Resource allocation and performance metrics analysis in cooperative cellular networks

Mohamad Maaz

Thèse

THESE INSA Rennes
sous le sceau de l'Université européenne de Bretagne
pour obtenir le titre de
DOCTEUR DE L'INSA DE RENNES
Spécialité : Électronique et Télécommunications

présentée par
Mohamad MAAZ
ECOLE DOCTORALE : MATISSE
LABORATOIRE : IETR

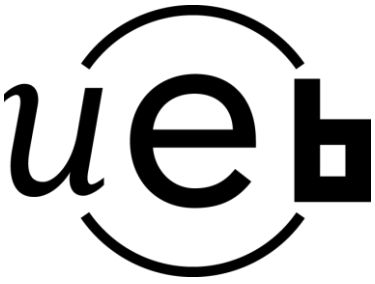
Allocation de ressource et analyse des critères de performance dans les réseaux cellulaires coopératifs

Thèse soutenue le 03.12.2013
devant le jury composé de :

Xavier LAGRANGE
Professeur à TELECOM Bretagne / *Président*
Daniel ROVIRAS
Professeur au CNAM Paris / *Rapporteur*
Charly POULLIAT
Professeur à l'INP - ENSEEIHT Toulouse / *Rapporteur*
Charlotte LANGLAIS
Maître de conférence à TELECOM Bretagne / *Examinatrice*
Philippe MARY
Maître de conférence à l'INSA de Rennes / *Co-encadrant*
Maryline HELARD
Professeur à l'INSA de Rennes / Directrice de thèse

Resource allocation and performance metrics analysis in cooperative cellular networks

Mohamad MAAZ



Avec le support de

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my advisor Philippe Mary for his fundamental role in my doctoral work. I would like to thank you for all I have learned from you. I really appreciate your guidance, assistance, and expertise that enabled me to accomplish the attributed task. I will forever be thankful to you for the constructive and fruitful discussions that leaded to this valuable work. Your attitude to research inspired me how to always look ahead and never give up to any scientific problem I could face. You guided me how to be a rigorous, serious and an independent thinker. I'm really a lucky person to have the opportunity to work with you!

In addition, I would like to gratefully and sincerely thank my thesis director, Professor Maryline Hélard. It was a great opportunity to do my doctoral program under your guidance and to learn from your research expertise. Thank you for the scientific advices and many insightful discussions and suggestions. Thank you for providing me an excellent atmosphere for doing research. Your impressive insight and technical experience had a great impact on me.

I would also like to thank my committee members, Maître de conférence Charlotte Langlais, Professor Daniel Roviras, Professor Charly Poulliat and Professor Xavier Lagrange for evaluating my doctoral work. I'm very grateful to you for the spent time and valuable feedback on this dissertation. Thank you for the brilliant comments and suggestions that inspired me a lot of ideas.

A special thank goes to the working team at the IETR laboratory. It is the most wonderful team I ever seen. I'm thankful to everyone of you. I would also like to thank Matthieu Crussière, Yvan Kokar, Ming Liu, Jean-Yves Baudais and Yasset Oliva.

Most importantly, I'm very grateful to my parent. You deserve as much thanks as anyone. I would like to thank you for the endless patience and encouragement. I'm deeply indebted for your love, caring and support throughout my entire life.

Last, but certainly not least, I must acknowledge with tremendous and deep thanks Alyona Karavaeva for her kindness and encouragement. You were always a great support in all my struggles and frustrations during my doctoral work. Your presence has been the most meaningful and wonderful thing in my life!

Table of contents

Acronyms

ACK	Acknowledgment
AF	Amplify-and-Forward
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
CC	Chase Combining
CF	Compress-and-Forward
CRC	Cyclic Redundancy Check
CSI	Channel State Information
dB	Decibel
DF	Decoded-and-Forward
FEC	Forward Error Correction
HARQ	Hybrid-Automatic Repeat Request
IP	Internet Protocol
KKT	Karush-Kuhn-Tucker
KPI	Key Performance Indicator
LDPC	Low-density Parity-Check
LLR	Log Likelihood Ratio
MAC	Medium Access Control
MIMO	Multiple-Input-Multiple-Output
MRC	Maximum Ratio Combing
MS	Mobile Station
MCS	Modulation and Coding schemes
NACK	Non-Acknowledgment
OFDMA	Orthogonal Frequency Division Multiple Access
PER	Packet Error Rate
pdf	Probability Density Function
QoS	Quality-of-Service
RS	Relay Station
SISO	Single-Input-Single-Output
SNR	Signal-to-Noise Ratio
STBC	Space-Time Block Code
TS	Time Slot
VAAs	Virtual Antenna Arrays
WSN	Wireless Sensor Networks
4G	4-th Generation

Notations

γ	Instantaneous signal-to-noise-ratio
$\overline{\gamma}$	Average signal-to-noise ratio
γ_{ij}	Instantaneous signal-to-noise-ratio in the $i-j$ link
$\overline{\gamma}_{ij}$	Average signal-to-noise-ratio in the $i-j$ link
h_{ij}	Channel fading in the $i-j$ link
κ	Path-loss exponent
$f(\gamma)$	Instantaneous packet error rate
P_{ij}	Average packet error rate in the $i-j$ link
$\Pr(.)$	Probability operator
$p_\gamma(\gamma)$	Probability density function of γ
$\boldsymbol{g} = [g_k]_{1\times K}$	A K length vector of element g_k
$\boldsymbol{G} = [g_{km}]_{K\times M}$	$K \times M$ matrix of element g_{km}
$[g^n_{km}]_{K\times M\times N}$	$K \times M \times N$ hyper-matrix with element g^n_{km}
N_F	Number of subcarriers
$\mathcal{D}_m$	Delay constraint of user m
$\overline{N}_t$	Average total delay
η_E	Energy efficiency
η_T	Throughput efficiency
$B(x, y)$	Beta function
$\lim$	Limit operator
$\max\{x, y\}$	The maximum between x and y
$\min\{x, y\}$	The minimum between x and y
$\mathbb{E}$	Mathematical expectation
$\mathbb{R}$	Set of real numbers
p_{ij}	Allocated power in the $i-j$ link
p^n_{ij}	Allocated power on the n-th subcarrier in the $i-j$ link
ρ_m	Throughput constraint of user m
B_w	Bandwidth
p_l	Path-loss
N_0	Noise spectral density
σ^2	Noise variance
E_b	Bit energy
d_r	Relay distance
θ_r	Relaying angle
L	Number of information bits in a packet
R_0	Code rate

$Q(.)$	Successful decoding probability
$N_{\max}$	Maximum Number of allowed retransmissions
R_b	Transmission bit rate
f_c	Carrier frequency
P_t	Transmission power
C	Channel capacity
$L(n)$	Number of transmitted bits till the instant n
R_0	Mother code rate
P_{tot}	Total power constraint
$\mathbb{P}(O)$	Outage probability
d_o	Diversity order
r_o	Multiplexing order
$\mathcal{L}$	The Lagrangian
$[.]^+$	$\max\{0, .\}$
d_g	Duality gap
P_{tx}	Transmitter circuitry power
P_{rx}	Receiver circuitry power
$\nabla(.)$	Gradient operator
L_{ack}	Number of bits in ACK packet
E_{tx}	Transmitter energy consumption
E_{rx}	Receiver energy consumption
E_{ack}	Energy consumption due to N/ACK packet
sup	Supremum
inf	Infinimum

List of figures

Résumé en Français

Dans les systèmes de communications sans fil, la transmission de grandes quantités d'information et à faible coût énergétique sont les deux principales questions qui n'ont jamais cessé d'attirer l'attention de la communauté scientifique au cours de la dernière décennie.

Récemment, il a été démontré que la communication coopérative est une technique intéressante notamment parce qu'elle permet d'exploiter la diversité spatiale dans les canaux sans fil. Cette technique assure une communication robuste et fiable, une meilleure qualité de service (QoS) et rend le concept de coopération prometteur pour la 4G, ainsi que pour les futures générations de systèmes cellulaires. Typiquement, les exigences de QoS sont le taux d'erreurs paquet, le débit et le délai. Ces métriques sont impactées par le délai, induit par les mécanismes de retransmission Hybrid-Automatic Repeat-Request (HARQ) inhérents à la réception d'un paquet erroné et qui induit un retard sur la qualité de service demandée. En revanche, les mécanismes de retransmission créent une diversité temporelle. Par conséquent, l'adoption conjointe de la communication coopérative et des protocoles HARQ pourrait s'avérer avantageuse pour la conception de schémas cross-layer.

Nous proposons tout d'abord une nouvelle stratégie de maximisation de débit total dans un réseau cellulaire hétérogène où les utilisateurs ont différentes contraintes de débit. Nous introduisons un algorithme qui alloue la puissance optimale à la station de base (BS) et aux relais, qui à chaque utilisateur attribue de manière optimale les sous-porteuses et les relais. Nous calculons le débit maximal atteignable ainsi que le taux d'utilisateurs sans ressources dans le réseau lorsque le nombre d'utilisateurs actifs varie. Nous comparons les performances de notre algorithme à ceux de la littérature existante, et montrons qu'un gain significatif est atteint sur la capacité globale. Cependant cette stratégie optimale en termes de débit, n'est généralement pas efficace en termes d'énergie.

Dans un second temps, nous analysons théoriquement le taux d'erreurs paquet, le délai de bout-en-bout ainsi que l'efficacité de débit des réseaux HARQ coopératifs, dans le canal à évanouissements par blocs. Dans le cas des canaux à évanouissement lents, le délai moyen (sur les évanouissements) du mécanisme HARQ n'est pas pertinent à cause de la non-ergodicité du processus. Ainsi, nous nous intéressons plutôt à la probabilité de coupure de délai moyen (sur un canal AWGN) en présence d'évanouissements lents. La probabilité de coupure de délai n'a jamais été traitée dans la littérature et est pourtant de première importance pour les applications sensibles au délai. Nous proposons une forme analytique

de la probabilité de coupure permettant de se passer de longues simulations numériques. De plus, nous considérons une taille de paquet finie ainsi qu'une modulation et un schéma de codage donné (MCS), ce qui rend ces analyses réalistes d'un point de vue pratique. Des simulations numériques viennent corroborer la validité de ces résultats analytiques.

Dans la suite de notre travail, nous faisons une analyse théorique de l'efficacité énergétique examinée en bits/joule dans les réseaux HARQ coopératifs. En se basant sur ces analyses, nous résolvons un problème de minimisation de l'énergie dans les réseaux cellulaires coopératifs en liaison descendante. Dans ce problème, chaque utilisateur possède une contrainte de délai moyen à satisfaire de telle sorte que la contrainte sur la puissance totale du système soit respectée. La BS est supposée n'avoir qu'une connaissance des moyennes statistiques du canal mais aucune information de l'état instantané du canal. L'algorithme de minimisation permet d'attribuer à chaque utilisateur la meilleure station-relai, sa puissance (si la coopération est décidée) ainsi que la puissance optimale de la BS afin de satisfaire la contrainte de délai de chaque utilisateur. Les simulations montrent qu'en termes de consommation d'énergie, les techniques assistées par relais surpassent nettement les transmissions directes, dans tout système limité en délai.

En conclusion, les travaux proposés dans cette thèse peuvent promettre d'établir des règles fiables pour l'ingénierie et la conception des futures générations de systèmes coopératifs cellulaires énergétiquement efficaces.

Abstract

In wireless systems, transmitting large amounts of information with low energetic cost are two main issues that have never stopped drawing the attention of the scientific community during the past decade.

Later, it has been shown that cooperative communication is an appealing technique that exploits spatial diversity in wireless channel. Therefore, this technique certainly promises robust and reliable communications, higher quality-of-service (QoS) and makes the cooperation concept attractive for 4G and beyond cellular systems. Typically, the QoS requirements are the packet error rate, throughput and delay. These metrics are affected by the delay, where each erroneous packet is retransmitted several times according to Hybrid-Automatic Repeat-Request (HARQ) mechanism inducing a delay on the demanded QoS but a temporal diversity is created. Therefore, adopting jointly cooperative communications and HARQ mechanisms could be beneficial for designing cross-layer schemes.

First, a new rate maximization strategy, under heterogeneous data rate constraints among users is proposed. It allocates the optimal power at the base station (BS) and relays, assigns subcarriers and selects relays. The achievable data rate is investigated as well as the average starvation rate in the network when the load, i.e. the number of active users in the network, increases. This algorithm is compared to the existing literature by means of extensive simulations. Although this resource allocation strategy is optimal in terms of throughput, it does not consider the energetic issues.

Second, in block fading channel, theoretical analyses of the packet error rate, end-to-end delay and throughput efficiency in relay-assisted HARQ networks are provided. In slow fading channels, the average delay of HARQ mechanisms w.r.t. the fading states is not relevant due to the non-ergodic process of the fading channel. The delay outage is hence invoked to deal with the slow fading channel and is defined as the probability that the average delay w.r.t. AWGN channel exceeds a predefined threshold. This criterion has never been studied in literature, although being of importance for delay sensitive applications in slow fading channels. Then, an analytical form of the delay outage probability is proposed which might be very useful to avoid long simulations time. These analyses consider a finite packet length and a given modulation and coding scheme (MCS) which leads to study the performance of practical systems. The obtained analytical results are corroborated by means of numerical simulations.

Third, a theoretical analysis of the energy efficiency (bits/joule) in relay-assisted HARQ networks is provided. Based on this analysis, an energy minimization problem in multiuser relay-assisted downlink

cellular networks is investigated. Each user has an average delay constraint to be satisfied such that a total power constraint in the system is respected. The BS is assumed to have only knowledge about the average channel statistics but no instantaneous channel state information (CSI). Finally, an algorithm that jointly allocates the optimal power at BS, the relay stations and selects the optimal relay (if cooperation is decided) in order to satisfy the delay constrains of users is proposed. Simulations show the improvement in terms of energy consumption of relay-assisted techniques compared to non-aided transmission in delay-constrained systems. Hence, the work proposed in this thesis can give useful insights for engineering rules in the design of the next generation energy-efficient cellular systems.

Résumé long de la thése en français

Introduction générale

Dans les systèmes de communications sans fil, la transmission de grandes quantités d'information et à faible coût énergétique sont les deux principales questions qui n'ont jamais cessé d'attirer l'attention de la communauté scientifique au cours de la dernière décennie. Fondamentalement la qualité d'une communication radio est liée au canal de propagation. Le signal émis subit un évanouissement à petite échelle qui se caractérise par une fluctuation aléatoire qui affecte son amplitude et également un évanouissement à grande échelle qui se caractérise par la perte de trajet et le masquage. Ces évanouissements ont des impacts sévères sur la puissance du signal reçu. Ainsi, pour améliorer la qualité des signaux reçus, il est important de concevoir un système de communication fiable et robuste.

En 2000, les travaux menés par Dohler [1] et Laneman [2] ont montré que les communications coopératives sans fil exploitent la diversité spatiale et peuvent assurer une communication robuste et fiable, une extension de la couverture et une meilleure qualité de service (QoS) ce qui rend le concept de coopération prometteur pour la 4G, ainsi que pour les futures générations de systèmes cellulaires. Typiquement, les exigences de QoS sont le taux d'erreurs paquet, l'efficacité de débit et le délai de transmission. Ces métriques sont impactées par la retransmission des paquets erronés grâce aux mécanismes de retransmission Hybrid-Automatic Repeat-Request (HARQ). Néanmoins, ces mécanismes de retransmission permettent d'exploiter la diversité temporelle. De plus, cette diversité temporelle permet d'émettre des paquets d'information à base puissance. Grâce à ces avantages, l'adoption de la technique HARQ conjointement avec une communication coopérative pourraient s'avérer avantageuse pour la conception de schémas cross-layer. Toutes ces améliorations des systèmes imposent de caractériser théoriquement les performances des systèmes afin de gérer plus efficacement les algorithmes de ressources radio.

Dans les réseaux cellulaires, le nombre des utilisateurs mobiles ne cesse d'augmenter. De plus, l'adoption des techniques coopératives rend le problème d'allocation des ressources particulièrement difficile en raison du grand nombre de ressources qui peuvent être partagées dans le système, comme la puissance, les sous-porteuse, le temps et les relais. Il est donc important de proposer des nouvelles stratégies d'allocation des ressources qui exploitent la majorité des dégrés de liberté disponibles et assurer la QoS destinée à chaque utilisateur.

Depuis quelques années, la communauté scientifique s'est rendu compte qu'il y a un énorme gaspillage d'énergie dans les réseaux sans fil. Cela est dû à l'augmentation du nombre de noeuds qui peuvent participer à la transmission de données vers un même terminal mobile. De plus, la communication sur de courtes distances (en utilisant les multi-sauts) rend la consommation d'énergie des circuits (partie statique) du même ordre de grandeur que celle utilisée pour la transmission des données.

Ainsi, pour des raisons écologiques et économiques, plusieurs travaux comme le projet EARTH ont eu pour but de réduire la consommation globale d'énergie dans un réseau d'accès radio d'au moins 50 %. L'amélioration technologique des composants du réseau et l'utilisation des techniques de communications innovantes pourraient être des solutions pragmatiques pour cet objectif. Désormais, deux critères doivent être respectés pour proposer des stratégies d'optimisation des ressources :

- Satisfaire une QoS pour chaque utilisateur.
- Diminuer ou optimiser la consommation totale d'énergie dans l'ensemble du réseau.

Par conséquent, un compromis entre la QoS requise et la consommation d'énergie existe. En outre, les critères ci-dessus sont devenus essentiels pour les futures normes sans fil qui visent à concevoir des protocoles cross-layer qui optimisent les ressources globales disponibles tout en respectant des contraintes QoS pour chaque utilisateur. En particulier, la communication à relais est considérée comme une technique intéressante pour répondre à ces critères.

Chapitre 1 : Etat de l'art et outils mathématiques

Le chapitre présente l'état de l'art sur les communications coopératives ainsi que les outils d'optimisation nécessaires à la lecture de la thèse. Nous rappelons l'intérêt des communications coopératives et à relais notamment en ce qui concerne le gain offert sur l'atténuation en espace libre ainsi que le gain de diversité spatial. Nous introduisons également la capacité du canal avec relais en bits/s/Hz pour un canal à évanouissements. Cette capacité est utilisée dans les problèmes d'allocation des ressources. Ensuite, comme les techniques de type HARQ ont été adoptées dans plusieurs standards de communications, nous présentons les différents mécanismes de retransmissions existant dans l'état de l'art :

ARQ simple : Chaque paquet à transmettre est constitué des bits d'information et du CRC (Cyclic Redundancy Check). L'émetteur envoie un paquet d'informations et attend le message ACK (Acknowledgment) émis par le récepteur. Si le paquet d'information est décodé avec succès, un accusé de réception positif est envoyé par le récepteur et un autre paquet est transmis. Dans le cas contraire, un paquet NACK (Non-acknowledgment) est envoyé. Pour chaque NACK, l'émetteur retransmet le même paquet et le mécanisme de retransmission continue jusqu'à ce que le nombre maximal de retransmissions $N_{\max}$ soit atteint.

HARQ-I : Le principe de HARQ-I est similaire à celui de l'ARQ simple, mais chaque paquet d'information est encodé par un code correcteur d'erreur (FEC) avant d'être transmis. A la destination, le paquet est décodé indépendamment des précédents paquets erronés.

HARQ-CC : Dans cette technique, chaque paquet reçu erroné est combiné avec les précédents paquets par la technique 'Chase Combining' [3].

HARQ-IR : La technique HARQ-IR est considérée comme étant le plus sophistiqué des mécanismes HARQ parce qu'un schéma de codage spécifique est incorporé au mécanisme de retransmission. Ainsi, après l'encodage d'un paquet d'information, un code mère est engendré et divisé en plusieurs sous-mots de code de différent rendement. Pour chaque paquet reçu de manière erronée un autre sous-code est retransmis et concaténé à la réception avec les mots précédents puis décodé. Ainsi, à chaque retransmission, le rendement du code décroit et la probabilité de décodage s'améliore.

Les techniques HARQ offrent un gain de diversité temporelle supplémentaire et peuvent être utilisées conjointement avec les communications coopératives et donc permettre une amélioration de la performance du système de communication. Les métriques de performance étudiées dans la thèse sont :

PER (taux d'erreurs paquet) : Il est défini comme étant égal au rapport entre le nombre de paquets d'information reçus erronés et le nombre de paquets d'information transmis.

Efficacité du débit : Elle est définie comme le rapport entre le nombre de bits d'information reçus correctement et le nombre de bits transmis.

Délai : Il est défini comme le nombre moyen de retransmissions utilisées par paquet dans un certain environnement. Dans ces travaux, nous le mesurons en nombre de slots temporels.

Efficacité énergétique (bits/joule) : Elle est définie comme le rapport entre le nombre de bits d'informations reçus correctement et l'énergie moyenne totale consommée.

Afin d'étudier l'aspect énergétique des communications coopératives avec HARQ, nous introduisons un modèle de la consommation énergétique existant dans l'état de l'art et que nous utiliserons tout au long de la thèse. Ce modèle est constitué d'une partie statique qui est reliée à la consommation électrique des équipements utilisés et une partie dynamique qui considère la consommation due à la transmission dans le canal de propagation.

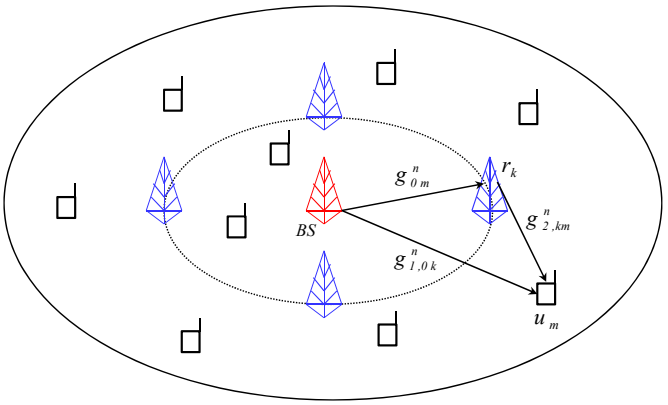


FIGURE 1 – Réseau OFDMA à relais

Chapitre 2 : Allocation de ressources dans les réseaux OFDMA cellulaires avec relayage

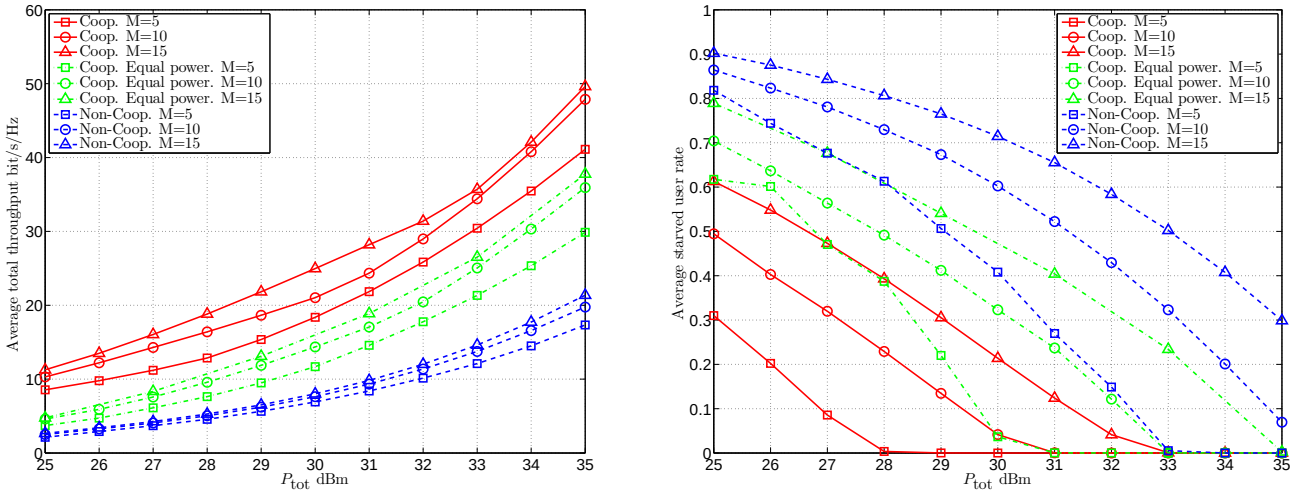
Dans ce chapitre, nous étudions un problème d'allocation des ressources dans les réseaux cellulaires avec la technique d'accès OFDMA et en utilisant des relais tel que décrit sur la figure 1. En utilisant la capacité du canal, nous proposons de maximiser le débit de transmission global dans la cellule sachant que chaque utilisateur à une contrainte de QoS en bits/s/Hz à satisfaire. Le réseau est contraint par une puissance totale de transmission ainsi qu'une bande passante limitée. L'objectif est alors d'allouer les puissances de manière optimale à la station de base (BS) et aux relais. Par ailleurs, selon l'état du canal, la technique d'optimisation permet d'attribuer à chaque utilisateur un certain nombre des sous-porteuses et de décider si la communication directe ou coopérative est préférable.

Nous montrons la convexité du problème d'optimisation et donc nous sommes assurés de trouver l'optimum global. Nous proposons ensuite un algorithme qui alloue les puissances optimales, et décide la communication directe avec la BS ou le relayage et qui attribue les sous-porteuses à chaque utilisateur.

Ensuite, nous comparons cet algorithme avec d'autres algorithmes proposés dans la littérature. L'algorithme proposé est plus performant par rapport à un algorithme qui utilise la communication directe et un algorithme qui utilise la communication coopérative sans faire l'allocation de puissance.

La figure 2a montre le débit total offert en fonction de la puissance de transmission P_{tot} totale dans un réseau cellulaire coopératif où les utilisateurs ont des contraintes de débit différentes, par exemple [1, 2, 2.5 , 3] bits/s/Hz. En considérant différent nombre des utilisateurs, comme par exemples M = 4, 8 et 12, nous remarquons que notre algorithme surpasse en efficacité spectrale ceux n'utilisent pas le relais et ceux ne faisant pas d'allocation de puissance.

De plus, la figure 2b, décrit le taux d'utilisateurs sans ressources en fonction de la puissance de transmission. L'algorithme proposé montre que le taux d'utilisateurs sans ressources est très bas par



(a) (b)

FIGURE 2 – Fig. 2a : L'efficacité spectrale moyenne en bits/s/Hz en fonction de la puissance total P_{tot} - Fig. 2b : Le taux d'utilisateurs moyen sans ressources en fonction de la puissance total P_{tot}

rapport à ces deux algorithmes.

Le résultat obtenu confirme que l'algorithme proposé est beaucoup plus performant par rapport à la littérature existante en termes de débit globale ainsi qu'au niveau du taux d'utilisateurs sans ressources.

Chapitre 3 : Analyse des métriques de performances de schémas HARQ

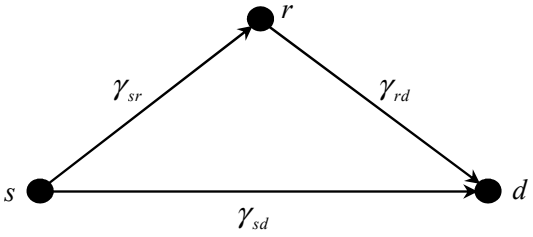


FIGURE 3 – Schéma de reliage

Le chapitre 3 est dédié à l'analyse des métriques de performance de schémas HARQ en communication directe et en communication coopérative avec le protocole 'decode-and-forward' (DF). En supposant que le schéma de transmission est composé d'une source s, un relai r et une destination d tels que représenté en figure 3 et en canal de Rayleigh nous définissons le protocole de retransmission suivant :

Protocole : La source s commence la transmission des paquets à la destination d. Si le paquet a été décodé avec succès par d, la destination envoie un message ACK à s et r et donc s commence à

transmettre un autre paquet. Sinon, d transmet un message NACK et s commence à retransmettre le même paquet jusqu'à ce que d ou r le décode. Ainsi, si r reçoit le paquet mais que d ne l'a pas encore reçu, le relai envoie un ACK à s et il commence à retransmettre le même paquet jusqu'à ce que d le reçoit ou que le nombre maximal de transmissions $N_{\max}$ soit atteint. Dans ce cadre, nous nous intéressons aux métriques de performance suivantes :

- PER
- Efficacité du débit
- Délai

Afin d'aborder ces analyses théoriques, nous commençons tout d'abord par l'analyse générale de ces métriques pour quelques schémas HARQ en fournissant des formes générales du PER, du délai et de l'efficacité de débit. Comme ces expressions ne dépendent que de la probabilité d'erreur à chaque instant de transmission, nous donnons une formule générale de la probabilité d'erreur à chaque instant pour les protocoles HARQ-I et HARQ-CC et aussi pour un type de modulation et un codage donné (MCS). De ce fait, le taux d'erreurs paquet, l'efficacité du débit ainsi que le délai s'en déduisent.

Ensuite, nous comparons les résultats théoriques obtenus avec des simulations et nous trouvons que notre analyse théorique permet de mesurer de manière précise les métriques de performance mentionnées ci-dessus.

Pour conclure, la technique de retransmission HARQ-CC avec relayage permet d'avoir les meilleurs PER, efficacité de débit et délai parmi tous les schémas HARQ étudiés. En effet, cette technique utilise conjointement la combinaison des paquets erronés et ainsi que le relayage et ainsi une diversité temporelle est exploitable. Un autre critère de choix lorsque l'on s'intéresse aux techniques HARQ et le délai induit par les retransmissions. Cependant lorsque le canal varie lentement, le délai moyen n'est plus très pertinent et la probabilité de coupure lui est préférée. Ainsi, pour des canaux à évanouissements lents, le canal peut être considéré comme gaussien pour une réalisation, le délai moyen est calculé dans ce cadre, et la probabilité de coupure de délai est calculée sous forme exacte en considérant la statistique d'évanouissement du canal. Nous montrons que les expressions théoriques trouvées constituent de très bonnes approximations de la probabilité de coupure évaluée en simulations.

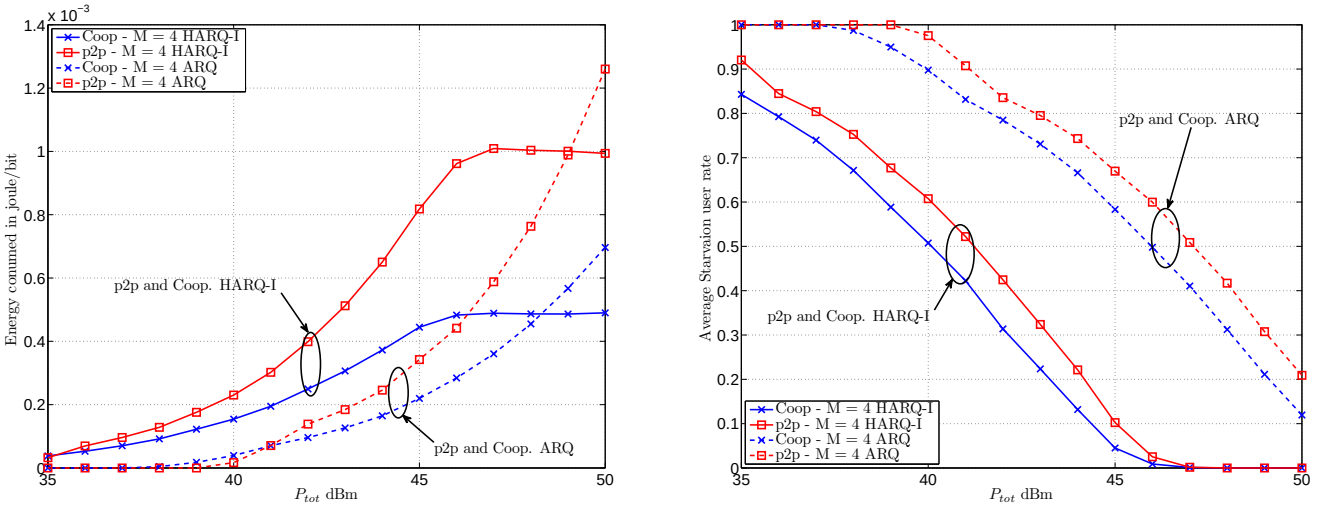
Chapitre 4 : Minimisation de la consommation énergétique dans les réseaux cellulaires

Ce chapitre considère les aspects énergétiques dans les réseaux cellulaires et il se compose de deux parties :

1. Nous analysons tout d'abord théoriquement l'efficacité énergétique des schémas HARQ-I et HARQ-CC en considérant les communications directes et les communications coopératives. Nous trouvons que la technique de retransmissions HARQ-CC avec relayage est énergétiquement efficace car elle

exploite la diversité temporelle ainsi que le gain sur l'affaiblissement en espace libre et donc elle conduit à une communication robuste avec une faible consommation énergétique.

2. Nous abordons ensuite un problème de minimisation de l'énergie dans les réseaux cellulaires coopératifs en liaison descendante avec multiutilisateurs. Le système est supposé avoir adopté la technique HARQ-I et chaque utilisateur a une contrainte de délai moyen à satisfaire sachant que la contrainte de puissance totale dans le réseau doit aussi être respectée. La BS est supposée avoir seulement une connaissance des statistiques moyennes du canal mais pas l'état du canal instantané (CSI). A première vue, le problème d'optimisation semble difficile à résoudre en raison de l'existence des termes exponentiels et polynômiaux dans la fonction objectif et dans les contraintes. La contribution de cette partie réside en trois points : i) Nous résolvons ce problème en introduisant de nouvelles contraintes et de nouvelles variables afin de le résoudre efficacement par le problème dual, ii) nous montrons que le problème de minimisation de l'énergie peut être facilement exprimé en fonction de contraintes quasi-convexes et une fonction objectif affine ce qui conduit à une solution optimale. Ensuite, nous montrons que la consommation d'énergie du système est minimum pour une contrainte de puissance totale élevée. iii) Enfin, nous proposons un algorithme qui alloue conjointement la puissance optimale à la BS et aux relais et qui sélectionne le relai le plus adapté (si la coopération est décidée) à chaque utilisateur afin de satisfaire la contrainte de délai. Les simulations effectuées montrent l'amélioration en termes de consommation énergétique des techniques coopératives par rapport à la transmission directe où les utilisateurs ont des contraintes de délais.



(a) (b)

FIGURE 4 – Fig. 4a : L'énergie totale consommée en joule/bit en fonction de la puissance totale P_{tot} - Fig. 4b : Le taux d'utilisateurs sans ressouces en fonction de la puissance totale P_{tot}. Le nombre d'utilisateurs M = 4 et les utilisateurs ont des contraintes de délai $\overline{\mathcal{D}}_m$ = [1.5 1.5 2.5 2.5] TS (Time slot)

En utilisant notre algorithme de minimisation de l'énergie, les figures 4a et 4b représentent l'énergie moyenne consommée en joule/bit dans un réseau cellulaire qui contient $M = 4$ utilisateurs en fonction de la contrainte de puissance totale disponible P_{tot} en dBm. Les utilisateurs ont des contraintes de délai comme $\overline{\mathcal{D}}_m$ =[1.5 1.5 2.5 2.5] slots. Nous comparons alors les schémas de communications directes et coopératives qui utilisent les protocoles de retransmissions ARQ simple et HARQ-I. Chaque paquet contient 102 bits d'information et le taux de codage adopté est 1/2 avec un type de modulation BPSK. On peut montrer que quand P_{tot} disponible augmente, la consommation d'énergie augmente aussi en raison de la diminution du taux moyen d'utilisateurs sans ressources. À près de $P_{tot} = 46$ dBm, le taux moyen d'utilisateurs sans ressources s'annule pour les réseaux de communications directes et coopératives qui adoptent le protocole de retransmission HARQ-I. Cependant, l'énergie moyenne consommée devient stable et atteint $0,44 \times 10^{-3}$ et 10^{-3} joule/bit pour respectivement les communications coopérative et directe. Ainsi, le réseau coopératif permet d'économiser de plus de 50 % d'énergie par rapport à la communication directe. Toutefois, en adoptant le protocole ARQ simple, la consommation d'énergie des deux systèmes est très élevée, car à $P_{tot} = 50$ dBm, la consommation moyenne d'énergie est environ $0,7 \times 10^{-3}$ et $1,25 \times 10^{-3}$ joule/bit pour respectivement les schémas coopératif et direct avec un taux d'utilisateurs sans ressources à 0,2 et 0,12. Ce grand taux d'utilisateurs sans ressources montre qu'il n'y a pas suffisamment de puissance pour satisfaire tous les utilisateurs.

Par conséquent, en comparant le protocole ARQ avec HARQ-I, nous pouvons affirmer que la consommation d'énergie et ainsi le taux d'utilisateurs sans ressources est beaucoup plus élevé. En outre, avec un schéma HARQ-I, moins de puissance est utilisée pour satisfaire tous les utilisateurs. Cela est dû au codage canal utilisé par HARQ-I qui augmente la probabilité de recevoir correctement le paquet à même SNR.

Conclusion

Dans cette thèse, nous avons proposé des algorithmes d'allocation des ressources efficaces dans les réseaux coopératifs cellulaires avec multiutilisateurs. Nous avons aussi fourni des analyses théoriques des métriques de performance pour les schémas cross-layer. Le travail présenté dans cette thèse est compatible avec les demandes de systèmes de télécommunications du futur. La contribution de cette thèse repose sur deux aspects principaux : i) Satisfaire aux exigences de qualité de service pour tous les utilisateurs mobiles avec ii) Une utilisation optimale de la puissance et une consommation énergétique minimale. Nous avons fourni une analyse théorique des métriques de performance des systèmes de retransmissions (HARQ) pour les communications directes et coopératives tels que le PER, l'efficacité de débit, le délai de bout-en-bout et l'efficacité énergétique. De plus, nous avons proposé une expression analytique de la probabilité de coupure pour le canal à évanouissement lent avec relais. Nous avons également proposé deux algorithmes d'allocation des ressources dans les réseaux cellulaires coopératif qui garantissent la qualité de service,

abordant deux aspects i.e. la maximisation du débit global et la minimisation de l'énergie totale.

Chapitre 1 : Tout au long de ce chapitre, nous avons introduit la communication coopérative et nous avons aussi donné un bref aperçu sur les techniques de retransmission HARQ qui sont incorporées dans les protocoles cross-layer. Nous avons également fourni les métriques de performance les plus importantes pour les techniques de communications avec les protocoles HARQ. Nous avons ensuite introduit le modèle de consommation énergétique adopté dans ce travail.

