg \max(J(\lambda))$

2.4 Adaptation des taux de transmission

L'objectif, à travers cette section, est de déterminer les politiques d'adaptation de taux qui maximisent le problème (2.20).

(2.19) peut s'écrire comme :

$$J(\lambda) = \max_R \max_{\rho_2} \dots \max_{\rho_K} \mathbb{E}_{\hat{\gamma}_1} \left\{ R(\hat{\gamma}_1) \left(1 - \mathbb{E}_{\gamma_1|\hat{\gamma}_1} \mathbb{E}_{\hat{\gamma}_2} \mathbb{E}_{\gamma_2|\hat{\gamma}_2} \dots \mathbb{E}_{\hat{\gamma}_K} \mathbb{E}_{\gamma_K|\hat{\gamma}_K} \{ \mathbb{I}(I_K < R(\hat{\gamma}_1)) \} \right) \right. \\ \left. + \lambda \mathbb{E}_{\gamma_1|\hat{\gamma}_1} \mathbb{E}_{\hat{\gamma}_2} \{ \rho_2(I_1, \hat{\gamma}_2, \hat{\gamma}_1) \} + \dots + \mathbb{E}_{\gamma_1|\hat{\gamma}_1} \mathbb{E}_{\hat{\gamma}_2} \mathbb{E}_{\gamma_2|\hat{\gamma}_2} \dots \mathbb{E}_{\hat{\gamma}_K} \mathbb{E}_{\gamma_K|\hat{\gamma}_K} \{ \rho_K(I_{K-1}, \hat{\gamma}_K, \hat{\gamma}_1) \} \right\} \quad (2.21)$$

En réorganisant les termes, (2.21) peut s'écrire via la programmation dynamique (PD) comme :

$$J_0(I_0) = \mathbb{E}_{\hat{\gamma}_1} \left\{ J_1(I_0, \hat{\gamma}_1) \right\}$$

$$J_1(I_0, \hat{\gamma}_1) = \max_{R(I_0, \hat{\gamma}_1)} \left\{ R(I_0, \hat{\gamma}_1) + \mathbb{E}_{\gamma_1|\hat{\gamma}_1} \mathbb{E}_{\hat{\gamma}_2} \{ J_2(I_1, \hat{\gamma}_2, \hat{\gamma}_1) \} \right\}$$

$$J_2(I_1, \hat{\gamma}_2, \hat{\gamma}_1) = \max_{\rho_2(I_1, \hat{\gamma}_2, \hat{\gamma}_1)} \left\{ \lambda \rho_2(I_1, \hat{\gamma}_2, \hat{\gamma}_1) + \mathbb{E}_{\gamma_2|\hat{\gamma}_2} \mathbb{E}_{\hat{\gamma}_3} \{ J_3(I_2, \hat{\gamma}_3, \hat{\gamma}_1) \} \right\} \quad (2.22)$$

...

$$J_K(I_{K-1}, \hat{\gamma}_K, \hat{\gamma}_1) = \max_{\rho_K(I_{K-1}, \hat{\gamma}_K, \hat{\gamma}_1)} \left\{ \lambda \rho_K(I_{K-1}, \hat{\gamma}_K, \hat{\gamma}_1) \right. \\ \left. - R(\hat{\gamma}_1) \mathbb{E}_{\hat{\gamma}_K|\hat{\gamma}_K} \{ \mathbb{I}(I_{K-1} + c_K \rho_K < R(\hat{\gamma}_1)) \} \right\}$$

où $J(\lambda) = J_1(I_0)$ et $I_0 = 0$ sont introduites uniquement pour la notation dans PD.

L'idée est de discrétiser la valeur de I_{k-1} en N valeurs entre $[0, R(\gamma_1)]$ ainsi que γ_k et γ_1 en L valeurs entre $[0, +\infty[$. On peut alors déterminer la valeur de $\rho_k(I_{k-1}, \hat{\gamma}_k, \hat{\gamma}_1)$ optimale sachant que $J_{k+1}(I_k, \hat{\gamma}_{k+1}, \hat{\gamma}_1)$ est connue.

Grâce à la formulation de la PD, l'optimisation de la fonction non convexe $J(\lambda)$ par rapport à $N^K L^{2K}$ valeurs se réduit à KNL^2 optimisations unidimensionnelles.

Le principe est de commencer à résoudre la dernière récursion qui peut s'écrire comme :

$$\begin{aligned} J_K(I_{K-1}, \hat{\gamma}_K, \hat{\gamma}_1) &= \max_{\rho_K(I_{K-1}, \hat{\gamma}_K, \hat{\gamma}_1)} \left\{ \lambda \rho_K(I_{K-1}, \hat{\gamma}_K, \hat{\gamma}_1) - R(\hat{\gamma}_1) F_{\gamma_K | \hat{\gamma}_K} \left(2^{\frac{R(\hat{\gamma}_1) - I_{K-1}}{\rho_K}} - 1 \right) \right\} \\ &= \max_{\rho_K(I_{K-1}, \hat{\gamma}_K, \hat{\gamma}_1)} \{ \theta_K(\rho_K) \} \end{aligned} \quad (2.23)$$

Une fois (2.23) résolue et les valeurs de J_K connues pour les différents échantillons de $I_{K-1}, \hat{\gamma}_K$ et $\hat{\gamma}_1$, on passe à l'optimisation de $\rho_{K-1}, \rho_{K-2}, \dots, \rho_2$ et finalement R .

Notons que la valeur du multiplicateur de Lagrange λ dépend de toute l'optimisation, ainsi λ est une fonction de la distribution des variables aléatoires $\gamma_1, \gamma_2, \dots, \gamma_K$ et $\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_K$. En d'autres termes, même si la dépendance de la distribution de γ_K et $\hat{\gamma}_K$ est explicite dans l'optimisation de J_K , la dépendance par rapport aux variables $\gamma_1, \gamma_2, \dots, \gamma_{K-1}$ et $\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_{K-1}$ est implicite (à travers le paramètre λ).

La résolution de (2.23) est différente des autres étapes de la PD car la dérivée par rapport à ρ_K est simple. Ainsi, la valeur optimale de ρ_K vérifie $\frac{d\theta_K(\rho_K)}{d\rho_K} = 0$:

$$\lambda + R \frac{R - I_{K-1}}{\rho_K^2} 2^{\frac{R - I_{K-1}}{\rho_K}} \ln(2) p_{\gamma_K | \hat{\gamma}_K} \left(2^{\frac{R - I_{K-1}}{\rho_K}} - 1 \right) = 0 \quad (2.24)$$

Comme le second terme de (2.24) est toujours positif ou nul, on s'intéresse alors aux valeurs de λ négatives. Quand λ est positif, la solution optimale est donnée par $\rho_K^* \rightarrow \infty$ qui n'est pas une solution réalisable et donc non pertinente. D'autre part, quand la dérivée est strictement négative, la solution optimale est donnée par $\rho_K^* = 0$. Dans ce cas, $J_K = -R$.

Les figures (2.2) et (2.3) représentent le scénario suivant. On suppose que le schéma IR-HARQ avec $K = 3$ est implémenté. Le récepteur estime l'état du canal et envoie à l'émetteur via le canal d'acquittement la valeur de $\hat{\gamma}_1$. Prenons par exemple $\hat{\gamma}_1 = 9dB$. L'émetteur commence la première transmission avec un taux $R = 2.8$ b/symb comme le

montre la figure (2.2) présentant, pour chaque valeur de $\hat{\gamma}_1$ donnée, le taux R qui maximise le débit utile.

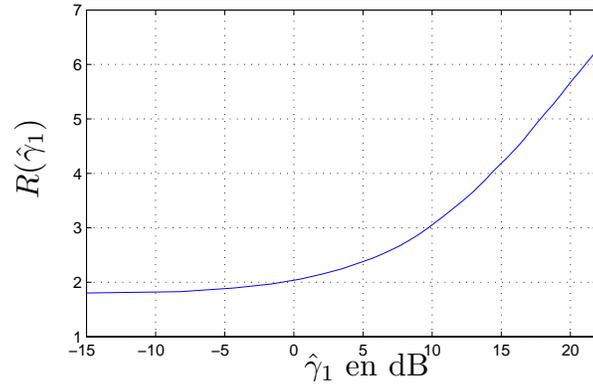


FIGURE 2.2 – $R(\hat{\gamma}_1)$ pour $K = 3$; $\bar{\gamma} = 5dB$; $\delta = 0.5$; $\lambda = -1.6$.

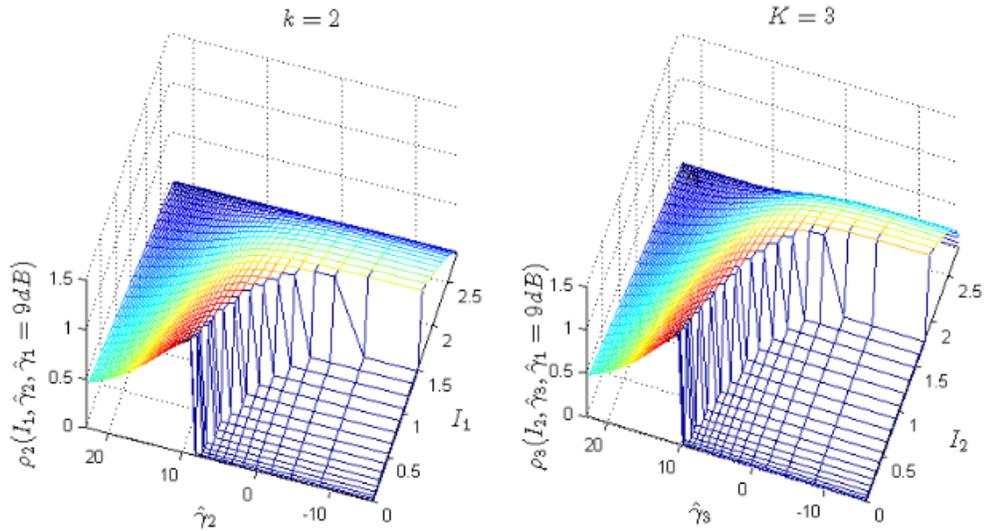


FIGURE 2.3 – De gauche à droite, on représente $\rho_2(I_1, \hat{\gamma}_2, \hat{\gamma}_1 = 9 dB)$ et $\rho_3(I_2, \hat{\gamma}_3, \hat{\gamma}_1 = 9 dB)$, pour $\bar{\gamma} = 5 dB$; $K = 3$; $\bar{\gamma} = 9 dB$; $\delta = 0.5$; $R(\hat{\gamma}_1 = 9 dB) = 2.8 b/symb$; $\lambda = -1.6$.