Chapitre 2 : Le chapitre 2 s'est focalisé sur un problème d'optimisation des ressources qui a pour but de maximiser le débit total dans un réseau cellulaire coopératif OFDMA. La principale nouveauté de ce chapitre, est qu'un nouvel algorithme d'allocation de ressource a été proposé. Il optimise conjointement la puissance à allouer, attribue des sous-porteuses et sélectionne un relais pour chaque utilisateur sachant que la qualité de service est assurée. Par ailleurs, nous montrons que cet algorithme est plus performant que ceux de la littérature existante en termes de débit global et aussi par rapport au taux d'utilisateurs sans ressources.

Chapitre 3 : Dans ce chapitre, nous nous sommes intéressés à l'étude des principaux critères de performances des techniques de retransmission HARQ-I et HARQ-CC, i.e. le PER, le délai moyen, l'efficacité de débit, l'efficacité énergétique et aussi la probabilité de coupure de délai. Nous avons également fourni une analyse théorique de la probabilité de coupure de délai pour les systèmes coopératifs avec HARQ-I dans des canaux à évanouissement lent, ce qui n'a jamais été fait dans la littérature. Par conséquent, ces analyses peuvent être utilisées pour évaluer de manière précise et rapide les métriques de performance des protocoles cross-layer dans les systèmes de communication avec relais.

Chapitre 4 : La contribution du chapitre 4 repose sur l'analyse des aspects énergétiques des protocoles HARQ dans les réseaux de communication coopératifs. Nous avons tout d'abord analysé théoriquement l'efficacité énergétique pour différents protocoles de retransmission HARQ. Ensuite, un problème de minimisation énergétique dans un réseau coopératif qui adopte la technique HARQ est étudié. Ce problème tient compte des aspects énergétiques statiques et dynamiques dans la fonction de coût et la puissance totale dans la fonction de contrainte sachant que la qualité de service désignée par le délai doit être garanti pour tous les utilisateurs. La contribution consiste en la proposition d'un algorithme d'allocation de ressource minimisant le cout énergétique sous contrainte de délai minimal pour les utilisateurs tout en considérant un protocole HARQ et une sélection optimale de relais.

Perspectives

Il y a encore plusieurs problèmes ouverts qui pourraient être abordés dans les travaux futurs :

- L'analyse théorique des protocoles HARQ suppose que les paquets ACK/NACK soient décodés sans erreurs avec un retard négligeable. Mais en réalité, il est possible d'avoir un faux message ACK. Par conséquent, il serait important d'examiner l'impact de canaux de retour non fiable.
- L'analyse théorique des métriques de performance du système HARQ avec redondance incrémentale (HARQ-IR) n'est pas étudiée dans cette thèse. L'étude de la probabilité d'erreur à chaque instant de transmission est une tâche difficile dans ce cas en raison de la grande complexité de la relation entre le SNR et la probabilité d'erreur dans les protocoles de type HARQ-IR. La levée de ce verrou ouvrirait la voie à l'étude du PER, le délai moyen, le débit et l'efficacité énergétique.
- Dans notre problème de minimisation de l'énergie, nous avons considéré un modèle de consommation énergétique linéaire. Toutefois, cette linéarité n'est plus garantie si on tient compte de la consommation de la partie de traitement numérique du signal. Par conséquent, un modèle de consommation énergétique plus précis dans les problèmes d'optimisation pourrait être étudié.
- Les algorithmes d'allocation des ressources proposés pourraient être mis en oeuvre sur une plateforme expérimentale pour évaluer l'intérêt de l'analyse théorique en conditions réelles.
- Dans les réseaux cellulaires, l'augmentation rapide du nombre d'utilisateurs par cellule et les déploiements de relais peuvent induire des interférences dans les cellules adjacentes. Par conséquent proposer des algorithmes efficaces pour réduire l'interférence dans les schémas de retransmission HARQ pourrait être un sujet de recherche future.

Publications

[1]- M. Maaz, P. Mary, and M. Helard, *"Delay Outage Probability in Block Fading Channel and Relay-Assisted Hybrid-ARQ Network,"* accepted to IEEE wireless communications letters, 2013, WCL2013-0686. (Early access DOI 10.1109/WCL.2013.113013.130686)

[2]- M. Maaz, P. Mary, and M. Helard, *" Energy Efficiency Analysis in Relay Assisted Hybrid-ARQ Communications,"* in PIMRC. IEEE, 2012, pp.2263-2268.

[3]- M. Maaz, P. Mary, and M. Helard, *" Resource Allocation for QoS Aware Relay-Assisted OFDMA Cellular Networks,"* 2011 Third International Workshop on Cross Layer Design (IWCLD), 2011, pp. 1-6.

[4]- M. Maaz, P. Mary, and M. Helard, *"Energy Minimization in HARQ-I Relay-Assisted Networks with Delay-limited Users,"*. (en cours de préparation)

Introduction

Context

In wireless communications, high data rate requirement with low energy cost is the main issue that has never stopped drawing the attention of the scientific community during the past decade. Fundamentally, wireless signal experiences small scale fading which is characterized by a random fluctuation of its amplitude or even large scale fading which is induced by the path-loss and shadowing and have a severe impact on the average signal power. Therefore, to combat channel fading, the design of reliable and efficient cross-layer schemes is an essential need.

Recently, it has been shown that cooperative communication is an appealing technique that exploits spatial diversity in wireless channel. This technique consists of a deployment of a group of intermediate relays between two communicating nodes that may assist during data transmission or a group of cooperating transmitting and receiving nodes. Therefore, it enables reliable communications, extends cell coverage and hence can be considered as an attractive concept for 4G and beyond cellular systems. On the other hand, the Hybrid-Automatic Repeat-Request (HARQ) mechanism is more and more envisaged for delay tolerant applications and allows to increase the system diversity by retransmitting each erroneous packet until it is correctly received. It also permits to communicate at low transmission power. Therefore, adopting jointly HARQ protocols with cooperative communications can enhance the efficiency of future cellular systems.

Typically, the performance of cross-layer scheme is measured in terms of the following metrics: the packet error rate (PER), average throughput and delay. Therefore, providing a theoretical analysis of these metrics gives a lower or upper bound on the demanded Quality-of-Service (QoS) and allows to save simulation time. Moreover, it allows to design efficient resource allocation policies that consider jointly the physical layer resources and the HARQ mechanisms.

In modern multiuser systems, the major problem of importance is how to optimally allocate the available resources between users. This is due to the large number of resources which can be shared in the system, e.g. power, bandwidth, relays, etc. Indeed, the traditional resource allocation policies focusing on the global throughput maximization, are generally inefficient in terms of energy consumption. These policies give the priority to users with good channel conditions to access the radio network. Hence, this

leads us to ask a question: Do these policies are able to support an exponential traffic growth volume? Therefore, it would be better to guarantee the average data rate requirement for every user rather than maximizing the global network data rate.

For ecologic and economic reasons, the emphasis of future wireless standards' efforts is to reduce the energy wastage on the core network to at least 50 % and also up to 90 % by 2020. This is the ambition of the European project EARTH [1] and the European research initiative GreenTouch. This waste is induced from the energy consumption of network equipments and also from the abuse of the available radio resources. Moreover, the deployment of multihop wireless technology in future cellular networks enables to communicate over short distances and hence leads the equipments' energy consumption to have the same order as that consumed for data transmission. Thereby, considering all these sources of energy consumption during the scheduling process and incorporating new communication technologies such as the HARQ with cooperative techniques could be a viable solution for this task.

To meet the cross-layer design criteria, it is of great interest to investigate and propose effective network management algorithms that reduce energy waste and improve energy efficiency of mobile cellular systems, without impacting the demanded QoS neither the cell traffic load.

Aim and organization

This dissertation aims to design resource optimization policies in muli-user cooperative networks. In order to optimally allocate resources, we study two types of resource optimization problems: 1) The sum rate maximization and 2) the aggregate energy minimization in multi-user cooperative network. Such policies have to guarantee the QoS and insure an optimal resource assignments between users according to a prevailing channel sate information (CSI) or channel statistics. In literature, we find a complete state of the art about the HARQ protocols but the works tackling their performance analytically are rather few. Hence, we provide a theoretical analysis of several performance metrics of HARQ schemes in fading channels, such as the PER, average throughput, end-to-end delay and energy efficiency.

This dissertation is organized as follows: chapter 1 provides a state of the art on cooperative techniques and gives a mathematical background for solving optimization problems. Chapter 2 tackles a rate maximization problem in Orthogonal Frequency Division Multiple Access (OFDMA) multi-user relay-assisted network. In chapter 3, the performance of several QoS metrics in HARQ relay-assisted network are theoretically quantified. Chapter 4 deals with the energy efficiency of HARQ relay-assisted networks and with an energy minimization problem in multi-user context. Finally, a general conclusion and some perspectives are drawn.

In more details, Chapter 1 shed light on several issues that must be considered on modern communications. Throughout this chapter, the benefits of relay-assisted networks are presented. Then, we provide a

1. Energy Aware Radio and neTwork tecHnologies

brief overview about the HARQ retransmission schemes that are incorporated in cross-layer protocols and the most prominent performance metrics such as: the PER, throughput, delay and outage probability. We also focus on the energetic issues by introducing the energy efficiency. Finally, some basic tools for optimization problems are presented.

The main results of this thesis are summarized as follows:

Chapter 2 concentrates on a throughput maximization problem in multi-user relay-assisted OFDMA downlink heterogeneous cellular network. In more details, users have different QoS data rate constraints in bits/s/Hz to be satisfied. The objective and constraint functions are expressed as function of the information theoretic capacity, where the global constraints are the total transmission power and bandwidth. A jointly optimal power allocation, subcarriers and relays assignment algorithm is proposed.

The major contribution of Chapter 3 lies on the theoretical analyses of the performance metrics of HARQ-I and HARQ-II of type Chase combining in relay-assisted networks. These analyses consider a practical scenario with a certain modulation and coding scheme (MCS) and finite packet length. The mostly prominent metrics such as the PER, delay and throughput are tightly derived in closed forms. In the same chapter and based on the derived delay closed form, the delay outage probability in block fading channels and relay-assisted decoded-and-forward (DF) HARQ-I based systems is theoretically analyzed.

Finally, Chapter 4 analyzes the energetic issues of HARQ networks. In a first time, the energy efficiency for different HARQ schemes in relay-assisted network is derived in closed form. Thereafter, considering a non reliable communication, i.e. non-zero PER, an energy minimization problem in multi-user relay-assisted downlink cellular networks is also investigated. This problem considers the circuitry energy consumption in addition to the allocated power in the cost function and the total power in the constraint function, provided that the QoS constraint designated by the delay is guaranteed for every user. The base station (BS) is assumed to have only a knowledge about the average channel statistics but no instantaneous CSI. Finally, an algorithm that jointly allocates the optimal powers at BS and relays and selects the optimal relay (if cooperation is decided) in order to satisfy the delay constrained users is proposed.

Publications

[1]- M. Maaz, P. Mary, and M. Helard, *"Delay Outage Probability in Block Fading Channel and Relay-Assisted Hybrid-ARQ Network,"* accepted to IEEE wireless communications letters, 2013, WCL2013-0686. (Early access DOI 10.1109/WCL.2013.113013.130686)

[2]- M. Maaz, P. Mary, and M. Helard, *" Energy Efficiency Analysis in Relay Assisted Hybrid-ARQ Communications,"* in PIMRC. IEEE, 2012, pp.2263-2268.

[3]- M. Maaz, P. Mary, and M. Helard, *" Resource Allocation for QoS Aware Relay-Assisted OFDMA Cellular Networks,"* 2011 Third International Workshop on Cross Layer Design (IWCLD), 2011, pp. 1-6.

[4]- M. Maaz, P. Mary, and M. Helard, *"Energy Minimization in HARQ-I Relay-Assisted Networks with Delay-limited Users,"*. (in preparation)

CHAPTER

1 State of the Art and Mathematical Background

1.1 Cooperative Communications

1.1.1 General Context

Basically, communication theory dates back to 1948, referring to Shannon's brilliant paper "A Mathematical Theory of Communication" [4]. This historical achievement has opened a new branch of science refereed as communication and information theory. In his elegant paper, Shannon introduced the channel capacity which is an upper bound on the number of transmitted bits that can be reliably sent in wireless Gaussian noisy channel in a given time duration for a given bandwidth. Thereafter, based on Shannon's capacity, the scientific community started developing this theorem and the data demand started increasing gradually.

Hence-after, a lot of time and rigorous work have been spent where the main purpose of which was to boost the channel capacity. It has been shown that Multiple-Input-Multiple-Output (MIMO) can be considered as an appealing communication technique that promises to meet the data demand requirements and improve robustness against fading. In their landmark work, Telatar [5] and Foschini [6] have shown that the capacity of MIMO scheme exceeds the one provided by the Single-Input-Single-Output (SISO). The penalty price of this scheme is that it adds more complexity at the transmitter and receiver sides and hence more additional signal processing is needed. But this complexity can be disregarded due to its capacity gain benefits. The capacity gain of MIMO channels is hence higher if the transmit antenna elements are spatially uncorrelated otherwise this gain degrades. Thereafter, a great deal of research effort has been focused on designing the so called space-time block codes (STBC) for MIMO systems. A simple transmit diversity scheme has been proposed by Alamouti [7] where two complex data streams of two antenna elements are space-time encoded and transmitted on the same bandwidth. As a result, it has been shown that this simple scheme achieves full diversity for 2×1 system. A more generic approach for designing optimal and quasi optimal STBC codes are proposed by Tarokh [8].

At the beginnings of 2000 and based on the MIMO and relaying schemes, Dohler [1] presented a novel communication technique which allows the deployment of MIMO schemes assuming that the antenna elements are geographically spaced. His approach allows the mutual communication between mobile terminals to create a virtual MIMO scheme. It is considered that the base station (BS) is equipped with several antenna arrays in which information data streams for different mobile terminals are space time encoded and transmitted to mobile terminals. Each terminal decodes its own information data stream

and relays the other data streams. Thus, this provides additional spatial diversity in the cell. This innovative concept is termed as Virtual Antenna Arrays (VAAs) [1].
Another excellent work has been exposed in parallel by Laneman [2]. In this framework, Laneman provided an information theoretic analysis of the outage probability in cooperative wireless networks. The system protocol is assumed to have several terminals that jointly transmit by means of a distributed space-time coded scheme. He demonstrated that this protocol achieves a full spatial diversity and can be efficiently used for higher spectral efficiencies.
Therefore, the aforementioned two baseline works retained a pragmatic approach from the point of view of system deployment and they could be certainly considered as appealing techniques for higher spectral efficient schemes. Originally, the idea of cooperation was first suggested by Van de Meulen [9] in 1968. Thereafter, Cover and Gamal [10] provided a first rigorous information theoretic analysis of the capacity in non-faded Gaussian relay channel and derived an achievable lower bound of the general relay channel capacity. In [10], it has been shown that the capacity of relay channel exceeds the one achieved by the direct scheme. An original work in the area of cooperative communications has been presented by Sendonarias *et al.* [11], where the purpose of which is to show that mobile terminals cooperation not only enhances the channel capacity but also increases the system robustness against fading. Thereafter, the cooperative diversity for multiple-access fading channels with relaying has been examined by the same authors in [12, 13].
Thereby, based on [1] and [2], many rigorous research topics have developed the idea of cooperative communications due to the offered substantial gains. The mostly important works related to this dissertation are summarized below.

In multiuser communications, the major important problem is how to optimally allocate resources between users while guaranteeing certain quality-of-service (QoS). The authors in [14] proposed a cross-layer scheduling scheme for orthogonal frequency division multiple access (OFDMA) wireless systems with heterogeneous delay requirements among users. They proposed an algorithm that allows a jointly optimal power allocation and subcarrier assignment such that the global data rate is maximized. However, relay-assisted communications were not considered by the authors.
The works related to [15, 16] dealt with optimization problems that aim to maximize the total throughput in multiuser OFDMA decode-and-forward (DF) relay-assisted network. In contrast, the QoS constraint has not been considered in the optimization procedure.
In [17], the authors studied a multiuser multi-service scheduling scheme in two-hops OFDMA systems. The power allocation and the joint subcarrier-relay selection are investigated in order to optimize a global utility function. However, carrying out two algorithms increases the complexity of the resource allocation procedure. Moreover, the proposed global utility function is built by aggregating several other cost functions, i.e. rate, delays, jitter and packet loss rate, assuming that they are independent from each other.
In [18], a global throughput maximization problem for cooperative OFDMA systems is investigated. Two

classes of services are considered, i.e. best effort and rate-constrained services. The authors proposed a joint relay selection and subcarrier allocation algorithm taking into account heterogeneous service constraints but without considering power control. In [19], the sum rate maximization in OFDM and DF relaying scheme has been investigated. However, the multiuser scenario with multiple relays was not considered and the QoS constraints have not been taken into account. In [20], a power minimization problem for uplink relay-assisted network has been studied. Each user is assumed to have a target rate constraint to be satisfied. However, this work did not focus on subcarriers allocation for each user. Therefore, a resource optimization policy that jointly optimizes power, assigns subcarriers and selects optimal relays is essential for multiuser QoS aware relay-assisted OFDMA networks.

Since a few years ago, the scientific community has realized that there is a tremendous waste of energy in the network core [21, 22]. This is due to the increase in the number of nodes that may participate in data transmission to the same mobile terminal and hence the equipments energy consumption increases. Moreover, the delay induced from the retransmission mechanism of erroneous packets leads to a higher energy consumption. Thereby, research started to put its emphasis on how to reduce the aggregate energy consumed for: data transportation, data generation and data processing at both transmitter and receiver sides. For ecologic and economic reasons, several works such as EARTH [1] project [21, 22] aim to reduce the overall energy consumption in a radio access network to at least 50%. Therefore, refining the energy consumption of the network components and utilizing innovative communication techniques could be a realistic solution for this task.

Many research topics in literature tried to model the energy consumption of the whole entities in cellular network. In [23–27], it has been demonstrated that the energy model comprises of two parts: the static one, consisting of the energy consumed for data processing, cooling, power supply *etc* and the dynamic one which is related to the energy consumed for data transmission. Once the energy consumption of the whole equipments is evaluated, the scheduler optimizes resources by considering these energetic issues.

Thereafter, several communication schemes enhancing the energy consumption have been proposed: for instance, femto cells or cell zooming deployment [28–30], cell cooperation [31,32], relay deployment [22] and BS idling [33–35] *etc*. These protocols are briefly exposed without providing any energy optimization policy.

Nowadays, energy minimization under QoS constraint is an important problem which plays a key role in green cellular communications [36, 37]. The traditional resource allocation policies focus on the throughput maximization, are generally inefficient in terms of energy consumption [37, 38] and references therein. Hence, an important effort has been recently granted toward energy efficient and reliable communications. In literature, we find works on energy minimization based on information theoretical approaches [39–41], i.e. considering reliable communications. The design of techniques in multiuser con-

1. EARTH: Energy Aware Radio and neTwork tecHnologies

text to minimize the energy consumption under non reliable communications, i.e. non zero packet error rate (PER), is a challenging task for the scientific community.

Several works dealing with the energy analysis have been produced in literature. The authors in [40] investigated the limit of the energy per bit consumption over an additive white Gaussian noise (AWGN) channel with relay. The authors in [42, 43] have dealt with wireless sensor networks (WSN) in which the energy consumption per node is the first criterion to optimize. In the above references, the authors studied the energy efficiency and the delay of virtual-MIMO WSN. The authors in [44, 45] analyzed the energy-delay performance in multi-hop WSN. In these works, the authors provided an analytical framework for the energy consumption in a network. They derived a Pareto front for the energy delay tradeoff defining two distinct regions: an achievable energy consumption region for a certain delay constraint and an unachievable energy consumption region.
In [46], the authors studied the end-to-end throughput and the energy efficiency of opportunistic and multi-hop routing protocols over linear wireless networks for fixed and optimal transmission rate. In [47] the authors studied the energy efficiency for cooperative and non cooperative hybrid-automatic repeat request (HARQ) schemes by minimizing the energy consumed for a single user with outage probability constraint. In [48], the authors studied the energy efficiency for several relaying schemes with HARQ under outage probability and delay constraints. It is also convenient to note that an important amount of work on WSN is provided, in which the energy consumption is of crucial importance [44]. However, all works above have not considered jointly the resource allocation strategy in multiuser cellular context with HARQ schemes under QoS constraint. The main challenge comes from the need of a closed-form expression of the PER with HARQ in order to express the QoS requirements [49].

Henceforth, from the overlying literature, two criteria have to be respected when proposing any scheduling policy:

- The higher QoS requirement for every user.
- The lower aggregate energy consumption in the whole network.

Therefore, a trade-off between the QoS requirement and the energy consumption exists. Moreover, the above criteria became essential for future wireless standards which aim to design cross-layer protocols that optimize the overall available resources and consider jointly the QoS and the energetic issues. In particular, cooperative communication is considered as an appealing technique that could be utilized to address these issues.

1.1.2 Cooperative Communications Protocols

According to literature, the relaying protocols can be classified into two families, the *transparent* and *regenerative* relaying protocols [50].

In the transparent relaying, very simple operations are usually performed without any signal processing, for instance, a simple amplification or phase rotation of the received wave form. The most prominent

protocol belongs to the transparent relaying is the amplify-and-forward protocol (AF).

Amplify and Forward (AF): In this protocol, the detected signal at the relay is amplified, frequency translated and retransmitted. The amplification factor can be constant or variable. Variable factor depends on the instantaneous channel state information (CSI) at the relay node. Thus, the main drawback of this protocol is that the CSI must be instantaneously available at the relay node. Moreover, a small channel realization leads to a higher amplification factor, which in turn leads to a strong noise amplification.

In the regenerative protocols, information stream or waveform samples are modified. Thus, the immediate price to pay is an increase of the transceiver complexity. However, this complexity is neglected regarding the enabled performance gain [50]. The most commonly used regenerative protocols in literature are:

Compress and Forward (CF): This protocol is not yet fully explored in literature. This introduces some form of source coding on the signal samples at the relay, in which a compressed version of a signal is sent to the destination.

Decode and Forward (DF): This communication strategy is considered as a very appealing scheme and a concurrent to AF protocol. The relay captures the transmitted signal by the source, processes it by demodulation and decoding, regenerates it by re-encoding and modulation, then retransmits it to the intended destination. It is known to have an optimal performance w.r.t. several metrics, such as error rate [50].

Because they outperform the transparent relaying protocols, the regenerative protocols are adopted in this dissertation and more specifically we will dwell on DF protocol since it enables a higher performance w.r.t. typical metrics such as error rate.

1.1.3 Gains of Cooperative Communications

Being incorporated in future wireless standards [51], cooperative communications certainly enable several gains compared to the traditional communication schemes. The mostly important gains of which are listed below:

Path-loss gain: It is inherent by nature that the path-loss has a severe long-term impact on the propagated signal. It has a non-linear behavior w.r.t. the separation distance between the two communicating nodes. Thus, the received signal-to-noise ratio (SNR) is path-loss dependent. More precisely, the SNR is inversely proportional to the propagation distance d, i.e.

$$\text{SNR} \propto \frac{1}{d^{\kappa}} \tag{1.1}$$

where $\kappa \geq 2$ is the path-loss exponent. Thereby, splitting the transmission path between two communicating nodes into several paths definitely yields to an SNR gain. For instance, assume that $n-1$ nodes

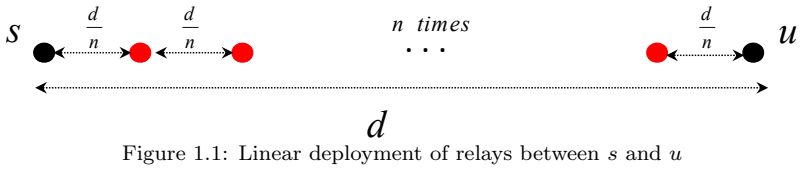


Figure 1.1: Linear deployment of relays between s and u

($n \geq 2$) are deployed along the direct path between the source and destination, fig. 1.1. Hence, this leads to n hops and the SNR gain associated to this scheme is given by

$$\frac{\frac{(p/n)}{\left(\frac{d}{n}\right)^{\kappa}} + \cdots + \frac{(p/n)}{\left(\frac{d}{n}\right)^{\kappa}}}{\left(\frac{p}{d^{\kappa}}\right)} = n^{\kappa} \tag{1.2}$$

which enables a gain of $10\kappa \log_{10}(n) \geq 10\kappa \log_{10}(2)$ dB. Where p is the allocated power for direct communication. More specifically, for a single-hop network, i.e. $n = 2$, the path-loss gain is $10\kappa \log_{10}(2)$ dB. Hence, from this simple example it is obvious that the path-loss gain is the prime reason of utilizing relaying communications.

Since it is inspired from the MIMO systems and it is considered as a virtual spatially distributed MIMO scheme, cooperative communications have certainly diversity and multiplexing gains:

Diversity gain: Deployment of one or several relays between two communicating nodes provides several copies of the same transmitted signal at the destination, which in turn combats the channel fading. As a result, a diversity gain is obtained and the probability of having deep channel fading at the destination decreases. Therefore, the diversity order d_o is upper bounded by [52]:

$$d_o \leq -\lim_{\mathrm{SNR}\to+\infty} \frac{\log_2\left(P(\mathrm{SNR})\right)}{\log_2(\mathrm{SNR})} \tag{1.3}$$

where P is the error probability. Asymptotically, if the equality in eq. (1.3) holds, P can be written as:

$$P(\mathrm{SNR}) \propto \mathrm{SNR}^{-d_o} \tag{1.4}$$

Assume that only one relay is used, then in addition to direct transmission, another copy of the signal is provided by the relay and hence the diversity order is $d_o = 2$. As a consequence, a power saving of about 3 dB is obtained.

Multiplexing gain: This gain stems from the fact that multiplexing independent data streams at each transmit antenna element in MIMO scheme yields to a multiplexing gain. As a result, deployment of several relays and forming a virtual distributed MIMO scheme leads to a multiplexing gain and hence a

higher transmission rate R. Thus, the multiplexing gain r_o for any system is upper bounded by [52]:

$$r_o \leq \lim_{\text{SNR}\to+\infty} \frac{R(\text{SNR})}{\log_2(\text{SNR})} \tag{1.5}$$

Asymptotically, the achievable rate R in bits per channel use is directly proportional to $\log_2(\text{SNR})$ and is given by:

$$R(\text{SNR}) \propto r_o \log_2(\text{SNR}) \tag{1.6}$$

For fast fading channels, the multiplexing gain is defined by $r_o = \min(n_t, n_r)$, with n_t and n_r are respectively the number of transmit and receive antenna elements [50]. Hence, introducing an intermediate relay and multiplexing data streams at both source and relay, yields to a multiplexing gain of 2 and thus the communication rate is doubled for a fixed SNR.

1.1.4 Channel Capacity

Direct communication: In his pioneering work, Shannon derived the channel capacity as the maximum of the mutual information $I(X,Y)$ between two random variables X and Y. It is defined as [53]:

$$C = \max_{p_X(X)} I(X,Y) \tag{1.7}$$

with $p_X(X)$ being the pdf of X.
Consider a Gaussian channel of input random variable X of power p and output $Y = X + N$, where $N \sim \mathcal{N}(0,\sigma^2)$ is a zero mean AWGN noise of variance σ^2 (average noise power). He showed that the maximum of (1.7) can be obtained if and only if X is Gaussian too, leading to $C = \frac{1}{2}\log_2\left(1+\frac{p}{\sigma^2}\right)$ per channel use. Moreover, for band-limited signal, the normalized channel capacity is given by:

$$C = \log_2\left(1+\frac{p}{\sigma^2}\right) \text{ bits/s/Hz} \tag{1.8}$$

with $\frac{p}{\sigma^2}$ being the received SNR.
Consider that a complex symbol X is transmitted over a fading channel with complex coefficient h, then the received signal Y is:

$$Y = hX + N \tag{1.9}$$

For a fixed channel fading h, the channel is AWGN with received SNR $= \frac{p|h|^2}{\sigma^2}$. Therefore, the instantaneous channel capacity is defined by:

$$C = \log_2\left(1+\frac{p|h|^2}{\sigma^2}\right) \text{ bits/s/Hz} \tag{1.10}$$

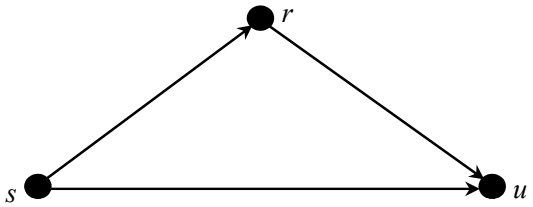


Figure 1.2: Relay-assisted scheme

Cooperative communication: Fig. 1.2 depicts a relay-assisted scheme consisting of two communicating nodes s and u and an intermediate relay r that may assist during the data transmission. In this scheme, each transmitting and receiving node is equipped with single antenna element. A two-hop DF relaying scheme is adopted and the transmission protocol consists of two phases:

- Phase 1: s transmits a symbol directly to u and the relay r may capture the transmitted information.
- Phase 2: The relay r processes the information data by decoding and re-encoding, then retransmit to u.

The channel is assumed to be frequency flat and constant during both transmission phases. Consider that h_{sd}, h_{sr} and h_{su}, the small-scale complex channel fading coefficients in the $s-u$, $s-r$ and $r-u$ links respectively. Moreover, $p_{l_{su}}$, $p_{l_{sr}}$ and $p_{l_{ru}}$ being respectively their associated path-loss coefficients. We assume an AWGN $n_r \sim \mathcal{CN}\left(0, \sigma_r^2\right)$ at the relay r and $n_u \sim \mathcal{CN}\left(0, \sigma_u^2\right)$ at the mobile user u.
The received signals at u and r in the first time slot are respectively:

$$y_{1_{su}} = \sqrt{p_{1_{su}} p_{l_{1_{su}}}} h_{1_{su}} x_{1_{su}} + n_u, \tag{1.11}$$

$$y_{1_{sr}} = \sqrt{p_{1_{sr}} p_{l_{1_{sr}}}} h_{1_{sr}} x_{1_{sr}} + n_r, \tag{1.12}$$

where $x_{1_{su}}$ ($x_{1_{sr}}$) being the transmitted symbol from s to u (r) in the first time slot. Moreover $p_{1_{su}}$ ($p_{1_{sr}}$) is the allocated power in the $s-u$ ($s-r$) link.
In the second time slot, the received information at u due to s or r transmissions are respectively:

$$y_{2_{su}} = \sqrt{p_{2_{su}} p_{l_{2_{su}}}} h_{2_{su}} x_{2_{su}} + n_u, \tag{1.13}$$

$$y_{2_{ru}} = \sqrt{p_{2_{ru}} p_{l_{2_{ru}}}} h_{2_{ru}} x_{2_{ru}} + n_u. \tag{1.14}$$

where $x_{2_{su}}$ ($x_{2_{ru}}$) is the transmitted data symbol from s to u (r to u) in the second slot. Moreover, $p_{2_{su}}$ ($p_{2_{ru}}$) is its associated allocated power. Since the channel realization is constant over two time slots, we have $h_{1_{su}} = h_{2_{su}} = h_{su}$ and $p_{su} = p_{1_{su}} = p_{2_{su}}$. Hence, the achievable data rate between s and u is:

$$C_{su} = \log_2(1 + g_{su} p_{su}), \tag{1.15}$$

where $g_{su} = \frac{|h_{su}|^2 p_{l_{su}}}{\sigma_u^2}$ being the channel gain in the $s-u$ link. Moreover, if the source s uses the relay r, the achievable data rate at r in the first time slot is:

$$C_{1_{sr}} = \log_2(1 + g_{1_{sr}} p_{1_{sr}}), \tag{1.16}$$

where $g_{1_{sr}} = \frac{|h_{1_{sr}}|^2 p_{l_{1_{sr}}}}{\sigma_r^2}$ is the channel gain in the $s-r$ link . Moreover, the data rate experienced by the user u due to $r-u$ link can be written as:

$$C_{2_{ru}} = \log_2(1 + g_{1_{su}} p_{1_{su}} + g_{2_{ru}} p_{2_{ru}}), \tag{1.17}$$

where $g_{1_{su}} = \frac{|h_{1_{su}}|^2 p_{l_{1_{su}}}}{\sigma_u^2}$ and $g_{2_{ru}} = \frac{|h_{2_{ru}}|^2 p_{l_{2_{ru}}}}{\sigma_u^2}$ are respectively the channel gains in the $s-u$ and $r-u$ links. Since two time slots are needed to decode the sent information through the relay, the reliable achievable data rate at u is [54]:

$$\frac{1}{2} \min\{C_{1_{sr}}, C_{2_{ru}}\} \text{ bits/s/Hz} \tag{1.18}$$

Therefore, the above capacity derivations for direct and relay-assisted schemes consider that the communication in each link is reliable, i.e. if the transmission rate is less than the maximum achievable rate (capacity), then each receiving node is able to decode successfully the transmitted information.

1.2 Cross-layer Protocols

Satisfying higher QoS and maintaining lower energy consumption are the main key performance indicators (KPIs) for designing wireless networks. Moreover, creating diversity in wireless network may combat the channel fading and enhance the system QoS.

Recently, it has been demonstrated that HARQ are very appealing protocols that induce a temporal diversity. These protocols combine the FEC (Forward error correction) coding with the retransmission mechanism and can achieve higher system performance at low SNR with only an acknowledgment (ACK) feedback from the receiver side. Moreover, cooperative communications have exhibited the gains provided by the MIMO systems. Therefore, it is of great interest to jointly incorporate cooperative communications and HARQ schemes in cross-layer protocols to satisfy the above KPI requirements.

Hybrid-ARQ Retransmission Protocols

As well documented throughout literature, there exists a whole gamut of HARQ retransmission schemes depending on the transmitter and receiver side processing. The mostly prominent HARQ protocols are listed below:

Simple ARQ: In this retransmission mechanism, each transmitted packet is constituted of the information bits and the CRC header (Cyclic Redundancy Check). The transmitter sends a data packet and

waits for the ACK message from the receiver side. If the data packet has been successfully decoded, a positive ACK is sent by the receiver and another MAC packet is transmitted. Otherwise, a negative ACK (NACK) is sent. For each NACK, the transmitter retransmits the same packet and the retransmission mechanism continues until the maximum number of allowed retransmissions $N_{\max}$ is achieved.

HARQ-I: This HARQ scheme incorporates the forward error correction (FEC) with the retransmission mechanism, where each retransmitted packet containing the CRC header is encoded with a coding rate R_0. The principle of this retransmission protocol is the same as the simple ARQ. Moreover, it depends on the coding rate R_0 and the chosen FEC, e.g. convolutional, turbocode, LDPC codes, *etc.* The main drawback of this scheme is that the redundancy bits are added at each retransmitted packet and hence a lower throughput and higher energy consumption are induced. For instance, under good channel conditions, i.e. high SNR, the throughput of this system achieves R_0 fraction of the throughput realized by the simple ARQ, since for each packet redundancy bits are added due to the FEC. Fig. 1.3 illustrates the retransmission protocol refereed to HARQ-I. The protocol concept was termed as HARQ-I since hybrids due to FEC are added to each retransmitted packet.

Since the simple ARQ and HARQ-I have limited performances, more sophisticated HARQ schemes have been proposed. These schemes are termed as HARQ-II and the most prominent protocols are introduced below:

HARQ-CC: The idea of this retransmission mechanism has been introduced in [3]. The concept was termed as HARQ of type Chase combining (HARQ-CC). In this scheme, each transmitted packet constituted of the CRC header is encoded by an FEC of rate R_0. In the first transmission attempt, the receiver processes and decodes the packet, then it checks the CRC header. If the packet is successfully decoded, a positive ACK is sent to the transmitter. Otherwise, a NACK is sent and the incorrect received packet is stored in a buffer rather than dropped. Thus, the same packet is retransmitted and at the receiver side, it is combined with the previously erroneous packet by means of maximum ratio combing (MRC) or softly combined by adding LLRs (Log-likelihood ratios) of the same bit, then it enters the decoding process. This retransmission mechanism continues until the packet is successfully decoded or $N_{\max}$ is achieved. Therefore, for a packet of L information bits, the transmission rate of the same packet at the instant n is $R_0\frac{L}{n}$. The main advantage of this scheme, is that each retransmitted packet experiences a temporal diversity and thus additional information for each bit is added before being decoded. Hence, this enables an increase in the probability of correct decision. A general scheme for HARQ-CC retransmission protocol is depicted in Fig. 1.4 for $N_{\max} = 3$.