Si un NACK est reçu, l'émetteur se base sur le CSI pour connaître le taux de la transmission suivante. Ainsi dépendamment de la valeur de I_1 et $\hat{\gamma}_2$, il transmet le second

sous-mot de code avec un taux $R_2 = \frac{R(\hat{\gamma}_1)}{\rho_2(I_1, \hat{\gamma}_2, \hat{\gamma}_1)}$. Les différentes valeurs de ρ_2 pour la valeur de $\hat{\gamma}_1 = 9dB$ sont illustrées dans la figure (2.3) pour $k = 2$.

Notons que les valeurs de I_1 qui nous intéressent sont entre 0 et $R(\hat{\gamma}_1) = 2.8 b/symb$. Lorsque $I_1 > R(\hat{\gamma}_1) = 2.8 b/symb$, un décodage correct a eu lieu au niveau du récepteur et donc $\rho_2 = 0$ ainsi que $\rho_3 = 0$.

On remarque que, pour certaines valeurs de I_1 et $\hat{\gamma}_2$, la valeur de ρ_2 optimale est nulle ce qui dénote l'abandon du paquet. En effet, si l'information mutuelle cumulée au niveau du récepteur est inférieure à un certain seuil $I_{th}(\hat{\gamma}_2)$ (par exemple, $I_{th}(\hat{\gamma}_2 = 0) = 1.5 b/symb$) et que l'état du canal est également inférieur à un autre seuil $\hat{\gamma}_{th}(I_1)$ (par exemple, $\hat{\gamma}_{th}(I_1 = 0) = 8.34 dB$), l'émetteur décide de ne pas transmettre et d'abandonner ainsi le paquet. Quand l'émetteur a uniquement la possibilité d'effectuer K transmissions et qu'en plus le canal est mauvais et la valeur de l'information mutuelle cumulée est faible, assurer un décodage correct au niveau du récepteur correspond à transmettre, dans ces conditions, avec un taux de transmission très faible (c.à.d. un sous-mot de code très long) ce qui pénalise le temps moyen de transmission d'où l'abandon du paquet.

Dans le cas où un second NACK est reçu, l'émetteur passe à la troisième et dernière transmission de notre scénario : De la même manière, dépendamment de la valeur de I_2 et $\hat{\gamma}_3$ pour $\hat{\gamma}_1 = 9dB$, l'émetteur transmet avec un taux $R_3 = \frac{R(\hat{\gamma}_1)}{\rho_3(I_2, \hat{\gamma}_3, \hat{\gamma}_1)}$ où les valeurs de $\rho_3(I_2, \hat{\gamma}_3, \hat{\gamma}_1 = 9dB)$ sont données par la figure (2.3) pour $K = 3$. On note que la même stratégie de l'abandon du paquet est adoptée.

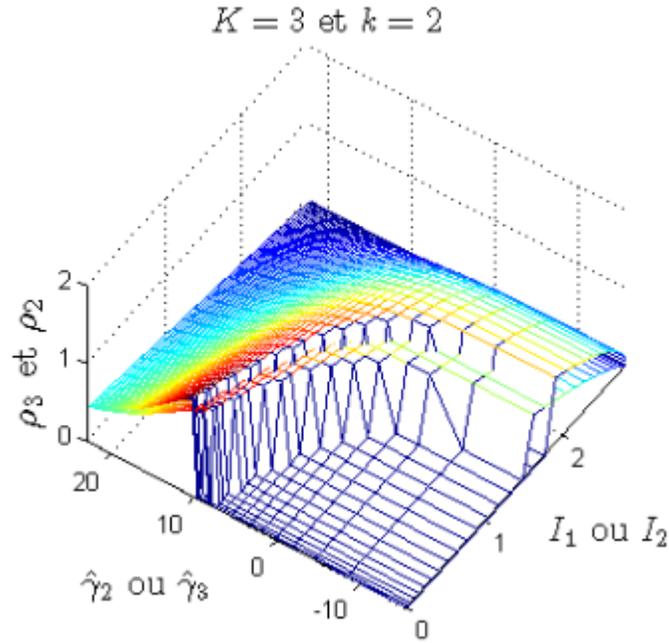


FIGURE 2.4 – La superposition de ρ_2 et ρ_3 pour $\bar{\gamma} = 5$ dB ; $K = 3$; $\bar{\gamma} = 9$ dB ; $\delta = 0.5$; $R(\hat{\gamma}_1 = 9$ dB) = 2.8 b/symb ; $\lambda = -1.6$.

La figure (2.4) représente la superposition de $\rho_3(I_2, \hat{\gamma}_3, \hat{\gamma}_1 = 9$ dB) (au dessus) et $\rho_2(I_1, \hat{\gamma}_3, \hat{\gamma}_1 = 9$ dB) (en dessous de ρ_3). En dehors de la stratégie de l'abandon du paquet, on note que les valeurs de ρ_3 sont supérieures à celles de ρ_2 pour une même valeur de $(\hat{\gamma}_k, I_{k-1})$, et donc le taux des sous-mots de code diminue avec k . En effet, la redondance introduite par IR-HARQ augmente avec k pour tenter de corriger les erreurs survenues lors de la transmission afin de décoder correctement le paquet et arrêter les retransmissions. Cependant, on remarque que l'occurrence de l'abandon du paquet augmente au fil des retransmissions. Comme les taux diminuent avec k , le temps moyen de transmission augmente considérablement à chaque retransmission. Ainsi, continuer à transmettre dans des conditions défavorables (en terme de $\hat{\gamma}_k$ et I_{k-1}) est contre productif en terme de débit utile d'où l'abandon précoce du paquet au fil des k .

2.5 Calcul du débit utile

Le calcul du débit utile n'est pas nécessaire pour la détermination des taux de transmission optimaux. Cependant, on en a besoin pour pouvoir comparer les performances de IR-HARQ avec adaptation des taux, allocation des taux et taux fixe.

L'expression du débit utile (2.16) peut également être écrite comme :

$$\eta = \frac{\mathbb{E}_{\hat{\gamma}_1} \{R(\hat{\gamma}_1)(1 - F_{I_K|\hat{\gamma}_1}(R(\hat{\gamma}_1)))\}}{1 + \mathbb{E}_{\hat{\gamma}_1} \left\{ \sum_{k=2}^K \mathbb{E}_{I_{k-1}|\hat{\gamma}_1} \left\{ \mathbb{E}_{\hat{\gamma}_k} \{ \rho_k(I_{k-1}, \hat{\gamma}_1, \hat{\gamma}_k) \} \right\} \right\}} \quad (2.25)$$

Ainsi, on a besoin des expressions des différentes densités de probabilité $p_{I_{k-1}|\hat{\gamma}_1}$ pour pouvoir calculer η . La fonction de répartition de I_1 sachant $\hat{\gamma}_1$ est donnée par :

$$\begin{aligned} F_{I_1|\hat{\gamma}_1}(x) &= Pr\{c_1 < x|\hat{\gamma}_1 = y\} \\ &= Pr\{\gamma_1 < 2^x - 1|\hat{\gamma}_1 = y\} \\ &= F_{\gamma_1|\hat{\gamma}_1}(2^x - 1) \end{aligned} \quad (2.26)$$

Après dérivation par rapport à x , la densité de probabilité de I_1 sachant $\hat{\gamma}_1$ est donné par :

$$p_{I_1|\hat{\gamma}_1}(x) = \ln(2)2^x p_{\gamma_1|\hat{\gamma}_1}(2^x - 1) \quad x = 1..R(\hat{\gamma}_1 = y) \quad (2.27)$$

où $p_{\gamma_1|\hat{\gamma}_1}$ est la loi chi-carré non centrée définie dans (2.8).

Pour $k \geq 2$, la densité de probabilité de I_k sachant $\hat{\gamma}_1$ est donnée par :

$$p_{I_k|\hat{\gamma}_1}(x) = \int_0^\infty p_{I_k|\hat{\gamma}_k=z,\hat{\gamma}_1}(x) p_{\hat{\gamma}_k}(z) dz \quad (2.28)$$

En tenant en compte la stratégie de l'abandon du paquet, la loi de répartition de I_k sachant $\hat{\gamma}_1$ et $\hat{\gamma}_k$ est donnée par :

$$\begin{aligned} F_{I_k|\hat{\gamma}_k,\hat{\gamma}_1}(x) &= Pr\{I_{k-1} + \rho_k c_k < x|\hat{\gamma}_k = z; \hat{\gamma}_1 = y\} \\ &= \begin{cases} F_{I_{k-1}|\hat{\gamma}_1}(x) & x < I_{th}(z) \\ \int_{I_{th}(z)}^x F_{c_k|\hat{\gamma}_k=z} \left(\frac{x-t}{\rho_k(t,z,y)} \right) p_{I_{k-1}}(t) dt & I_{th}(z) < x < R(y) \end{cases} \end{aligned} \quad (2.29)$$

Ainsi, après dérivation, on obtient la densité de probabilité suivante :

$$p_{I_k|\hat{\gamma}_k,\hat{\gamma}_1}(x) = \begin{cases} p_{I_{k-1}|\hat{\gamma}_1}(x) & x < I_{th}(z) \\ \int_{I_{th}(z)}^x \frac{1}{\rho_k} p_{c_k|\hat{\gamma}_k=z} \left(\frac{x-t}{\rho_k(t,z,y)} \right) p_{I_{k-1}}(t) dt & I_{th}(z) < x < R(y) \end{cases} \quad (2.30)$$

En utilisant les expressions des $p_{I_k|\hat{\gamma}_1}$, on peut calculer l'expression du débit utile donnée par (2.25).

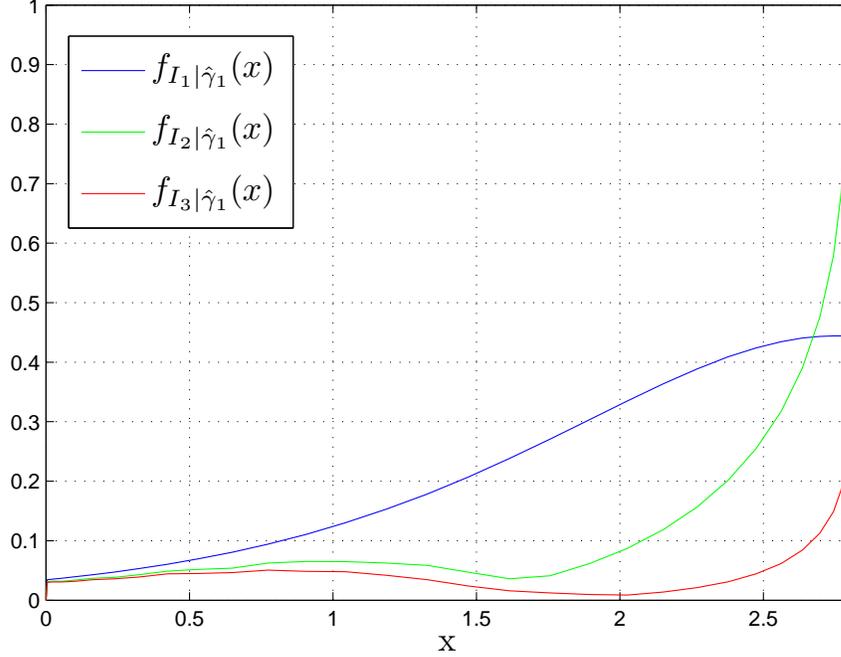


FIGURE 2.5 – La densité de probabilité du décodage erroné après une, deux et trois transmissions conditionnées sur $\hat{\gamma}_1 = 9 \text{ dB}$; Pour $\bar{\gamma} = 5 \text{ dB}$; $K = 3$; $\bar{\gamma} = 9 \text{ dB}$; $\delta = 0.5$; $R(\hat{\gamma}_1 = 9 \text{ dB}) = 2.8 \text{ b/symb}$; $\lambda = -1.6$.

La figure (2.5) représente $p_{I_k|\hat{\gamma}_1=9}(x) \text{ dB}$, $k = 1, 2$ et 3 pour les mêmes conditions du scénario traité dans la section précédente. Les densités de probabilité sont en fonction de $x = 0 \dots R(\hat{\gamma}_1) = 2.8 \text{ b/s}$ (c.à.d. lorsqu'un NACK est reçu au niveau de l'émetteur). On remarque que, plus k augmente, plus la moyenne de la densité de probabilité $p_{I_k|\hat{\gamma}_1}(x)$ augmente. Comme $F_{I_k}(R(\hat{\gamma}_1)) = \int_0^{R(\hat{\gamma}_1)} p_{I_k|\hat{\gamma}_1}(x) dx$ représente la probabilité de ne pas décoder le paquet correctement après k transmissions, cette probabilité diminue considérablement avec l'augmentation de k . $F_{I_k}(R(\hat{\gamma}_1))$ passe de 0.7286 à 0.4425 après juste une retransmission et la probabilité de coupure pour $\hat{\gamma}_1 = 9 \text{ dB}$ est égale à 0.2206.

2.6 Comparaison entre débit utile avec adaptation de taux, allocation de taux ou taux fixe

Rappelons tout d'abord les deux cas limites du débit utile. Lorsque la puissance est constante, la valeur maximale du débit utile est la capacité ergodique qu'on a définie dans (1.12). La limite inférieure est, quant à elle, donnée par le débit utile pour les schémas sans HARQ ($K = 1$) et sans CSIT. Cette expression a été présentée dans (1.11). Le débit utile avec adaptation de taux (η_{AT}) est comparé à celui à taux fixe (η_{FT}) [24] et à celui avec allocation des taux (η_{VT}) [36] pour $K = 2$.

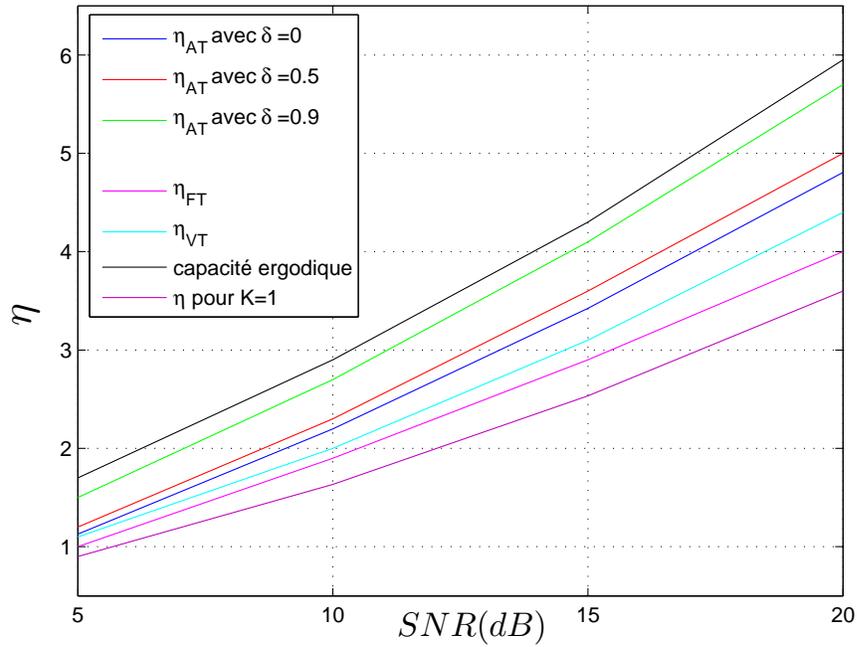


FIGURE 2.6 – Comparaison entre débit utile avec adaptation de taux, allocation de taux ou taux fixe pour IR-HARQ avec $K = 2$.

Comme attendu, le schéma à taux fixe présente le plus mauvais débit utile pour les différentes valeurs du SNR. On remarque que pour un SNR faible, le gain en débit utile pour η_{AT} et $\delta = 0$ est faible voir inexistant par comparaison à celui de η_{VT} . Le gain n'est appréciable que pour les valeurs de δ proche de 1 (c.à.d. si on connaît presque parfaitement

l'état du canal). Par contre, pour les grands SNR, même si $\delta = 0$, le débit utile de η_{AT} est considérablement supérieur à celui à allocation de taux. Plus δ augmente plus le débit utile se rapproche de la capacité ergodique jusqu'à l'atteindre pour $\delta = 1$.

2.7 Conclusion

Dans ce chapitre, on a présenté un système implémentant le schéma IR-HARQ avec CSIT retardé et on a répondu à la question suivante : comment peut-on utiliser cette information obsolète pour améliorer les performances du système en terme de débit utile ?

On a formulé notre problème d'optimisation qui se base sur une adaptation des différents taux de transmission suivant l'information sur I_{k-1} , $\hat{\gamma}_k$ et $\hat{\gamma}_1$. Grâce à une récursion PD, notre problème s'est réduit à K maximisations unidimensionnelles et on a pu ainsi déterminer les valeurs R, ρ_2, \dots, ρ_K optimales. Les valeurs obtenues du débit utile en fonction du SNR montrent un gain appréciable par comparaison au débit utile avec allocation des taux ou taux fixes surtout pour les grands SNR.

Dans le chapitre suivant, on présente une adaptation de puissance afin de répondre à un problème de minimisation de la probabilité de coupure.

Chapitre 3

Adaptation des puissances de transmission pour les schémas HARQ-IR avec CSIT

3.1 Introduction

Après l'étude de l'adaptation des taux, on se concentre à travers ce chapitre sur l'étude du schéma IR-HARQ tronqué avec adaptation de puissance. Comme les taux de transmission sont fixes, on s'intéressera à la probabilité de coupure comme critère de performance. Le débit utile optimal pourra être déduit directement à partir de la probabilité de coupure optimale. Cependant, cette métrique ne reflète pas la variation du délai avec l'adaptation de puissance, c'est pourquoi on étudiera également la *capacité effective* avec IR-HARQ et adaptation de puissance.

Ce chapitre est organisé comme suit : Dans la section 3.2, on présente l'étude de l'adaptation de puissance dans l'objectif de minimiser la probabilité de coupure. Dans les sous-sections 3.2.1, 3.2.2 et 3.2.3 on présente respectivement le modèle du système étudié, le problème d'optimisation de la probabilité de coupure et enfin une comparaison entre probabilité de coupure avec adaptation de puissance, allocation de puissance et puissance fixe

La section 3.3 expose le problème d'optimisation de la capacité effective pour IR-HARQ. Après la définition du modèle du système dans 3.3.1, on détermine l'expression de la capacité effective pour les schémas HARQ tronqués dans la sous-section 3.3.2. On présente, dans la sous-section 3.3.3, le problème de la maximisation de la capacité effective par adaptation de

puissance. Enfin, une comparaison entre la capacité effective avec adaptation de puissance, allocation de puissance et puissance fixe est présentée dans 3.3.4.

3.2 Adaptation des puissances de transmission pour la minimisation de la probabilité de coupure

3.2.1 Modélisation du système

Tout comme pour le chapitre 2, on considère le schéma IR-HARQ tronqué avec un maximum de K transmissions par paquet. Les bits d'information sont divisés en paquets de longueur N_b . Chaque paquet est codé en un mot de code contenant N_s symboles x_1, x_2, \dots, x_{N_s} indépendants et aléatoires suivant la loi gaussienne complexe et centrée. Ces symboles sont regroupés en K sous-mots de code X_1, X_2, \dots, X_K .

Par contre, contrairement au système étudié au cours du chapitre précédent, le nombre de bits N_b est constant et les sous-mots de code sont de même longueur $N_{s,1} = N_{s,2} = \dots = N_{s,K}$. Comme les longueurs des transmissions sont fixes, le mode de transmission ne pose plus de problèmes et l'adoption des trames regroupant différents sous-mots de code n'est plus justifiée. Le système de transmission avec des segments fixes de durée T est optimal dans ce cas. Ainsi, chaque transmission occupe un seul segment et présente un taux de transmission $R = \frac{N_b}{T}$ b/s.

On suppose que le signal reçu est donné par :

$$Y_k = h_k \cdot X_k + z_k \quad (3.1)$$

où :

- z_k représente le bruit gaussien centré de variance égale à 1,
- $h_k > 0$ représente l'évanouissement du canal qu'on suppose constant durant la transmission d'un sous-mot de code, mais variant indépendamment d'une transmission à une autre. On note $\gamma_k = |h_k|^2$ le gain du canal à la $k^{\text{ième}}$ transmission. On suppose que h_k suit une loi de Rayleigh, ainsi γ_k suit la loi exponentielle donnée par (2.6),
- X_k est le signal émis de puissance P_k . Cette dernière peut varier d'une transmission à une autre.