HARQ-IR: The protocol concept of this scheme is termed as HARQ of type incremental redundancy (HARQ-IR). It is considered as the most sophisticated HARQ protocol since a specific coding scheme must be incorporated with the retransmission mechanism. In this process, a CRC header is added to each L information bits and the resultant data packet is encoded by a rate-compatible punctured FEC

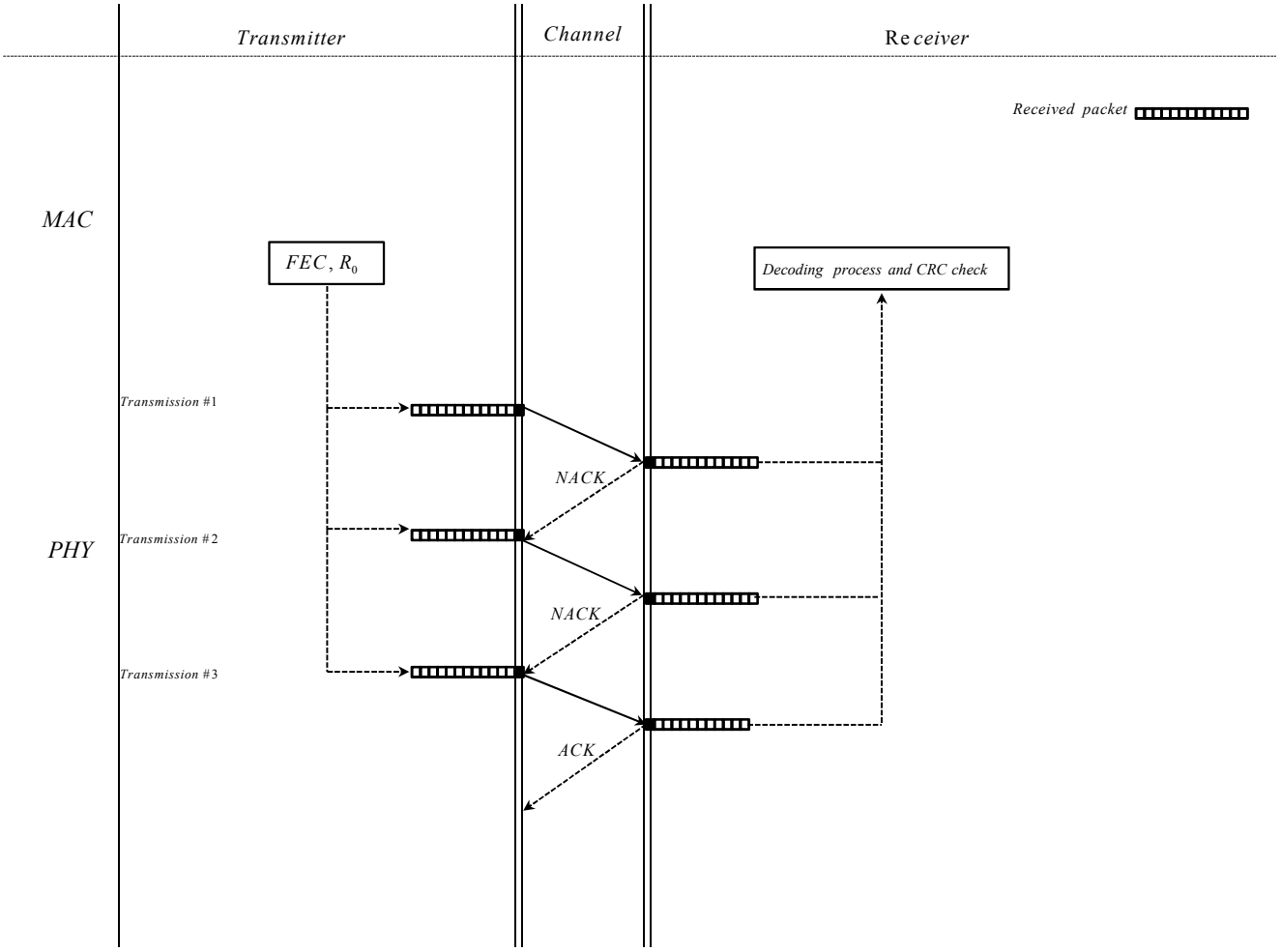


Figure 1.3: HARQ-I retransmission protocol where R_0 is the used coding rate

encoder that generates a systematic mother code of rate R_0. By mean of a systematic puncturer, this mother code is split into $N_{\max}$ sub-codes, each of $L(n)$ bits to be associated to the n-th retransmission instant. Thereby, the retransmission mechanism of this HARQ protocol is described as follows:
In the first transmitted packet, the information packet and some redundancy bits with the CRC header are sent. By passing through the noisy and fading channel toward the receiver side, the MAC packet is processed, depunctured and decoded (the puncturing method is known at the receiver), then the CRC header is checked. Thus, if the information packet has been successfully decoded, an ACK is sent to the transmitter and another packet is transmitted. If the decoder fails to extract the correct information, it stores this packet and sends a NACK packet to the transmitter that begins the retransmission process.

In contrast to the HARQ-CC, the HARQ-IR starts retransmitting redundancy bits to the receiver rather than retransmitting the same packet. These redundancy bits have to be combined with the

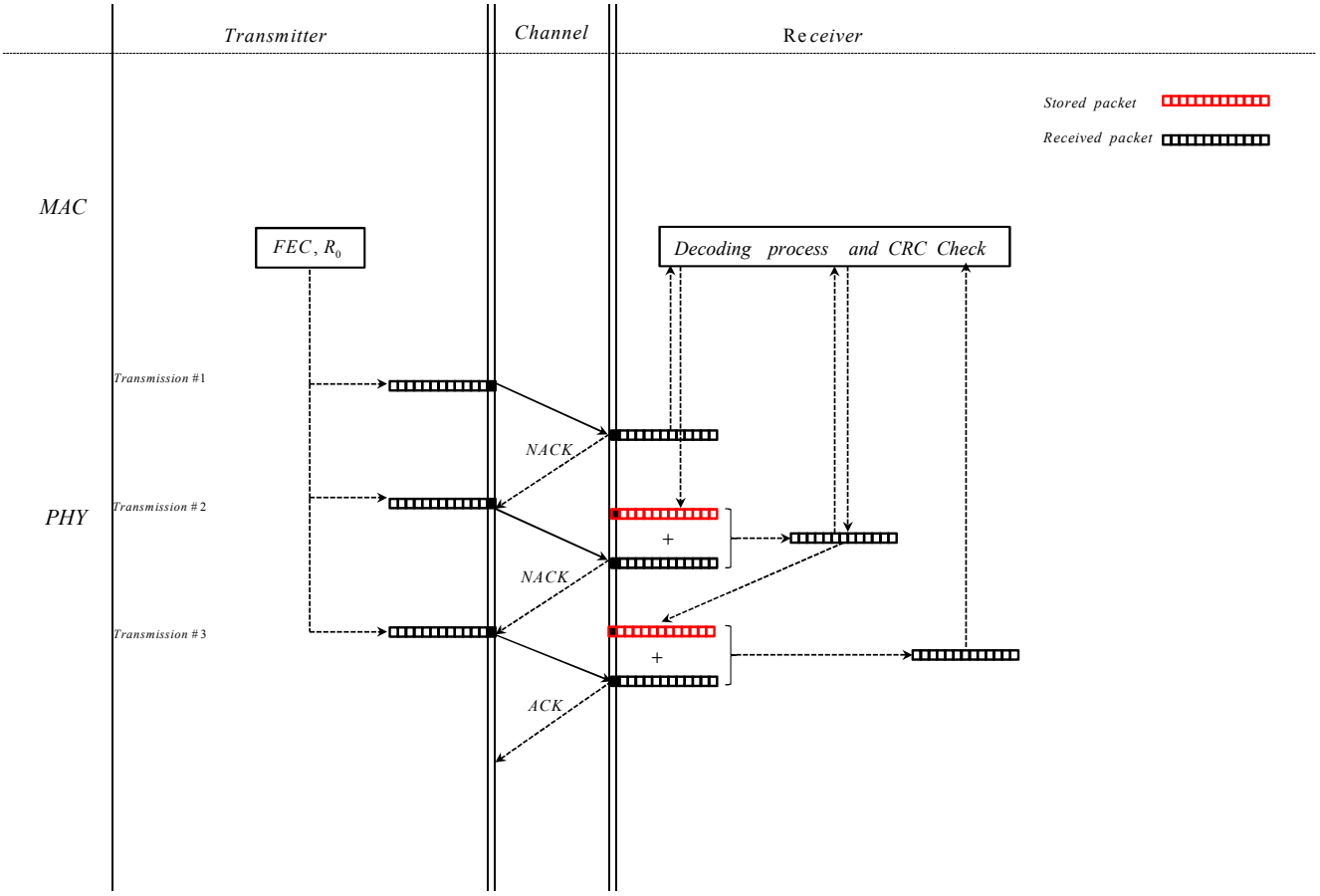


Figure 1.4: HARQ-CC retransmission protocol for $N_{\max} = 3$. Each retransmitted packet is combined with the previously erroneous packets

previously stored packets. This scenario continues until $N_{\max}$ is reached or the packet is correctly decoded. Hence, at each retransmission instant, the coding rate decreases (for instance, the coding rate at the instant n becomes $\frac{L}{\sum_{i=1}^{n} L(i)}$) and the decoding capability at the receiver side increases. This technique is more throughput efficient compared to aforementioned HARQ schemes since only redundancy bits are transmitted if required. In Fig. 1.5, the HARQ-IR process is depicted for $N_{\max} = 3$, where at each instant redundancy bits are added to the information packet. The main drawback of the HARQ-IR is that if the first packet was in deep fading then it will be lost and hence the retransmitted redundancy bits will not help to successfully decode the information packet. Furthermore, it depends on the utilized puncturing method and the number of retransmissions of this scheme depends on the code length. Another technical inconvenience of this scheme is that at each retransmission instant, the code length increases and thus a big buffer is needed at the receiver side.

Several families of rate-compatible punctured codes have been designed and could be useful for the

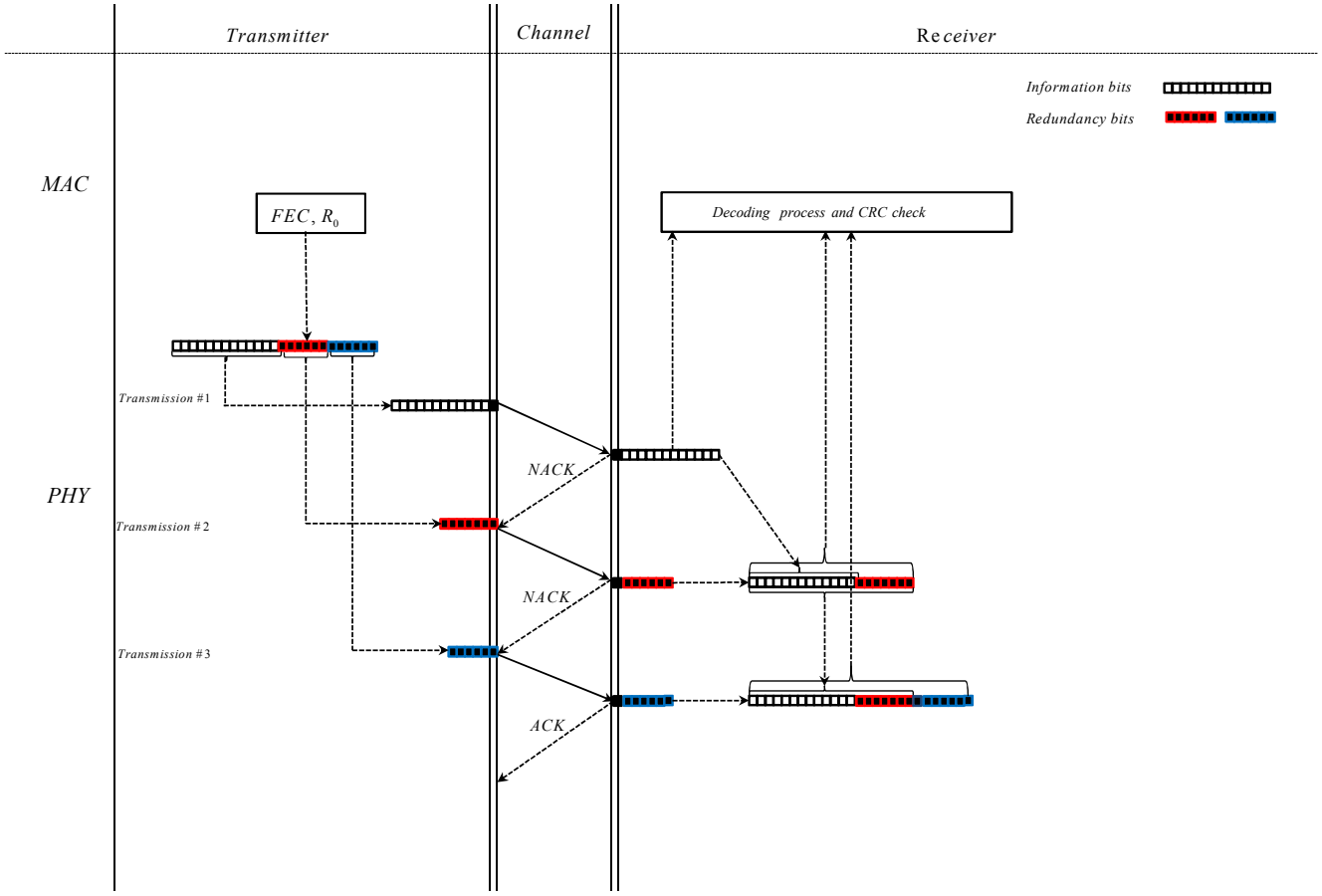


Figure 1.5: HARQ-IR retransmission protocol for $N_{\max} = 3$. Each retransmitted redundancy packet is added to the previously erroneous packets

HARQ-IR scheme, for instance, [55–57] for rate compatible punctured convolutional codes (RCPC), [58] for punctured turbocodes, [59–61] for LDPC codes.

HARQ-III: In this scheme, the HARQ-CC and HARQ-IR are jointly used in the retransmission mechanism. HARQ-III uses the HARQ-IR in the first n retransmissions and at the instant $n+1$ the retransmission scheme flips to HARQ-CC.

It is worthwhile to mention that in this dissertation, we restrict our analysis on the HARQ at the MAC packet level and more specifically on the HARQ-I and HARQ-CC protocols. In literature it could be found the analysis of HARQ schemes at the IP packet level (each IP packet contains several MAC packets). A more detailed recent and rigorous description of all HARQ schemes can be found in the dissertation [62] and [49].

1.3 Performance Metrics of Wireless Systems

Channel capacity: The channel capacity C is an information theoretic QoS. It is an upper bound on the achievable data rate in bits/s/Hz that can be reliably transmitted over a wireless noisy channel. Channel capacity has been adopted in several resource allocation policies, where optimization problems are expressed as function of it. For a pre-known CSI, the scheduler adapts the sufficient power and also assigns for every user certain number of subcarriers, time slots, and selects the optimal path (in multi-hop networks), such that the QoS is satisfied.

In principle, this generic approach can be justified by assuming that there exists some capacity approaching codes, such as LDPC codes or turbo-codes.

Packet Error Rate: The packet error rate (PER) is one of the most prominent performance metrics in wireless system. It is the probability that an error occurs due to bursty channel noise or to a deep fading that may occur. If at least only one bit is in error then the whole packet is dropped. Assuming that the instantaneous PER for any MCS over AWGN channel is defined as $P(\gamma)$ with γ being the instantaneous SNR. Then, the average PER is defined as:

$$\text{PER} = \int_0^\infty P(\gamma) p_\gamma(\gamma) d\gamma \tag{1.19}$$

where $p_\gamma(\gamma)$ being the probability density function (pdf) of γ.

Average Delay: Delay or latency is another important KPI in wireless systems. It is induced from the retransmission mechanism of erroneous packet due to non-reliable communications. At the MAC level, the average delay is defined as the average number of round-trips consumed per transmitted MAC packet. It is measured in time slots (TSs) or in seconds and each transmitted packet has a fixed number of retransmissions $N_{\max}$. Therefore, the average end-to-end delay $\overline{N}_t$ between two communicating nodes can be expressed as:

$$\overline{N}_t = \lim_{\tau \to \infty} \left(\frac{1}{\tau} \right) \sum_{k=1}^{\tau} N(k) \quad \text{TS} \tag{1.20}$$

where $N(k) \leq N_{\max}\ \forall k$ being the instantaneous delay induced from the k-th transmitted packet.

Throughput Efficiency: Once the error probability at any retransmission instant is calculated, the throughput efficiency (η_T) can be deduced. This quantity gives an insight on how much successful delivered information bits per certain quantity of transmitted bits using particular retransmission mechanism.

Hence, η_T can be defined as:

$$\eta_T \triangleq \frac{\text{Total number of successfully received bits}}{\text{Total number of transmitted bits}} \tag{1.21}$$

Another alternative definition of η_T is provided according to reward theorem [63]:

$$\eta_T = \frac{\mathbb{E}[\mathcal{R}]}{\mathbb{E}[\mathcal{B}]} \tag{1.22}$$

where $\mathcal{R}$ is the random reward equals to L bits for successfully decoded packet and zero for decoding failure. Moreover, $\mathcal{B}$ denotes the number of transmitted bits between two successfully received packets. Thus, η_T can be re-written as:

$$\eta_T = \frac{L(1-\text{PER})}{\mathbb{E}[\mathcal{B}]} \tag{1.23}$$

Outage probability: The outage probability $\mathbb{P}(O)$ is by definition, the probability that a given function $f(\gamma)$ of continuous random variable γ is below or exceeds certain threshold θ. If the channel does not vary quickly, then the temporal average of the random process is not equal to its statistical mean and hence the outage probability is preferred. Therefore, the outage probability of $f(\gamma)$ is defined as:

$$\mathbb{P}(O) \triangleq \Pr(f(\gamma) \lesseqgtr \theta) \tag{1.24}$$

with Pr(.) designates the probability operator.

Energy efficiency: Of augmenting importance is to quantify the energy efficiency (η_E) for a given communication scheme. It is defined as the average number of successfully received bits over the total energy consumed and its unit of measure is bits/joule. Hence, η_E is defined as:

$$\eta_E \triangleq \frac{\text{Average number of correctly received bits}}{\text{Average energy consumed}} \quad \text{bits/joule} \tag{1.25}$$

Recently, the energy efficiency has been introduced in cross-layer schemes as new QoS metric and considered as a design criterion of wireless systems [22]. It is well known that η_E is time and power dependent. Therefore, the average delay due to the retransmission mechanism has an impact on this quantity. Moreover, the spatial diversity gain offered by the relaying schemes certainly decreases the transmission power in wireless links. Hence, it would be important to analyze the energy efficiency in HARQ relay-assisted schemes. However, the impact of terminals' energy consumption has to be considered. It comprises of the circuitry energy consumption at the transmitter and receiver sides [22]. This issue is of particular importance in future wireless systems where the distances between terminals are becoming smaller and hence the circuitry energy consumption has the same order or greater than the energy consumed for data

transmission.

1.4 Energy Consumption Model

Generally, the consumed energy at the transmitter side E_{tx} stems from two sources: the first source of energy consumption is due to RF signal generation, which is mostly dependent on the number of generated bits in a packet, the adopted MCS and also the allocated power for data transmission P_t. The second source of energy consumption is due to the circuitry electronic components, i.e. filters and amplifiers *etc.* This energy cost is assumed to be fix. Therefore, the consumed energy at the transmitter side E_{tx} is quantified as follows [45, 64]:

$$E_{tx} = \frac{L}{R_b R_0}.(P_{tx} + \beta_{amp}.P_t) \tag{1.26}$$

where L is the number of information bits in a packet. R_0 and R_b are respectively the coding and modulation rates. P_{tx} and β_{amp} are respectively the circuitry power and the amplification factor at the transmitter side [45, 64].

At the receiver side, the consumed energy E_{rx} depends also on the modulation and coding rates and the circuitry energy consumption:

$$E_{rx} = \frac{L}{R_b R_0}.P_{rx} \tag{1.27}$$

where P_{rx} is the circuitry power at the receiver side.
The decoding energy consumption is neglected since it is complicated to be modeled. However, it has certainly an effect on the consumed energy since iterative algorithms are usually used in the decoding strategy, e.g. LDPC, turbocodes.

In order to announce the correct or failure reception of data packet, an ACK packet is sent from the receiver to transmitter side. Hence, an additional energy is consumed at both the transmitter that generates an ACK packet and the receiver that processes it. Therefore, the consumed energy for transmitting an ACK packet E_{ACK} is defined as [45]:

$$E_{ACK} = \tau_{ack}.(E_{tx} + E_{rx}) \tag{1.28}$$

with $\tau_{ack} = \frac{L_{ack}}{L}$ being the ratio between the number of bits in an ACK packet L_{ack} and the number of bits in a data packet L.

Henceforth, the aggregate consumed energy E_{ag} for one data packet round-trip is given by:

$$\begin{aligned} E_{ag} &= E_{tx} + E_{rx} + E_{ACK} \\ &= \frac{L(1+\tau_{ack})}{R_b R_0} \cdot (P_{tx} + P_{rx}) + \frac{L(1+\tau_{ack})}{R_b R_0} \cdot \beta_{amp}.P_t \end{aligned} \tag{1.29}$$

Thereby, it is worthwhile to recognize that the energy consumption model in (1.29) is a linear function w.r.t. P_t and consists of two parts: the first is static, which depends on the circuitry power P_{tx}, P_{rx} and the adopted transmission scheme. The second is dynamic, which depends on the allocated power P_t and the used MCS.

1.5 Mathematical Optimization Tools

1.5.1 Standard Form

An optimization problem can be written in standard form as follows [65]:

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f_0(\boldsymbol{x}) \\ \text{subject to:} \quad & f_i(\boldsymbol{x}) \leq 0 \quad \forall i = 1, \cdots m \\ & h_i(\boldsymbol{x}) = 0 \quad \forall i = 1, \cdots p, \end{aligned} \tag{1.30}$$

where the aim of which is to find an optimization vector $\boldsymbol{x} \in \mathbb{R}^n$ that minimizes the objective or the cost function $f_0(\boldsymbol{x})$. Moreover, $\boldsymbol{x}$ must satisfy the inequality constraints $f_i(\boldsymbol{x}) \leq 0$ for all $i = 1, \cdots, m$ and the equality constraints $h_i(\boldsymbol{x}) = 0$ for all $i = 1, \cdots, p$. Therefore, the optimization variable $\boldsymbol{x} \in \mathbf{D}$, where $\mathbf{D}$ is defined as:

$$\mathbf{D} = \bigcap_{i=0}^{m} D_{f_i} \cap \bigcap_{i=1}^{p} D_{h_i}, \tag{1.31}$$

with D_{f_i} and D_{h_i} designate respectively the domains of the functions f_i and h_i. Hence, the problem in (1.30) is feasible if there exists $\boldsymbol{x} \in \mathbf{D}$ such that $f_i(\boldsymbol{x}) \leq 0$ for all $i = 1, \cdots, m$ and $h_i(\boldsymbol{x}) = 0$ for all $i = 1, \cdots, p$. Otherwise this problem is infeasible. Thus, the optimal value of (1.30) is defined as:

$$q^* = \inf \{ f_0(\boldsymbol{x}) | f_i(\boldsymbol{x}) \leq 0,\ i = 1, \cdots, m,\ h_i(\boldsymbol{x}) = 0,\ i = 1, \cdots, p \} \tag{1.32}$$

1.5.2 Optimal Solution

Generally, an optimization problem has no analytic solution. However, for convex problems, it can be certainly known that the obtained solution is optimal. Therefore, a problem is said to be convex if and only if the following optimality conditions are satisfied:

- The objective function $f_0(\boldsymbol{x})$ must satisfy the convexity property

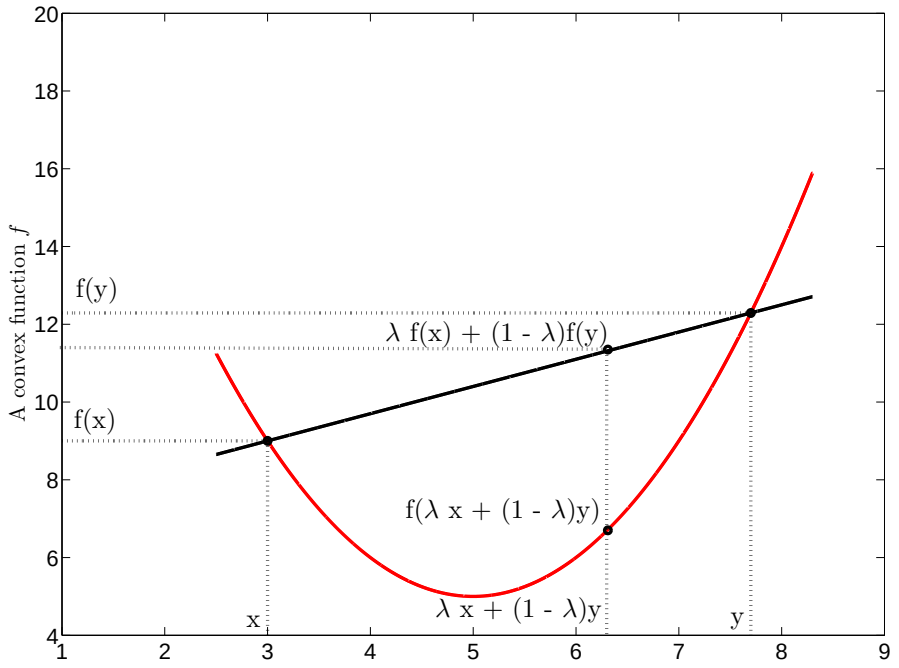


Figure 1.6: A convex function f

- The inequality constraint functions are convex
- The equality constraints must also preserve the convexity property

Convexity property(1): A function $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ is said to be convex if the following condition is satisfied:

$$f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) \tag{1.33}$$

$\forall\ \boldsymbol{x},\ \boldsymbol{y} \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$.

Geometrically, the above property means that the line segment between $(\boldsymbol{x}, f(\boldsymbol{x}))$ and $(\boldsymbol{y}, f(\boldsymbol{y}))$ lies above the function f for any $\boldsymbol{x} \neq \boldsymbol{y}$ and $0 < \lambda < 1$ (See fig. 1.6 for one dimensional function).

Convexity property(2): Suppose that $f(\boldsymbol{x})$ is twice differentiable, hence the Hessian of f at $\boldsymbol{x}$, $\nabla^2 f(\boldsymbol{x})$ is written as:

$$\nabla^2 f(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial^2 x_n} \end{pmatrix}$$

Therefore $f(\boldsymbol{x})$ is convex if and only if $\boldsymbol{D}$ is convex and its Hessian is semi-definite, i.e. $\nabla^2 f(\boldsymbol{x}) \succeq 0$ for all $\boldsymbol{x} \in \boldsymbol{D}$ [65].
It is worthwhile to notice that if $\nabla^2 f(\boldsymbol{x}) \preceq 0$ for all $\boldsymbol{x} \in \boldsymbol{D}$, then the function f is a concave function, i.e. $-f$ is convex. Furthermore, the optimal solution of (1.30) can be obtained if the optimality conditions on the objective and constraint functions are satisfied.

Now, let us consider that the optimality conditions on the optimization problem are satisfied. The Lagrangian associated to (1.30) $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ can be written as a weighted sum of the constraint functions as follows:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \nabla f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x}) \tag{1.34}$$

with μ_i and ν_i being the Lagrange multipliers related respectively to the constraints $f_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$. Moreover $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_m]^{\mathrm{T}}$ and $\boldsymbol{\nu} = [\nu_1, \cdots, \nu_p]^{\mathrm{T}}$ are the Lagrange multiplier vectors. The Lagrange multipliers signify the importance of the constraint functions, for instance, if any Lagrange multiplier equals to zero then the constraint related to this multiplier is neglected. Moreover, as it increases, the constraint function becomes more important. Therefore, the optimal solution of the optimization problem in (1.30) is obtained if there exists $(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$ such that the following conditions are satisfied:

$$\begin{aligned} f_i(\boldsymbol{x}^*) &\leq 0, \quad i = 1, \cdots, m \\ h_i(\boldsymbol{x}^*) &= 0, \quad i = 1, \cdots, p \\ \mu_i^* &\geq 0, \quad i = 1, \cdots, m \\ \mu_i^* f_i(\boldsymbol{x}^*) &= 0, \quad i = 1, \cdots, m \\ \nabla \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu^*}, \boldsymbol{\nu^*}) &= 0, \end{aligned}$$

The above conditions are termed as Karush-Kuhn-Tucker (KKT) conditions.

1.5.3 Lagrange Dual Optimization

The idea of the Lagrange duality is to take the Lagrangian function $\mathcal{L}$ in (1.34) and transform the optimization problem into a non constrained one. In this case, the vectors $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_m]^{\mathrm{T}}$ and $\boldsymbol{\nu} = [\nu_1, \cdots, \nu_p]^{\mathrm{T}}$ are termed as Lagrange dual variables. We define the Lagrange dual function g as the minimum of the Lagrangian function over $\boldsymbol{x}$:

$$g(\boldsymbol{\mu}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x} \in \boldsymbol{D}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\nu}) \tag{1.35}$$

Since $g(\boldsymbol{\mu}, \boldsymbol{\nu})$ is an affine function w.r.t. $(\boldsymbol{\mu}, \boldsymbol{\nu})$, then it is concave even if the optimization problem in (1.30) is not convex. Hence, for any $\boldsymbol{\mu} \geq 0$ and any $\boldsymbol{\nu}$, the dual function g is a lower bound of the optimal

solution q^*:

$$g(\boldsymbol{\mu}, \boldsymbol{\nu}) \leq q^* \tag{1.36}$$

Therefore, the optimization problem in (1.30) is equivalent to:

$$\sup_{(\boldsymbol{\mu},\boldsymbol{\nu})} g(\boldsymbol{\mu}, \boldsymbol{\nu}) = \sup_{(\boldsymbol{\mu},\boldsymbol{\nu})} \inf_{\boldsymbol{x}\in\boldsymbol{D}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\nu}) \tag{1.37}$$

Hence (1.37) is called a dual problem, where the aim of which is to find $(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$ that minimizes the duality gap $d_g = g(\boldsymbol{\mu}^*, \boldsymbol{\nu}^*) - q^*$. Moreover, if the optimization problem is convex, then the optimal solution can be obtained (i.e. $d_g = 0$) by solving directly the dual problem in (1.37).

1.6 Conclusion

This chapter covered the basic materials utilized in this dissertation. We started by introducing the existing literature related to our work. We shed light on innovative techniques that could be beneficial for modern communications. We figured out that cooperative communications could be an appealing technique for future communication standards. This is due to the substantial gains it offers this scheme, i.e. spatial and diversity gains.

We realized that QoS and energy consumption issues must be jointly considered when designing cross-layer concepts. Since, it is well known that communication over short multi-hops enhances the data service. In the other hand, increasing the number of cooperating nodes in the cell may lead to an energy waste. Therefore, novel resource optimization policies that jointly optimize the energy consumption and consider the demanded service are essential for wireless systems.

Throughout this chapter, we gave a brief overview about the HARQ retransmission schemes that are incorporated in cross-layer protocols. We also provided the most prominent performance metrics for communication schemes with HARQ protocols. With this in mind, we focused on the energy consumption by introducing another performance metric which is the energy efficiency. Finally, some basic tools for optimization problems are presented.

CHAPTER

2 Resource Allocation for QoS Aware Relay-Assisted OFDMA Cellular Networks

2.1 Introduction

Since the first works on cooperative concepts for wireless communications at the beginning of the century [1, 54, 66], cooperative communications for wireless cellular systems have never stopped drawing the attention of the scientific community during the past decade [67]. Indeed, the promise of higher data rates, lower bit error rates (BER) or coverage enhancement makes this concept attractive for 4G and beyond cellular systems. The efforts were first mainly dedicated to obtain general performance bounds on capacity, error and outage probabilities for two-hop AF or DF relaying schemes [54, 66, 68] and references therein.

The resource allocation problem in relaying systems is a particularly difficult issue due to the large number of resources which can be shared in the system, e.g. power, subchannels, relays, etc. In this chapter, we consider a single-cell network using OFDMA scheme as well as multiple fixed relays facilitating the communication between the BS and the mobile stations (MS). We are interested in finding a jointly optimal resource allocation policy that allocates optimal power, assign subcarriers and propose a relay selection strategy in order to maximize the global data rate while keeping a low starvation rate for the users. Indeed, the maximization of the global throughput often leads to starve users without sufficiently good channel conditions. The network is assumed to be constrained in power, bandwidth and also each user has a rate constraint to be satisfied. To the best of our knowledge, there is no work that addresses a jointly optimal power, subcarrier and relay allocation in OFDMA cellular systems with DF relaying scheme, under data rate constraints.

2.2 System Model

We consider a downlink cellular network consists of a BS, K relay stations (RS) r_k with $k \in \{1, \cdots, K\}$ and M mobile users u_m with $m \in \{1, \cdots, M\}$. A two-hop DF relaying technique is adopted as depicted in Fig. 2.1. We consider an OFDMA system with N_F orthogonal subcarriers that can be shared between users. The relay-assisted communication can be decomposed in two phases:

- Phase 1: BS transmits some subcarriers of the OFDMA symbol directly to users and some others to selected RSs that will assist users in the second transmission phase.

- Phase 2: Selected RSs transmit the received subcarriers to users that need assistance.

On the n-th subcarrier, the instantaneous channel coefficients between the BS and the m-th mobile user, the BS and the k-th RS, the k-th RS and the m-th mobile user are respectively $h^n_{1,0m}$, $h^n_{1,0k}$, $h^n_{2,km}$ for the small-scale fading and $l^n_{1,0m}$, $l^n_{1,0k}$, $l^n_{2,km}$ for the pathloss effect. Moreover $h^n_{i,km} \sim \mathcal{CN}(0,1)$ and the subscript 0 denotes the base station. Moreover, we assume an AWGN on the n-th subcarrier at the relays such as $n^n_r \sim \mathcal{CN}\left(0,\sigma^2_r\right)$ and at the mobile stations such as $n^n_m \sim \mathcal{CN}\left(0,\sigma^2_m\right)$. The channel is assumed to be invariant during both transmission phases i.e. $i = 1, 2$, which is a reasonable assumption for slow moving users and hence the instantaneous channel state information (CSI) of all relays and users are assumed to be known at the base station. The cell radius is equal to R and the relays are located around the BS each at a distance $R/2$.

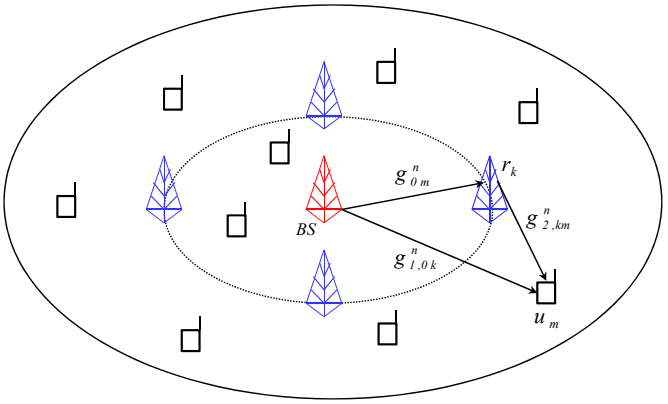


Figure 2.1: System model: Multi-user OFDMA cooperative cellular network

Let us consider a transmission between BS and the user m on the subcarrier n (as in fig. 2.1). In the first time slot, the received signal at the user m or at the relay k are respectively expressed by:

$$y^n_{1,0m} = \sqrt{p^n_{1,0m} l^n_{1,0m}} h^n_{1,0m} x^n_{1,0m} + n^n_m, \tag{2.1}$$

$$y^n_{1,0k} = \sqrt{p^n_{1,0k} l^n_{1,0k}} h^n_{1,0k} x^n_{1,0k} + n^n_r, \tag{2.2}$$

where $x^n_{1,0m}$, $(x^n_{1,0k})$ is the transmitted data symbol from the BS to the m-th user (k-th relay) on the n-th subcarrier in the first time slot. Moreover $p^n_{1,0m}$ $(p^n_{1,0k})$ is the power on the n-th subcarrier on the link BS-MS (BS-RS respectively). In the second time slot and the n-th subcarrier, the received information at the user m due to the relay transmission or due to the direct transmission are given by:

$$y^n_{2,0m} = \sqrt{p^n_{2,0m} l^n_{2,0m}} h^n_{2,0m} x^n_{2,0m} + n^n_m, \tag{2.3}$$

$$y^n_{2,km} = \sqrt{p^n_{2,km} l^n_{2,km}} h^n_{2,km} x^n_{2,km} + n^n_m. \tag{2.4}$$

Where $x^n_{2,km}$ ($x^n_{2,0m}$) is the data symbol on the n-th subcarrier transmitted from the k-th relay to the m-th user in the second time slot (respectively from the BS to the m-th user), while $p^n_{2,km}$ ($p^n_{2,0m}$) is the allocated power to the n-th subcarrier for the RS-MS link (BS-MS respectively), in the second time slot. Since the channel is constant over two time slots we have $h^n_{1,0m} = h^n_{2,0m} = h^n_{0m}$ and $p^n_{0m} = p^n_{1,0m} = p^n_{2,0m}$. The achievable data rate between BS and the user m in the direct link is:

$$C^n_{0m} = \log_2(1 + g^n_{0m} p^n_{0m}), \tag{2.5}$$

where $g^n_{0m} = |h^n_{0m}|^2 l^n_{0m}/\sigma^2_m$. If the BS uses the relay k, the achievable data rate at relay k in the first time slot is expressed as:

$$C^n_{1,0k} = \log_2(1 + g^n_{1,0k} p^n_{1,0k}), \tag{2.6}$$

where $g^n_{1,0k} = |h^n_{1,0k}|^2 l^n_{1,0k}/\sigma^2_r$. The data rate experienced by the user m due to the RS-MS hop can be written as:

$$C^n_{2,km} = \log_2(1 + g^n_{2,km} p^n_{2,km} + g^n_{1,0m} p^n_{1,0k}), \tag{2.7}$$

where $g^n_{2,km} = |h^n_{2,km}|^2 l^n_{2,km}/\sigma^2_m$ and $g^n_{1,0m} = |h^n_{1,0m}|^2 l^n_{1,0m}/\sigma^2_m$. Since two time slots are needed to decode the information sent through a relay, the achievable data rate of the user m is [54]:

$$\frac{1}{2} \min \left\{ C^n_{1,0k}, C^n_{2,km} \right\} \tag{2.8}$$

By using the relay-assisted mode, the data rate at user m is maximum if

$$C^n_{1,0k} = C^n_{2,km} \tag{2.9}$$

Let us denote p^n_{km} the overall power consumed in both phases, it can be expressed as a function of $p^n_{1,0k}$ and $p^n_{2,km}$:

$$p^n_{km} = p^n_{1,0k} + p^n_{2,km} \tag{2.10}$$

Inserting (2.10) in (2.9), $p^n_{1,0k}$ and $p^n_{2,km}$ can be expressed w.r.t. p^n_{km} as follows:

$$p^n_{1,0k} = \frac{g^n_{2,km}}{g^n_{1,0k} + g^n_{2,km} - g^n_{0m}} p^n_{km}, \tag{2.11}$$

$$p^n_{2,km} = \frac{g^n_{1,0k} - g^n_{0m}}{g^n_{1,0k} + g^n_{2,km} - g^n_{0m}} p^n_{km}, \tag{2.12}$$

where $g^n_{1,0k} > g^n_{0m}$. The equivalent channel gain g^n_{km} between the BS and the user m using the relay k can be expressed as:

$$g^n_{km} = \frac{g^n_{2,km} g^n_{1,0k}}{g^n_{1,0k} + g^n_{2,km} - g^n_{0m}}. \tag{2.13}$$

$$\max_{\boldsymbol{s},\boldsymbol{p}} \sum_{k=0}^{K} \sum_{m=1}^{M} \sum_{n=1}^{N_F} s_{km}^{n} C_{km}^{n} \tag{2.16}$$

subject to:

$$\begin{aligned}
&(c_1) \quad \sum_{k=0}^{K} \sum_{m=1}^{M} \sum_{n=1}^{N_F} s_{km}^{n} p_{km}^{n} \leq P_{\text{tot}} \\
&(c_2) \quad \sum_{k=0}^{K} \sum_{n=1}^{N_F} s_{km}^{n} C_{km}^{n} \geq \rho_m \quad \forall m \in \{1, \cdots, M\} \\
&(c_3) \quad \sum_{k=0}^{K} \sum_{m=1}^{M} s_{km}^{n} = 1 \quad \forall n \in \{1, \cdots, N_F\} \\
&(c_4) \quad s_{km}^{n} \in \{0, 1\} \quad \forall (k, m, n) \\
&(c_5) \quad p_{km}^{n} \geq 0 \quad \forall (k, m, n)
\end{aligned}$$

Hence, the maximum achievable rate over the n-th subcarrier can be expressed in terms of g_{km}^{n} and p_{km}^{n}:

$$C_{km}^{n} = \frac{1}{2} \log_2(1 + p_{km}^{n} g_{km}^{n}) \quad \forall k = 1 \ldots K. \tag{2.14}$$

According to eqs. (2.5) and (2.14), the achievable rate on the subcarrier n, under direct or indirect transmission can be written as:

$$C_{km}^{n} = \frac{(1 + \delta(k))}{2} \log_2(1 + p_{km}^{n} g_{km}^{n}) \quad \forall k = 0, \cdots, K, \tag{2.15}$$

where $\delta(k)$ is a Dirac operator which equals to one if and only if $k = 0$.