L'adaptation de la puissance se basera sur CSIT retardé. Cette information sera transmise sur le canal d'acquiescement en plus de l'acquiescement ACK-NACK. L'objectif de cette adaptation de puissance est de minimiser la probabilité de coupure qui, dans ce cas, est donnée par :

$$f_K = Pr\{I_K < R\} = Pr\left\{\sum_{k=1}^K \log_2(1 + P_k \gamma_k) < R\right\} \quad (3.2)$$

Dans les cas pratiques, la ressource de la puissance est limitée, principalement au niveau de l'équipement usager. Une contrainte sur la puissance moyenne transmise à long terme doit être imposée. En se basant sur le théorème de Renewal-Reward, défini dans le premier chapitre dans la sous-section 1.3.1, la contrainte sur la puissance moyenne est donnée par :

$$P_{avg} = \frac{\mathbb{E}(P)}{\mathbb{E}(T)} < \bar{P} \quad (3.3)$$

où :

- $\mathbb{E}(P)$ est la puissance moyenne transmise,
- $\mathbb{E}(T)$ est le nombre moyen de segments nécessaires pour compléter la transmission du paquet,
- \bar{P} est la limite qu'on impose sur la puissance moyenne à long terme.

Comme adoptée précédemment, l'adaptation de puissance se fera sur la base de l'information mutuelle cumulée au niveau du récepteur et d'une estimation du gain du canal. Ainsi, P_k dépend des deux informations I_{k-1} et $\hat{\gamma}_k$: $P_k(I_{k-1}, \hat{\gamma}_k)$.

P_{avg} s'écrit alors comme :

$$P_{avg} = \frac{\mathbb{E}\{\sum_{k=1}^K P_k(I_{k-1}, \hat{\gamma}_k)\}}{1 + \sum_{k=1}^{K-1} f_k} \quad (3.4)$$

Notre objectif, dans la sous-section suivante, est de minimiser la probabilité de coupure par le biais de l'adaptation de la puissance des transmissions et d'évaluer le potentiel du CSI retardé à diminuer cette métrique par comparaison à celle avec une puissance fixe et avec allocation de puissance.

3.2.2 Optimisation de la probabilité de coupure

Le problème d'optimisation de la probabilité de coupure avec une adaptation de puissance et une contrainte sur la puissance moyenne à long terme est donné par :

$$J(\lambda) = \min_{P_1, \dots, P_K} f_K \quad \text{s.t.} \quad P_{avg} < \bar{P} \quad (3.5)$$

Pour simplifier l'optimisation, on choisit $\bar{P} = 1$. Ce choix n'affecte en rien notre analyse qui sera abordée de manière similaire pour d'autres valeurs de \bar{P} .

Soit λ un multiplicateur de Lagrange, alors le problème de minimisation de (3.5) s'écrit comme :

$$J(\lambda) = \min_{P_1, \dots, P_K} f_K + \lambda(P_{avg} - 1) \quad (3.6)$$

En se basant sur la formule du P_{avg} (3.4), on obtient :

$$J(\lambda) = \min_{P_1, \dots, P_K} f_K + \lambda \left(\mathbb{E} \left\{ \sum_{k=1}^K P_k(I_{k-1}, \hat{\gamma}_k) \right\} - \sum_{k=1}^{K-1} f_k \right) \quad (3.7)$$

Comme f_k peut être exprimée en fonction des espérances sur γ et $\hat{\gamma}$:

$$\begin{aligned} f_k &= \mathbb{E}_{\hat{\gamma}_1} \mathbb{E}_{\gamma_1 | \hat{\gamma}_1} \mathbb{E}_{\hat{\gamma}_2} \mathbb{E}_{\gamma_2 | \hat{\gamma}_2} \dots \mathbb{E}_{\hat{\gamma}_k} \mathbb{E}_{\gamma_k | \hat{\gamma}_k} \{ \mathbb{I}(I_k < R) \} \\ &= \mathbb{E}_{\hat{\gamma}_1} \mathbb{E}_{\gamma_1 | \hat{\gamma}_1} \mathbb{E}_{\hat{\gamma}_2} \mathbb{E}_{\gamma_2 | \hat{\gamma}_2} \dots \mathbb{E}_{\hat{\gamma}_k} \left\{ F_{\gamma_k | \hat{\gamma}_k} \left(\frac{2^{R-I_{k-1}} - 1}{P_k} \right) \right\} \end{aligned} \quad (3.8)$$

$J(\lambda)$ peut alors s'écrire, en appliquant la programmation dynamique, comme suit :

$$\begin{aligned} J_0(I_0) &= \mathbb{E}_{\hat{\gamma}_1} \{ J_1(I_0, \hat{\gamma}_1) \} \\ J_1(I_0, \hat{\gamma}_1) &= \min_{P_1(I_0, \hat{\gamma}_1)} -\lambda F_{\gamma_1 | \hat{\gamma}_1} \left(\frac{2^{R-I_0} - 1}{P_1} \right) + \lambda P_1 + \mathbb{E}_{\gamma_1 | \hat{\gamma}_1} \{ \mathbb{E}_{\hat{\gamma}_2} \{ J_2(I_1, \hat{\gamma}_2) \} \} \\ J_2(I_1, \hat{\gamma}_2) &= \min_{P_2(I_1, \hat{\gamma}_2)} -\lambda F_{\gamma_2 | \hat{\gamma}_2} \left(\frac{2^{R-I_1} - 1}{P_2} \right) + \lambda P_2 + \mathbb{E}_{\gamma_2 | \hat{\gamma}_2} \{ \mathbb{E}_{\hat{\gamma}_3} \{ J_3(I_2, \hat{\gamma}_3) \} \} \\ &\dots \\ J_K(I_{K-1}, \hat{\gamma}_K) &= \min_{P_K(I_{K-1}, \hat{\gamma}_K)} \lambda P_K + F_{\gamma_K | \hat{\gamma}_K} \left(\frac{2^{R-I_{K-1}} - 1}{P_K} \right) \end{aligned} \quad (3.9)$$

On commence par la détermination de la P_K optimale pour différents échantillons de I_{k-1} entre $[0..R]$ et $\hat{\gamma}_K$ entre $[0..+\infty[$. Les valeurs de I_{K-1} entre $[0..R]$ correspondent aux valeurs pour lesquelles le paquet n'a pu être décodé correctement après $K-1$ transmissions. P_K optimale vérifie :

$$\lambda - \frac{2^{R-I_{K-1}} - 1}{P_K^2} P_{\gamma_K | \hat{\gamma}_K} \left(\frac{2^{R-I_{K-1}} - 1}{P_K} \right) = 0 \quad (3.10)$$

Une fois que P_K est déterminée pour différentes valeurs de I_{k-1} et $\hat{\gamma}_K$, $J_K(I_{k-1}, \hat{\gamma}_K)$ est connue, on peut alors passer au calcul du P_{K-1} optimale et ainsi de suite jusqu'à P_1 . Précisons que toutes ces étapes sont effectuées pour un λ fixe. Une fois que toutes les valeurs de P_k , $k = 1..K$ sont déterminées, on calcule la probabilité de coupure. Puis on recommence toutes ces étapes pour une valeur différente de $\lambda > 0$ jusqu'à obtenir la valeur minimale de f_K tout en vérifiant la contrainte sur la puissance moyenne à long terme. Rappelons que notre optimisation est convexe en terme de λ puisqu'on traite le problème dual et non le problème primal.

La figure (3.1) représente les valeurs optimales des puissances P_2 et P_1 pour un schéma IR-HARQ tronqué avec $K = 2$. P_2 est une fonction de I_1 et $\hat{\gamma}_2$. Comme $I_0 = 0$, P_1 dépend uniquement de $\hat{\gamma}_1$.

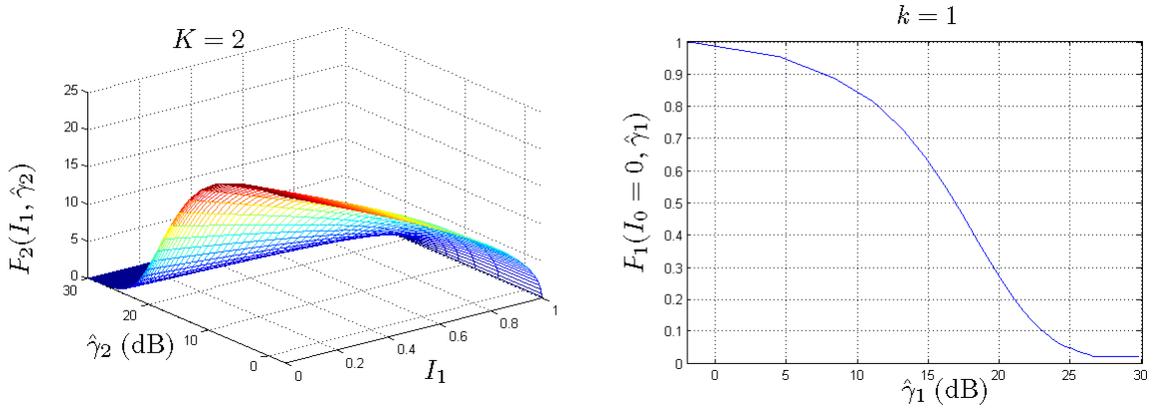


FIGURE 3.1 – De gauche à droite, on représente $P_2(I_1, \hat{\gamma}_2)$ et $P_1(I_0 = 0, \hat{\gamma}_1)$ optimales, pour $\bar{\gamma} = 15$ dB ; $K = 2$; $\delta = 0.5$; $R = 1$ b/s ; $\lambda = -1.25 \cdot 10^{-4}$.

On remarque que, pour le premier sous-mot de code $k = 1$, la puissance maximale est de 1 (attribuée quand $\hat{\gamma}_1$ est faible). Alors que la valeur maximale de la puissance attribuée à la deuxième transmission pour les mêmes conditions du canal est de 22.2721. Cette énorme différence est due à la grande diminution de la probabilité de transmission du sous-mot de code numéro 2. En effet, la transmission du premier sous-mot de code se fait avec une probabilité égale à 1. Si le paquet n'a pas été décodé, on passe à la deuxième transmission. Ainsi la probabilité d'occurrence de la deuxième transmission est égale à f_1 . Dans notre scénario, elle est donnée par $f_1 = 0.0410$. Comme la probabilité de transmission du deuxième sous-mot de code est faible, on peut lui attribuer une plus grande puissance pour diminuer la

probabilité de coupure tout en vérifiant la contrainte sur la puissance moyenne à long terme.