2.3 Problem Formulation and Solution

According to the instantaneous CSI and the required QoS, the MAC layer performs the scheduling for the downlink transmission by allocating the power, subcarriers and relays for every user that needs assistance, in order to maximize the global data rate. The aim is hence to solve the data rate maximization problem under total power and rate constraints by jointly allocating the power, subcarriers and relays for each downlink transmission.

2.3.1 Problem Formulation

The resource allocation problem can be formulated as in (2.16). Where s_{km}^{n} is the subcarrier allocation index. If $s_{km}^{n} = 1$, the subcarrier n is allocated to the link k-m and no subcarrier is allocated to the link if $s_{km}^{n} = 0$, i.e. constraint (c_4). The expression in (c_1) formulates the overall power constraint P_{tot}

on the whole cell, i.e. the consumed powers at all RSs and BS. The constraint (c_2) ensures that the overall achievable rate for each user m must be above or equal to its rate constraint ρ_m in bits/s/Hz. The constraint (c_3) states that the subcarrier n is dedicated only to one user and one relay. The constraint (c_5) implies that the allocated power on the subcarrier n must be positive.
For each CSI matrix $\boldsymbol{G} = [g_{km}^n]_{K\times M\times N_F}$, this optimization problem aims to allocate an optimal power matrix $\boldsymbol{P} = [p_{km}^n]_{K\times M\times N_F}$ and assign a binary matrix $\boldsymbol{S} = [s_{km}^n]_{K\times M\times N_F}$ such that the rate constraint vector $\boldsymbol{\rho} = [\rho_1, \cdots, \rho_M]^T$ and the total power constraint are satisfied.

2.3.2 Problem Convexity and Optimal Solution

Problem Convexity: The above problem is a combination of a continuous optimization and an integer programming problem. It cannot be solved with a computationally efficient algorithm due to its combinatorial nature. However, by relaxing the constraint associated to the integer variable s_{km}^n, i.e. (c_4), by considering that $0 \leq s_{km}^n \leq 1$, this problem can be solved [14]. Moreover, considering a change of variable $\bar{p}_{km}^n = s_{km}^n p_{km}^n$ [14], the cost function of this problem is a sum of functions that each has the following form:

$$g(x, y) = x \log\left(1 + \frac{y}{x}\right), \tag{2.17}$$

where $g(x, y)$ is called the perspective function of $\log(1 + x)$. Since $\log(1 + x)$ is a concave function, then $g(x, y)$ is also a concave function [65, pp. 89] as it can be noticed in fig. 2.2. Moreover, the cost function of this optimization problem is a sum of concave functions which remains a concave function. By the same methodology, the constraint function (c_2) can be proved to be also a concave function $\forall m \in \{1, \cdots, M\}$. Moreover, the constraints (c_1) and (c_3) are affine functions. Thus, this problem is a convex optimization problem and an optimal solution can be obtained.

Optimal Solution: Since the optimization problem in (2.16) is convex, taking the gradient of the Lagrangian function $\mathcal{L}$ w.r.t. $\bar{p}_{km}^n$, s_{km}^n and writing the KKT conditions [65], this problem can be solved efficiently. The Lagrangian function $\mathcal{L}$ associated to this problem can be expressed as:

$$\begin{aligned}\mathcal{L} = &\sum_{m=1}^{M}(1+\gamma_m)\sum_{k=0}^{K}\sum_{n=1}^{N_F} s_{km}^n \frac{1+\delta(k)}{2}\log_2\left(1+\frac{\bar{p}_{km}^n g_{km}^n}{s_{km}^n}\right)\\ &-\mu\left(\sum_{k=0}^{K}\sum_{m=1}^{M}\sum_{n=1}^{N_F}\bar{p}_{km}^n - P_{\text{tot}}\right) - \sum_{m=1}^{M}\gamma_m\sum_{k=0}^{K}\sum_{n=1}^{N_F}\rho_m\\ &+\sum_{n=1}^{N_F}\alpha_n\sum_{k=0}^{K}\sum_{m=1}^{M}(s_{km}^n - 1),\end{aligned} \tag{2.18}$$

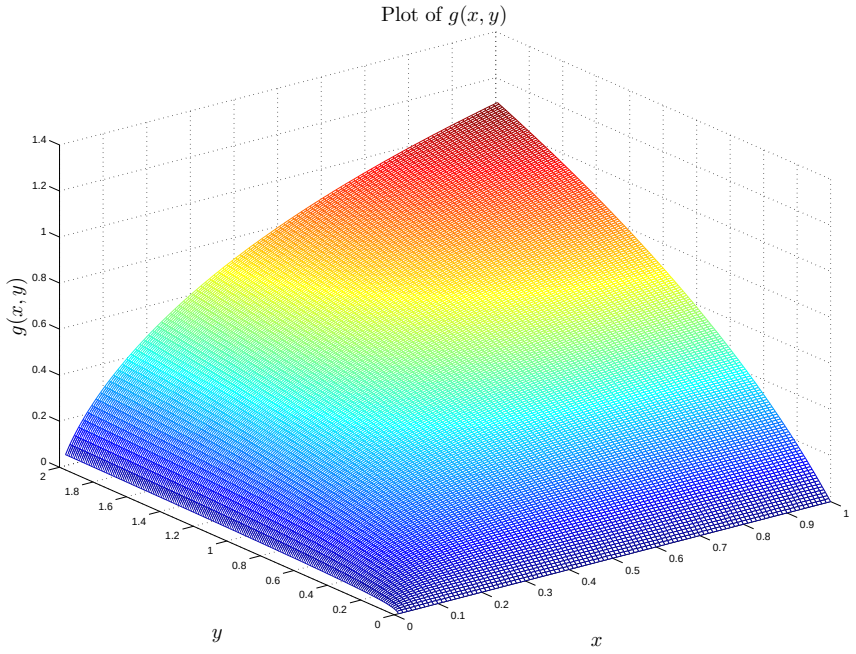


Figure 2.2: Concavity of the function $g(x, y)$

where μ, $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_M]^T$ and $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_{N_F}]^T$ are the Lagrange multiplier vectors related respectively to the constraints (c_1), (c_2) and (c_3).
Moreover, the KKT conditions associated to the above optimization problem are:

$$\begin{aligned}
\mu \left(\sum_{k=0}^{K} \sum_{m=1}^{M} \sum_{n=1}^{N_F} \bar{p}_{km}^n - P_{\text{tot}} \right) &\leq 0, \\
\gamma_m \left(\rho_m - \sum_{k=0}^{K} \sum_{n=1}^{N_F} s_{km}^n \tfrac{1+\delta(k)}{2} \log_2 \left(1 + \tfrac{\bar{p}_{km}^n g_{km}^n}{s_{km}^n} \right) \right) &\leq 0, \quad m = 1, \cdots, M \\
\gamma_m &\geq 0, \quad m = 1, \cdots, M \\
\mu &\geq 0, \\
\nabla \mathcal{L} &= 0,
\end{aligned}$$

Therefore, applying the gradient of $\mathcal{L}$ w.r.t. $\bar{p}_{km}^n$, we get the following:

$$\nabla_{\bar{p}_{km}^n} \mathcal{L} = \frac{\frac{1}{2} s_{km}^n (1 + \delta(k))(1 + \gamma_m) \frac{g_{km}^n}{s_{km}^n}}{1 + \frac{g_{km}^n}{s_{km}^n} \bar{p}_{km}^n} - \mu \tag{2.19}$$

for $n = 1 : N_F$

$$(k,m)^* = \arg\max_{(k,m)} \frac{(1+\delta(k))(1+\gamma_m)}{2}\left[\log_2\left(\frac{(1+\delta(k))(1+\gamma_m)}{2\mu}g_{km}^n\right)\right]^+ - \mu\left[\frac{(1+\delta(k))(1+\gamma_m)}{2\mu} - \frac{1}{g_{km}^n}\right]^+$$

$$s_{km}^n = \begin{cases} 1 & \forall\ (k,m) = (k,m)^* \\ 0 & \text{otherwise} \end{cases}$$

end

(2.21)

Hence, solving $\nabla_{\bar{p}_{km}^n}\mathcal{L} = 0$ w.r.t. p_{km}^n, the optimal allocated power $p_{km}^{n^*}$ on the subcarrier n for a couple (k,m) is found to be equal to:

$$p_{km}^{n^*} = \begin{cases} \left[\frac{(1+\delta(k))(1+\gamma_m)}{2\mu} - \frac{1}{g_{km}^n}\right]^+ & \text{if } s_{km}^n = 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.20}$$

where the operator $[x]^+$ stands for $\max\{0, x\}$.

Moreover, differentiating $\mathcal{L}$ with respect to s_{km}^n we get the following:

$$\nabla_{s_{km}^n}\mathcal{L} = \frac{(1+\delta(k))(1+\gamma_m)}{2}\left(\log_2\left(1 + g_{km}^n\frac{\bar{p}_{km}^n}{s_{km}^n}\right) - s_{km}^n\left(\frac{\frac{\bar{p}_{km}^n}{(s_{km}^n)^2}g_{km}^n}{1+\frac{\bar{p}_{km}^n}{s_{km}^n}g_{km}^n}\right)\right) - \alpha_n \tag{2.22}$$

According to problem assumptions, if $p_{km}^n = 0$ then $s_{km}^n = 0$. However, for $s_{km}^n \neq 0$, we have:

$$\nabla_{s_{km}^n}\mathcal{L}\begin{cases} > 0 & \text{if } s_{km}^n = 1 \\ = 0 & \text{if } 0 < s_{km}^n < 1 \end{cases} \tag{2.23}$$

where (2.23) can be interpreted as follows:

- $\nabla_{s_{km}^n}\mathcal{L} = 0$ signifies that the maximum is occurred for $0 < s_{km}^n < 1$.
- In the other hand, at the boundary, i.e. $s_{km}^n = 1$, $\nabla_{s_{km}^n}\mathcal{L} \neq 0$. Hence, if $\nabla_{s_{km}^n}\mathcal{L} > 0$ then $s_{km}^n = 1$.

Thereby, substituting eq. (2.20) in (2.22) and since α_n is common for all (k,m) couples, a subcarrier allocation strategy for each (k,m) pair can be obtained thanks to the routine described in eq. (2.21) at the top of the next page.

The expressions given in (2.21) and (2.20) are similar to the ones obtained in [14]. However, they account for the relay selection strategy which can be performed jointly with the subcarrier allocation.

Combining equations (2.11), (2.12) and (2.20), the powers to be allocated in the first and the second hop can be readily deduced. Looking at the equation of the subcarrier allocation strategy in (2.21), we can figure out a trade-off between the throughput and its associated power i.e. the left hand side equation

signifies the throughput and the second hand side equation characterizes the allocated power. The power allocation in eq. (2.20) can be realized as a multi-level water-filling [14].

2.4 Joint Optimal Power Allocation, Relay selection and Subcarriers Assignment (JPRS)

Based on the expressions given in eq. (2.21) and (2.20), we propose a global resource allocation algorithm that jointly optimizes the power, assigns subcarriers and selects the relays, presented in Algorithm 1.

Algorithm 1 Joint Power, Subcarrier and Relay Selection algorithm (JPRS)

1: $\mu \leftarrow \mu_0$, $U = \{1, \cdots, M\}$
2: $\gamma_m^+ \leftarrow 0$, $\gamma_m^- \leftarrow 0$, $r_{e_m}^- \leftarrow -\rho_m \ \forall m \in U$
3: **while** $r_{e_m}^- < 0, \forall m$ **do**
4: $\quad \gamma_m^+ \leftarrow \gamma_m^+ + \delta$
5: $\quad$ Apply (2.21)
6: $\quad r_{e_m}^- \leftarrow \sum_{k=0}^{K} \sum_{n=1}^{N_F} s_{km}^n C_{km}^n - \rho_m$
7: **end while**
8: $r_{e_m}^+ \leftarrow +\rho_m \ \forall m \in U$
9: **while** $r_{e_m}^+ > 0, \forall m$ **do**
10: $\quad \gamma_m^- \leftarrow \gamma_m^- - \delta$
11: $\quad r_{e_m}^+ \leftarrow \sum_{k=0}^{K} \sum_{n=1}^{N_F} s_{km}^n C_{km}^n - \rho_m$
12: **end while**
13: Find $\gamma_m^* \in [\gamma_m^-, \gamma_m^+]$ such that $f_m(\gamma_m, \mu_0) = 0 \ \forall \ m$
14: $P_{\text{tot}_{\text{req}}} \leftarrow \sum_{m=1}^{M} \sum_{k=0}^{K} \sum_{n=1}^{N_F} s_{km}^n p_{km}^n$
15: **if** $P_{\text{tot}_{\text{req}}} > P_{\text{tot}}$ **then**
16: $\quad P_{m_{\text{req}}} \leftarrow \sum_{k=0}^{K} \sum_{n=1}^{N_F} s_{km}^n p_{km}^n \quad \forall m \in U$
17: $\quad m^* \leftarrow \arg\min_m \frac{\rho_m}{P_{m_{\text{req}}}}$
18: $\quad$ Update U by removing the user m^*
19: **else**
20: $\quad$ Find $P(\boldsymbol{\gamma}^*, \mu) = 0$, else
21: $\quad$ Update μ and go to step 2
22: **end if**

The joint power, subcarrier and relay selection algorithm (JPRS) attempts to search the optimal Lagrange multipliers μ and γ_m. For an initialized $\mu = \mu_0$, it starts by finding upper and lower bounds of γ_m, i.e. γ_m^+ and γ_m^- described by the while loops at steps 3 and 9. The rate of all users is first

set to zero, i.e. $r_{e_m}^- = -\rho_m$ (step 2) and the upper bound γ_m^+ is increased until all users satisfy their rate constraint, i.e. $r_{e_m}^- > 0$. For this purpose, for each subcarrier, the optimal pair (k, m) is selected according to eq. (2.21) and the gap between the rate of each user and its rate constraint is updated (step 5 and 6). According to the allocated subcarriers and relays for every user, we search γ_m^- in order to get $r_{e_m}^+ < 0 \ \forall m \in \{1, \cdots, M\}$. Once the interval $[\gamma_m^-, \gamma_m^+]$ is defined, the optimal γ_m^* is found by solving the equation $f_m(\gamma_m^*, \mu_0) = 0$ with:

$$f_m(\gamma_m^*, \mu_0) = \sum_{k=1}^{K} \sum_{n=1}^{N_F} s_{km}^n \log_2 \left[\frac{(1+\delta(k))(1+\gamma_m^*) g_{km}^n}{2\mu_0} \right]^+ - \rho_m \quad \forall m \in \{1, \cdots, M\} \tag{2.24}$$

where eq. (2.24) can be solved by applying the bisection method on the interval $[\gamma_m^-, \gamma_m^+]$. In a second time and from the obtained $\gamma_m^* \ \forall \ m$, the algorithm searches for the optimal μ^* by first computing the power required to achieve the data rate constraints determined before (step 14). If the required power is available, the algorithm updates μ and restarts updating γ_m for every user until finding $(\boldsymbol{\gamma}^*, \mu^*)$ such that $P(\boldsymbol{\gamma}^*, \mu^*) = 0$, with

$$P(\boldsymbol{\gamma}^*, \mu^*) = \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{n=1}^{N_F} s_{km}^n \left[\frac{(1+\delta(k))(1+\gamma_m^*)}{2\mu^*} - \frac{1}{g_{km}^n} \right]^+ - P_{\text{tot}} \tag{2.25}$$

If the required power exceeds the total power constraint of the network, it implies that all constraints cannot be satisfied at the same time. In this case, the user with the lower spectral efficiency per Watt allocated is removed (step 17) and the algorithm restarts until all users be satisfied or removed.

2.5 Numerical results

In this section, the performances of our algorithm in terms of global achievable data rate and average starved user rate, i.e. the number of users without resources over the total number of users, are investigated. We consider a circular cell of radius $R = 1$ km and four relays uniformly deployed around the base station at a distance $R/2$ with the same angle between them. The system bandwidth is 1.4 MHz and an OFDMA symbol is composed of 64 subcarriers. The noise spectral density is -155 dBm/Hz. The pathloss gain in dB is $l_{km}^n|_{\text{dB}} = 139.90 + 34.41.\log_{10}(d)$ where d is the distance in km between the base station or any relays to the user m [69]. We also consider heterogeneous classes of users having different data rate requirements, i.e. $[r_1, r_2, r_3, r_4] = [1\ 2\ 2.5\ 3]$ bits/s/Hz. We assume that users are positioned around the BS at R. In the following simulations, the global data rate and the starved user rate are averaged over 10000 channel realizations. Our allocation algorithm JPRS is compared with two other strategies, i.e. a power-subcarrier non cooperative algorithm [14] labeled "Non-Coop" in the figures and a cooperative algorithm without power allocation labeled "Equal power" in [18].

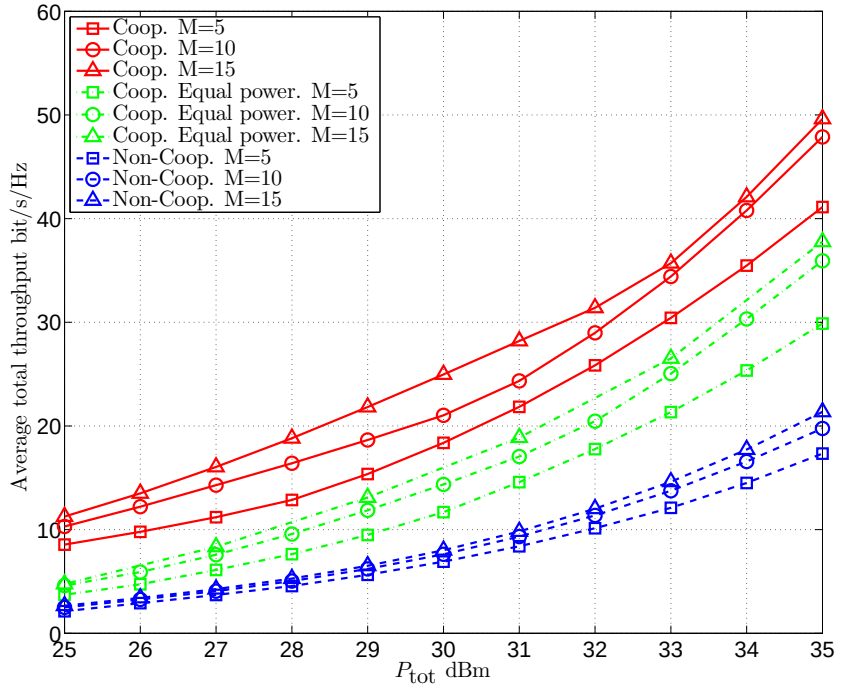


Figure 2.3: Average total throughput versus P_{tot}

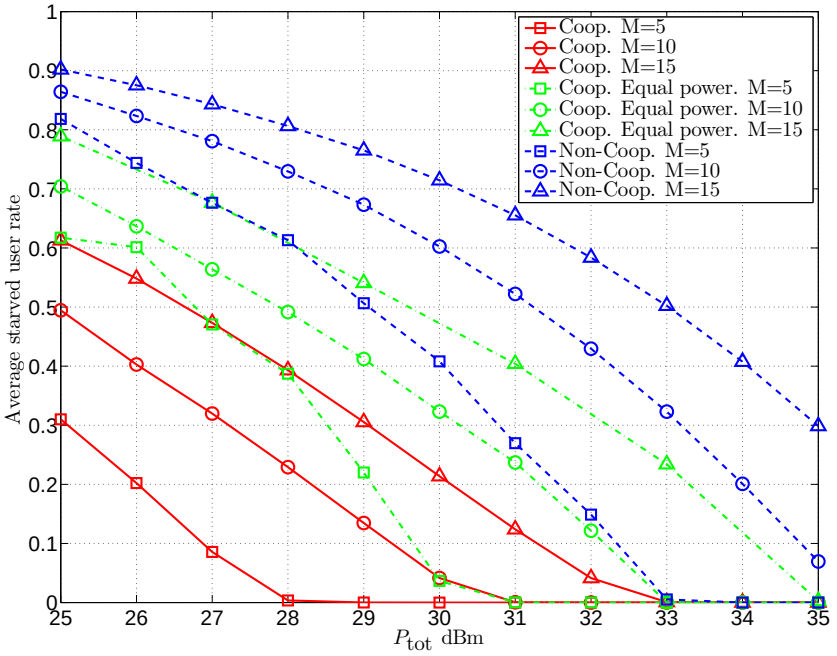


Figure 2.4: Average starved user rate versus P_{tot}

2.5.1 Throughput performance

Figures 2.3 and 2.4 show the average total throughput and the average starved user rate in the cell versus P_{tot} for $M = 5$, 10 and 15 users respectively. The JPRS algorithm outperforms the non-cooperative case and an algorithm without power allocation in terms of achievable data rate as well as average starved user rate. For instance in Fig. 2.3, for a global data rate about 20 bits/s/Hz and $M = 15$ users, the JPRS algorithm has about 3 dB power saving compared to the cooperative algorithm in [18] and more than 6 dB compared to a non-cooperative scheme. In Fig. 2.4 for $M = 5$, we notice that a minimum power of $P_{\text{tot}} = 28$ dBm is needed to satisfy all users, while the algorithm in [18], the required power is about 31 dBm and for a non-cooperative technique the required power is 33 dBm. These results show that the number of satisfied users can be significantly improved using this relaying technique compared to the non-cooperative case and can be used as design guidelines for future wireless cellular systems.

The average total throughput and average starved user rate are investigated according to the cell load, i.e. number of users, in Fig. 2.5 and 2.6 respectively, for $P_{\text{tot}} = 30$ and 35 dBm. In this case, all users are assumed to have the same data rate requirement, i.e. $\rho_m = r_2 = 2$ bits/s/Hz for all $m \in \{1, \cdots, M\}$. One can notice, as M is increasing the overall throughput of the cell is globally increasing but slowly compared to the gap in total average throughput when the available power increases, e.g. from 30 to 35 dBm. For $M = 20$ and $P_{\text{tot}} = 35$ dBm, the JPRS outperforms the algorithm in [18] about 10 bits/s/Hz, due to the power allocation offered by JPRS algorithm. In Fig. 2.6, the average starved user rate obtained with the JPRS algorithm is well below the two other algorithms. For instance, with 20 users and P_{tot}=35 dBm, all users satisfy their rate constraint using the JPRS algorithm while under non-cooperation, about 47% of users are starved. Moreover, the algorithm in [18] allows to decrease the starvation rate about 17%. By increasing the number of users to 30 and using our algorithm, 15% of users are not satisfied while the average starved user rate for non-cooperative scheme increases above 62% and is about 43% with the algorithm in [18]. The results show that our algorithm is more efficient in terms of total spectral efficiency and average starved user rate, mainly because our algorithm exploits all the available resources, including power, contrary to the algorithm in [18].

2.5.2 Power efficiency

Considering that all mobile users have the same QoS demand, fig. 2.7 studies the average minimum required power to satisfy all users versus the throughput constraint for M = 5 and 10. For a rate constraint of 2 bits/s/Hz and M=5, the average minimum power needed is about 26.5 dBm for cooperation and 32 dBm for non cooperation. Therefore, this gives an insight about the substantial power gain enabled by cooperation compared to non cooperation. Furthermore, as the throughput constraint increases, the minimum required power increases logarithmically in both cases.

Fig. 2.8 depicts the average minimum required power to satisfy all users versus relays position for

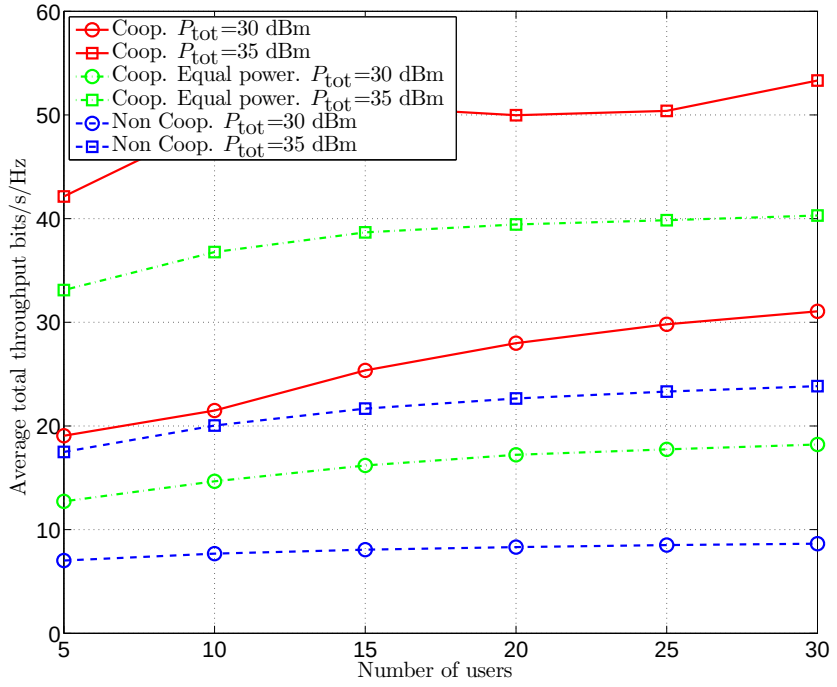


Figure 2.5: Average total throughput versus M

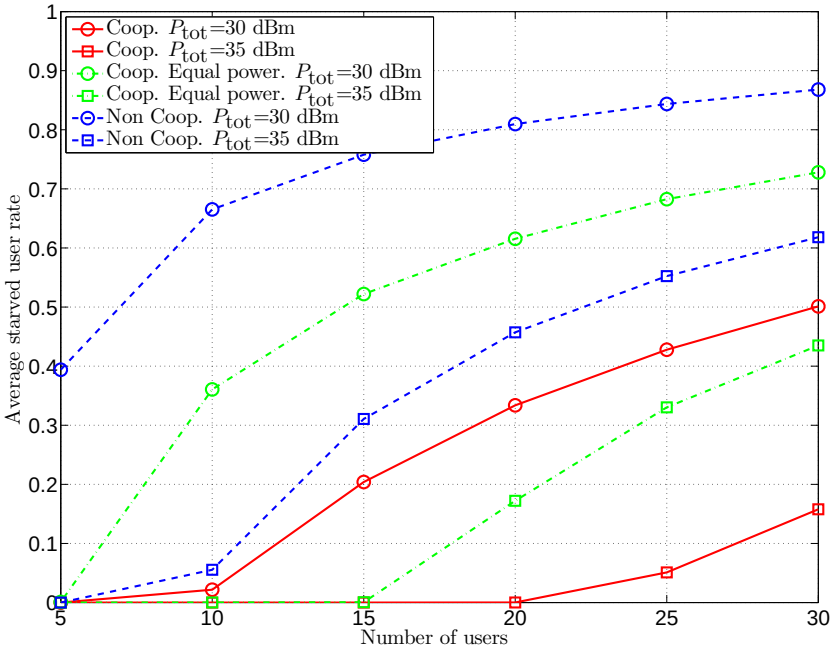


Figure 2.6: Average starved user rate versus M

M = 5 and 10. Assume also that all users have the same QoS rate constraints that could be 2 or 3 bits/s/Hz. It can be noticed that if the relays are near to BS, a high power is needed to satisfy all users. Since, at this position the direct communication is preferable. For instance, at a distance of 100 m, $M = 5$ and r_2=2 bits/s/Hz, the minimum power is 31 dBm. This power decreases to 26 dBm at 600 m. Moreover, its starts increasing as the relays position is beyond 600 m to achieve about 28 dBm at 900 m. Furthermore, changing number of users or the QoS demand, the required power increases. However, this power is always minimum at 600 m. Hence, at 600 m a minimum power consumption is achieved. Since when BS transmits, it makes a decision whether direct transmission or cooperation is preferable leading to a robust equivalent link. In this case, the relays must be nearer to the users than the BS in order to have better RS-u_m links $\forall$ m.

Therefore, these analyses provide an overview about the importance of optimal relays deployment in cellular network and show the better relays position that leads to power efficient communications.

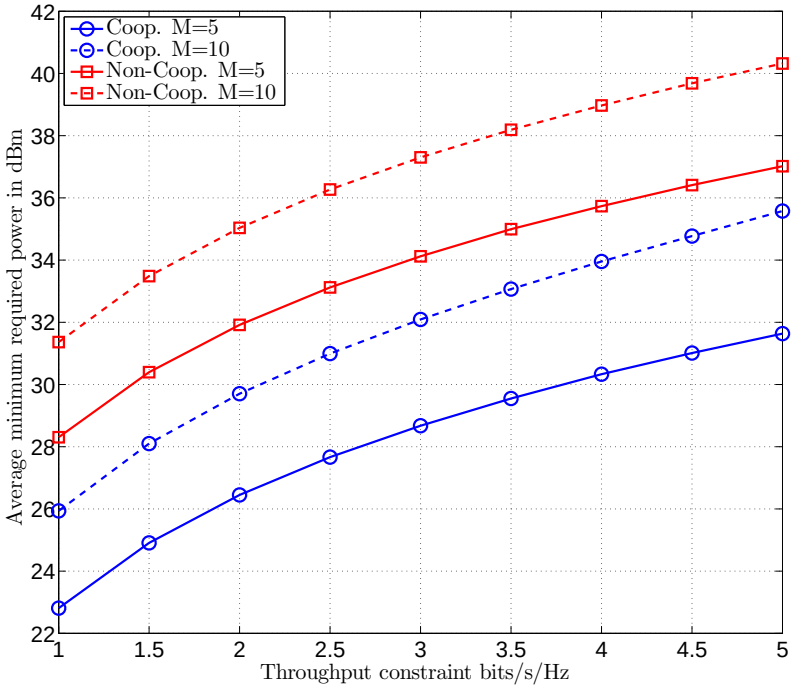


Figure 2.7: Minimum required power versus throughput constraint for M = 5 and 10

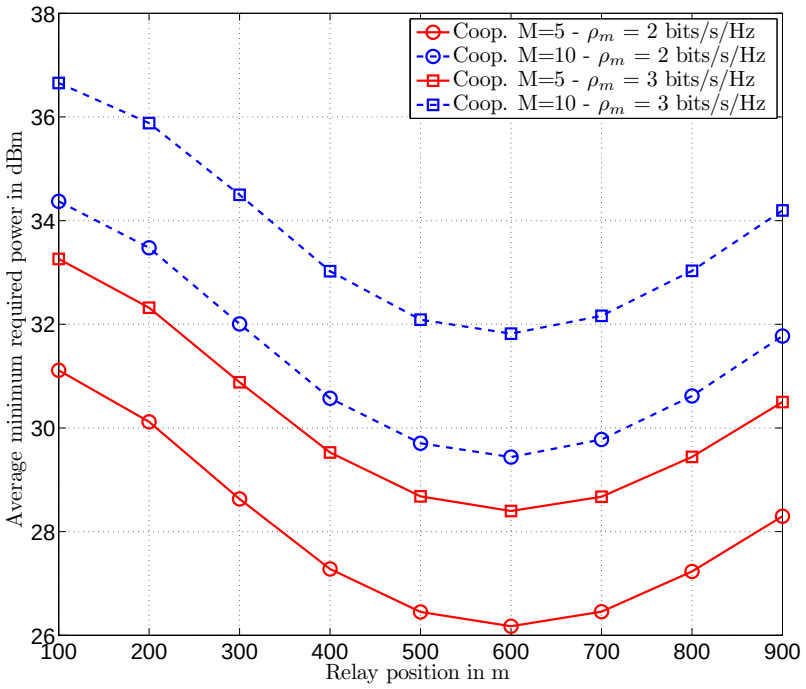


Figure 2.8: Minimum required power versus Relays position for M = 5 and 10

2.6 Conclusions and Remarks

In this chapter, we studied a throughput maximization problem for multi-user downlink OFDMA-based relay-assisted networks. We proposed a resource allocation algorithm that jointly allocates the optimal power in each link, assigns subcarriers and selects an optimal relay for every user if needed, under heterogeneous data rate constrained users. The system is also assumed to have a total power constraint and to be band-limited. Furthermore, we have studied the average starved user rate in relay-assisted network under optimal resource allocation and shown that the number of starved users is far below the starvation rate of non cooperative networks. Moreover, our proposed algorithm outperforms the algorithm in [18] in terms of total average throughput and number of starved users. It achieves about 50 bits/s/Hz where the two other strategies, i.e. a power-subcarrier non cooperative algorithm [14] delivers 20 bits/s/Hz and a cooperative algorithm without power allocation in [18] offers about 38 bits/s/Hz.

It is worthwhile to notice that, this optimization problem considers reliable links, where each user is able to decode its own data information if the transmission rate is below the channel capacity. Hence, we have several remarks:

- The proposed algorithm allocates resources from the information theoretic point of view. Therefore, it would be essential to allocate resources by considering a practical MCS and finite packet lengths.
- The retransmission mechanisms of the erroneous packets is not considered.
- This algorithm allocates all the available power.
- It maximizes optimally the overall throughput in the network without considering the static energy consumption neither at BS nor at RS.

Thereby, in the next chapters, we focus on the theoretical analysis of HARQ schemes in relay-assisted networks by considering a given MCS and finite packet lengths. Moreover, by taking into account the energy consumption at each node and for a given practical MCS, we minimize the overall energy consumption with HARQ mechanisms at the MAC layer such that the QoS demand is guaranteed.

CHAPTER

3 Performance Metrics Analysis of Hybrid-ARQ Schemes

3.1 Introduction

In the preceding chapter, we focused on a resource optimization problem considering a reliable communication in each link. However, in reality, an erroneous packet could be retransmitted several times according to an ACK feedback from the receiver side. Moreover, modern communication standards try to design cross-layer schemes that jointly optimize the available resources and considers the HARQ techniques. Therefore, this chapter is dedicated to the performance analysis of HARQ schemes in *direct* and *relay-assisted* communications.

Typically, any HARQ communication scheme can be analyzed according to the following key performance metrics:

- Packet error rate
- Average delay
- Throughput efficiency

In order to tackle these theoretical analysis, we firstly start by analyzing the aforementioned metrics for any HARQ scheme. As the performance metrics are error probability dependent, we derive a closed form expression of the error probability at any retransmission instant for HARQ-I and HARQ-CC and also for a given MCS. Thereby, the packet error rate (PER), average delay ($\overline{N}_t$), throughput efficiency (η_T) or other performance metrics can be deduced. Furthermore, after deriving the average delay, the delay outage probability in quasi-static fading channels can be analyzed.

3.2 System Model and Protocol

System model: Consider a relay-assisted scheme depicted in fig. 3.1 consisting of a source s, a relay r and a destination d. We denote by h_{ij} the complex channel realization in the $i-j$ link, where $ij \in \{sd, sr, rd\}$. We consider a block fading channel, in which the channel coefficient h_{ij} is assumed to be constant along the packet duration and vary at each retransmission instant. We assume a Rayleigh fading channel in each link, where $|h_{ij}|^2$ follows an exponential distribution and $\gamma_{ij} = p_{ij}\,|h_{ij}|^2\,/\sigma_j^2$ is the corresponding instantaneous SNR, p_{ij} is the received power in the $i-j$ link and σ_j^2 is the AWGN noise variance at j, where $j \in \{r, d\}$. Thus, the probability density function (pdf) of γ_{ij} is hence $p_{\gamma_{ij}}(\gamma_{ij}) = \frac{1}{\overline{\gamma}_{ij}} \exp\left(\frac{\gamma_{ij}}{\overline{\gamma}_{ij}}\right)$, with $\overline{\gamma}_{ij} = \mathbb{E}[\gamma_{ij}] = p_{ij}/\sigma_j^2$.

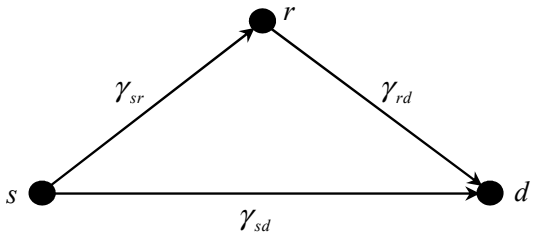


Figure 3.1: Relay-assisted scheme

Protocol: In this communication scheme, the two-hop DF relaying technique is adopted, i.e. r may start transmitting if and only if it has successfully decoded the information. The HARQ retransmission protocol is described as follows:

- In the first instant, s transmits a data packet to the intended destination d and the relay r is hearing data transmission. If d has successfully received the packet, it sends an ACK to s and r. Then, the relay would not cooperate and a new packet is transmitted by s.
- If d has not successfully received the packet, it sends a NACK packet to s and r. Then, s starts retransmitting the same packet until d or r decodes it.
- If r has successfully decoded and the destination d does not yet, r sends an ACK to s. Thereby, s stop transmitting and r starts relaying the same packet until d receives it or the maximum number of transmissions $N_{\max}$ is attained.
- If $N_{\max}$ is attained, then the current packet is dropped and a new packet is to be transmitted.