La stratégie de l'abandon du paquet n'est pas adoptée pour $\bar{\gamma} = 15 \text{ dB}$ mais elle peut l'être pour d'autres valeurs du $\bar{\gamma}$ faibles comme le montre la figure suivante :

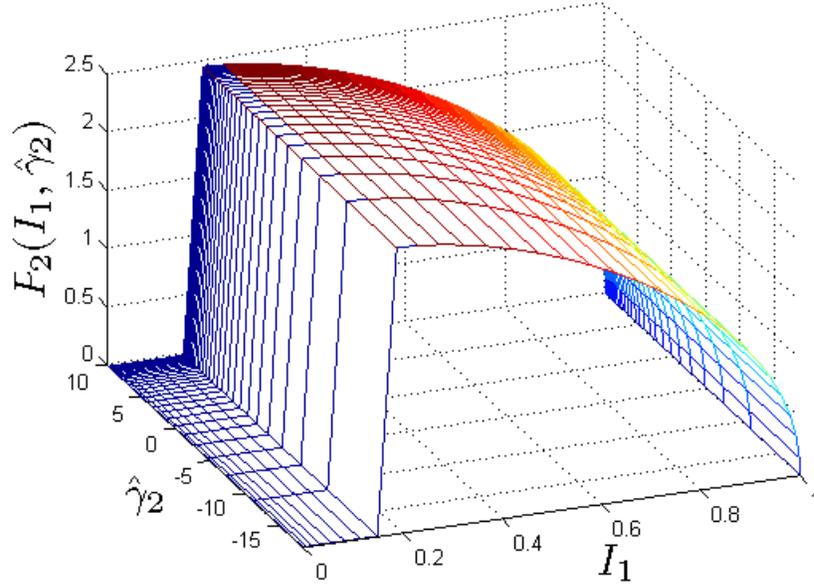


FIGURE 3.2 – $P_2(I_1, \hat{\gamma}_2)$ optimale, pour $\bar{\gamma} = -5 \text{ dB}$; $K = 2$; $\delta = 0$; $R = 1 \text{ b/s}$; $\lambda = -0.23$.

Remarquons, de plus, que les valeurs de P_2 sont invariables en fonction de $\hat{\gamma}_2$ car on a choisi $\delta = 0$. Ceci représente le cas où aucune corrélation n'existe entre l'estimation du canal et la valeur actuelle du canal ou, tout simplement, le cas où aucune information sur l'état du canal n'est disponible au niveau de l'émetteur. L'unique information dont dispose l'émetteur est celle de I_{k-1} . Donc, l'optimisation des puissances se fait sur une unique information : tous les $J_k(I_{k-1})$ dépendent uniquement de I_{K-1} . Dans ce cas, la distribution du canal se réduit à la loi exponentielle caractéristique de l'évanouissement de Rayleigh.

3.2.3 Comparaison entre probabilité de coupure avec adaptation de puissance, allocation de puissance ou puissance fixe

On n'a pas besoin de déterminer explicitement les expressions de p_{I_k} . Cependant, on est amené à les déterminer quand on veut comparer la probabilité de coupure avec adaptation de puissance f_K^{AP} , allocation de puissance f_K^{VP} et puissance fixe f_K^{FP} .

Comme p_{I_k} s'écrit :

$$p_{I_k}(x) = \int_0^\infty p_{I_k|\hat{\gamma}_k}(x; y) dy = \int_0^\infty p_{I_k|\hat{\gamma}_k}(x) p_{\hat{\gamma}_k}(y) dy, \quad (3.11)$$

on commence par déterminer la densité de répartition de $p_{I_1|\hat{\gamma}_1}$:

$$\begin{aligned} F_{I_1|\hat{\gamma}_1}(x) &= Pr\{I_0 + \log_2(1 + P_1\gamma_1) < x | \hat{\gamma}_1 = y\} = Pr\left\{\gamma_1 < \frac{2^x - 1}{P_1(I_0, \hat{\gamma}_1 = y)} | \hat{\gamma}_1 = y\right\} \\ &= F_{\gamma_1|\hat{\gamma}_1}\left(\frac{2^x - 1}{P_1}\right) \end{aligned} \quad (3.12)$$

Après dérivation par rapport à x , on obtient :

$$p_{I_1|\hat{\gamma}_1}(x) = \frac{2^x}{P_1} p_{\gamma_1|\hat{\gamma}_1}\left(\frac{2^x - 1}{P_1}\right) \quad (3.13)$$

Une fois qu'on a déterminé p_{I_1} , on passe au calcul de p_{I_k} , $k = 2, 3 \dots$ puis K .

La fonction de répartition de $F_{I_k|\hat{\gamma}_k}$, $k \geq 2$, est donnée par :

$$F_{I_k|\hat{\gamma}_k}(x) = Pr\{I_{k-1} + \log_2(1 + P_k(I_{k-1}, \hat{\gamma}_k)\gamma_k) < x | \hat{\gamma}_k = y\} \quad (3.14)$$

d'où,

$$F_{I_k|\hat{\gamma}_k}(x) = \begin{cases} F_{I_{k-1}}(x) & \text{si } x < I_{th}(y) \\ F_{I_{k-1}}(I_{th}(y)) + \int_{I_{th}(y)}^x F_{\gamma_k|\hat{\gamma}_k=y}\left(\frac{2^{x-z}-1}{P_k(z,y)}\right) p_{I_{k-1}}(z) dz & \text{sinon} \end{cases} \quad (3.15)$$

Ainsi, $p_{I_k|\hat{\gamma}_k}$ est donnée par :

$$p_{I_k|\hat{\gamma}_k}(x) = \begin{cases} p_{I_{k-1}}(x) & \text{si } x < I_{th}(y) \\ \int_{I_{th}(y)}^x \frac{2^{x-z}}{P_k(z,y)} p_{\gamma_k|\hat{\gamma}_k=y}\left(\frac{2^{x-z}-1}{P_k(z,y)}\right) p_{I_{k-1}}(z) dz & \text{sinon} \end{cases} \quad (3.16)$$

Finalement, la probabilité de coupure f_K^{AP} , qui est égale à $F_{I_K}(R)$, est donnée par :

$$\begin{aligned} f_K^{\text{AP}} &= \\ & \int_{y=0}^\infty \left\{ F_{I_{K-1}}(I_{th}(y)) + \int_{z=I_{th}(y)}^R \frac{2^{R-z}}{P_K(z,y)} p_{\gamma_K|\hat{\gamma}_K=y}\left(\frac{2^{R-z}-1}{P_K(z,y)}\right) p_{I_{K-1}}(z) dz \right\} p_{\hat{\gamma}_K}(y) dy \end{aligned} \quad (3.17)$$

La probabilité de coupure minimale obtenue par notre schéma d'adaptation de puissance (f_K^{AP}) est représentée dans la figure (3.3) et comparée aux deux modèles suivants : Le cas de puissance constante (f_K^{FP}) et le cas d'allocation de puissance (f_K^{VP}) étudié dans [4]. Ces trois schémas sont comparés pour IR-HARQ avec $K = 2$ et $K = 3$.

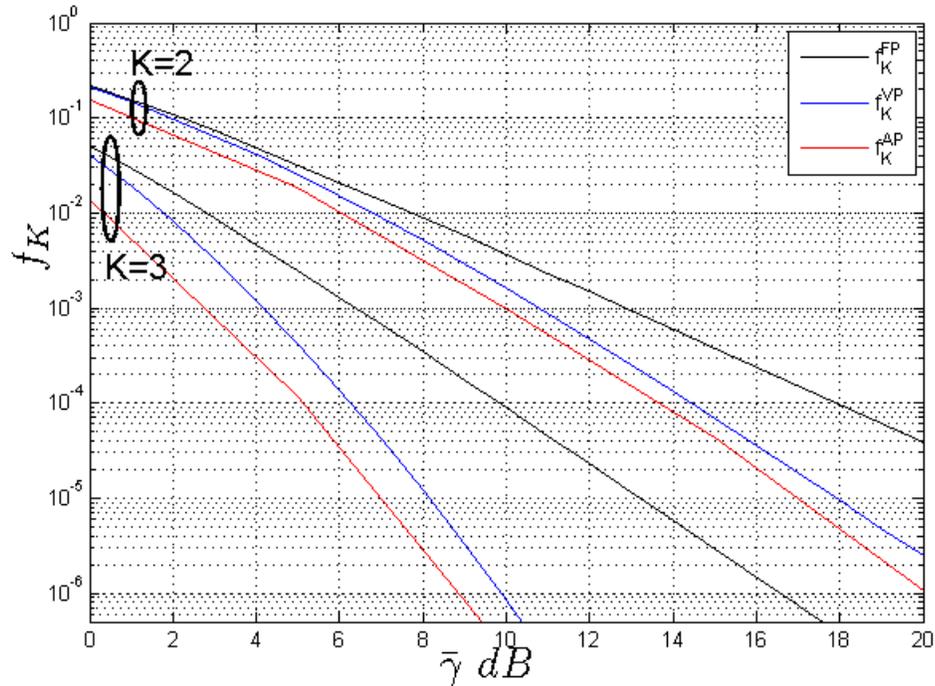


FIGURE 3.3 – La probabilité de coupure pour l'adaptation de puissance (f_K^{AP}), allocation de puissance (f_K^{VP}) et puissance fixe (f_K^{FP}). Pour $K = 2$, $K = 3$, $R = 1$ et $\delta = 0.5$.

On note que, pour les faibles SNR et $K = 2$, les trois schémas sont presque identiques. La différence n'apparaît que pour les SNR élevés. Notre schéma d'adaptation est meilleur que les deux autres schémas en terme de probabilité de coupure. Cependant, le gain de f_K^{AP} par rapport à f_K^{VP} est faible (de l'ordre de 0.3 dB en SNR moyen par transmission). Il est donc plus simple et moins coûteux en acquittements d'utiliser l'allocation de puissance.

Pour $K = 3$, le schéma HARQ avec adaptation de puissance est plus performant que les autres schémas HARQ avec puissance fixe et allocation de puissance. Pour une probabilité de coupure égale à 10^{-4} , la solution optimale présente un gain de 1dB en SNR moyen par transmission par rapport à la solution d'allocation.

Notons que la notion de délai n'est pas traduite par la probabilité de coupure, il est donc intéressant d'étudier d'autres critères de performance tel que la capacité effective avec adaptation de puissance pour avoir une idée globale sur la performance des systèmes étudiés.

3.3 Adaptation de la puissance des transmissions pour la maximisation de la capacité effective

Dans cette section, on s'intéresse à l'étude de l'adaptation de puissance pour les schémas IR-HARQ tronqués afin de maximiser la capacité effective. Ce critère, défini dans le premier chapitre, présente une mesure du débit utile avec une pondération sur le critère du délai.

3.3.1 Modélisation du système

On adopte un modèle d'étude identique à celui présenté dans la sous-section (3.2.1). Les puissances des sous-mots de code seront adaptées sur la base du CSIT. Dans ce cas, le CSIT contient uniquement la valeur de l'information mutuelle cumulée au niveau du récepteur (cas de $\delta = 0$). On a omis l'information sur le gain du canal pour simplifier les notations, mais une analyse similaire à celle qu'on présente dans cette section peut être facilement faite pour inclure $\hat{\gamma}_k$, $k = 1..K$.

3.3.2 Expression de la capacité effective pour les schémas HARQ

Pour pouvoir déterminer l'expression de la capacité effective donnée par (1.7), on modélise notre système d'étude par la machine d'état suivante :

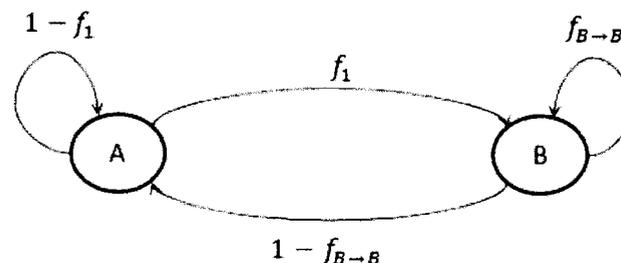


FIGURE 3.4 – Diagramme à états finis pour IR-HARQ tronqué.

où,

- l'état $S = A$ représente l'évènement «transmettre le premier sous-mot de code».
- l'état $S = B$ représente l'évènement «effectuer une retransmission».

Ce diagramme modélise le schéma suivant : IR-HARQ commence par transmettre le premier sous-mot de code. Si un ACK est reçu, le schéma HARQ recommence à nouveau et transmet le premier sous-mot de code du paquet suivant. Ainsi, cette probabilité de passer de l'état A à l'état A traduit la réception d'un ACK après la première transmission et elle est donnée par :

$$1 - f_1 = 1 - Pr\{I_1 < R\} = Pr\{I_1 > R\} \quad (3.18)$$

Cependant si un NACK est reçu, le schéma HARQ effectue une retransmission. Ainsi, on passe de l'état A à l'état B . Cette probabilité de transition traduit la réception d'un NACK après la première transmission et elle est donnée par :

$$f_1 = Pr\{I_1 < R\} \quad (3.19)$$

Maintenant qu'on est dans l'état B , deux scénarios se présentent. Si l'émetteur reçoit un NACK alors une seconde retransmission est nécessaire. La probabilité de transition de l'état B vers l'état B est donné par la probabilité de se retrouver à l'état B à $t + 1$ sachant qu'on est à l'état B à l'instant t :

$$f_{B \rightarrow B} = Pr\{S_{t+1} = B | S_t = B\} = \frac{Pr\{S_{t+1} = B \wedge S_t = B\}}{Pr\{S_t = B\}} \quad (3.20)$$

La probabilité de passer de B vers B est égale à la probabilité de se retrouver à la 3^{ième}, 4^{ième}, ... ou K ^{ième} transmission et donc dépend du numéro de la transmission. Après la K ^{ième} transmission et indépendamment de l'acquittement reçu, le schéma HARQ est réinitialisé et le système se trouve à l'état A . Ceci découle du schéma tronqué de IR-HARQ. Ainsi, $f_{B \rightarrow B}$ s'écrit comme :

$$f_{B \rightarrow B} = \frac{\sum_{k=3}^K Pr\{Tr_{t+1} = k\}}{\sum_{k=2}^K Pr\{Tr_t = k\}} \quad (3.21)$$

où, Tr_t traduit le numéro de la transmission à t .

L'objectif est maintenant de déterminer $\pi_k = Pr\{Tr_t = k\}$:

$$\pi_k = Pr\{Tr_t = k | Tr_{t-1} = k - 1\} Pr\{Tr_{t-1} = k - 1\} = \frac{f_k}{f_{k-1}} \pi_{k-1} \quad (3.22)$$

Par récurrence, on obtient :

$$\pi_k = \frac{f_k}{f_1} \pi_1 \quad (3.23)$$

Ainsi, (3.21) devient :

$$f_{B \rightarrow B} = \frac{\frac{\pi_1}{f_1} \sum_{k=3}^K f_k}{\frac{\pi_1}{f_1} \sum_{k=2}^K f_k} \quad (3.24)$$

Finalement,

$$f_{B \rightarrow B} = \frac{\sum_{k=3}^K f_k}{\sum_{k=2}^K f_k} \quad (3.25)$$

Pour le second scénario, c.à.d. lorsque l'émetteur, qui est dans l'état B , reçoit un ACK, le schéma HARQ est réinitialisé à nouveau et on passe à l'état A pour effectuer la première transmission du paquet suivant. Cette probabilité est donnée par :

$$f_{B \rightarrow A} = 1 - f_{B \rightarrow B} = 1 - \frac{\sum_{k=3}^K f_k}{\sum_{k=2}^K f_k} \quad (3.26)$$

Pour calculer $E_c(\theta)$, on doit déterminer M ainsi que $\Phi(\theta)$ définis dans (1.7). Pour le diagramme d'état décrit par (3.4), la matrice du transition est donnée par :

$$M = \begin{bmatrix} 1 - f_1 & f_1 \\ 1 - f_{B \rightarrow B} & f_{B \rightarrow B} \end{bmatrix} \quad (3.27)$$

et

$$\Phi(\theta) = \begin{bmatrix} e^{-\theta R} & 0 \\ 0 & 1 \end{bmatrix} \quad (3.28)$$

le taux de service est défini par R pour une première transmission et par 0 pour les retransmissions. En effet, un paquet de taux R passe de la file d'attente au schéma HARQ lorsqu'un nouveau paquet est à transmettre, ainsi la file d'attente est vidée avec un taux de R b/s lorsque le système est à l'état A . Le schéma HARQ génère les différents sous-mot de code et se charge des retransmissions. Durant les retransmissions, aucun paquet n'est retiré du buffet d'où un taux de libération de 0 b/s.

Finalement, la capacité effective est donnée par :

$$E_c(\theta) = -\frac{1}{\theta} \log \left\{ \frac{1}{2} \left((1-f_1)e^{-\theta R} + f_{B \rightarrow B} + \sqrt{\left((1-f_1)e^{-\theta R} + f_{B \rightarrow B} \right)^2 - 4 \left((1-f_1)f_{B \rightarrow B}e^{-\theta R} - (1-f_{B \rightarrow B})f_1e^{-\theta R} \right)} \right) \right\} \quad (3.29)$$

3.3.3 Maximisation de la capacité effective

L'idée est de maximiser $E_c(\theta)$ pour un f_1 donné et ensuite de, chercher la valeur de f_1 (ou de manière équivalente la valeur de P_1) qui permet de vérifier la contrainte sur la puissance moyenne à long terme ($P_{avg} < \bar{P}$) et maximise la capacité effective.

Ainsi, pour un f_1 fixe, on dérive $E_c(\theta)$ par rapport à $f_{B \rightarrow B}$. La dérivée obtenue est strictement négative. Donc, la valeur maximale de $E_c(\theta)$ pour un f_1 fixe est donnée par la valeur minimale de $f_{B \rightarrow B}$. Notre problème initial de maximisation de la capacité effective devient alors un problème de minimisation de $f_{B \rightarrow B}$. Notons que la diminution de la probabilité de retransmission revient à diminuer le temps moyen de transmission et donc de diminuer le délai.

Le nouveau problème à résoudre est donné par l'expression suivante :

$$\min_{P_1, \dots, P_K} \frac{\sum_{k=3}^K f_k}{\sum_{k=2}^K f_k} \quad \text{s.t.} \quad \frac{\mathbb{E}\{\sum_{k=1}^K P_k(I_{k-1})\}}{1 + \sum_{k=1}^{K-1} f_k} < \bar{P} \quad (3.30)$$

Soit λ et β deux multiplicateurs de Lagrange. Alors (3.30) s'écrit comme :

$$\min_{P_1, \dots, P_K} \sum_{k=3}^K f_k + \beta f_k \sum_{k=2}^K f_k + \lambda \mathbb{E} \left\{ \sum_{k=1}^K P_k(I_{k-1}) \right\} - \lambda \sum_{k=1}^{K-1} f_k \quad (3.31)$$

Par programmation dynamique, le problème de minimisation s'écrit comme :

$$\begin{aligned} J_2(I_1) &= \min_{P_2(I_1)} \lambda P_2 + (\beta - \lambda) F_{\gamma_2} \left(\frac{2^{R-I_1} - 1}{P_2} \right) + \mathbb{E}_{\gamma_2} \{ J_3(I_2) \} \\ J_3(I_2) &= \min_{P_3(I_2)} \lambda P_3 + (\beta - \lambda + 1) F_{\gamma_3} \left(\frac{2^{R-I_2} - 1}{P_3} \right) + \mathbb{E}_{\gamma_3} \{ J_4(I_3) \} \\ J_4(I_3) &= \min_{P_4(I_3)} \lambda P_4 + (\beta - \lambda + 1) F_{\gamma_4} \left(\frac{2^{R-I_3} - 1}{P_4} \right) + \mathbb{E}_{\gamma_4} \{ J_5(I_4) \} \\ &\dots \\ J_K(I_{K-1}) &= \min_{P_K(I_{K-1})} \lambda P_K + (\beta + 1) F_{\gamma_K} \left(\frac{2^{R-I_{K-1}} - 1}{P_K} \right) \end{aligned} \quad (3.32)$$

On détermine ainsi P_K, P_{K-1}, \dots, P_2 et enfin la valeur de P_1 qui permet de satisfaire la contrainte sur la puissance moyenne à long terme et de maximiser $E_c(\theta)$.

Une fois qu'on a déterminé les puissances optimales $P_1 \dots P_k$, on calcule les probabilités de décodage incorrect après k transmissions f_k , $k = 1 \dots K$. Enfin, en utilisant (3.25) et (3.29), on détermine la capacité effective optimale pour différentes valeurs de θ .

3.3.4 Comparaison entre capacité effective avec adaptation de puissance, allocation de puissance ou puissance fixe

La figure (3.5) représente une comparaison de la capacité effective avec notre schéma d'adaptation de puissance (E_c^{PA}), allocation de puissance (E_c^{PV}) et puissance fixe (E_c^{PF}) pour IR-HARQ avec $K = 3$. Comme le cas de capacité effective avec allocation de puissance n'a pas été traité dans la littérature, on dérive E_c^{PV} pour $K = 3$ dans l'annexe A.