Before starting our analysis, the following system assumptions are considered:

- The ACK/NACK packets are assumed to be error-free with a negligible delay which is a reasonable assumption considering low rate and short length ACK/NACK packets [70].
- The CRC check in each packet is also error-free.

3.3 Packet error rate

Consider that the transmission has started at the instant $n = 1$ and each packet has a limited number of retransmissions $N_{\max}$. Moreover, let us define the following two events:
$\overline{A}_n$: A NACK message is received at the instant n
A_n : A positive ACK is received at the instant n
In this case, the end-to-end PER between two communicating nodes is defined as

$$\text{PER} \triangleq \Pr\left\{\bigcap_{n=1}^{N_{\max}} \overline{A}_n\right\} \tag{3.1}$$

Moreover, occurring an error at the instant n signifies that an error has also been occurred at the instants $1, 2 \cdots, n-1$, i.e. $\overline{A}_{n-1} \subset \overline{A}_n$. Therefore, $\bigcap_{n=1}^{N_{\max}} \overline{A}_n = \overline{A}_{N_{\max}}$ and the PER in (3.1) is written as:

$$\mathrm{PER} = \Pr\left\{\overline{A}_{N_{\max}}\right\} \tag{3.2}$$

Therefore, according to [71], another alternative equivalent form of PER can be written as function of the successful decoding probability:

$$\begin{aligned} \mathrm{PER} &= 1 - \Pr\left\{\bigcup_{n=1}^{N_{\max}} D_n\right\} \\ &= 1 - \sum_{n=1}^{N_{\max}} \Pr\{D_n\} \end{aligned} \tag{3.3}$$

since $\Pr\{D_n\} = \Pr\left\{\bigcap_{n=1}^{n-1} \overline{A}_n \bigcap A_n\right\} = \Pr\left\{\overline{A}_{n-1} \bigcap A_n\right\}$ and $D_m \cap D_n = \phi \ \forall m \neq n$.
Therefore, using the fact that [72]

$$\Pr\left\{\overline{A}_{n-1} \bigcap A_n\right\} = \Pr\left\{\overline{A}_{n-1}\right\} - \Pr\left\{\overline{A}_n\right\} \tag{3.4}$$

with $\Pr\left\{\overline{A}_0\right\} = 1$, the equivalency between eqs. (3.2) and (3.3) can be easily verified.

Direct communication: According to (3.2), the PER in the case of direct communication between s and d is written as:

$$\mathrm{PER} = P_{sd}\left(N_{\max}\right) \tag{3.5}$$

with $P_{sd}(n) \triangleq \Pr\left(\overline{A}_n\right)$ being the average error probability in fading channel in the $s-d$ link at the instant n. It depends on the adopted MCS.

Relay-assisted communication: Consider now that an intermediate relay r is deployed between two communicating nodes s and d as depicted in Fig. 3.1. Therefore, according to the above HARQ protocol, the PER in the case of cooperation can be defined as:

$$\mathrm{PER} \triangleq 1 - \Pr\{A_s \cup A_r\} \tag{3.6}$$

with the events A_s and A_r are respectively defined as:
A_s: d successfully decodes due to s transmissions.
A_r: d successfully decodes due to r transmissions.

Moreover, A_s and A_r are two disjoint events ($A_s \cap A_r = 0$), then the PER in (3.6) can be defined as:

$$\mathrm{PER} = 1 - (\Pr\{A_s\} + \Pr\{A_r\}) \tag{3.7}$$

$\Pr\{A_s\}$ can be expressed as:

$$\Pr\{A_s\} = \sum_{n=1}^{N_{\max}} Q_s(n) \tag{3.8}$$

with $Q_s(n)$ being the probability that d has successfully decoded at the instant n due to s transmissions when r has failed to decode the packet at the instant $n-1$. Therefore $Q_s(n)$ is written as:

$$\begin{aligned} Q_s(n) &= P_{sr}(n-1)Q_{sd}(n) \\ &= P_{sr}(n-1)\Big(P_{sd}(n-1) - P_{sd}(n)\Big) \end{aligned} \tag{3.9}$$

where $P_{sr}(n-1)$ is the probability that r has not decoded at the instant $n-1$ and $P_{sr}(0) = 1$. $Q_{sd}(n)$ being the decoding probability at d due to s transmission at the instant n.
Assuming now, that the relay r has successfully decoded at the instant n' and the destination d has not decoded yet. In this case, r starts retransmitting the same packet at the instant $n'+1$ and the probability $\Pr\{A_r\}$ can be written as:

$$\Pr\{A_r\} = \sum_{n'=1}^{N_{\max}-1} \sum_{n=n'+1}^{N_{\max}} Q_r(n, n') \tag{3.10}$$

where $Q_r(n, n')$ is the probability that d has decoded at the instant n due to the transmissions of r that has decoded at the instant n' ($n > n'$). Therefore, $Q_r(n, n')$ can be written as:

$$Q_r(n, n') = Q_{sr}(n')\Big(P_{rd}(n-1, n') - P_{rd}(n, n')\Big) \tag{3.11}$$

where $Q_{sr}(n') = P_{sr}(n'-1) - P_{sr}(n')$ is the decoding probability at r at the instant n'. Moreover, $P_{rd}(n, n')$ is the error probability at d at the instant $n > n'$ due to r that has decoded at the instant n' and $P_{rd}(n', n') = P_{sd}(n')$. The probabilities $P_{sd}(n')$ and $P_{sr}(n')$ are respectively, the error probabilities in the $s-d$ and $s-r$ links at the instant n'. Finally, in the case of cooperation, the PER can be written as:

$$\mathrm{PER} = 1 - \left(\sum_{n=1}^{N_{\max}} Q_s(n) + \sum_{n'=1}^{N_{\max}-1} \sum_{n=n'+1}^{N_{\max}} Q_r(n, n')\right) \tag{3.12}$$

where $Q_s(n)$ and $Q_r(n, n')$ depend on the probabilities $P_{sd}(n)$, $P_{sr}(n)$ and $P_{rd}(n, n')$ at the instants n and n'.

3.4 Average delay

Direct communication: The average delay in time slots (TSs) between s and d can be written as follows:

$$\begin{aligned}\overline{N}_t &= \sum_{n=1}^{N_{\max}} n.Q_{sd}(n) + N_{\max}P_{sd}(N_{\max}) \\ &= 1 + \sum_{n=1}^{N_{\max}-1} P_{sd}(n)\end{aligned} \tag{3.13}$$

where $Q_{sd}(n) = P_{sd}(n-1) - P_{sd}(n)$.

Relay-assisted communication: In relay-assisted communications, the average delay depends on the delay induced due to s transmissions, i.e. $\overline{N}_{t_s}$, and r transmissions, i.e. $\overline{N}_{t_r}$. Therefore, the total average delay for the case of cooperation can be written as:

$$\overline{N}_t = \overline{N}_{t_s} + \overline{N}_{t_r} \tag{3.14}$$

The average delay when s is transmitting is defined as:

$$\begin{aligned}\overline{N}_{t_s} = &\sum_{n=1}^{N_{\max}-1} n\Big(Q_{sd}(n)P_{sr}(n-1) + P_{sd}(n)Q_{sr}(n)\Big) \\ &+ N_{\max}\Big(Q_{sd}(N_{\max})P_{sr}(N_{\max}-1)\Big) \\ &+ N_{\max}\Big(P_{sr}(N_{\max}-1)P_{sd}(N_{\max})\Big)\end{aligned} \tag{3.15}$$

The average delay when r is transmitting can be expressed as:

$$\overline{N}_{t_r} = \sum_{n'=1}^{N_{\max}-1} Q_{sr}(n')\Bigg(\sum_{n=n'+1}^{N_{\max}} (n-n')Q_{rd}(n,n') + (N_{\max}-n')\ P_{rd}(N_{\max},n')\Bigg) \tag{3.16}$$

where n' is the instant when r has decoded the packet. $Q_{rd}(n,n')$ is the decoding probability at d due to r transmissions at the instant n when r has decoded at the instant n' and d has not. $P_{rd}(N_{\max},n')$ is the probability of not decoding at the destination after $N_{\max}$ retransmissions when r has started retransmission at the instant n'. $Q_{rd}(n)$ is written as $Q_{rd}(n) = P_{rd}(n-1,n') - P_{rd}(n,n')$. Therefore eq. (3.16) can be simplified to:

$$\overline{N}_{t_r} = \sum_{n'=1}^{N_{\max}-1} Q_{sr}(n') \sum_{n=n'}^{N_{\max}-1} P_{rd}(n,n') \tag{3.17}$$

3.5 Throughput efficiency

According to eq. (1.21), the throughput efficiency η_T can be defined as:

$$\eta_T = \frac{L(1 - \text{PER})}{\mathbb{E}[\mathcal{B}]} \tag{3.18}$$

with L being the number of information bits in a packet. $\mathbb{E}[\mathcal{B}]$ is the average number of transmitted bits until the correct reception of the packet or $N_{\max}$ is reached.

Direct communications: Using eq. (3.5), the PER can be calculated. Moreover, assuming that $L(n)$ is the number of transmitted bits until the instant n, the average number of transmitted bits can be calculated as follows:

$$\mathbb{E}[\mathcal{B}] = \sum_{n=1}^{N_{\max}} L(n).Q_{sd}(n) + L(N_{\max})P_{sd}(N_{\max}) \tag{3.19}$$

Hence, according to (3.18), η_T can be readily deduced provided that $P_{sd}(n)$ is evaluated for any instant n.

Relay-assisted communication: The average number of transmitted bits in the case of relay-assisted communications can be split into two terms as follows:

$$\mathbb{E}[\mathcal{B}] = \mathbb{E}[\mathcal{B}_s] + \mathbb{E}[\mathcal{B}_r] \tag{3.20}$$

where $\mathbb{E}[\mathcal{B}_s]$ and $\mathbb{E}[\mathcal{B}_r]$ are respectively, the average number of transmitted bits due to s and r transmissions. Therefore, $\mathbb{E}[\mathcal{B}_s]$ can be written as follows:

$$\begin{aligned}\mathbb{E}[\mathcal{B}_s] = &\sum_{n=1}^{N_{\max}-1} L(n)\Big(Q_{sd}(n)P_{sr}(n-1) + P_{sd}(n)Q_{sr}(n)\Big) \\ &+ L(N_{\max})\Big(Q_{sd}(N_{\max})P_{sr}(N_{\max}-1)\Big) \\ &+ L(N_{\max})\Big(P_{sr}(N_{\max}-1)P_{sd}(N_{\max})\Big)\end{aligned} \tag{3.21}$$

Moreover, $\mathbb{E}[\mathcal{B}_r]$ can be expressed as:

$$\mathbb{E}[\mathcal{B}_r] = \sum_{n'=1}^{N_{\max}-1} Q_{sr}(n')\Bigg(\sum_{n=n'+1}^{N_{\max}} L(n-n')Q_{rd}(n,n') + L(N_{\max}-n')\ P_{rd}(N_{\max},n')\Bigg) \tag{3.22}$$

Thereby, using eqs. (3.12), (3.21) and (3.22), η_T can be deduced for relay-assisted communications.
To best of our knowledge, the above derived performance metrics of PER, average delay and throughput

efficiency can be applied for any HARQ scheme in *direct* and *relay-assisted* communications. Furthermore, these performance metrics depend on the error probabilities $P_{sd}(n)$, $P_{sr}(n)$ and $P_{rd}(n, n')$ at the instants n and n'. Since the error probability of HARQ-IR is intractable, we restrict our analysis on ARQ, HARQ-I and HARQ-CC schemes.

Special case: In both cases (direct and relay-assisted communications), if the number of bits is constant at each retransmission instant i.e. $L(n) = \frac{L}{R_0}\ \forall n$, then $\mathbb{E}[\mathcal{B}] = \frac{L}{R_0}.\overline{N}_t$ and hence η_T is:

$$\eta_T = R_0 \frac{(1 - \text{PER})}{\overline{N}_t}, \tag{3.23}$$

which is the case of ARQ, HARQ-I and HARQ-CC schemes.

3.6 Analysis for a given MCS

3.6.1 Error probability model

In order to keep a tractable form of the theoretical analysis, the instantaneous error probability in AWGN channel for one transmission can be obtained by curve fitting as:

$$f(\gamma) = \begin{cases} a\ \exp(-g\gamma) & \text{if } \gamma \geq \gamma_M \\ 1 & \text{otherwise} \end{cases} \tag{3.24}$$

where a, g and $\gamma_M = \ln(a)/g$ are curve fitting parameters. If we denote the simulated SNRs and their corresponding PERs by $\gamma_1, \gamma_2, \cdots, \gamma_l$ and $p_1, p_2, \cdots, p_l$ respectively, the two parameters a and γ_M are calculated such that the relative quadratic error,

$$\sum_{i=1}^{l} \frac{(p_i - f(\gamma_i))^2}{p_i^2} \tag{3.25}$$

is minimized [73]. These parameters can be changed according to the packet length and the adopted MCS [74].

3.6.2 ARQ and HARQ-I

The average error probability for one transmission in the $i-j$ link (P_{ij}) can be written as:

$$\begin{aligned} P_{ij} &= \int_0^\infty f(\gamma)\, p_\gamma(\gamma)\, d\gamma \\ &= 1 - \left(\frac{g\overline{\gamma}_{ij}}{1+g\overline{\gamma}_{ij}}\right) \exp\left(-\frac{\gamma_M}{\overline{\gamma}_{ij}}\right) \end{aligned} \tag{3.26}$$

where $ij \in \{sd, sr, rd\}$.

From the principles of ARQ and HARQ-I, the decoding probability at the instant n is independent from the preceding instants. Therefore, for the case of cooperation, using eqs. (3.9) and (3.11), $Q_s(n)$ and $Q_r(n, n')$ can be respectively written as:

$$Q_s(n) = (1 - P_{sd})(P_{sd}P_{sr})^{n-1} \tag{3.27}$$

$$Q_r(n, n') = P_{sr}^{n'-1}(1 - P_{sr})P_{sd}^{n'}(P_{rd}^{n-1}(1 - P_{rd})) \tag{3.28}$$

where P_{sd}, P_{sr} and P_{rd} are computed as in eq. (3.26). As defined in eq. (3.7), the PER in relay-assisted communication depends on $\Pr\{A_s\}$ and $\Pr\{A_r\}$ that are respectively defined in eqs. (3.8) and (3.10). Therefore, inserting eq. (3.27) in (3.8), $P\{A_s\}$ can be written in closed form as follows:

$$\begin{aligned} \Pr\{A_s\} &= (1 - P_{sd}) \sum_{n=1}^{N_{\max}} (P_{sr}P_{sd})^{n-1} \\ &= (1 - P_{sd}) \left(\frac{1 - (P_{sr}P_{sd})^{N_{\max}}}{1 - P_{sr}P_{sd}}\right) \end{aligned} \tag{3.29}$$

and from eqs. (3.10) and (3.28), $P\{A_r\}$ can be computed as:

$$\begin{aligned} \Pr\{A_r\} &= P_{sd}\,(1 - P_{sr})\,(1 - P_{rd}) \sum_{n'=1}^{N_{\max}-1} \sum_{n=n'+1}^{N_{\max}} [P_{sd}P_{sr}]^{n'-1} P_{rd}^{n-1} \\ &= P_{sd}\,(1 - P_{sr}) \left[P_{rd} \left(\sum_{n'=1}^{N_{\max}-1} (P_{sd}P_{sr}P_{rd})^{n'-1}\right) - P_{rd}^{N_{\max}} \left(\sum_{n'=1}^{N_{\max}-1} (P_{sd}P_{sr})^{n'-1}\right)\right] \\ &= P_{sd}\,(1 - P_{sr}) \left[P_{rd} \frac{1 - (P_{sd}P_{sr}P_{rd})^{N_{\max}-1}}{1 - P_{sd}P_{sr}P_{rd}} - P_{rd}^{N_{\max}} \frac{1 - (P_{sd}P_{sr})^{N_{\max}-1}}{1 - P_{sd}P_{sr}}\right] \end{aligned} \tag{3.30}$$

and hence the PER can be written in closed form according to eqs. (3.29) and (3.30).
Moreover, the average delay in HARQ-I based relaying network can also be written in closed form. By

replacing eq. (3.27) in eq. (3.15), the average delay $\overline{N}_{t_s}$ is given by:

$$\begin{aligned}\overline{N}_{t_s} &= (1-P_{sd}P_{sr})\sum_{n=1}^{N_{\max}-1} n(P_{sd}P_{sr})^{(n-1)} + N_{\max}(P_{sd}P_{sr})^{(N_{\max}-1)} \\ &= \frac{1-(P_{sd}P_{sr})^{N_{\max}-1}}{1-P_{sd}P_{sr}} + (P_{sd}P_{sr})^{(N_{\max}-1)} \\ &= \frac{1-(P_{sd}P_{sr})^{N_{\max}}}{1-P_{sd}P_{sr}}\end{aligned} \tag{3.31}$$

Furthermore, the average delay $\overline{N}_{t_r}$ can be computed by plugging $Q_{sr}(n') = (1-P_{sr})P_{sr}^{n'-1}$ and $P_{rd}(n,n') = P_{sd}^{n'}P_{rd}^{n-n'}$ in eq. (3.17) as follows:

$$\begin{aligned}\overline{N}_{t_r} &= \frac{P_{sr}}{(1-P_{sr})}\left(\sum_{n'=1}^{N_{\max}-1}\left(\frac{P_{sd}P_{sr}}{P_{rd}}\right)^{n'}\sum_{n=n'}^{N_{\max}-1}P_{rd}^{n}\right) \\ &= \frac{1-P_{sr}}{P_{sr}(1-P_{rd})}\left(\sum_{n'=1}^{N_{\max}-1}(P_{sd}P_{sr})^{n'} - P_{rd}^{N_{\max}}\sum_{n'=1}^{N_{\max}-1}\left(\frac{P_{sd}P_{sr}}{P_{rd}}\right)^{n'}\right) \\ &= \frac{1-P_{sr}}{P_{sr}(1-P_{rd})}\left(\frac{P_{sd}P_{sr}-(P_{sd}P_{sr})^{N_{\max}}}{1-P_{sd}P_{sr}} - P_{rd}^{N_{\max}}\left(\frac{\frac{P_{sd}P_{sr}}{P_{rd}}-(\frac{P_{sd}P_{sr}}{P_{rd}})^{N_{\max}}}{1-\frac{P_{sd}P_{sr}}{P_{rd}}}\right)\right)\end{aligned} \tag{3.32}$$

Therefore, the total delay can be written in closed form as a sum of $\overline{N}_{t_s}$ and $\overline{N}_{t_r}$ given respectively in eqs. (3.31) and (3.32). Moreover, η_T defined in (3.23) can also be deduced in closed form by using eqs. (3.31) and (3.32).

3.6.3 HARQ-CC

As mentioned in the preceding sections, all performance metrics depend on the error probabilities $P_{sd}(n)$, $P_{sr}(n)$ and $P_{rd}(n,n')$ at the instants n and n'. Hence, for HARQ-CC, the error probabilities at any instant have to be derived.

Direct communication: In point-to-point communication (p2p), the error probability $P(n)$ in HARQ-CC at the instant n can be calculated as [75]

$$P(n) = \int_0^\infty \cdots \int_0^\infty f(\gamma_1)f(\gamma_1+\gamma_2)\cdots f(\gamma_1+\cdots+\gamma_n)\times p_{\gamma_{sd}}(\gamma_1)p_{\gamma_{sd}}(\gamma_2)\cdots p_{\gamma_{sd}}(\gamma_n)d\gamma_1 d\gamma_2\cdots d\gamma_n \tag{3.33}$$

where $p_{\gamma_{sd}}(\gamma_n) = \frac{1}{\bar{\gamma}_{sd}}\exp\left(-\frac{\gamma_n}{\bar{\gamma}_{sd}}\right)$ is the pdf of the instantaneous SNR at the instant n, i.e. γ_n. $\bar{\gamma}_{sd}$ is the average SNR between s and d. Moreover, detecting an error at the instant n, signifies that at the instants $1,2,\cdots,n-1$ an error has certainly occurred. Therefore, $f(\gamma_1)f(\gamma_1+\gamma_2)\cdots f(\gamma_1+\cdots+\gamma_{n-1}) \approx 1$ and

hence, eq. (3.33) can be approximated by

$$P(n) \approx \int_0^\infty \int_0^\infty \cdots \int_0^\infty f(\gamma_1+\gamma_2+\cdots+\gamma_n) \times p_{\gamma_{sd}}(\gamma_1)p_{\gamma_{sd}}(\gamma_2)\cdots p_{\gamma_{sd}}(\gamma_n)d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{3.34}$$

This approximation is tight enough to measure the performance metrics as we will see later. Following similar steps as done in [75], (3.34) can be split into three parts as:

$$P(n) = A_n + \sum_{m=1}^{n-1} B_{n,m} + C_n \tag{3.35}$$

Where A_n, $B_{n,m}$ and C_n are respectively written as follows:

$$A_n = \int_0^{\gamma_M} \int_0^{\gamma_M-\gamma_1} \cdots \int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n-1}} p_{\gamma_{sd}}(\gamma_1)p_{\gamma_{sd}}(\gamma_2)\cdots p_{\gamma_{sd}}(\gamma_n)d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{3.36}$$

$$B_{n,m} = \int_0^{\gamma_M} \cdots \int_0^{\gamma_M-\gamma_1-\cdots-\gamma_{m-1}} \int_{\gamma_M-\gamma_1-\cdots-\gamma_m}^\infty \cdots \int_0^\infty a\exp\big(-g(\gamma_1+\cdots+\gamma_n)\big) \times p_{\gamma_{sd}}(\gamma_1)p_{\gamma_{sd}}(\gamma_2)\cdots p_{\gamma_{sd}}(\gamma_n)d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{3.37}$$

$$C_n = \int_{\gamma_M}^\infty \int_0^\infty \cdots \int_0^\infty a\exp\big(-g(\gamma_1+\cdots+\gamma_n)\big)p_{\gamma_{sd}}(\gamma_1)p_{\gamma_{sd}}(\gamma_2)\cdots p_{\gamma_{sd}}(\gamma_n)d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{3.38}$$

Therefore, after some cumbersome calculations, A_n, $B_{n,m}$ and C_n can be expressed as:

$$A_n = 1 - \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)\sum_{k=0}^{n-1}\frac{1}{k!}\left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^k \tag{3.39}$$

$$B_{n,m} = \left(\frac{{\gamma_M}^m}{m!}\right)\left[\frac{1}{1+g\bar{\gamma}_{sd}}\right]^{n-m}\left(\frac{1}{\bar{\gamma}_{sd}}\right)^m \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \tag{3.40}$$

$$C_n = \left[\frac{1}{1+g\bar{\gamma}_{sd}}\right]^n \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \tag{3.41}$$

The derivation details of A_n, $B_{n,m}$ and C_n are reported in the appendices A.1, A.2 and A.3 respectively. Hence, substituting eqs. (3.39), (3.40) and (3.41) in (3.35), $P(n)$ can be calculated.

Relay-assisted communication: Let us assume the case of relay-assisted communication in which the relay r can help for the communication between s and d. The derivations of $P_{sd}(n)$ and $P_{sr}(n)$ can

be deduced from the preceding section. However, $P_{rd}(n,n')$ can be expressed as:

$$P_{rd}(n,n') = \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n' \text{ folds}} \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n-n' \text{ folds}} f(\gamma_1) f(\gamma_1+\gamma_2)\cdots \quad \times f(\gamma_1+\gamma_2+\cdots+\gamma_n) \times p_{\gamma_{sd}}(\gamma_1)\cdots p_{\gamma_{sd}}(\gamma_{n'}) \times p_{\gamma_{rd}}(\gamma_{n'+1})\cdots p_{\gamma_{rd}}(\gamma_n) d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{3.42}$$

Following the similar approximation method adopted previously, eq. (3.42) can be written as:

$$P_{rd}(n,n') \approx \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n' \text{ folds}} \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n-n' \text{folds}} f(\gamma_1+\gamma_2+\cdots+\gamma_n) \times p_{\gamma_{sd}}(\gamma_1)\cdots p_{\gamma_{sd}}(\gamma_{n'}) \times p_{\gamma_{rd}}(\gamma_{n'+1})\cdots p_{\gamma_{rd}}(\gamma_n) d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{3.43}$$

Moreover, using [75], eq. (3.43) can be split into three parts as follows:

$$P_{rd}(n,n') = A(n,n') + \sum_{m=1}^{n-1} B(n,m,n') + C(n,n') \tag{3.44}$$

where $A(n,n')$ is written as:

$$A(n,n') = \int_0^{\gamma_M} \cdots \int_0^{\gamma_M-\gamma_1\cdots\gamma_{n-1}} p_{\gamma_{sd}}(\gamma_1)\cdots p_{\gamma_{sd}}(\gamma_{n'}) p_{\gamma_{rd}}(\gamma_{n'+1})\cdots p_{\gamma_{rd}}(\gamma_n) d\gamma_1 \cdots d\gamma_n \tag{3.45}$$

Thus, from appendix A.4, $A(n,n')$ is proved to be written as:

$$A(n,n') = \varphi_1 - \varphi_2, \tag{3.46}$$

with φ_1 and φ_2 are respectively given by:

$$\varphi_1 = 1 - \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \sum_{k=0}^{n'-1} \frac{1}{k!} \left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^k \tag{3.47}$$

$$\varphi_2 = \frac{1}{\Gamma(n')} \left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^{n'} \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{rd}}\right) \sum_{k=0}^{n-n'-1} \frac{1}{k!} \left(\frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^k Z_1(k) \tag{3.48}$$

where $Z_1(k)$ is:

$$Z_1(k) = \int_0^1 \exp\left(\left(\frac{\gamma_M}{\bar{\gamma}_{rd}} - \frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \cdot x\right) x^{n'-1} (1-x)^k dx \tag{3.49}$$

Using the binomial theorem $(1-x)^k = \sum_{i=0}^{k} \binom{k}{i}(-1)^i x^i$, $Z_1(k)$ can be written as [76, eq. (11) §2.33, pp. 108]:

$$Z_1 = \sum_{i=0}^{k} \binom{k}{i} (-1)^i \frac{(n'+i-1)!}{\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^{(n'+i)}} \times \left[1 - \exp\left(-\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)\right) \sum_{j=0}^{n'+i-1} \frac{\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^j}{j!}\right] \tag{3.50}$$

Moreover, $B(n, n', m)$ can be written as follows:

$$B(n,n',m) = \int_0^{\gamma_M} \cdots \int_0^{\gamma_M - \gamma_1 - \cdots \gamma_{m-1}} \int_{\gamma_M - \cdots \gamma_m}^{\infty} \int_0^{\infty} \cdots \int_0^{\infty} p_{\gamma_{sd}}(\gamma_1) \cdots p_{\gamma_{sd}}(\gamma_{n'}) \times p_{\gamma_{rd}}(\gamma_{n'+1}) \cdots p_{\gamma_{rd}}(\gamma_n) \times a \exp\left(-g\left(\gamma_1 + \cdots + \gamma_n\right)\right) d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{3.51}$$

Therefore, from appendix A.5, if $n' \geq m+1$, $B(n, n', m)$ becomes:

$$B(n,n',m) = \frac{1}{m!}\left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^m \left[\frac{1}{1+g\bar{\gamma}_{rd}}\right]^{n-n'} \left[\frac{1}{1+g\bar{\gamma}_{sd}}\right]^{n'-m} \times \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \tag{3.52}$$

If $n' < m+1$, we have

$$B(n,n',m) = \left(\frac{1}{\bar{\gamma}_{sd}}\right)^{n'} \left(\frac{1}{\bar{\gamma}_{rd}}\right)^{n-n'} \left[\frac{1}{g + \frac{1}{\bar{\gamma}_{rd}}}\right]^{n-m} \times \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{rd}}\right) \frac{\gamma_M^m}{(m-n')!\Gamma(n')} Z_2 \tag{3.53}$$

where Z_2 is given by:

$$Z_2 = \int_0^1 (1-x)^{m-n'} x^{n'-1} \exp\left(-\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right) x\right) dx \tag{3.54}$$

Applying binomial theorem, Z_2 can be obtained using the integral and series identities in [76, eq. 11, pp. 108]:

$$Z_2 = \sum_{j=0}^{m-n'} \binom{m-n'}{j} (-1)^j \times \int_0^1 x^{j+n'-1} \exp\left(-\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right) x\right) dx \tag{3.55}$$

$$= \sum_{j=0}^{m-n'} \binom{m-n'}{j} (-1)^j \frac{(j+n'-1)!}{\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^{(j+n')}} \times \left[1 - \exp\left(-\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)\right) \sum_{i=0}^{j+n'-1} \frac{\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^i}{i!}\right] \tag{3.56}$$

Finally $C(n, n')$ is written as follows:

$$C(n,n') = \underbrace{\int_{\gamma_M}^{\infty} \cdots \int_{0}^{\infty}}_{n' \text{folds}} \underbrace{\int_{0}^{\infty} \cdots \int_{0}^{\infty}}_{n-n' \text{folds}} p_{\gamma_{sd}}(\gamma_1) \cdots p_{\gamma_{sd}}(\gamma_{n'}) p_{\gamma_{rd}}(\gamma_{n'+1}) \cdots p_{\gamma_{rd}}(\gamma_n) \times a \exp\left(-g(\gamma_1 + \cdots + \gamma_n)\right) d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{3.57}$$

After calculations, $C(n, n')$ is quantified in closed form as:

$$C(n,n') = \left[\frac{1}{1+g\bar{\gamma}_{sd}}\right]^{n'} \left[\frac{1}{1+g\bar{\gamma}_{rd}}\right]^{n-n'} \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \tag{3.58}$$

Hence, inserting (3.46), (3.53) and (3.58) in (3.44), we can calculate $P_{rd}(n, n')$ and the probabilities $P_{sd}(n)$ and $P_{sr}(n)$ are calculated from (3.35). Therefore, the PER, average delay and η_T for HARQ-CC is easily deduced for any given MCS in direct and relay-assisted communications.

3.7 Numerical analysis

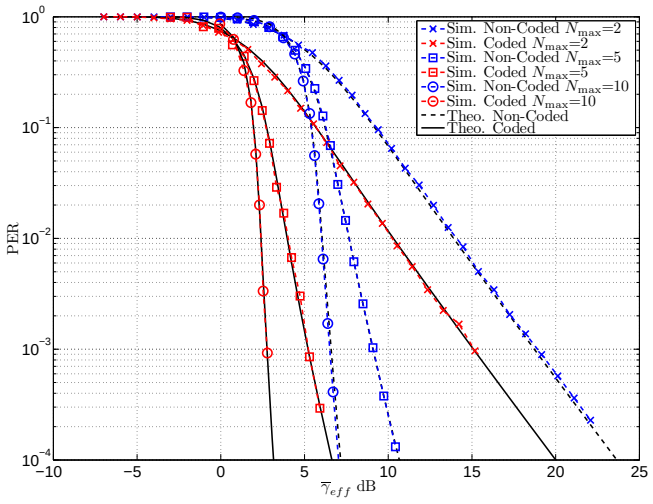
In our simulation settings, we consider that s communicates with d either by a direct transmission or by the assistance of the relay r that is located at a distance $d_r = 500$ m from BS. The angle between the line joining s to d and s to r is $\theta_r = \pi/6$ radians. The other system parameters are detailed in table 3.1. The propagation path-loss model in dB is $139.90 + 34.41.\log_{10}(d_{ij})$ in the $i-j$ link, where d_{ij} is the distance in km [69]. The cases where $N_{\max} = [2, 5, 10]$ TSs are compared and the adopted modulation scheme is BPSK. Moreover, we consider two cases: the coded and the non-coded schemes by means of LDPC code of rate $R_0 = \frac{1}{2}$, where the curve fitting parameters for both schemes are shown in table 3.2.

Parameters	Description	value
N_0	noise power spectral density	-155 dBm/Hz
L	number of information bits	102
R_b	transmission bit rate	150 kbps
f_c	carrier frequency	2.4 GHz

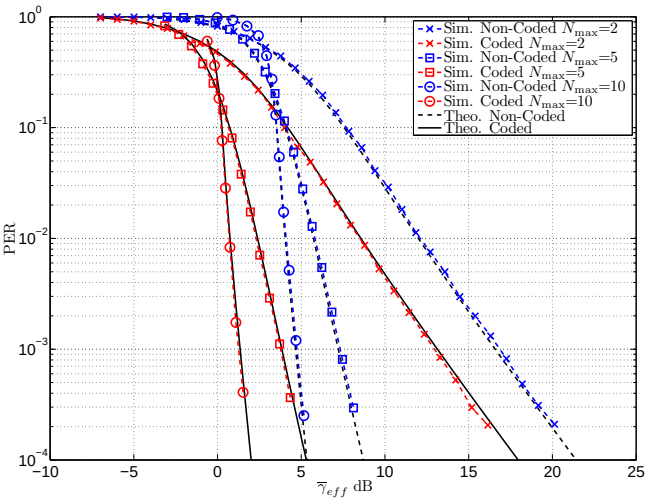
Table 3.1: System model parameters

MCS	R_c	a	g	$\gamma_M = \log(a/g)$
BPSK	1	8.117	0.998	2.0966
BPSK (LDPC)	1/2	4116.88	6.680	1.2458

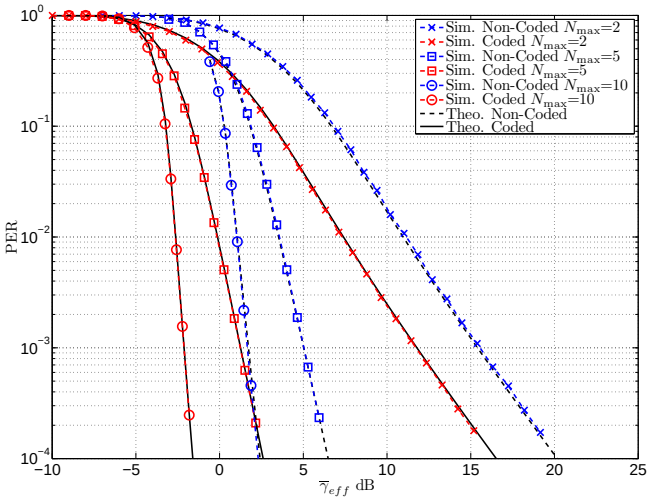
Table 3.2: Curve fitting parameters



(a) p2p HARQ-CC scheme



(b) Cooperative HARQ-I scheme



(c) Cooperative HARQ-CC scheme

Figure 3.2: PER for coded and non-coded p2p and cooperative HARQ schemes versus $\bar{\gamma}_{eff}$ for $N_{\max} = 2, 5$ and 10

3.7.1 PER versus SNR

Figs. 3.2a, 3.2b and 3.2c show the theoretical and the simulated PER versus the average effective SNR that is defined as $\bar{\gamma}_{eff} = \bar{\gamma}_{sd} \cdot \overline{N}_t$, for p2p HARQ-CC, cooperative HARQ-I and cooperative HARQ-CC schemes respectively.

In the three communication schemes, it can be noticed that our theoretical approximations of PER for

coded and non-coded schemes match tightly the simulated ones and hence this confirms our theoretical findings. We can further observe that increasing the number of allowed retransmissions, the packet error rate enhances significantly, i.e. the system reliability increases. Moreover, comparing coded scheme to non-coded one, we can notice that the coding gain is always between 3.5 and 4 dB for all schemes. Therefore, $N_{\max}$ has an influence on the diversity gain.

Considering a non-coded scheme and at $N_{\max} = 2$, the effective SNR $\bar{\gamma}_{eff}$ to obtain a PER $= 10^{-3}$ is about 18.72, 16.5 and 15.4 dB for p2p HARQ-CC, cooperative HARQ-I and cooperative HARQ-CC schemes respectively. Hence, cooperative HARQ-CC provides a gain of about 3.32 dB and 1.18 dB w.r.t. p2p HARQ-CC and cooperative HARQ-I respectively. This shows that cooperative HARQ-CC has the better performance in terms of PER. Indeed, packet combining enables a substantial time diversity compared to simple HARQ-I that drops the erroneous packets. Furthermore, cooperation offers a significant spatial diversity w.r.t. to p2p communication. For $N_{\max} = [2, 5, 10]$, the average SNRs required to achieve a PER $= 10^{-3}$ are summarized in table 3.3.

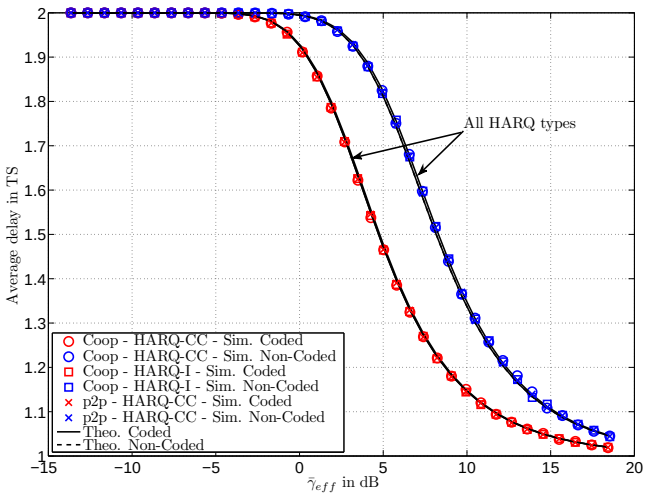
From table 3.3, it can be noticed that the higher $N_{\max}$, the better SNR gain is. Moreover, a significant gain is induced from cooperative HARQ-CC compared to p2p HARQ-CC. This is due to the path-loss gain that leads to a power saving. Furthermore, the cooperative HARQ-I outperforms the p2p HARQ-CC scheme. Thus, the path-loss gain induced by cooperation is more important than the temporal diversity offered by the p2p HARQ. However, using cooperation and comparing HARQ-I to HARQ-CC, a substantial SNR gain can also be achieved due to the packet combining with the previous erroneous packets and hence it increases the probability of successful decoding of each transmitted packet.

MCS	$N_{\max}$	2	5	10
BPSK - $R_c = 1$	Coop. CC vs. p2p CC	3.32	4.08	4.87
	Coop. CC vs. Coop. I	1.18	2.21	2.93
	Coop. I vs. p2p CC	2.14	1.87	1.94
BPSK - $R_c = \frac{1}{2}$	Coop. CC vs. p2p CC	2.84	3.96	4.88
	Coop. CC vs. Coop. I	0.87	2.59	3.47
	Coop. I vs. p2p CC	1.97	1.37	1.41

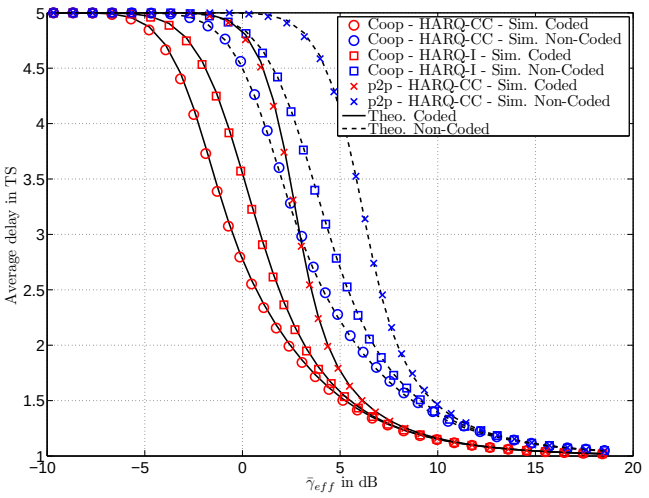
Table 3.3: SNR gain in dB for different schemes at PER $= 10^{-3}$

Finally, we can conclude the following:

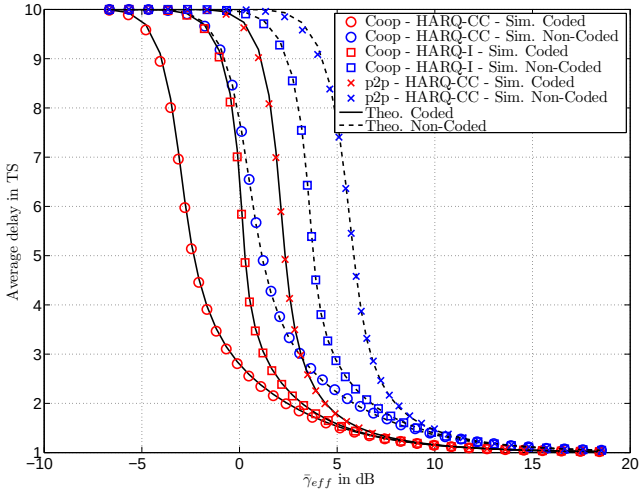
- As $N_{\max}$ increases, the system reliability and the SNR gain enhance significantly.
- Cooperation enhances the system reliability.
- Cooperative HARQ-CC outperforms p2p HARQ-CC and cooperative HARQ-I.



(a) $N_{\max} = 2$



(b) $N_{\max} = 5$



(c) $N_{\max} = 10$

Figure 3.3: Theoretical and simulated average delay in time slots (TS) versus γ_{eff} in dB for coded and non-coded cooperative HARQ-I, cooperative HARQ-CC and p2p HARQ-CC schemes and for $N_{\max} = [2, 5, 10]$

3.7.2 Average delay versus γ_{eff}

Figs. 3.3a, 3.3b and 3.3c display the theoretical and the simulated average delay in TS versus γ_{eff} for p2p HARQ-CC, cooperative HARQ-I and cooperative HARQ-CC schemes, respectively for $N_{\max} =$ 2, 5 and 10. The theoretical approximation matches perfectly the simulated one. As expected, as γ_{eff} increases, the average delay of the system decreases and it converges to 1 at high SNR.

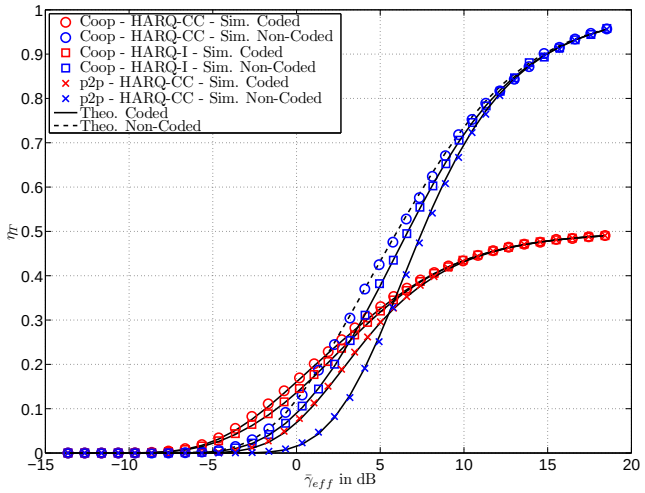
For $N_{\max} = 2$, fig. 3.3a shows that the average delay induced by the three schemes is approximately the same. However, introducing a channel code, the average delay enhances and we have a gain of about 3.5 dB in the transition region. But, the average delay is also the same for all schemes. Moreover at high γ_{eff}, the SNR gain between coded and non-coded schemes degrades significantly.
Now, increasing $N_{\max}$ to 5 or 10 TSs (figs. 3.3b and 3.3c), it can be stated that a substantial gain is enabled by the cooperative HARQ-CC scheme compared to cooperative HARQ-I and p2p HARQ-CC schemes. Moreover, the simple HARQ-I has a better performance w.r.t. p2p HARQ-CC. With reference to figs. 3.3a, 3.3b and 3.3c it can be clearly noticed that the average delay decreases sharply for all schemes as $N_{\max}$ increases. Therefore, cooperative HARQ-CC has the better performance in terms of the average delay, since it jointly exploits the spatial and temporal diversities offered by the relaying and packet combining techniques respectively.

3.7.3 η_T versus γ_{eff}

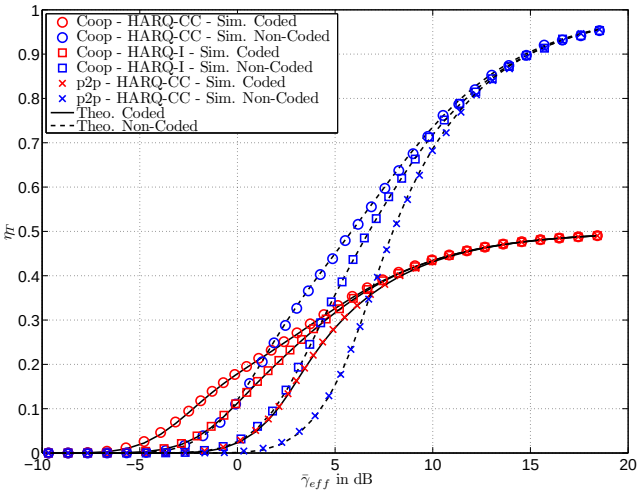
Figs. 3.4a, 3.4b and 3.4c depict the throughput efficiency η_T versus γ_{eff} respectively for $N_{\max} = 2$, 5 and 10. It can be clearly seen that our theoretical analysis are also tight enough to compute η_T. Moreover, the cooperative HARQ-CC outperforms the p2p HARQ-CC and the cooperative HARQ-I for both coded and non-coded schemes. Now comparing the coded scheme to the non-coded one, the coding gain can be seen at low SNR region. However, if one built a system operating at $N_{\max} = 2$ and if throughput efficiency is less than 0.2, the coded cooperative HARQ-CC scheme enables better throughput efficiency. Furthermore, if the requested throughput is greater than 0.2, the coded scheme has a negative influence on the throughput efficiency and hence a higher code rate must be used. However, for cooperative HARQ-I and p2p HARQ-CC schemes, the coded curve intersects with the non-coded one respectively at a $\gamma_{eff} = 3$ and 3.5 dB which also correspond respectively to a throughput efficiency of 0.25 and 0.32.
Moreover, increasing $N_{\max}$ to 5 or 10 respectively in figs. 3.4b and 3.4c, the enabled system throughput

Scheme	$N_{\max}$	(SNR , η_T)
p2p HARQ-CC	2	(5.7 , 0.32)
	5	(6.8 , 0.36)
	10	(6.85 , 0.36)
Coop. HARQ-I	2	(3.1 , 0.25)
	5	(4.25 , 0.29)
	10	(4.5 , 0.31)
Coop. HARQ-CC	2	(2 , 0.22)
	5	(1.5 , 0.22)
	10	(1.4 , 0.22)

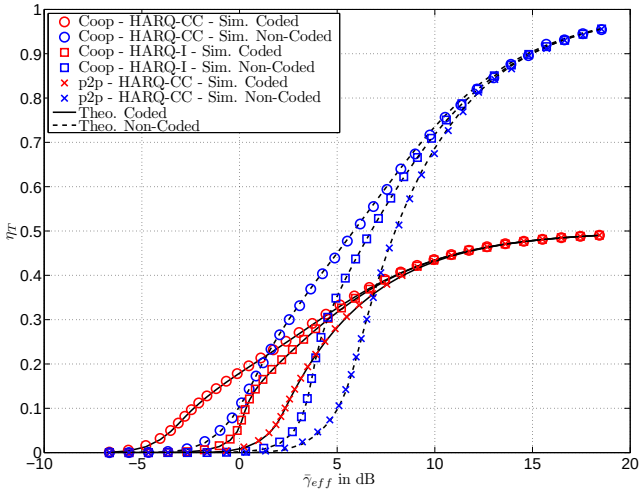
Table 3.4: Intersection points between coded and non-coded HARQ schemes η_T



(a) $N_{\max} = 2$



(b) $N_{\max} = 5$



(c) $N_{\max} = 10$

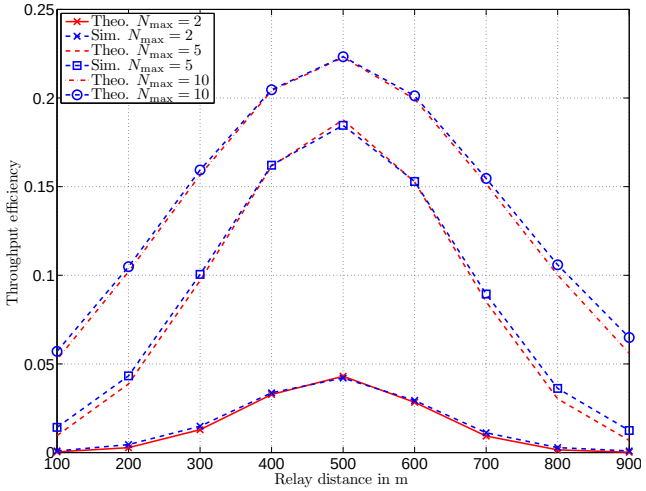
Figure 3.4: Theoretical and simulated throughput efficiency (η_T) versus γ_{eff} in dB for coded and non-coded Cooperative HARQ-I, Cooperative HARQ-CC and p2p HARQ-CC schemes and for $N_{\max} = [2, 5, 10]$

enhances significantly and table 3.4 shows the intersection points between coded and non-coded curves for the same scheme. It can be further shown that if the demanded throughput is greater than the one provided in table 3.4 then the coded scheme may be detrimental for the system throughput efficiency.

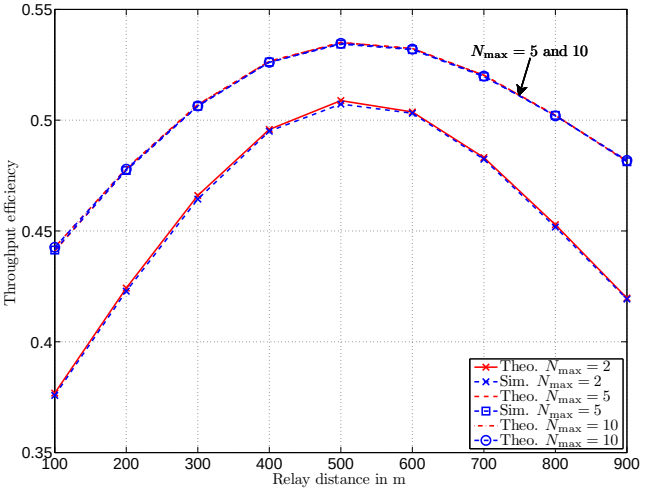
This drastic degradation in the throughput efficiency shows that coded HARQ is code rate dependent, in which at high SNR, the achievable throughput efficiency is equal to the coding rate R_0. Since, by

referring to eq. (3.18) and at high SNR, the PER tends to zero and $\overline{N}_t$ tends to 1 and hence $\eta_T \to R_0$. However, the throughput efficiency for non-coded scheme reaches to 1 at high SNR. Thus, this shows the main drawback of HARQ schemes with fixed coding rate which imposes a limitation on the system throughput efficiency.

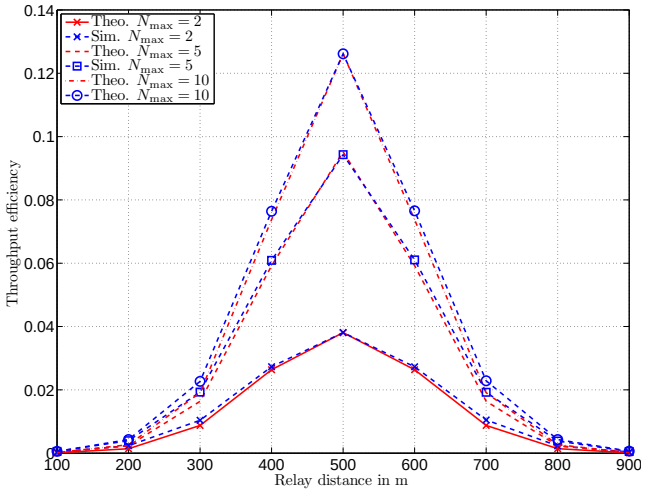
3.7.4 η_T versus d_r



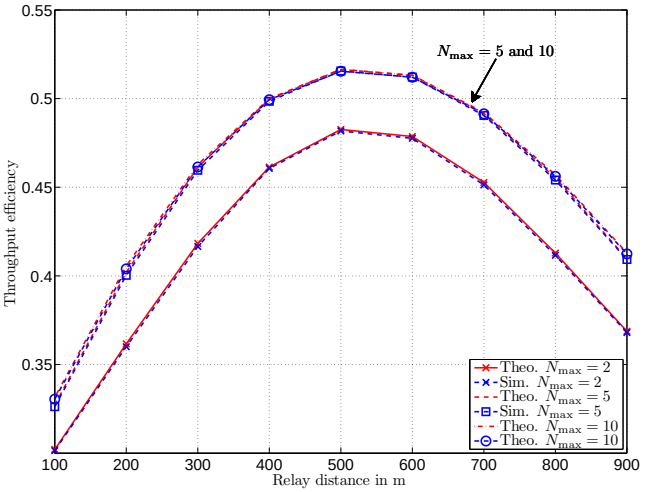
(a) HARQ-CC scheme of $P_t = 35$ dBm



(b) HARQ-CC scheme of $P_t = 45$ dBm



(c) HARQ-I scheme of $P_t = 35$ dBm



(d) HARQ-I scheme of $P_t = 45$ dBm

Figure 3.5: Theoretical and simulated η_T versus relay position d_r for non-coded cooperative HARQ-I and cooperative HARQ-CC schemes. The transmission powers are $P_t = 35$ and 45 dBm and $N_{\max} \in [2, 5, 10]$

In this section, we consider a non-coded BPSK modulation with s, r and d are collinear. The destination d is fixed at a distance 1000 m from s. The relay distance d_r is assumed to vary in the interval $[100, 900]$ m. Therefore, fig. 3.5 displays the theoretical and the simulated throughput efficiency η_T w.r.t. d_r for HARQ-CC and HARQ-I and $P_t = 35$ and 45 dBm. At $P_t = 35$ dBm the throughput efficiency reaches its maximum at $d_r = 500$ m for both schemes. However, increasing $N_{\max}$, η_T increases to achieve 0.12 and 0.25 at $N_{\max} = 10$ for HARQ-I and HARQ-CC schemes respectively. Thus, using HARQ-CC, η_E is doubled and hence HARQ-CC has a better performance compared to HARQ-I.

Now, increasing P_t to 45 dBm, η_T achieves also its maximum at $d_r = 500$ m but it has a smooth variation compared to the sharp variation observed in figs. 3.5a and 3.5c, i.e. for $P_t = 35$ dBm. Moreover, the maximum of η_T that can be achieved at $N_{\max} = 5$ or 10 is 0.54 and 0.52 for HARQ-CC and HARQ-I respectively. Thus the throughput gain of HARQ-CC relative to HARQ-I degrades. Even though, increasing $N_{\max}$ to 10 has no more influence on the throughput gain of the same scheme. Since, at high transmission power, the probability of decoding successfully a packet at the first instant increases and hence the HARQ-I and HARQ-CC have the same performance.

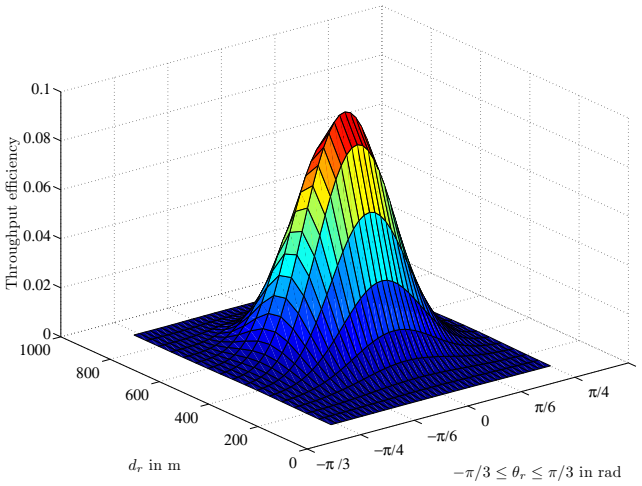
In short, we can conclude that in low and medium transmission power regions, the HARQ-CC scheme has a better performance compared to HARQ-I. However, at high transmission power, this gain is no longer achievable. Moreover, choosing the better $N_{\max}$ and allocating the optimal transmission power P_t could be beneficial for system throughput efficiency.

3.7.5 η_T versus d_r and θ_r

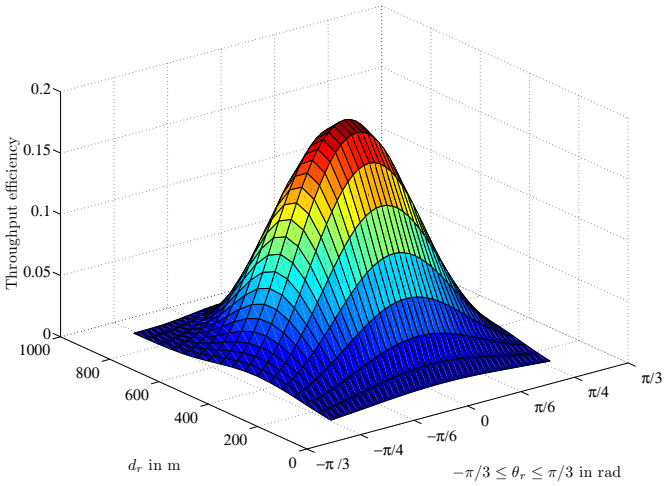
Consider that the transmission power is $P_t = 35$ dBm and the system operates at $N_{\max} = 5$. Using our theoretical analysis, figs. 3.6a and 3.6b display η_T versus the relay distance d_r in m and the relaying angle θ_r in rad for HARQ-I and HARQ-CC respectively. From figs. 3.6a and 3.6b, it can be seen that for both HARQ schemes, the maximum of η_T can be achieved if $(d_r, \theta_r) = (500, 0)$, i.e. s, r and d are collinear and r is in the middle. However, the obtained throughput efficiency is 0.08 and 0.15 for HARQ-I and HARQ-CC respectively.
Since the delivered η_T is very low, the transmission power P_t is changed to 45 dBm and η_T for the HARQ-I and HARQ-CC are displayed in figs. 3.6c and 3.6d respectively. Therefore, the delivered throughput for both schemes becomes more flat and also the maximum achieved throughput is about 0.5 for HARQ-I scheme and 0.55 for HARQ-CC scheme. Moreover, the minimum achieved throughput for HARQ-CC is about 0.45 at the border positions, in which the HARQ-I scheme delivers about 0.3.

Now, in order to compare HARQ-CC to HARQ-I scheme, figs. 3.7a and 3.7b depict the relative gain of HARQ-CC scheme compared to HARQ-I scheme respectively for $P_t = 35$ and 45 dBm. The relative gain is defined as follows:

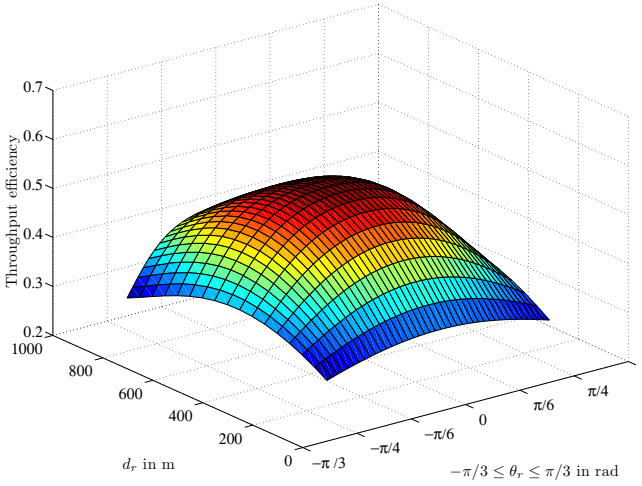
$$G_{\eta_T} = \frac{\left|\eta_{T_{(\text{HARQ-CC})}} - \eta_{T_{(\text{HARQ-I})}}\right|}{\eta_{T_{(\text{HARQ-CC})}}} \times 100\% \tag{3.59}$$



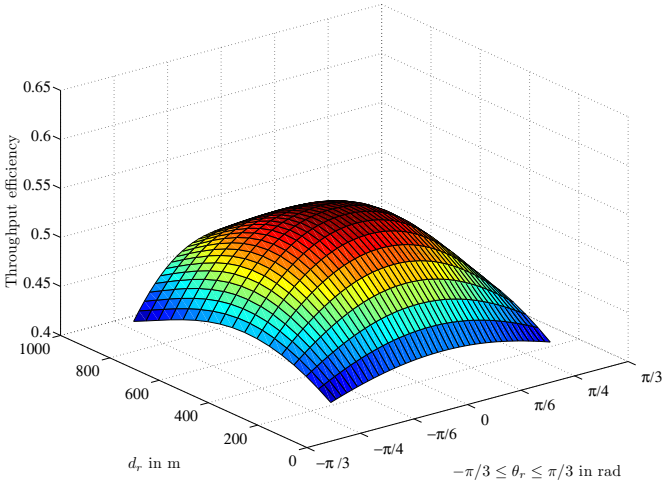
(a) HARQ-I scheme of $P_t = 35$ dBm



(b) HARQ-CC scheme of $P_t = 35$ dBm



(c) HARQ-I scheme of $P_t = 45$ dBm

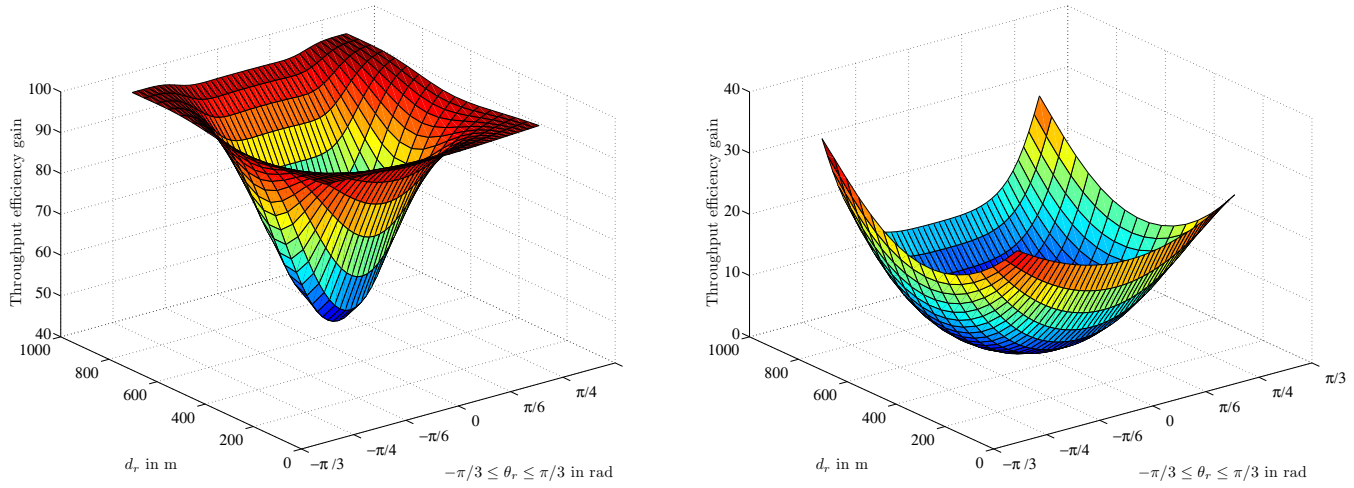


(d) HARQ-CC scheme of $P_t = 45$ dBm

Figure 3.6: η_T for cooperative HARQ-I and HARQ-CC schemes versus relay position d_r and relaying angle θ_r, for $N_{\max} = 5$ and $P_t = 35$ and 45 dBm

From fig. 3.7a, we observe that if the relay is located at $(d_r, \theta_r) = (500, 0)$, the obtained relative gain is about 50%. As the relay becomes faraway from this position the relative gain of HARQ-CC increases to achieve a gain of more that 90 % compared to HARQ-I scheme. Moreover, at $P_t = 45$ dBm (fig. 3.7b), the minimum achieved gain is about 10% and is at the position $(d_r, \theta_r) = (500, 0)$. However, at the border positions, the achieved relative gain exceeds 30%.

Therefore, by increasing the transmission power, the relative gain degrades significantly. Since at high



(a) $P_t = 35$ dBm

(b) $P_t = 45$ dBm

Figure 3.7: η_T Relative Gain between cooperative HARQ-CC and HARQ-I schemes versus relay position d_r and relaying angle θ_r, for $N_{\max} = 5$ and $P_t = 35$ and 45 dBm

transmissions power, the probability of successfully receiving a packet at the first transmission instants increases in both schemes and hence the throughput gain of HARQ-CC compared to HARQ-I decreases. Once again, we can conclude that HARQ schemes are useful at low and medium transmission power ranges and HARQ-CC has a significant relative gain compared to HARQ-I.

3.8 Outage Probability in HARQ Networks

Slow variation of channel fading during data communication makes the average of a given performance metric useless, since it yields to different values for each channel realization. Thus, the concept of outage is meaningful, which quantifies the probability that a given performance metric could not be met. Therefore, closed form representations of the outage probability for a given service is needed. In addition, the evaluation of the outage probability by simulations is very costly in time and analytical studies as proposed in this work might be very useful to save simulation time.

In this section, the QoS of interest is the probability that the average delay for transmitting a packet, due to numerous retransmissions, exceeds a threshold θ_m due to quasi-static channel. The average delay is understood as the average delay in presence of AWGN for a given channel realization. In most practical operating conditions, the number of allowed retransmissions for one packet is sufficiently large to consider that it is infinite without impacting too much the system performance. For quasi-static channel, in which the channel state varies slowly in time, the channel can be considered as Gaussian for one channel realization and hence the average delay can be computed in this framework, but outage is later invoked

to take into account the slow variation of the channel.

To the best of our knowledge, there is no work dealing with the delay outage probability considering relay-assisted communications and HARQ mechanisms.

3.8.1 System model and end-to-end delay

System model: In this section, we consider a quasi-static channel, where the complex channel realizations h_{sd}, h_{sr} and h_{rd} are assumed to be constant during the transmission of one block and vary for each successfully received block. We assume that each block contains a large number of packets which enables to analyze the average delay. The system descriptions and protocol are given in section 3.2.

The instantaneous PER $P(\gamma_{ij})$ in the link $i-j$ can be efficiently approximated by:

$$P(\gamma_{ij}) = \begin{cases} \frac{1}{1+\exp\left(a(\gamma_{ij}-\gamma_{th})\right)} & \text{if } \gamma_{ij} > 0 \\ 1 & \text{if } \gamma_{ij} = 0 \end{cases} \quad \forall ij \in \{sd, sr, rd\} \tag{3.60}$$

with a and γ_{th} are curve fitting parameters depending on the MCS.

End-to-end delay: Consider that the HARQ-I scheme is adopted, i.e the decoding events at the instants n and $n-1$ are assumed to be independent at each node. Moreover, in most practical conditions the number of allowed retransmissions, $N_{\max}$, is large enough to neglect the packet loss and hence can be considered as infinite [75]. Therefore, in the case of a direct transmission with d, the average delay per received packet is [75, 77]:

$$\overline{N}_s^d = \sum_{n=1}^{+\infty} n Q_s^d(n) \tag{3.61}$$

where $Q_s^d(n)$ is the successful decoding probability at the instant n which is equal to the probability that d does not decode at the instants $1, 2, \cdots, n-1$ but decodes at n. Hence, $Q_s^d(n) = P(\gamma_{sd})^{n-1}(1-P(\gamma_{sd}))$ and using [76, eq. (2) §0.231, pp. 8], we obtain:

$$\overline{N}_s^d = \frac{1}{1-P(\gamma_{sd})} \tag{3.62}$$

Let us consider that the relay r has been selected to help d. The average delay induced when s transmits is:

$$\overline{N}_s^c = \sum_{n=1}^{+\infty} n Q_s^c(n) \tag{3.63}$$

with $Q_s^c(n)$ being the successful decoding probability at d or r at the instant n due to s transmissions. The correct (or not) decoding events at d and r are independent, hence the error probability at any

instant is $P(\gamma_{sd})P(\gamma_{sr})$. Moreover, the error events are independent from one time slot to another; $Q_s^c(n)$ is hence given by $Q_s^c(n) = (1 - P(\gamma_{sd})P(\gamma_{sr}))(P(\gamma_{sd})P(\gamma_{sr}))^{n-1}$ and using [76, eq. (2) §0.231, pp. 8], we have:

$$\overline{N}_s^c = \frac{1}{1 - P(\gamma_{sd})P(\gamma_{sr})}. \tag{3.64}$$

If at the instant n', the user has not received the packet and the relay r has successfully decoded it, s stops transmitting the packet, while r continues to forward the packet to d. Thus, the average delay induced when r transmits is:

$$\overline{N}_r^c = \left(\sum_{n'=1}^{\infty} Q_{sr}^c(n') \times \sum_{n=n'+1}^{\infty} (n - n') Q_{rd}^c(n) \right) \tag{3.65}$$

where $Q_{sr}^c(n')$ is the successful decoding probability at r at n' and considering that d has not decoded. $Q_{rd}^c(n)$ is the successful decoding probability at d due to r transmissions at the instant $n > n'$. Hence, $Q_{sr}^c(n') = P(\gamma_{sd})^{n'} P(\gamma_{sr})^{(n'-1)} (1 - P(\gamma_{sr}))$ and $Q_{rd}^c(n) = P(\gamma_{rd})^{(n-n')}(1-P(\gamma_{rd}))$ (r starts transmission at the instant $n' + 1$). Using (3.65) and [76, eq. (2) §0.231, pp. 8]:

$$\overline{N}_r^c = \left(\frac{P(\gamma_{sd})(1-P(\gamma_{sr}))}{1-P(\gamma_{sd})P(\gamma_{sr})} \right) \left(\frac{1}{1-P(\gamma_{rd})} \right) \tag{3.66}$$

The end-to-end average delay (in AWGN channel) per successfully received packet at d is:

$$\overline{N}^c = \overline{N}_s^c + \overline{N}_r^c \tag{3.67}$$

3.8.2 Outage probability derivation

The delay outage probability $\mathbb{P}(O)$ is defined as the probability that the delays of direct and relay-assisted communications exceed a threshold θ_m and is given by:

$$\mathbb{P}(O) \stackrel{\triangle}{=} \mathbb{P}\left(\overline{N}_s^c + \overline{N}_r^c > \theta_m; \overline{N}_s^d > \theta_m \right) \tag{3.68}$$

Theorem 1. *In quasi-static Rayleigh fading channels and relay-assisted communications using DF protocol, the delay outage probability $\mathbb{P}(O)$ can be efficiently approximated by the following expressions:*

$$\mathbb{P}(O) = \mathbb{P}^{(1)}(O_1) + \mathbb{P}^{(1)}(O_2) + \mathbb{P}^{(2)}(O_1) + \mathbb{P}^{(2)}(O_2) \tag{3.69}$$

with

$$\mathbb{P}^{(1)}(O_1) = G\left((\overline{\gamma}_{sd})^{-1}\right) - e^{-\frac{\gamma_{th}}{\overline{\gamma}_2}} (\alpha_1)^{\frac{1}{a\overline{\gamma}_2}} I_0 \tag{3.70}$$

$$\mathbb{P}^{(2)}(O_1) + \mathbb{P}^{(2)}(O_2) = \begin{cases} 1 - \exp\left(-\frac{p_{\gamma_{sd}}{}^{-1}(\alpha_2)}{\overline{\gamma}_{sd}} \right) & \text{if } \theta_m < 2 \\ 0 & \text{otherwise} \end{cases} \tag{3.71}$$

	λ_1	λ_2	λ_3
Eq. (3.73a)	$1-\alpha_1$	$\frac{\exp(-a\gamma_{th})}{a(1-\alpha_1)}$	$\frac{1}{a(1-\alpha_1)}$
Eq. (3.73b)	$\frac{1}{\frac{1}{P(\gamma_m)}-\frac{1}{\theta_m-1}}-1$	$\frac{\lambda_1}{a}\exp\left(a(\gamma_m-\gamma_{th})\right)$	$\frac{\lambda_1}{a}\exp\left(-a\gamma_{th}\right)$
Eq. (3.73c)	$g(\gamma_m)$	$\frac{g(\gamma_m)}{1-g(\gamma_m)}$	$\frac{\exp\left(-a\gamma_{th}\right)}{a\lambda_2(1-\lambda_2)}$

Table 3.5: Taylor expansion parameters of eqs. (3.73a), (3.73b) and (3.73c)

$$\mathbb{P}^{(1)}(O_2) = \exp\left(-\frac{\gamma_{th}}{\overline{\gamma}_{sr}}\right)I_1 - I_2\left(\frac{(\theta_m-1)^{\frac{1}{\overline{\gamma}_{rd}}}}{a\overline{\gamma}_{sr}}\right)\times\exp\left(-\left(\frac{\gamma_{th}}{\overline{\gamma}_{sr}}+\frac{\gamma_{th}}{\overline{\gamma}_{rd}}\right)\right)\beta\left(\frac{1}{a\overline{\gamma}_{rd}}+1,\frac{1}{a\overline{\gamma}_{sr}}\right) \quad (3.72)$$

where I_0, I_1 and I_2 are approximated by:

$$I_0 \approx (\lambda_1)^{\frac{-1}{a\overline{\gamma}_{sr}}}\left[\left(1-\frac{\lambda_2}{\overline{\gamma}_{sr}}\right)G\left(\frac{1}{\overline{\gamma}_{sd}}\right)-\frac{\lambda_3\exp\left(-a\gamma_{th}\right)}{(\overline{\gamma}_{sr})(a\overline{\gamma}_{sd}-1)}G\left(\frac{1}{\overline{\gamma}_{sd}}-a\right)\right] \quad (3.73a)$$

$$I_1 \approx (\lambda_1)^{\frac{-1}{a\overline{\gamma}_{sr}}}\left[(1-\lambda_2)G\left(\frac{1}{\overline{\gamma}_{sd}}\right)-\frac{\lambda_3}{\overline{\gamma}_{sr}(a\overline{\gamma}_{sd}-1)}G\left(\frac{1}{\overline{\gamma}_{sd}}-a\right)\right] \quad (3.73b)$$

$$\begin{aligned} I_2 \approx{}& (\lambda_2)^{\frac{1}{a\overline{\gamma}_{sr}}}(\lambda_1)^{\frac{1}{a\overline{\gamma}_{rd}}}\left[\left(1-\lambda_3\left(\frac{(1-\lambda_2)}{\overline{\gamma}_{sr}}+\frac{1}{\overline{\gamma}_{rd}}\right)\exp(a\gamma_m)\right)\right. \\ &\left.\times G\left(\frac{1}{\overline{\gamma}_{sd}}\right)-\left(\frac{\lambda_3}{a\overline{\gamma}_{sd}-1}\right)\left(\frac{(1-\lambda_2)}{\overline{\gamma}_{sr}}+\frac{1}{\overline{\gamma}_{rd}}\right)G\left(\frac{1}{\overline{\gamma}_{sd}}-a\right)\right] \end{aligned} \quad (3.73c)$$

with $\alpha_1=\frac{\theta_m-1}{\theta_m}$, $\alpha_2=\min\{\theta_m-1,1\}$, $\gamma_m=\frac{P^{-1}_{\gamma_{sd}}(\alpha_1)+P^{-1}_{\gamma_{sd}}(\alpha_2)}{2}$, $G(x)$ and the inverse of the PER $P^{-1}_{\gamma_{ij}}$ are respectively defined by

$$G(x)\stackrel{\triangle}{=}\exp\left(-P^{-1}_{\gamma_{sd}}(\alpha_2)x\right)-\exp\left(-P^{-1}_{\gamma_{sd}}(\alpha_1)x\right)$$

$$P^{-1}_{\gamma_{ij}}=\begin{cases}\frac{1}{a}\ln\left(\frac{1}{P(\gamma_{ij})}-1\right)+\gamma_{th} & \text{if } P(\gamma_{ij})<1\\ 0 & \text{if } P(\gamma_{ij})=1\end{cases}\quad\forall ij$$

and the variables, λ_1, λ_2 and λ_3 are defined table 3.5.