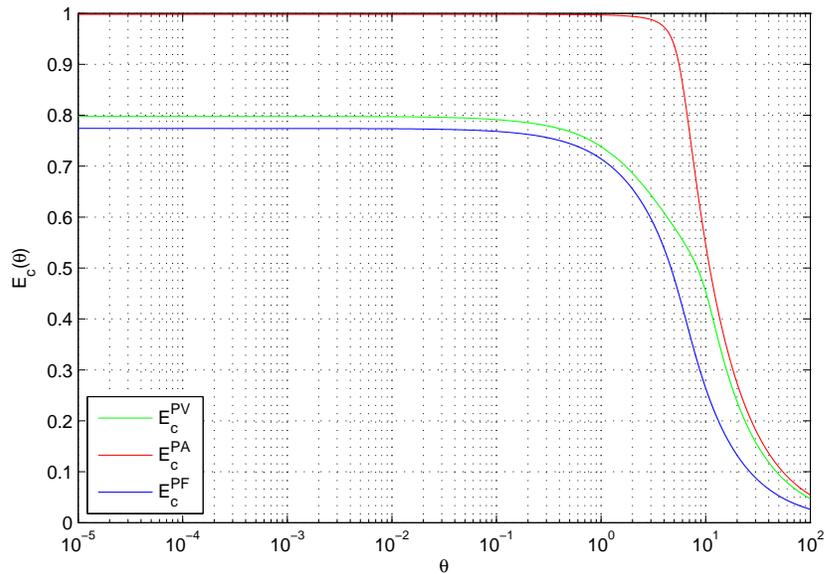


FIGURE 3.5 – La capacité effective pour l'adaptation de puissance (E_c^{PA}), allocation de puissance (E_c^{PV}) et puissance fixe (E_c^{PF}). Pour $\bar{\gamma} = 5 \text{ dB}$, $K = 3$ et $R = 1$.

Notre schéma d'adaptation de puissance présente la capacité effective optimale par

comparaison à celle de l'allocation de puissance et celle à puissance fixe. Le gain apporté par l'adaptation en terme de $E_c(\theta)$ est significatif pour les valeurs faibles à moyennes de θ .

Notre schéma optimal converge vers le débit utile pour {IR-HARQ, $K = 3$ et adaptation de puissance} lorsque $\theta \rightarrow 0$, c.à.d. lorsqu'aucune contrainte sur le délai n'est imposée.

Si par contre, aucun retard n'est toléré c.à.d. $\theta \rightarrow \infty$, la capacité effective des trois schémas, adaptation de puissance, allocation de puissance et puissance fixe, tend vers 0. En effet, même si on utilise les puissances optimales, le mieux qu'on puisse faire lorsque $\theta \rightarrow \infty$ est une capacité effective nulle pour un canal avec évanouissement de Rayleigh. Peu importe la puissance allouée aux transmissions, tant qu'il n'y pas d'autres techniques d'optimisation, un canal avec un évanouissement de Rayleigh ne peut pas supporter une contrainte très stricte sur la QoS. Ceci s'explique par le fait qu'une capacité de coupure nulle n'est pas réalisable pour les canaux avec évanouissement de Rayleigh [38].

3.4 Conclusion

Dans l'objectif d'améliorer les performances des systèmes avec HARQ et CSIT, on a présenté dans ce chapitre une optimisation de deux métriques de performance, la probabilité de coupure et la capacité effective, basée sur l'adaptation de puissance. Grâce à une récursion PD, on a pu déterminer les valeurs P_1, P_2, \dots, P_K optimales pour minimiser la probabilité de coupure puis pour maximiser la capacité effective. Le schéma IR-HARQ avec adaptation de puissance présente toujours les meilleurs performances par comparaison à l'allocation de puissance et puissance fixe. Le gain est d'autant plus appréciable lorsqu'un nombre de transmissions $K \geq 3$ est implémenté ou lorsqu'une contrainte faible à moyenne est imposée sur le délai.

Conclusion générale

Le travail présenté dans ce mémoire porte sur l'adaptation des paramètres de l'émetteur quand le schéma Incremental Redundancy HARQ tronqué est utilisé. L'étude a été effectuée pour les canaux indépendants et avec évanouissement de Rayleigh. Une estimation de l'état du canal et/ou de l'information mutuelle cumulée au niveau du récepteur sont disponibles au niveau de l'émetteur. En exploitant ce CSI retardé, on a pu proposer des schémas d'adaptations du IR-HARQ.

Dans le premier chapitre, on a présenté un problème de maximisation du débit utile basé sur l'adaptation des taux de transmission. Grâce à la formulation du problème avec la programmation dynamique, on a pu passer du problème de maximisation à K variables à un problème de K maximisations à une variable chacune. Ainsi, on a pu déterminer les taux optimaux des transmissions. On a remarqué l'adoption de la stratégie de l'abandon du paquet dans les solutions optimales. Finalement, le schéma HARQ avec adaptation des taux introduit une amélioration du débit utile par comparaison au système HARQ avec allocation de taux ou à taux fixe.

Dans le second chapitre, on a étendu cette méthode d'adaptation des taux aux puissances des retransmissions des schémas IR-HARQ tronqués. On a ainsi présenté le problème de minimisation de la probabilité de coupure avec une contrainte sur la puissance moyenne et on a déterminé les puissances optimales des transmissions. On a remarqué que la stratégie de l'abandon du paquet n'est adoptée que pour les faibles SNR. Pour des grands SNR, les puissances optimales sont continues en fonction de l'estimation du canal et de l'information mutuelle cumulée. Cependant, le gain en terme de probabilité de coupure par rapport à l'allocation de puissance n'est pas significatif pour $K = 2$. Le gain commence à être appréciable pour $K = 3$.

Finalement, on a introduit la notion de capacité effective dans le dernier chapitre. Cette métrique est pertinente quand on s'intéresse à des services sensibles aux délais. On a présenté au début la dérivation et la détermination de la formule exacte de la capacité effective pour

les schémas HARQ. On a ensuite exposé le problème d'optimisation de cette métrique sur la base de l'adaptation des puissances. La capacité effective optimale est comparée à celle avec allocation de puissance et puissance fixe. On a également présenté et résolu le problème d'allocation de puissance pour $K = 3$ qui n'a pas été traité auparavant dans la littérature. Le gain en E_c apporté par notre schéma optimal est significatif par rapport aux autres schémas pour une contrainte faible à moyenne sur θ . Cependant, il est à noter que la capacité effective optimale tend vers zéro lorsqu'une contrainte très stricte sur le délai est imposée. En effet, une capacité de coupure nulle n'est pas réalisable pour les canaux avec évanouissement de Rayleigh. C'est le mieux qu'on puisse faire quand aucune autre technique d'optimisation n'est implémentée pour contrecarrer l'effet de l'évanouissement.

Ces travaux de mémoire consacrés à l'adaptation des paramètres de l'émetteur ont suscité plusieurs questions qui méritent d'être traitées dans des travaux de recherche futurs. Pour commencer, les optimisations développées dans ce mémoire pourraient être étendues à des systèmes d'étude plus généraux, principalement, étendre l'étude de la capacité effective à d'autres type d'évanouissement (Nakagami avec paramètre m) pour pouvoir estimer le gain, s'il existe, entre l'adaptation de puissance, l'allocation de puissance et la puissance fixe pour une contrainte stricte sur le délai. Cette extension des types d'évanouissement serait également intéressante lors de l'étude de la probabilité de coupure. Finalement, une adaptation jointe taux et puissance peut être également traitée afin de maximiser le débit utile ou la capacité effective et minimiser la probabilité de coupure.

Annexe A

On présente, dans cet annexe, la dérivation de la capacité effective pour le cas d'allocation de puissance et IR-HARQ tronqué. Cette analyse est effectuée pour $K = 3$ et utilisée afin de comparer la capacité effective résultante avec celle dérivée dans le troisième chapitre avec adaptation de puissance.

L'idée est de maximiser l'expression de la capacité effective donnée par (3.29) dans un schéma de communication sans CSIT. Le problème de maximisation de E_c (3.30) ne change pas. Seules les expressions de la probabilité de décodage incorrecte après k transmissions, notées f_k^{PV} , diffèrent de f_k . Pour un f_1^{PV} fixé, maximiser E_c correspond à minimiser $f_{B \rightarrow B}$. On obtient alors, pour $K = 3$, le problème de maximisation suivant :

$$\max_{P_2, P_3} \frac{f_2^{\text{PV}}}{f_3^{\text{PV}}} \quad (3.33)$$

sujet à la contrainte sur la puissance moyenne à long terme :

$$P_{\text{avg}} = \frac{P_1 + P_2 f_1^{\text{PV}} + P_3 f_2^{\text{PV}}}{1 + f_1^{\text{PV}} + f_2^{\text{PV}}} \leq \bar{P} = 1 \quad (3.34)$$

En utilisant l'approximation «Saddle Point» [39], f_k^{PV} pour $k = 2, 3$ peut s'écrire comme :

$$f_k^{\text{PV}} \approx \hat{f}_k^{\text{PV}} = \frac{v(k)}{P_k} f_{k-1}^{\text{PV}} \quad (3.35)$$

où,

$$v(k) = \frac{1}{\gamma} \left\{ 2^R \sum_{i=1}^k \frac{(-1)^{k-1}}{(i-1)!} [R \log(2)]^{i-1} + (-1)^k \right\} \quad (3.36)$$

En utilisant (3.35), (3.33) peut s'écrire comme :

$$\max_{P_3} \frac{P_3}{v(3)} \quad (3.37)$$

Comme on doit maximiser P_3 sujet à la contrainte (3.34), P_3 doit vérifier :

$$P_3 \leq \frac{1 + f_1^{\text{PV}} + f_2^{\text{PV}} - P_1 - P_2 f_1^{\text{PV}}}{f_2^{\text{PV}}} \quad (3.38)$$

Finalement, pour un f_1^{PV} donné (et donc un P_1 donné) on doit retrouver la valeur de P_2 qui maximise :

$$P_2 = \arg \max \frac{1 + f_1^{\text{PV}} + f_2^{\text{PV}} - P_1 - P_2 f_1^{\text{PV}}}{f_2^{\text{PV}}} \quad (3.39)$$

On résout (3.39) analytiquement en utilisant l'approximation Saddle Point (3.35) pour f_2^{PV} .