Proof. *The proof of Theorem 1 is given in Appendix B.* ■

3.8.3 Simulation results

In our simulations, BPSK modulation and LDPC code are adopted. Each packet contains 102 bits and the coding rate is $R_c=1/2$. In addition to block fading, we consider a free path-loss propagation model $p_{l_{ij}}=1/d^{\kappa}_{ij}$, where d_{ij} is the distance of the $i-j$ link $\forall ij\in\{sd,sr,rd\}$, $p_{l_{ij}}$ is the associated path-loss and $\kappa=2$ being the path-loss exponent. Assuming equal power at each transmitting node, then $\overline{\gamma}_{ij}\,|_{dB}=\overline{\gamma}_{sd}\,|_{dB}+10.\kappa.\log_{10}(d_{sd}/d_{ij})$ for $ij=\{sr,rd\}$. We consider that $d_{sr}=d_{rd}=d_{sd}/2$. The

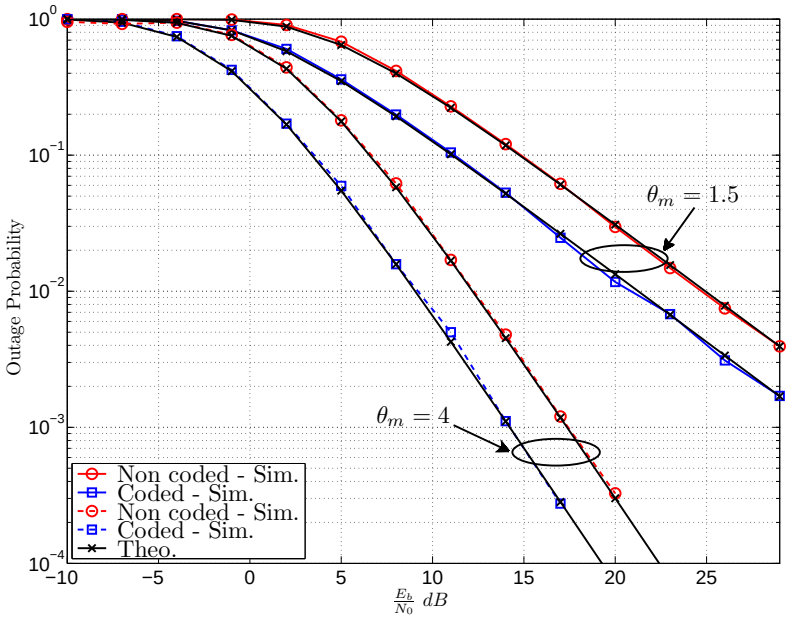


Figure 3.8: Outage probability $\mathbb{P}(O)$ vs. $\frac{E_b}{N_0}$ for $\theta_m = 1.5$ and 4 TS

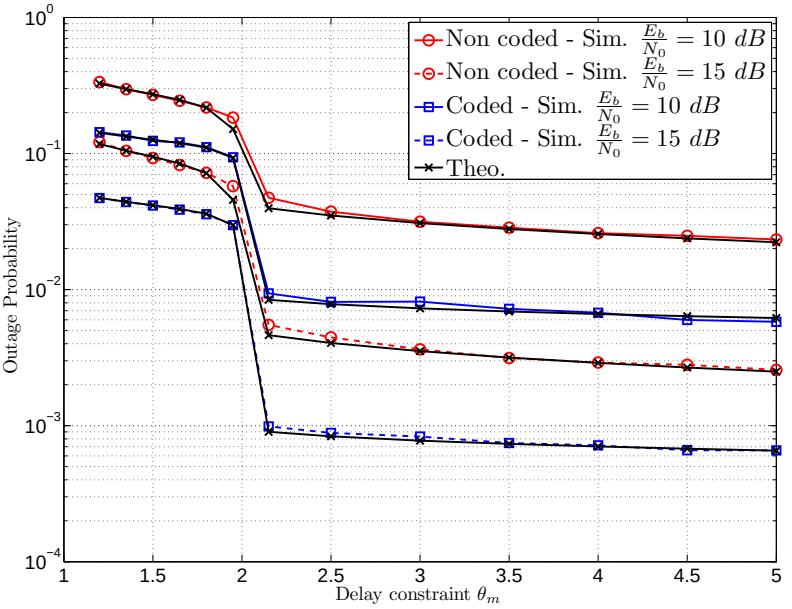


Figure 3.9: Outage probability $\mathbb{P}(O)$ vs. θ_m for $\frac{E_b}{N_0} = 10$ and 15 dB

PER fitting parameters are $a = 1.22, \gamma_{th} = 2.94$ and $a = 7.44, \gamma_{th} = 1.34$ for non coded and coded packets respectively.

Fig. 3.8 depicts the outage probability w.r.t. E_b/N_0 for $\theta_m \in \{1.5, 4\}$ time slots (TS). Our theoretical analysis, labeled as "Theo" in figures, are verified thanks to Monte Carlo simulations. The approximations made are sufficient to predict the real behavior of the delay outage probability, provided that fitting parameters γ_{th} and a are well known for the considered MCS. We can observe that as θ_m increases, the outage probability significantly decreases for coded and uncoded schemes.

Fig. 3.9 shows the variation of the outage probability w.r.t. the required QoS θ_m for $E_b/N_0 = 10$ and 15 dB. We can observe an interesting behavior on the outage probability which sharply decreases for $\theta_m > 2$. It corresponds to the fact that for $\theta_m < 2$ the probability to be in outage is very high due to the too much stringent requirement on the delay. Such a requirement implies that certain number of packets must be decoded in the first transmission instant, i.e. in the link $s - d$, reducing the diversity order of the system. On the contrary, for $\theta_m \geq 2$ a packet has two ways to reach the destination, i.e. the direct link, $s - d$, or the relay link, $s - r - d$, increasing the diversity order of the system and making the outage probability more governed by the received SNR. This behavior can be also interpreted with our theorem; if $\theta_m < 2$ then $\mathbb{P}^{(2)}(O_1) + \mathbb{P}^{(2)}(O_2) \neq 0$ in eq. (12) and for $\theta_m \geq 2$ then $\mathbb{P}^{(2)}(O_1) + \mathbb{P}^{(2)}(O_2) = 0$ reducing significantly the outage probability.

3.9 Conclusions and remarks

The contributions of this chapter are summarized as follows:

1. In block fading channel, different HARQ schemes in direct and relay-assisted networks have been analyzed. Closed form expressions of the packet error rate, end-to-end delay, throughput efficiency for p2p HARQ-CC, cooperative HARQ-I and cooperative HARQ-CC are provided.
2. In quasi-static fading channel, we provided theoretical analysis of the delay outage probability for relay-assisted HARQ-I based systems.

We think that the suggested theoretical analysis in 1 and 2 allow to save an important amount of simulation time. Hence, it might be useful to precisely and rapidly evaluate the performance metrics of future relay-assisted networks. Moreover, the derived expressions can be useful for all modulation and coding schemes.

Several remarks are also drawn:

1. These analyses focus on the QoS without invoking the energy consumption. Thus, what is about the energy efficiency of these HARQ schemes?

2. HARQ schemes are efficient at low and medium transmission powers. Hence, an optimal power allocation at the source and the relay is needed.

Therefore, in the next chapter we focus on the energy efficiency analysis of these HARQ schemes. Moreover, we study an energy minimization algorithm for multiuser relay-assisted HARQ based networks.

CHAPTER

4 Energy Efficient Relay-assisted Hybrid-ARQ Networks

4.1 Introduction

Quality of service (QoS) and energy consumption are two important metrics to evaluate the performance of wireless green cellular networks [36, 37]. Fundamentally, the energy consumption is power and time dependent. However, it has been shown that introducing intermediate relays between two communicating nodes enables to transmit over short distances and may enhance the QoS. Thus, exploiting the spatial diversity enabled by the relaying schemes leads to a power saving. Moreover, in non-reliable communications, i.e. non-zero PER, the delay induced by the retransmission mechanism of erroneous packets has certainly an impact on the energy consumption. Thus, exploiting the temporal diversity offered by the HARQ schemes could reduce energy consumption. Therefore, optimizing jointly the available resources and considering the HARQ mechanisms in relay-assisted networks could be energy efficient for cross-layer schemes.

Recently, it has been stated in [21, 22], that there is a tremendous waste of energy in the core network. This waste is due to the energy consumed from data transportation, data generation and data processing. This issue is of particular importance in future wireless systems where the distances between terminals are becoming smaller and hence the circuitry energy consumption has the same order of the energy consumed for data transmission. Therefore, these sources of energy consumption must be certainly considered when optimizing resources.

The traditional resource allocation policies, focusing on throughput maximization, are generally inefficient in terms of energy consumption (see [37, 38] and references therein) as it has been shown in chapter 2. Hence, an important effort has been recently granted toward energy efficient and reliable communications. On one hand, we find works on energy minimization based on information theoretical approaches [39–41]. On the other hand, most of works on cellular green networks focus on the energy minimization or efficiency with reliable communications, i.e. capacity limited links [78]. The design of techniques in multiuser (MU) context to minimize the energy consumption under non-reliable communications, is a challenging task for the scientific community to enhance future cellular networks.

In [47] the authors studied the energy efficiency for cooperative and non cooperative HARQ schemes by minimizing the energy consumed for a single user under outage probability constraint. In our previous work [77], the energy efficiency in relay-assisted HARQ scheme is analyzed for single network without considering any resource optimization. In [48], the authors studied the energy efficiency for DF, AF and

CF relaying schemes with HARQ under outage probability and delay constraints. It is also convenient to note an important amount of work for WSN in which the energy consumption is of crucial importance [44]. However, all works above did not jointly consider the resource allocation strategy in MU cellular context with HARQ schemes under QoS constraint. The main challenge comes from the need of a closed-form expression of the PER with HARQ in order to express the delay requirements [49].
The contribution of this chapter lies in two folds:

- In the first part, we provide theoretical analyses of the energy efficiency for different HARQ protocols in point-to-point (p2p) and relay-assisted communications.
- Based on these analyses, we propose a new resource allocation algorithm for MU relay-assisted HARQ-I network that aims to allocate powers at BS and RS and select the optimal relay that minimizes the overall energy consumption. The QoS of interest in this work is the end-to-end average delay, in time slots (TSs), needed to successfully receive a packet. Thus, a delay constraint for every user must be guaranteed. In addition, an overall power constraint must be respected. The resource scheduler at BS is assumed to have only a knowledge about the average channel statistics but no instantaneous CSI. In this work, we firstly try to give a formulation of the energy minimization problem by introducing another variables that permit to efficiently solve this problem by Lagrange dual optimization. Then, the convexity of this problem is proved leading to a global optimal solution. Finally, an energy minimization algorithm that allocates the optimal powers between users is proposed.

To the best of our knowledge, this is the first work that jointly considers an HARQ scheme and a given modulation and coding schemes (MCS) in multiuser relay-assisted network that aims to minimize the overall energy for delay constrained users.

4.2 Energy efficiency analysis

Based on the analysis of the error probability described in the previous chapter, the energy efficiency η_E in bits/joule is defined as:

$$\eta_E = L\frac{1-\text{PER}}{\mathbb{E}[E_{agg}]} \quad \text{bits/joule}, \tag{4.1}$$

where the PER is derived in chapter 3 and $\mathbb{E}[E_{agg}]$ is the average aggregate energy consumed during the retransmission mechanism. Let us assume that the average energy consumed due to relay-assisted communications can be split into two terms as follows:

$$\mathbb{E}\left[E_{agg}\right] = \mathbb{E}\left[E_s\right] + \mathbb{E}\left[E_r\right] \tag{4.2}$$

where $\mathbb{E}[E_s]$ and $\mathbb{E}[E_r]$ are the average energy consumed during the retransmission mechanism when s and r transmit respectively. Therefore, $\mathbb{E}[E_s]$ can be written as follows:

$$\begin{aligned}\mathbb{E}[E_s] = \sum_{n=1}^{N_{\max}-1} E_s(n)\Big(Q_{sd}(n)P_{sr}(n-1) + P_{sd}(n)Q_{sr}(n)\Big) \\ +E_s(N_{\max})\Big(Q_{sd}(N_{\max})P_{sr}(N_{\max}-1)\Big) \\ +E_s(N_{\max})\Big(P_{sr}(N_{\max}-1)P_{sd}(N_{\max})\Big)\end{aligned} \tag{4.3}$$

where $E_s(n)$ is the consumed energy till the instant n, when s transmits. Moreover, $\mathbb{E}[E_r]$ is expressed as:

$$\mathbb{E}[E_r] = \sum_{n'=1}^{N_{\max}-1} Q_{sr}(n')\Bigg(\sum_{n=n'+1}^{N_{\max}} E_r(n-n')Q_{rd}(n,n') + E_r(N_{\max}-n')\ P_{rd}(N_{\max},n')\Bigg) \tag{4.4}$$

where $E_r(n)$ is the consumed energy till the instant n, when r transmits. When s transmits, the two receiving nodes r and d consume both energy of their circuitry and for transmitting N/ACK messages. Therefore, considering HARQ-I and HARQ-CC, the consumed energy $E_s(n)$ at the instant n can be written as:

$$E_s(n) = n(E_{tx} + 2E_{rx} + 2E_{ACK}) \tag{4.5}$$

with E_{tx}, E_{rx} and E_{ACK} are respectively defined in eqs. (1.26), (1.27) and (1.28) in chapter 1 and reminded here:

$$E_{tx} = \frac{L}{R_b R_0}.(P_{tx} + \beta_{amp}.P_t) \tag{4.6a}$$

$$E_{rx} = \frac{L}{R_b R_0}.P_{rx} \tag{4.6b}$$

$$E_{ACK} = \tau_{ack}.(E_{tx} + E_{rx}) \tag{4.6c}$$

Moreover, for each transmitted packet from r, d sends N/ACK packets to s and r and hence $E_r(n)$ can be written as:

$$E_r(n) = n(E_{tx} + E_{rx} + 2E_{ACK}) \tag{4.7}$$

Therefore, from eqs. (4.3) and (4.4), $\mathbb{E}[E_r]$ and $[E_s]$ can be respectively written as:

$$\mathbb{E}[E_s] = (E_{tx} + 2E_{rx} + 2E_{ACK}).\overline{N}_s \tag{4.8a}$$

$$\mathbb{E}[E_r] = (E_{tx} + E_{rx} + 2E_{ACK}).\overline{N}_r \tag{4.8b}$$

where $\overline{N}_s$ and $\overline{N}_r$ are derived in chapter 3. Moreover, using the derived closed form of PER in chapter 3, the energy efficiency η_E in eq. (4.1) can be theoretically analyzed.

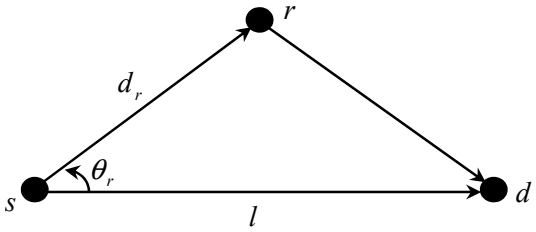


Figure 4.1: Relay-assisted scheme

Table 4.1: System model parameters

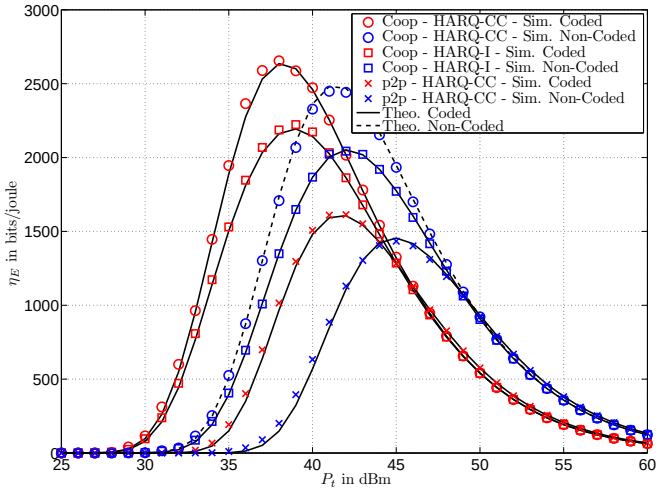
Parameters	Description	value
P_{tx}	transmitter circuitry power	20 dBm
P_{rx}	receiver circuitry power	20 dBm
τ	ACK ratio	0.08125

Numerical analysis

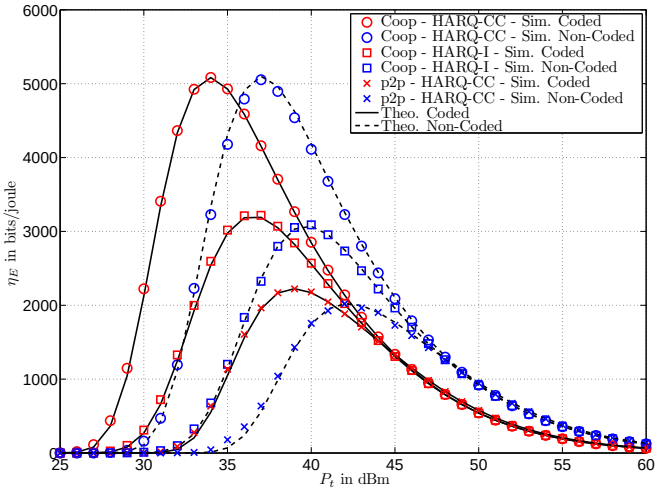
In our simulation settings, the system parameters adopted in chapter 3 are used and the other system parameters are tabulated in Table 4.1 [45]. We consider a relay-assisted scheme as depicted in fig. 4.1 where the destination d is located at a distance $l = 1000$ m from s. The relay r is also assumed to be located at a distance $d_r = l/2$ from s with $\theta_r = \pi/6$ rad, fig. 4.1. The schemes p2p HARQ-CC, Cooperative HARQ-I and Cooperative HARQ-CC are compared for $N_{\max} = [2, 5, 10]$ and for coded and non-coded schemes, otherwise mentioned.

Energy efficiency η_E versus P_t: Figs. 4.2a, 4.2b and 4.2c depict the theoretical and simulated energy efficiency η_E in bits/joule versus the transmission power P_t for $N_{\max} = 2$, 5 and 10 respectively. First of all, we notice that our theoretical analyses are tight enough to evaluate η_E for different HARQ schemes. We remark that η_E has an interesting behavior, where for a given power range it increases (called range 1) to achieve a maximum, then it starts to decrease (called range 2). Therefore, in range 1, the increase of η_E w.r.t. P_t is due to the fast decrease in the average delay that leads to a lower energy consumption. Moreover, we can notice that there is an optimal power that enables a maximum energy efficiency. However, after achieving a maximum, the decreasing behavior of η_E is due to the high increase in the transmission power and to the average delay that reaches to a steady state. In other words, referring to eq. (4.1) and at high P_t, the PER tends to zero and the average delay tends to 1. Thus, the energy efficiency became inversely proportional to P_t, i.e. $\eta_E \propto \frac{1}{P_t}$.

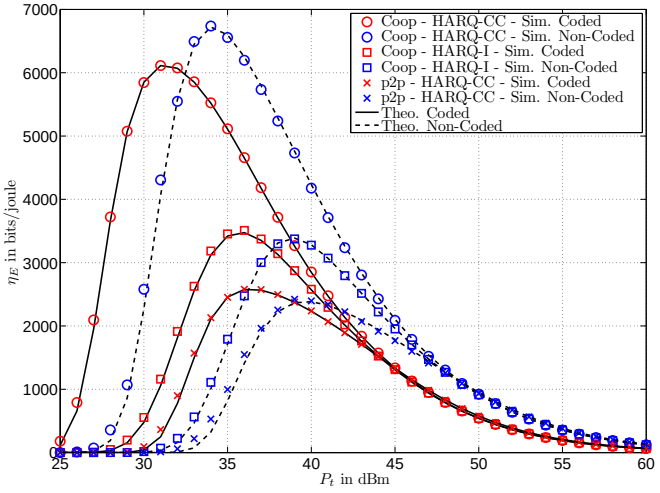
Consider a non-coded scheme and $N_{\max} = 2$, the maximum energy efficiency achieved by p2p HARQ-CC is about 1450 bits/joule, Coop. HARQ-I is 2000 bits/joule and Coop. HARQ-CC is nearly 2500 bits/joule. Then, introducing a coded scheme, it can be noticed that there is a power shifting in all



(a) $N_{\max} = 2$



(b) $N_{\max} = 5$



(c) $N_{\max} = 10$

Figure 4.2: Theoretical and simulated η_E in bits/joule versus P_t in dBm for coded and non-coded Cooperative HARQ-I, Cooperative HARQ-CC and p2p HARQ-CC schemes and for $N_{\max} = [2, 5, 10]$

HARQ schemes. In range 1, η_E is more energy efficient compared to the non-coded one. In contrast, at high transmission power (range 2), the non-coded schemes have better energy efficiency. Indeed, at high P_t, the probability of successfully receiving a packet at the first transmission instant increases for both schemes, however the consumed energy by the coded one stays $1/R_0$ of the energy consumed by the non-coded one.

Now increasing $N_{\max}$ to 10, the energy efficiency for all HARQ schemes enhances significantly to achieve a maximum of about 2400 bits/joule for p2p HARQ-CC, 3300 bits/joule for Coop. HARQ-I and 6800 bits/joule for Coop. HARQ-CC. We can further observe that the required power to achieve these maximums became lower as $N_{\max}$ increases. Therefore, this explains that HARQ schemes become more power and energy efficient as the number of allowed retransmissions increases. Indeed, increasing $N_{\max}$, enables more reliable transmissions and hence enhances η_E. However, this result is also governed by the allocated power, where at higher transmission power, η_E becomes the same for all $N_{\max}$.

In summary, we can conclude that the higher $N_{\max}$ is, the more power and energy efficient communications are for all HARQ types. Moreover, cooperative HARQ-CC protocol has the better performance compared to other protocols, since it jointly exploits the path-loss gain offered by the relaying scheme and also the temporal diversity enabled by the packet combining technique.

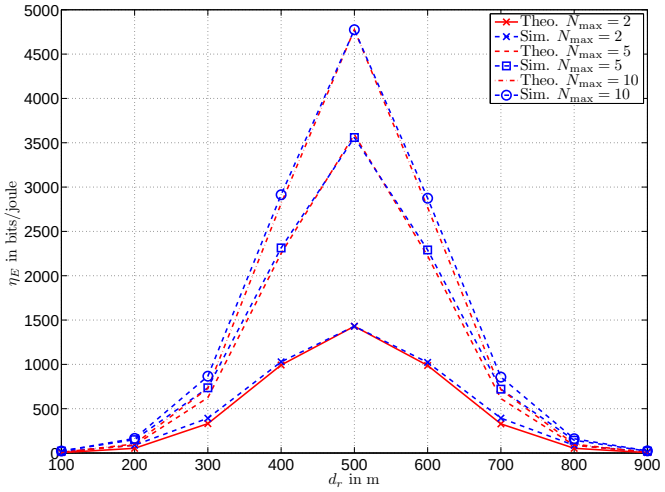
Energy efficiency η_E versus d_r: We consider a non-coded communication (the same conclusion can be drawn for coded communications) and that s, r and d are collinear and the relay distance d_r from s is varying in the interval $[100, 900]$ m.

For $N_{\max} = [2, 5, 10]$ and at $P_t = 35$ and 45 dBm, the variation of η_E in bits/joule w.r.t. d_r in m is depicted in figs. 4.3a and 4.3b for HARQ-I and figs. 4.3c and 4.3d for HARQ-CC respectively. Firstly, we can observe that the energy efficiency has the same behavior as the throughput efficiency observed in chapter 3 and at $d_r = 500$ m, η_E achieves its maximum. Thus, this gives an insight about the optimal relay deployment that leads to the best energy efficiency.

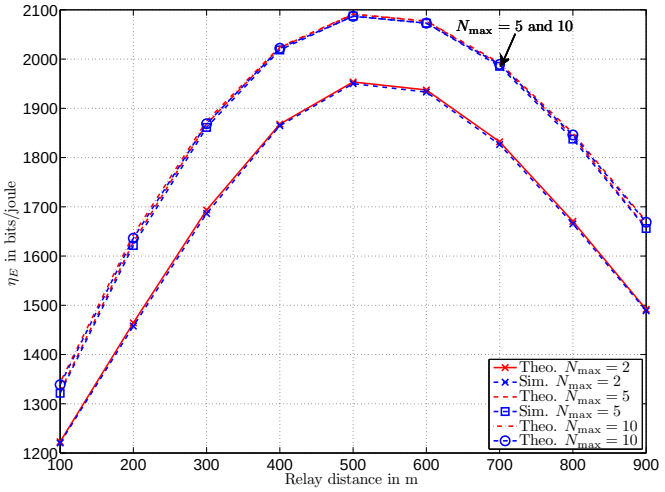
For a desired $N_{\max} = 2$ and at $P_t = 35$ dBm, η_E provided by HARQ-I and HARQ-CC are slightly different. This shows that the number of allowed retransmissions are not sufficient to observe the gain of HARQ-CC w.r.t. HARQ-I. Moreover, the obtained energy efficiency is very low since there is no enough transmission power that leads to an efficient transmission. However, increasing $N_{\max}$ to 5 or 10, η_E enhances significantly for both schemes with η_E delivered by HARQ-CC is doubled compared to HARQ-I. Thus, increasing the number of allowed retransmissions, compensates the lack of transmission power and hence enhances the temporal diversity enabling higher energy efficiency. Moreover, at low transmission power, we can conclude that the energy efficiency becomes more impacted by the relay position in which near $d_r = 500$ m, the energy efficiency is very high. However, at border relay positions it degrades significantly.

For $P_t = 45$ dBm and $N_{\max} = 2$, the energy efficiency in both schemes enhances. In contrast, allowing $N_{\max}$ to be 5 or 10 has a minimal influence on η_E in both HARQ schemes. Higher P_t makes the transmitted packet to be received at the first transmission instants and hence the HARQ mechanisms are useless.

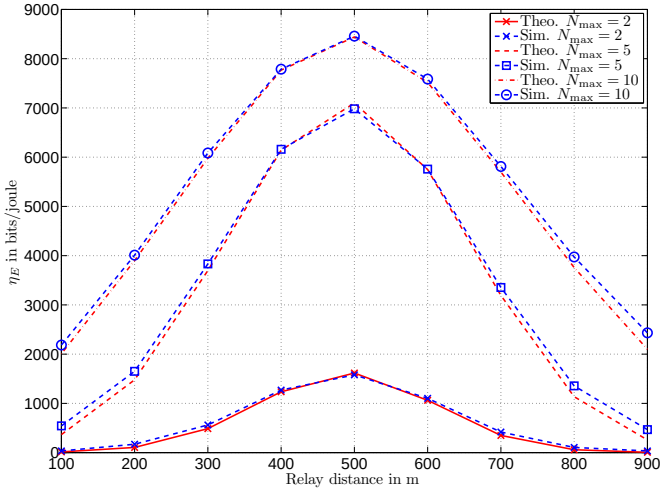
In conclusion, adopting HARQ-CC scheme, deploying optimally the relay and choosing carefully the maximum number of allowed retransmissions and the allocated power enable energy efficient communi-



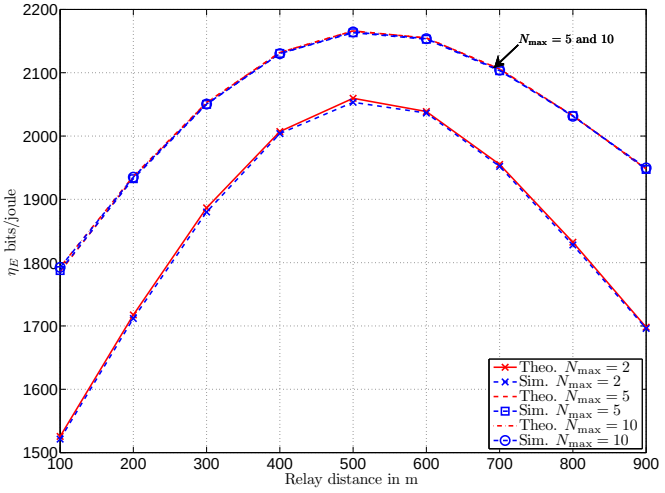
(a) HARQ-I scheme of $P_t = 35$ dBm



(b) HARQ-I scheme of $P_t = 45$ dBm



(c) HARQ-CC scheme of $P_t = 35$ dBm

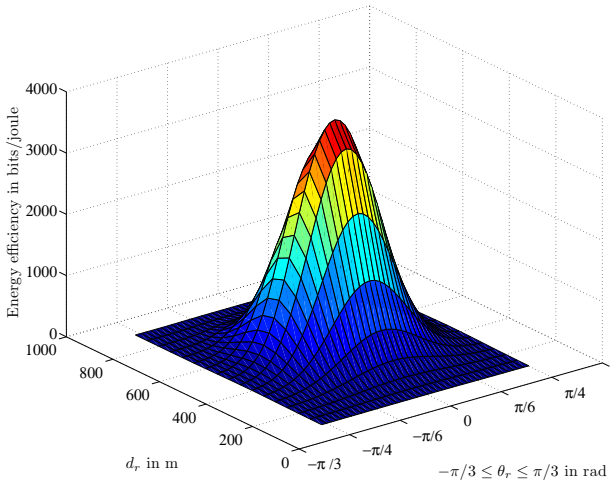


(d) HARQ-CC scheme of $P_t = 45$ dBm

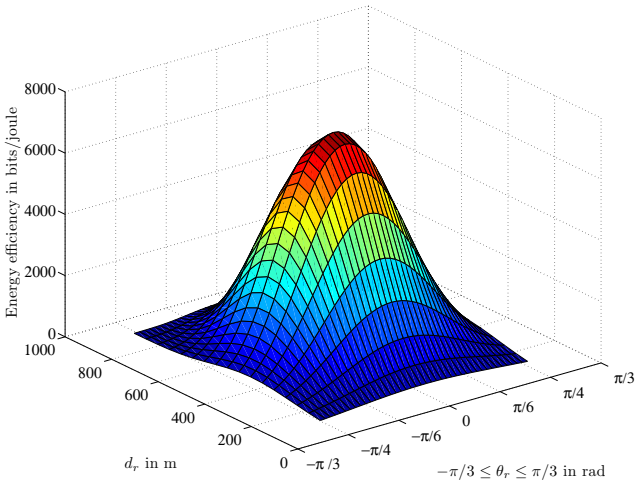
Figure 4.3: Theoretical and simulated η_E versus relay position d_r for non-coded cooperative HARQ-I and cooperative HARQ-CC schemes. The transmission powers are $P_t = 35$ and 45 dBm and $N_{\max} \in [2, 5, 10]$

cation.

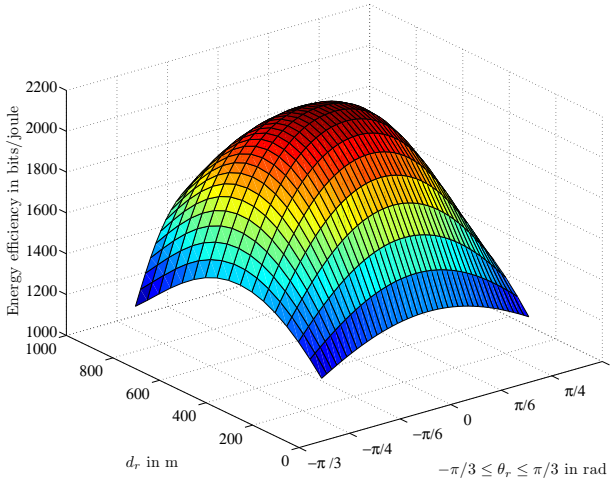
Energy efficiency η_E versus d_r and θ_r: Consider that a non-coded modulation is adopted and the system operates at $N_{\max} = 5$. At $P_t = 35$ dBm, figs 4.4a and 4.4b depict η_E in bits/joule versus relay position designated by (d_r,θ_r), for HARQ-I and HARQ-CC respectively. As the analysis of the throughput efficiency in chapter 3, η_E achieves its maximum at $(d_r, \theta_r) = (500, 0)$, i.e. s, r and d are collinear and



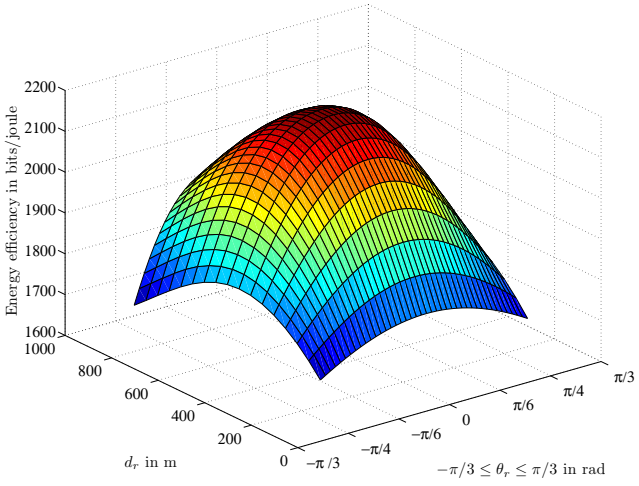
(a) HARQ-I scheme of $P_t = 35$ dBm



(b) HARQ-CC scheme of $P_t = 35$ dBm



(c) HARQ-I scheme of $P_t = 45$ dBm

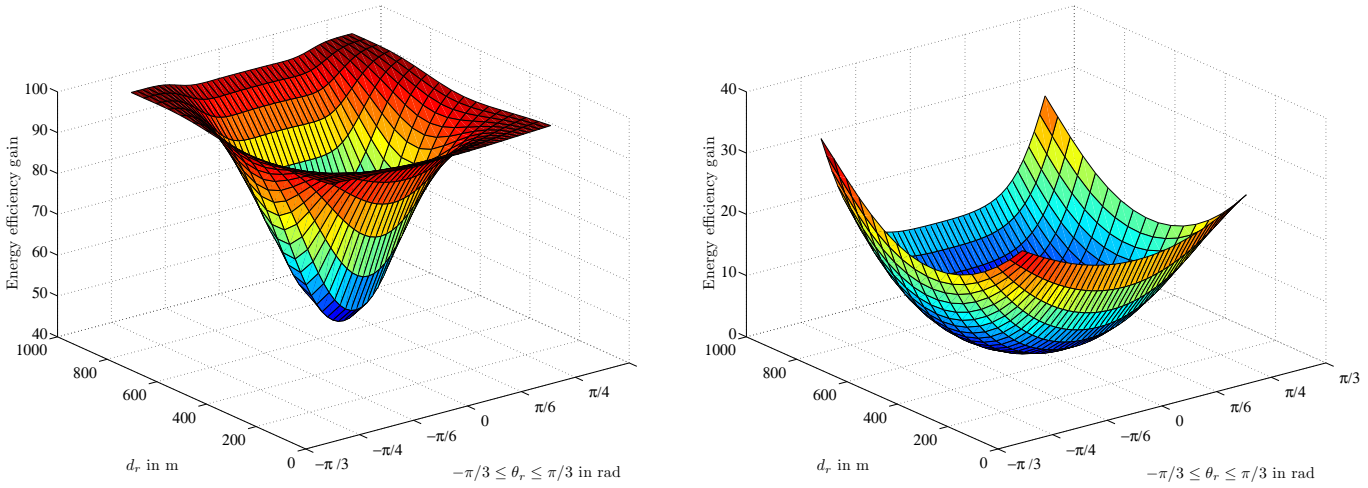


(d) HARQ-CC scheme of $P_t = 45$ dBm

Figure 4.4: η_E for cooperative HARQ-I and HARQ-CC scheme versus relay position d_r and relaying angle θ_r, for $N_{\max} = 5$ and $P_t = 35$ and 45 dBm

r is located at the middle. However, HARQ-I achieves 3000 bits/joule and HARQ-CC achieves about 6000 bits/joule. Moreover, as the relay becomes far from (500,0) position, the energy efficiency decrease drastically in both schemes. However, HARQ-I has a sharper form w.r.t. HARQ-CC. This significant gain of η_E is due to the packet combining technique which enhances the reliability for each retransmitted bit in the same packet.

For $P_t = 45$ dBm (Figs. 4.4c and 4.4d), a smooth variation on the energy efficiency can be observed.



(a) $P_t = 35$ dBm

(b) $P_t = 45$ dBm

Figure 4.5: η_E Relative gain between cooperative HARQ-CC and HARQ-I schemes versus relay position d_r and relaying angle θ_r, for $N_{\max} = 5$ and $P_t = 35$ and 45 dBm

However, the maximum on η_E is always achieved at $(d_r, \theta_r) = (500, 0)$ with about 1800 bits/joule and 2100 bits/joule for HARQ-I and HARQ-CC respectively. Therefore, η_E degrades significantly in both schemes and HARQ-CC has no substantial gain w.r.t. HARQ-I. This is due to the high transmission power that increases the energy consumption and hence degrades η_E.

To conclude, the lower the transmission power is, the higher the network geometry impacts the energy efficiency.

Relative gains: At $P_t = 35$ and 45 dBm, figs. 4.5a and 4.5b depict the relative gain in terms of energy efficiency for HARQ-CC compared to HARQ-I. Its clearly shown that at 35 dBm, the minimum relative gain is about 40% at $(d_r, \theta_r) = (500, 0)$. However, at border positions this gain enhances significantly to achieve about 90%. Moreover, at $P_t = 45$ dBm, the achieved relative gain at $(d_r, \theta_r) = (500, 0)$ is about 10%, where at border positions it enhances to about 30%.

As a summary, HARQ-CC enables a substantial gain for certain relay deployments and transmission powers. At low transmission power, the relative gain depends on the relay position. However, high transmission power makes the retransmission techniques inefficient in terms of energy and hence the relative gain of HARQ-CC w.r.t. HARQ-I degrades.

4.3 Energy minimization in HARQ-I relay-assisted networks

This section tackles an energy minimization problem in MU relay-assisted downlink cellular network. The system is assumed to use the HARQ-I protocol and each user has an average delay constraint to be satisfied. Moreover, the system has a total power constraint to be respected. The BS is assumed to have only a knowledge about the average channel statistics but no instantaneous CSI. The contribution of this section lies in three folds: i) We propose an approach for solving efficiently the energy minimization problem by applying the Lagrange dual method, ii) we show that this optimization problem can be conveniently expressed as function of quasi-convex constraints and affine objective function which leads to an optimal solution. Then we prove that the energy consumption of the system is minimal if and only if a high total power constraint is available. iii) Finally, an algorithm that jointly allocates the optimal power at BS, the RS and selects the optimal relay (if cooperation is decided) in order to satisfy the delay constrained users is proposed. The performed simulations show the improvement in terms of energy consumption of relay-assisted techniques compared to non-aided transmission in delay-constrained HARQ systems.