Finalement, la valeur de P_3 optimale est donnée par :

$$P_3 = \frac{1 + f_1^{\text{PV}} + f_2^{\text{PV}} - P_1 - P_2 f_1^{\text{PV}}}{f_2^{\text{PV}}} \quad (3.40)$$

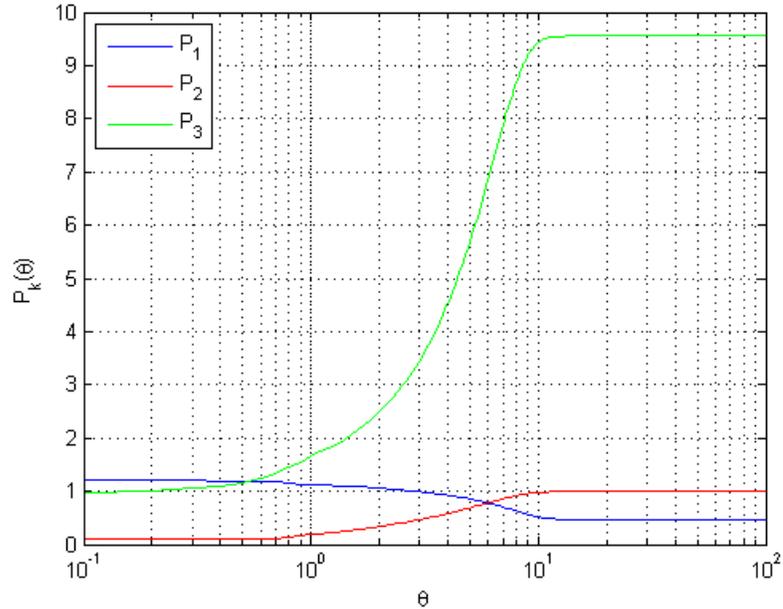


FIGURE 3.6 – P_1 , P_2 et P_3 optimales pour les schémas IR-HARQ avec allocation de puissance, $\bar{\gamma} = 5$ dB, $K = 3$ et $R = 1$.

La figure (3.6) représente les valeurs optimales des puissances de transmission pour les schémas IR- HARQ avec allocation de puissance et SNR moyen égal à 5dB. On remarque

que la puissance de la première transmission $k = 1$ diminue en fonction de θ ce qui implique une augmentation de f_1 et donc une diminution de la probabilité de décodage correct après la première transmission pour les grandes valeurs de θ .

Pour les retransmissions, $k = 2$ et $K = 3$, la tendance s'inverse. En effet, l'échec du décodage de la première transmission engendre des délais supplémentaires. Ainsi, la puissance augmente considérablement pour les grandes valeurs de θ au cours des retransmissions pour augmenter la probabilité d'un décodage correct et arrêter les retransmissions.

Notons de plus que les valeurs de P_3 sont supérieures à celles de P_2 , ce qui implique une valeur de f_3 beaucoup plus faible que f_2 . Ce résultat permet alors de maximiser le rapport $\frac{f_2}{f_3}$ et ainsi minimiser $f_{B \rightarrow B}$.

Même pour des valeurs très grandes de P_3 , la condition sur la puissance moyenne est vérifiée. Ceci s'explique par la diminution de la probabilité f_k^{PV} au cours des retransmissions. Ainsi, une puissance importante peut être allouée à P_3 tout en vérifiant la contrainte sur la puissance moyenne à long terme car l'occurrence d'une troisième transmission est faible par comparaison à l'occurrence d'une deuxième transmission.

Bibliographie

- [1] D. TSE et P. VISWANATH : *Fundamentals of Wireless Communication*. Cambridge University Press, 1 édition, 2005.
- [2] J. TANG et X. ZHANG : Quality-of-service driven power and rate adaptation over wireless links. volume 6, pages 3058 –3068, août 2007.
- [3] S. SESIA, I. TOUFIK et M. BAKER : *LTE the UMTS Long Term Evolution From Theory to Practice*. John Wiley and Sons, 2009.
- [4] T.V.K. CHAITANYA et E.G. LARSSON : Outage-optimal power allocation for hybrid ARQ with incremental redundancy. volume 10, pages 2069 –2074, jul 2011.
- [5] E. VISOTSKY, V. TRIPATHI et M. HONIG : Optimum ARQ design : a dynamic programming approach. *In Proc. IEEE International Symposium on Information Theory*, page 451, juin 2003.
- [6] E. UHLEMANN, L. K. RASMUSSEN, A. GRANT et P. WIBERG : Optimal incremental-redundancy strategy for type-II hybrid ARQ. *In Proc. IEEE International Symposium on Information Theory*, page 448, 2003.
- [7] J.-F. CHENG, Y.-P. WANG et S. PARKVALL : Adaptive incremental redundancy. *In IEEE Veh. Tech. Conf.*, pages 737–741, Orlando, Florida, USA, octobre 2003.
- [8] E. VISOTSKY, Y. SUN, V. TRIPATHI, M.L. HONIG et R. PETERSON : Reliability-based incremental redundancy with convolutional codes. volume 53, pages 987 – 997, juin 2005.
- [9] S. PFLETSCHINGER et M. NAVARRO : Adaptive HARQ for imperfect channel knowledge. *In 2010 International ITG Conference on Source and Channel Coding (SCC)*, pages 1–6, 2010.
- [10] D. TUNINETTI : On the benefits of partial channel state information for repetition protocols in block fading channels. volume 57, pages 5036 –5053, août 2011.

- [11] D.V. DJONIN, A.K. KARMOKAR et V.K. BHARGAVA : Joint rate and power adaptation for type-I hybrid ARQ systems over correlated fading channels under different buffer-cost constraints. volume 57, pages 421–435, janvier 2008.
- [12] D. P. BERTSEKAS : *Dynamic Programming and Optimal Control*, volume 1. Athena Scientific, 3 édition, 2005.
- [13] R. NEGI et J.M. CIOFFI : Delay-constrained capacity with causal feedback. volume 48, pages 2478 – 2494, septembre 2002.
- [14] J. H. BELL : Printing telegraph systems. volume XXXIX, pages 167–230, 1920.
- [15] J. MOORE : Constant-ratio code and automatic-rq on transoceanic hf radio services. volume 8, pages 72–75, 1960.
- [16] F. ADACHI, S. ITO et K. OHNO : Performance analysis of a time diversity ARQ in land mobile radio. volume 37, pages 177–183, 1989.
- [17] S. LIN et D. J. COSTELLO : *Error Control Coding : Fundamentals and Applications*. Englewood Cliffs, 1983.
- [18] D. CHASE : Code combining- a maximum-likelihood decoding approach for combining an arbitrary number of noisy packets. volume 33, pages 385–393, may 1985.
- [19] J. HAGENAUER : Rate-compatible punctured convolutional codes (rcpc codes) and their applications. volume 36, pages 389–400, apr 1988.
- [20] K. JAEHONG, H. WOONHAING, A. RAMAMOORTHY et S.W. MCLAUGHLIN : Design of rate-compatible irregular ldpc codes for incremental redundancy hybrid ARQ systems. *In Information Theory, 2006 IEEE International Symposium on*, pages 1139–1143, 2006.
- [21] G. I. DAVIDA et S. M. REDDY : *Forward-error correction with decision feedback*. Information and Control, September 1972.
- [22] A. LE DUC : *Performance closed-form derivations and analysis of Hybrid-ARQ retransmission schemes in a cross-layer context*. Ph.D. dissertation Télécom ParisTech Paris France, 2009.
- [23] C.J. LE MARTRET, A. LE DUC, S. MARCILLE et P. CIBLAT : Analytical performance derivation of hybrid ARQ schemes at IP layer. volume 60, pages 1305–1314, 2012.
- [24] G. CAIRE et D. TUNINETTI : The throughput of hybrid-ARQ protocols for the Gaussian collision channel. volume 47, pages 1971–1988, juillet 2001.

- [25] B. DEVILLERS, J. LOUVEAUX et L. VANDENDORPE : Bit and power allocation for goodput optimization in coded parallel subchannels with ARQ. volume 56, pages 3652–3661, 2008.
- [26] W. DAPENG et R. NEGI : Effective capacity : a wireless link model for support of quality of service. volume 2, pages 630–643, 2003.
- [27] C. CHENG-SHANG : Stability, queue length, and delay of deterministic and stochastic queueing networks. volume 39, pages 913–931, 1994.
- [28] S.B. WICKER : *Error Control Systems for Digital Communications and Data Storage*. Englewood Cliffs, 1995.
- [29] S. KALLEL : Analysis of a type II hybrid ARQ scheme with code combining. volume 38, pages 1133–1137, 1990.
- [30] C. JUNG-FU : Coding performance of hybrid ARQ schemes. volume 54, pages 1017–1029, 2006.
- [31] D. TUNINETTI : Transmitter channel state information and repetition protocols in block fading channels. In *IEEE Information Theory Workshop, ITW '07*, pages 505–510, California, USA, septembre 2007.
- [32] Q. LIU, S. ZHOU et G. B. GIANNAKIS : Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links. volume 3, pages 1746–1755, septembre 2004.
- [33] X. WANG, Q. LIU et G.B. GIANNAKIS : Analyzing and optimizing adaptive modulation coding jointly with ARQ for QoS-guaranteed traffic. volume 56, pages 710–720, mars 2007.
- [34] L. SZCZECINSKI, S. KHOSRAVIRAD, P. DUHAMEL et M. RAHMAN : Rate allocation and adaptation for incremental redundancy truncated HARQ, 2013.
- [35] N. NAKAGAMI : The m -distribution, a general formula for intensity distribution of rapid fading. In W. G. HOFFMAN, éditeur : *Statistical Methods in Radio Wave Propagation*. Oxford, England : Pergamon, 1960.
- [36] L. SZCZECINSKI, C. CORREA et L. AHUMADA : Variable-rate transmission for incremental redundancy hybrid ARQ. In *IEEE Global Telecommunications Conference, GLOBECOM 2010*, décembre 2010.
- [37] A. RIBEIRO et G.B. GIANNAKIS : Separation principles in wireless networking. volume 56, pages 4488–4505, septembre 2010.

- [38] G. CAIRE, G. TARICCO et E. BIGLIERI : Optimum power control over fading channels. volume 45, pages 1468–1489, 1999.
- [39] M. JABI, L. SZCZECINSKI et M. BENJILLALI : Accurate outage approximation of MRC receivers in arbitrarily fading channels. volume 16, pages 789–792, 2012.