4.3.1 System and delay models

System model: We consider a MU downlink relay-assisted network consisting of a BS, a set of relays $\mathcal{K} = \{r_1, \cdots r_K\}$ deployed around BS and a set of randomly distributed users $\mathcal{M} = \{u_1, \cdots u_M\}$. Users are assumed to be simultaneously served due to the orthogonal channelization which can be performed with OFDMA technique. Each receiving node is corrupted by an AWGN with variance σ^2. In addition to path-loss, we consider that each link suffers from a Rayleigh fading and the communication protocol between BS and users is described in Section 3.2 in chapter 3. Moreover, each user u_m is assumed to have an average delay constraint per received packet $\overline{\mathcal{D}}_m$ to be satisfied and the network has a total power constraint P_{tot} to be respected.

Delay model: In point-to-point communication, the average error probability P in Rayleigh fading channel can be written as in chapter 3 eq. (3.26) [1]:

$$P = 1 - \left(\frac{g\overline{\gamma}}{1+g\overline{\gamma}}\right)\exp\left(-\frac{\gamma_M}{\overline{\gamma}}\right) \tag{4.9a}$$

$$\approx 1 - \exp\left(-\left(\gamma_M + \frac{1}{g}\right)\frac{1}{\overline{\gamma}}\right) \tag{4.9b}$$

with a, g and $\gamma_M = \log(a)/g$ are the curve fitting parameters which depend on the adopted MCS.

1. At high $\overline{\gamma}$, (4.9a) is approximated by (4.9b). $\ln\left(\frac{g\overline{\gamma}}{1+g\overline{\gamma}}\right) \approx \frac{-1}{g\overline{\gamma}}$ and then $\frac{g\overline{\gamma}}{1+g\overline{\gamma}} \approx \exp\left(\frac{-1}{g\overline{\gamma}}\right)$

Considering a non truncated HARQ, i.e. $N_{\max} \to \infty$, the average delay per received packet in the case of direct transmission with u_m is defined as:

$$\overline{N}^d_{0,m} = \frac{1}{1 - P^d_{0,m}}, \tag{4.10}$$

where $P^d_{0,m} = 1 - \exp\left(-g^d_{0,m}/p^d_{0,m}\right)$ and $g^d_{0,m} = (\gamma_M + 1/g)\frac{\sigma^2}{p_{l_{0,m}}}$. $P^d_{0,m}$ and $p^d_{0,m}$ are respectively the average PER and the allocated power in the $(\text{BS} - u_m)$ link.

Let us consider that BS has selected the relay r_k to help the user u_m. Therefore, the allocated powers at BS and r_k are respectively $p^c_{0,k,m}$ and $p^c_{k,m}$. The PERs in the links $(\text{BS} - u_m)$, $(\text{BS} - r_k)$ and $(r_k - u_m)$ are respectively $P^c_{0,m} = 1 - \exp\left(-g^c_{0,m}/p^c_{0,k,m}\right)$, $P^c_{0,k,m} = 1 - \exp\left(-g^c_{0,k,m}/p^c_{0,k,m}\right)$ and $P^c_{k,m} = 1 - \exp\left(-g^c_{k,m}/p^c_{k,m}\right)$. Where $g^c_{0,m} = (\gamma_M + 1/g)\frac{\sigma^2}{p_{l_{0,m}}}$, $g^c_{0,k,m} = (\gamma_M + 1/g)\frac{\sigma^2}{p_{l_{0,k}}}$ and $g^c_{k,m} = (\gamma_M + 1/g)\frac{\sigma^2}{p_{l_{k,m}}}$ and $p_{l_{0,m}}$, $p_{l_{0,r}}$ and $p_{l_{k,m}}$ being the path-loss in the $(\text{BS} - u_m)$, $(\text{BS} - r_k)$ and $(r_k - u_m)$ links respectively. Therefore, under cooperation between BS and r_k, the average delay induced when BS is transmitting is given by eq. (3.64) in chapter 3 and reminded as follows:

$$\overline{N}^c_{0,k,m} = \frac{1}{1 - P^c_{0,m}P^c_{0,k,m}} \tag{4.11}$$

If at any instant, the destination has not received the packet and the relay r_k has successfully decoded it, BS stops transmitting the packet, while r_k continues to forward the packet to u_m. The average number of transmissions by r_k can be written as in eq. (3.66):

$$\overline{N}^c_{k,m} = \left(\frac{P^c_{0,m}(1 - P^c_{0,k,m})}{1 - P^c_{0,m}P^c_{0,k,m}}\right)\left(\frac{1}{1 - P^c_{k,m}}\right) \tag{4.12}$$

Hence, in the case of cooperation between BS and r_k, the average delay per successfully received packet at u_m is:

$$\overline{N}_{k,m} = \overline{N}^c_{0,k,m} + \overline{N}^c_{k,m} \tag{4.13}$$

Energy consumption model: In case of direct transmission with u_m, the energy consumed for transmitting one packet $E_1(p^d_{0,m})$ can be re-written as:

$$\begin{aligned} E_1(p^d_{0,m}) &= (E_{tx} + E_{rx} + E_{ACK}) \\ &= k_1 + k_2 p^d_{0,m} \end{aligned} \tag{4.14}$$

$$\sum_{m=1}^{M}\left(\left(1-\sum_{k=1}^{K}s_{k,m}\right)E_1\left(p_{0,m}^{d}\right)\overline{D}_{0,m}^{d}\left(p_{0,m}^{d}\right)+\sum_{k=1}^{K}s_{k,m}\Big(E_2\left(p_{0,k,m}^{c}\right)\overline{D}_{0,k,m}^{c}\left(p_{0,k,m}^{c}\right)+\overline{D}_{k,m}^{c}\left(p_{0,k,m}^{c},p_{k,m}^{c}\right)E_3\left(p_{k,m}^{c}\right)\Big)\right) \tag{4.18}$$

with $k_1 = \frac{2(1+\tau_{ack})L}{R_b R_0}p_c$, $k_2 = \frac{(1+\tau_{ack})L}{R_b R_0}\beta_{amp}$ and the circuitry power consumption $p_c = P_{tx} = P_{rx}$. Therefore, the average energy consumed per received packet is:

$$\overline{E}_{0,m}^{d} = E_1(p_{0,m}^{d})\overline{N}_{0,m}^{d} \tag{4.15}$$

For the case of cooperation between BS and a relay r_k, a N/ACK message is sent from both u_m or r_k for each transmitted packet from BS. Furthermore, when r_k is transmitting, a N/ACK message is sent from u_m. Hence, using eqs. (4.8a) and (4.8b), the consumed energy for transmitting one packet from BS or r_k are respectively given by:

$$E_2(p_{0,k,m}^{c}) = k_3 + k_4 p_{0,k,m}^{c} \tag{4.16a}$$

$$E_3(p_{k,m}^{c}) = k_5 + k_6 p_{k,m}^{c} \tag{4.16b}$$

where $k_3 = \left(\frac{2(1+\tau_{ack})L}{R_b R_0} + \frac{(1+2\tau_{ack})L}{R_b R_0}\right).p_c$, $k_4 = k_6 = \frac{(1+2\tau_{ack})L}{R_b R_0}\beta_{amp}$ and $k_5 = \frac{2(1+2\tau_{ack})L}{R_b R_0}p_c$. Therefore, the average energy consumed per received packet in relay-assisted mode is:

$$\overline{E}_{k,m}^{c} = E_2(p_{0,k,m}^{c})\overline{N}_{0,k,m}^{c} + E_3(p_{k,m}^{c})\overline{N}_{k,m}^{c} \tag{4.17}$$

4.3.2 Energy minimization problem

The overall energy consumption in the cell can be written as in eq. (4.18) at the top of page. The left most term represents the overall energy consumed when a group of users is selected to communicate directly with the BS, i.e. $\sum_{k=1}^{K} s_{k,m} = 0$ and $s_{k,m}$ is a binary assignment variable equals to 1 if the relay k is selected to help user m and zero otherwise. However, the right most term designates the energy consumed for all users that adopt the relay-assisted mode, i.e. $s_{k,m} = 1$.

Thereby, the optimization problem can be firstly formulated with the objective function in (4.18) and has the constraints (c_1), (c_2), (c_6) and (c_7) defined in the optimization problem in (4.19), where

$$\overline{D}_{0,m}^{d}(p_{0,m}^{d}) = \frac{1}{1-P_{0,m}^{d}} \tag{4.20}$$

$$\min_{\boldsymbol{\lambda_{k,m}}} \sum_{m=1}^{M} \left(\left(1 - \sum_{k=1}^{K} s_{k,m} \right) E_1(p^d_{0,m}) \overline{N}^d_{0,m} + \sum_{k=1}^{K} s_{k,m} \left(E_2(p^c_{0,k,m}) \overline{N}^c_{0,k,m} + E_3(p^c_{k,m}) \overline{N}^c_{k,m} \right) \right)$$

subject to:

$$\begin{aligned}
&(c_1)\ s_{k,m} = \{0,1\}\ \forall\,(k,m) \in \mathcal{K} \times \mathcal{M} \text{ and } \sum_{k=1}^{K} s_{k,m} = \{0,1\}\ \forall m \in \mathcal{M} \\
&(c_2)\ \left(1 - \sum_{k=1}^{K} s_{k,m} \right) \overline{N}^d_{0,m} + \sum_{k=1}^{K} s_{k,m} \left(\overline{N}^c_{0,k,m} + \overline{N}^c_{k,m} \right) \leq \overline{\mathcal{D}}_m\ \forall m \in \mathcal{M} \\
&(c_3)\ \overline{N}^d_{0,m} = \overline{D}^d_{0,m}(p^d_{0,m})\ \forall m \in \mathcal{M} \\
&(c_4)\ \overline{N}^c_{0,k,m} = \overline{D}^c_{0,k,m}(p^c_{0,k,m})\ \forall\,(k,m) \in \mathcal{K} \times \mathcal{M} \\
&(c_5)\ \overline{N}^c_{k,m} = \overline{D}^c_{k,m}(p^c_{0,k,m}, p^c_{k,m})\ \forall\,(k,m) \in \mathcal{K} \times \mathcal{M} \\
&(c_6)\ \sum_{m=1}^{M} \left(\left(1 - \sum_{k=1}^{K} s_{k,m} \right) p^d_{0,m} + \sum_{k=1}^{K} s_{k,m} (p^c_{0,k,m} + p^c_{k,m}) \right) \leq P_{tot} \\
&(c_7)\ p^d_{0,m},\ p^c_{0,k,m},\ p^c_{k,m} \geq 0
\end{aligned} \tag{4.19}$$

$$\overline{D}^c_{0,k,m}(p^c_{0,k,m}) = \frac{1}{1 - P^c_{0,m} P^c_{0,k,m}} \tag{4.21}$$

$$\overline{D}^c_{k,m}(p^c_{0,k,m}, p^c_{k,m}) = \left(\frac{P^c_{0,m}(1 - P^c_{0,k,m})}{1 - P^c_{0,m} P^c_{0,k,m}} \right) \left(\frac{1}{1 - P^c_{k,m}} \right) \tag{4.22}$$

The constraint (c_1) ensures that only one relay can be selected per user, i.e. $s_{k,m} = 1$ if the relay r_k has been selected to help the user u_m. If $\sum_{k=1}^{K} s_{k,m} = 0$, then direct communication is preferable. The constraint (c_2) ensures that the average delay is less than or equal to $\overline{\mathcal{D}}_m$ for every user u_m. The constraint (c_6) states that the overall allocated power (for BS and relays) is less than or equal to the power constraint P_{tot}. Constraint (c_7) ensures the positivity of the allocated powers.

This problem cannot be solved by optimizing directly over the variables $p^d_{0,m}$, $p^c_{0,k,m}$, $p^c_{k,m}$ and the integer variable $s_{k,m}$. This is due to the existence of exponential, fractional and linear terms that make this problem intractable. Thus we propose another formulation to solve this problem by optimizing over the variables $p^d_{0,m}$, $p^c_{0,k,m}$, $p^c_{k,m}$ and $s_{k,m}$ and considering that $\overline{N}^d_{0,m}$, $\overline{N}^c_{0,k,m}$ and $\overline{N}^c_{k,m}$ $\forall k,m$ are also variables satisfying the new equality constraints in (c_3), (c_4) and (c_5) respectively. Moreover, the constraints (c_3), (c_4) and (c_5) ensure respectively that the equalities in eqs. (4.20), (4.21) and (4.22) are satisfied.

Hence, the energy minimization under average delay and power constraints can be formulated as in

$$\begin{aligned}\mathcal{L} = &\sum_{m=1}^{M}\sum_{k=1}^{K} s_{k,m}\left[\left(E_2(p^c_{0,k,m})+\gamma_m\right)\overline{N}^c_{0,k,m} + \left(E_3(p^c_{k,m})+\gamma_m\right)\overline{N}^c_{k,m} - \left(E_1(p^d_{0,m})+\gamma_m\right)\overline{N}^d_{0,m}\right]\\ &+\sum_{m=1}^{M}\sum_{k=1}^{K}\left[\alpha^c_{0,k,m}\left(\overline{D}^c_{0,k,m}(p^c_{0,k,m})-\overline{N}^c_{0,k,m}\right)+\alpha^c_{k,m}\left(\overline{D}^c_{k,m}(p^c_{0,k,m},p^c_{k,m})-\overline{N}^c_{k,m}\right)\right]\\ &+\sum_{m=1}^{M}\alpha_{0,m}\left(\overline{D}^d_{0,m}(p^d_{0,m})-\overline{N}^d_{0,m}\right)+\mu\sum_{m=1}^{M}\left[\left(1-\sum_{k=1}^{K}s_{k,m}\right)p^d_{0,m}+\sum_{k=1}^{K}s_{k,m}(p^c_{0,k,m}+p^c_{k,m})\right]-\mu P_{tot}\\ &+\sum_{m=1}^{M}\phi_m\left(\sum_{k=1}^{K}s_{k,m}-1\right)+\sum_{m=1}^{M}\left[\left(E_1(p^d_{0,m})+\gamma_m\right)\overline{N}^d_{0,m}-\gamma_m\overline{\mathcal{D}}_m\right]\end{aligned} \tag{4.23}$$

(4.19). One can notice that formulating the optimization problem with the objective function as in eq. (4.18) is equivalent to (4.19) just by substituting the equality constraints (c_3), (c_4) and (c_5) in the objective function of (4.19). The problem is equivalent to find the optimal vector $\boldsymbol{\lambda_{k,m}} = \{\lambda_{k,m} : (k,m) \in \mathcal{K}\times\mathcal{M}\}$ with $\lambda_{k,m} = [s_{k,m}\ p^d_{0,m}\ p^c_{0,k,m}\ p^c_{k,m}\ \overline{N}^d_{0,m}\ \overline{N}^c_{0,k,m}\ \overline{N}^c_{k,m}]$. In order to deal with continuous constraints, the integer constraint on $s_{k,m}$, i.e. (c_2), is first relaxed assuming it could be a time sharing factor between 0 and 1, leading to $\sum_{k=1}^{K} s_{k,m} \leq 1\ \forall m = \{1,\cdots,M\}$.

4.3.3 Problem Convexity

Proposition 1. *If the communication with the relay is decided, i.e. for $g^c_{0,m} > g^c_{0,k,m}$, the optimization problem in (4.19) is convex.*

Proof. *The proof is given in appendix C.1.* ■

From proposition 1, we can state that if $g^c_{0,m} \leq g^c_{0,k,m}$, less power is needed if a direct communication with u_m is decided, which is a preferred solution. However, if $g^c_{0,m} > g^c_{0,k,m}$, cooperative or direct communication can be decided according to the binary assignment variable $s_{k,m}$.

4.3.4 Problem Solution

The Lagrangian $\mathcal{L}$ associated with this problem is expressed in (4.23). We refer to the vectors $\boldsymbol{\phi_m} = [\phi_1,\cdots,\phi_M]^T$, $\boldsymbol{\gamma_m} = [\gamma_1,\cdots,\gamma_M]^T$, $\boldsymbol{\alpha_{k,m}} = \left\{(\alpha_{0,m},\alpha^1_{0,m},\alpha^c_{k,m}) : (k,m)\in\mathcal{K}\times\mathcal{M}\right\}$ and μ as Lagrange multipliers associated respectively to the constraints from (c_1) to (c_6).
Therefore, the primal minimization problem can be expressed as function of the Lagrange dual function g as the infinimum of the Lagrangian over the so called primal variable $\boldsymbol{\lambda_{k,m}}$. The Lagrange dual function g associated with the primal problem in (4.19) is defined as:

$$g(\boldsymbol{\gamma_m},\boldsymbol{\mu},\boldsymbol{\alpha_{k,m}},\boldsymbol{\phi_m}) = \inf_{\boldsymbol{\lambda_{k,m}}} \mathcal{L}(\boldsymbol{\lambda_{k,m}},\boldsymbol{\gamma_m},\mu,\boldsymbol{\alpha_{k,m}},\boldsymbol{\phi_m}) \tag{4.24}$$

Since the dual function g is an infinimum of a family of affine functions of dual variables $(\boldsymbol{\gamma_m}, \boldsymbol{\mu}, \boldsymbol{\alpha_{k,m}}, \boldsymbol{\phi_m})$, then it is concave even if the optimization problem is not convex [65] and yields to a lower bound of the primal problem for any $\boldsymbol{\gamma_m}$ and $\mu \geq 0$.
The primal minimization problem in (4.19) can be expressed as function of the dual function:

$$\begin{aligned} &\max_{\boldsymbol{\gamma_m},\ \mu,\ \boldsymbol{\alpha_{k,m}},\ \boldsymbol{\phi_m}} \quad g(\boldsymbol{\gamma_m}, \mu, \boldsymbol{\alpha_{k,m}}, \boldsymbol{\phi_m}) \\ &\text{subject to:} \quad \boldsymbol{\gamma_m},\ \mu \geq 0 \end{aligned} \tag{4.25}$$

It is worthwhile to notice that (4.19) and (4.25) are equivalent to:

$$\begin{aligned} &\sup_{\boldsymbol{\gamma_m},\boldsymbol{\mu},\boldsymbol{\alpha_{k,m}},\boldsymbol{\phi_m}} \left(\inf_{\boldsymbol{\lambda_{k,m}}} \mathcal{L}(\boldsymbol{\lambda_{k,m}}, \boldsymbol{\gamma_m}, \mu, \boldsymbol{\alpha_{k,m}}, \boldsymbol{\phi_m}) \right) \\ &\text{subject to:} \quad \boldsymbol{\gamma_m},\ \mu \geq 0 \end{aligned} \tag{4.26}$$

4.3.5 Optimal power and delay allocation

According to (4.26), it is convenient to firstly find the infinimum of the Lagrangian function $\mathcal{L}$ in (4.23) over the set of primal variables $\boldsymbol{\lambda_{k,m}}$ and then find the supremum of g over the dual variables. Thus, the primal variables can be expressed as function of the dual variables.

Optimal power allocation: Applying the gradient operator over the Lagrangian function $\mathcal{L}$ w.r.t. $\overline{N}^d_{0,m}$, we get the following:

$$\nabla_{\overline{N}^d_{0,m}} \mathcal{L} = \left(1 - \sum_{k=1}^{K} s_{k,m}\right) \left(E_1\left(p^d_{0,m}\right) + \gamma_m\right) - \alpha_{0,m} \tag{4.27}$$

where at the optimal $\nabla_{\overline{N}^{d*}_{0,m}} \mathcal{L} = 0$.
Therefore, if $\sum_{k=1}^{K} s_{k,m} = 1$, i.e. $\exists\ s_{k,m}$ such that $s_{k,m} = 1$, then $\alpha_{0,m} = 0$. Moreover if $\sum_{k=1}^{K} s_{k,m} = 0$, i.e. there is no relay that can minimize the overall energy consumed when serving user m and $s_{k,m} = 0\ \forall k \in \mathcal{K}$, then a direct transmission is preferable and the optimal allocated power $p^{d*}_{0,m}$ for direct transmission is given by:

$$p^{d*}_{0,m} = \left[\frac{(\alpha_{0,m} - \gamma_m - k_1)}{k_2}\right]^+ \tag{4.28}$$

where $[.]^+ \triangleq \max\{0, .\}$.
Moreover, optimizing respectively over the variables $\overline{N}^c_{0,k,m}$ and $\overline{N}^c_{k,m}$, we get the following:

$$\nabla_{\overline{N}^c_{0,k,m}} \mathcal{L} = s_{k,m}\left(E_2\left(p^c_{0,k,m}\right) + \gamma_m\right) - \alpha^c_{0,k,m} \tag{4.29}$$

$$\nabla_{\overline{N}^{c}_{k,m}} \mathcal{L} = s_{k,m} \left(E_3 \left(p^{c}_{k,m} \right) + \gamma_m \right) - \alpha^{c}_{k,m} \tag{4.30}$$

and then at the optimal $\nabla_{\overline{N}^{c*}_{0,k,m}} \mathcal{L} = 0$ and $\nabla_{\overline{N}^{c*}_{k,m}} \mathcal{L} = 0$. Therefore, if $s_{k,m} = 0$, then $\alpha^{c}_{0,k,m} = 0$ and $\alpha^{c}_{k,m} = 0$. Moreover, consider that a relay k is selected to help user m, then $s_{k,m} = 1$. Hence, using respectively eqs. (4.29) and (4.30), the allocated powers when BS or relay k communicates with user m are respectively $p^{c^*}_{0,k,m}$ and $p^{c^*}_{k,m}$ and are given by:

$$p^{c^*}_{0,k,m} = \left[\frac{\left(\alpha^{c}_{0,k,m} - \gamma_m - k_3 \right)}{k_4} \right]^{+} \tag{4.31}$$

$$p^{c^*}_{k,m} = \left[\frac{\left(\alpha^{c}_{k,m} - \gamma_m - k_5 \right)}{k_6} \right]^{+} \tag{4.32}$$

Since, γ_m, k_1, k_3 and k_5 are positive, then $\alpha_{0,m} \geq \gamma_m + k_3$, $\alpha^{c}_{0,k,m} \geq \gamma_m + k_3$ and $\alpha^{c}_{k,m} \geq \gamma_m + k_5$.

Optimal delay: By optimizing $\mathcal{L}$ w.r.t. to the primal variable $p^{d}_{0,m}$, we get the following:

$$\nabla_{p^{d}_{0,m}} \mathcal{L} = \left(1 - \sum_{k=1}^{K} s_{k,m} \right) \left(k_2 \overline{N}^{d}_{0,m} + \mu \right) + \alpha_{0,m} \nabla_{p^{d}_{0,m}} \mathcal{F}^{d}_{0,m} \tag{4.33}$$

where $\mathcal{F}^{d}_{0,m} = \frac{1}{1 - P^{d}_{0,m}(p^{d}_{0,m})}$ and at the optimal $\nabla_{p^{d*}_{0,m}} \mathcal{L} = 0$. Hence, if $\sum_{k=1}^{K} s_{k,m} = 1$, then $\alpha_{0,m} \nabla_{p^{d}_{0,m}} \mathcal{F}^{d}_{0,m} = 0$. Since $\mathcal{F}^{d}_{0,m}$ is a strictly decreasing function, then $\nabla_{p^{d}_{0,m}} \mathcal{F}^{d}_{0,m} \neq 0$ and hence $\alpha_{0,m} = 0$. Moreover, assuming that $\sum_{k=1}^{K} s_{k,m} = 0$, i.e. no relay is selected to help user m, then the optimal delay $\overline{N}^{d}_{0,m}$ associated to user m is given by [2]:

$$\overline{N}^{d^*}_{0,m} = \max \left\{ 1, \left[\frac{\left(-\mu - \alpha_{0,m} \nabla_{p^{d}_{0,m}} \mathcal{F}^{d}_{0,m} \right)}{k_2} \right]^{+} \right\} \tag{4.34}$$

Moreover, by firstly applying the gradient of $\mathcal{L}$ over $p^{c}_{k,m}$, $\nabla_{p^{c}_{k,m}} \mathcal{L}$ is given as follows:

$$\nabla_{p^{c}_{k,m}} \mathcal{L} = s_{k,m} \left(k_6 \overline{N}^{c}_{k,m} + \mu \right) + \alpha^{c}_{k,m} \mathcal{F}_1 \mathcal{F}_2 \nabla_{p^{c}_{k,m}} \mathcal{F}_3 \tag{4.35}$$

with

$$\mathcal{F}_1(p^{c}_{0,k,m}) = \frac{1}{1 - P^{c}_{0,m}(p^{c}_{0,k,m}) P^{c}_{0,k,m}(p^{c}_{0,k,m})} \tag{4.36a}$$

$$\mathcal{F}_2(p^{c}_{0,k,m}) = P^{c}_{0,m}(p^{c}_{0,k,m})(1 - P^{c}_{0,k,m}(p^{c}_{0,k,m})) \tag{4.36b}$$

$$\mathcal{F}_3(p^{c}_{k,m}) = \frac{1}{(1 - P^{c}_{k,m}(p^{c}_{k,m}))} \tag{4.36c}$$

2. It is worthwhile to note that $\max\{1, .\}$ is applied to insure that $\overline{N}^{d}_{0,m} \geq 1$

and at the optimal $\nabla_{p^{c*}_{k,m}} \mathcal{L} = 0$.
Hence if $s_{k,m} = 0$, then $\alpha^c_{k,m} \mathcal{F}_1 \mathcal{F}_2 \nabla_{p^c_{k,m}} \mathcal{F}_3 = 0$. But since $\mathcal{F}_1 \neq 0$, $\mathcal{F}_2 \neq 0$ and $\mathcal{F}_3$ is a strictly decreasing function $\left(\nabla_{p^c_{k,m}} \mathcal{F}_3 \neq 0\right)$ then $\alpha^c_{k,m} = 0$. Moreover, if $s_{k,m} = 1$, then $\overline{N}^{c*}_{k,m}$ is given as follows:

$$\overline{N}^{c*}_{k,m} = \left[\frac{\left(-\mu - \alpha^c_{k,m} \mathcal{F}_1 \mathcal{F}_2 \frac{d\mathcal{F}_3}{dp^c_{0,k,m}}\right)}{k_6} \right]^+ \tag{4.37}$$

with $\alpha^c_{k,m} > 0$.
The gradient of $\mathcal{L}$ w.r.t. $p^c_{0,k,m}$ is:

$$\nabla_{p^c_{0,k,m}} \mathcal{L} = s_{k,m} \left(k_4 \overline{N}^c_{0,k,m} + \mu\right) + \alpha^c_{0,k,m} \nabla_{p^c_{0,k,m}} \mathcal{F}_1 + \alpha^c_{k,m} \nabla_{p^c_{0,k,m}} \left(\mathcal{F}_1 \mathcal{F}_2 \mathcal{F}_3\right) \tag{4.38}$$

with $\nabla_{p^c_{0,k,m}} \mathcal{L} = 0$ at the optimal.
Thus, if $s_{k,m} = 0$, then the relay r_k is not selected to help user u_m and hence

$$\alpha^c_{0,k,m} \nabla_{p^c_{0,k,m}} \mathcal{F}_1 + \alpha^c_{k,m} \nabla_{p^c_{0,k,m}} \left(\mathcal{F}_1 \mathcal{F}_2 \mathcal{F}_3\right) = 0 \tag{4.39}$$

Since, $\alpha^c_{k,m} = 0$ for $s_{k,m} = 0$, then $\alpha^c_{0,k,m} \nabla_{p^c_{0,k,m}} \mathcal{F}_1 = 0$ and hence $\alpha^c_{0,k,m} = 0$. Moreover, if $s_{k,m} = 1$, the optimal delay $\overline{N}^{c*}_{0,k,m}$ is given as:

$$\overline{N}^{c*}_{0,k,m} = \max \left\{ 1, \left[\frac{-\mu - \left(\alpha^c_{0,k,m} + \alpha^c_{k,m} \mathcal{F}_2 \mathcal{F}_3\right) \nabla_{p^c_{0,k,m}} \mathcal{F}_1}{k_4} - \frac{\alpha^c_{k,m} \mathcal{F}_1 \mathcal{F}_3 \nabla_{p^c_{0,k,m}} \mathcal{F}_2}{k_4} \right]^+ \right\} \tag{4.40}$$

4.3.6 Relay selection strategy

Our problem also consists of a relay selection strategy, i.e. determining the value of $s_{k,m}$ that leads to the minimum energy path. Hence, minimizing the primal problem w.r.t. $s_{k,m}$, we get the following:

$$\nabla_{s_{k,m}} \mathcal{L} = \mathcal{Z}_{k,m} + \phi_m \tag{4.41}$$

with

$$\mathcal{Z}_{k,m} = (E_3(p^c_{k,m}) + \gamma_m) \overline{N}^c_{k,m} + (E_2(p^c_{0,k,m}) + \gamma_m) \overline{N}^c_{0,k,m} - (E_1(p^d_{0,m}) + \gamma_m) \overline{N}^d_{0,m} \tag{4.42}$$

and

$$\nabla_{s_{k,m}} \mathcal{L}\big|_{\lambda^*_{k,m}} \begin{cases} = 0, & \text{if } s_{k,m} \in (0,1), \\ < 0, & \text{if } s_{k,m} = 1. \end{cases} \tag{4.43}$$

$\mathcal{Z}_{k,m}$ can be interpreted as a weighted difference between the energy consumption of a relay-assisted communication and a direct communication. Since ϕ_m is common to all relays, only the relay k with the smallest $\mathcal{Z}_{k,m}$ can be selected to help the user m. Hence, $s_{k,m} = 1$ if $\mathcal{Z}_{k,m} < 0$ and $s_{k,m} = 0$ if $\mathcal{Z}_{k,m} > 0$. Thus the decision rule is given as follows:

Algorithm 2 Relay selection strategy

for $m = 1 : M$ **do**
  if $\mathcal{Z}_{k,m} > 0 \ \ \forall \ \ k \in \mathcal{K}$ **then**
    $s_{k,m} = 0$
    $(k,m)^* = (0,m)$
  else if then
    $(k,m)^* = \arg\min_{(k,m)} \mathcal{Z}_{k,m}$
    $s_{k,m} = 1$
  end if
end for

4.3.7 Lagrange dual variables update

As it can be noticed, the optimization problem in (4.25) is an unconstrained maximization problem. Moreover for $g^c_{0,m} > g^c_{0,k,m}$, the solution is unique due to the convexity and quasi-convexity of the objective and the constraint functions (See appendix C.1).

It follows that the dual function is differentiable and the subgradient method ensures the convergence of the problem toward an optimal solution [79]. Therefore, the dual problem in (4.25) can be solved w.r.t. $\boldsymbol{\gamma_m}, \mu, \boldsymbol{\alpha_{k,m}}$ according to the following dual variables updates used in algorithm 3 in Section 4.3.8.

μ update: The Lagrange dual variable μ is updated according to eq. (4.44) at the top of page. Where t is the iteration index and $\beta_\mu \in]0,1[$ is a sufficiently small positive step-size. The update of μ stops if the power constraint (c_6) is satisfied.

γ_m update: For a given μ, γ_m is updated in parallel for every user m according to eq. (4.45) at the top of page until it satisfies its delay constraint. Where $\beta_\gamma \in]0,1[$ is a sufficiently small positive step-size. The dual variable $\gamma_m(t)$ converges to the optimal γ^*_m as $t \to \infty$.

Updates of $\alpha_{0,m}$: For every updated γ_m, every user m updates its Lagrange dual variable $\alpha_{0,m}$ as follows:

$$\alpha_{0,m}(t+1) = \left[\alpha_{0,m}(t) + \beta_\alpha \left(\frac{1}{1 - P^d_{0,m}(p^d_{0,m})} - \overline{N}^d_{0,m}\right)\right]^+ \tag{4.47}$$

$$\mu(t+1) = \left[\mu(t) + \beta_\mu \left(\sum_{m=1}^{M} \left(\left(1 - \sum_{k=1}^{K} s_{k,m}\right) p_{0,m}^{d} + \sum_{k=1}^{K} s_{k,m}(p_{0,k,m}^{c} + p_{k,m}^{c})\right) - P_{tot}\right)\right]^{+} \tag{4.44}$$

$$\gamma_m(t+1) = \left[\gamma_m(t) + \beta_\gamma \left(\left(1 - \sum_{k=1}^{K} s_{k,m}\right) \overline{N}_{0,m}^{d} + \sum_{k=1}^{K} s_{k,m} \left(\overline{N}_{0,k,m}^{c} + \overline{N}_{k,m}^{c}\right) - \overline{\mathcal{D}}_m\right)\right]^{+} \tag{4.45}$$

$$\begin{bmatrix} \alpha_{0,k,m}^{c}(t+1) \\ \alpha_{k,m}^{c}(t+1) \end{bmatrix} = \left[\begin{bmatrix} \alpha_{0,k,m}^{c}(t) \\ \alpha_{k,m}^{c}(t) \end{bmatrix} + \beta_\alpha \begin{bmatrix} \left(\frac{1}{1-P_{0,m}^{c}(p_{0,k,m}^{c})P_{0,k,m}^{c}(p_{0,k,m}^{c})} - \overline{N}_{0,k,m}^{c}\right) \\ \left(\frac{P_{0,m}^{c}(p_{0,k,m}^{c})(1-P_{0,k,m}^{c}(p_{0,k,m}^{c}))}{(1-P_{0,m}^{c}(p_{0,k,m}^{c})P_{0,k,m}^{c}(p_{0,k,m}^{c}))(1-P_{k,m}^{c}(p_{k,m}^{c}))} - \overline{N}_{k,m}^{c}\right) \end{bmatrix}\right]^{+} \tag{4.46}$$

$\beta_\alpha \in]0,1[$ is a sufficient small positive step size [79]. The update of the dual variable $\alpha_{0,m}(t)$ continues until the equality constraint (c_3) is satisfied and it converges to the optimal value $\alpha_{0,m}^{*}$.

Updates of $\alpha_{0,k,m}^{c}$ and $\alpha_{k,m}^{c}$: For every couple (k,m) and for a given γ_m, the Lagrange dual variables $\alpha_{0,k,m}^{c}$ and $\alpha_{k,m}^{c}$ are updated as in eq. (4.46). As it can be noticed, eqs. (4.40) and (4.37) depend on $\alpha_{0,k,m}^{c}$ and $\alpha_{k,m}^{c}$. Hence, $\alpha_{0,k,m}^{c}$ and $\alpha_{k,m}^{c}$ are updated in parallel for every couple (k,m) [79]. As a consequence, eqs. (4.37) and (4.40) are also updated in parallel. Therefore, the dual variables are updated until the equality constraints (c_4) and (c_5) are satisfied.

Thereafter, for the updated $\gamma_m(t)$ and from $\alpha_{0,m}^{*}$, $\alpha_{0,k,m}^{c^*}$ and $\alpha_{k,m}^{c^*}$, the resource scheduler performs the relay selection strategy defined in algorithm 2 until satisfying the delay constraint for every user m.

4.3.8 Energy Minimization Algorithm

The algorithm 3 jointly allocates powers at BS and RS and selects an optimal relay if needed. This algorithm starts by initializing the Lagrange multipliers μ, γ_m and $\alpha_{0,m}$, $\alpha_{0,k,m}^{c}$ and $\alpha_{k,m}^{c}$ $\forall(k,m)$ (from step 1 to 3). For each updated γ_m, the resource scheduler has to find the optimal $\alpha_{0,m}^{*}$, $\alpha_{0,k,m}^{c^*}$ and $\alpha_{k,m}^{c^*}$ (also in parallel) that satisfy respectively the equality constraints (c_3), (c_4) and (c_5) (steps 4 to 6). Thereby, it selects the relay k that has the minimal energy path or decides the direct transmission (step 8). Then, the Lagrange multiplier γ_m is updated in parallel for all users (step 9) until they satisfy their delay constraints with a relative error ϵ_m or $\gamma_m(t) = 0$ for certain users. If $\gamma_m(t) = 0$ for $t > 1$, this means that the user m has a delay $< \overline{\mathcal{D}}_m$. If the power constraint P_{tot} can satisfy all users, the algorithm stops. Otherwise, the Lagrange multiplier μ is updated (step 16) and the resource scheduler restarts allocation. Now, if the available power P_{tot} cannot satisfy all users and the *equality* in the delay constraint (c_2) holds for all users, then the problem is unfeasible. In this case, the user with the maximum energy-delay ratio

Algorithm 3 Joint Power and Relay Selection algorithm

1: Initialize: $\mu \leftarrow 0$
2: Initialize: $\gamma_m \leftarrow 0\ \forall m \in \mathcal{M}$
3: Initialize:

$$\begin{aligned} &\alpha_{0,m} > \gamma_m + k_1, \\ &\alpha^1_{k,m} > \gamma_m + k_3, \\ &\alpha^c_{k,m} > \gamma_m + k_5,\ \forall m \in \mathcal{M} \text{ and } k \in \mathcal{K} \end{aligned}$$

4: **while** (c_3), $(c_4)$and (c_5) not satisfied **do**
5: Update $\alpha_{0,m}$ by (4.47), $\alpha_{0,m}$ and $\alpha^c_{k,m}$ by (4.46) $\forall\,(k,m) \in \mathcal{K} \times \mathcal{M}$
6: **end while**
7: $t = 0$
8: Apply (2)
9: **while** $\exists m$ s.t. (c_1) not satisfied **do**
10: Update $\gamma_m(t)$ by (4.45)
11: $t \leftarrow t + 1$ and return to 3
12: **end while**
13: **if** (c_6) satisfied **then**
14: end
15: **else if** **then**
16: Update μ until (c_6) satisfied or problem unfeasible
17: Return to 2
18: **end if**
19: Substitute $p^{d^*}_{0,m}$, $p^{c^*}_{0,k,m}$, $p^{c^*}_{k,m}$ and $s^*_{k,m}$ $\forall (k,m) \in \mathcal{K} \times \mathcal{M}$ in (c_6)
20: **while** Problem is unfeasible **do**
21: Reject $u_m = \arg\max_m \{R_m : m \in \mathcal{M}\}$
22: $\mathcal{M} = \mathcal{M} - \{u_m\}$
23: Return to 1
24: **end while**

R_m is rejected (step 21) and the algorithm restarts allocation.

Theorem 2. *In delay limited systems, the higher the total power constraint is, the lower the total energy consumption in the system is induced.*

Proof. *See proof in Appendix C.2.* ■

Let us assume that each user has an energy-delay profile characterized by the plot in fig. 4.6. Therefore, fig. 4.6 depicts the total energy consumed (left side axis) in joules and the average delay (right side axis) versus the total power needed p_m for user u_m. Referring to this plot and as μ is updated by eq. (4.44), the power constraint becomes more stringent and the power that satisfies the inequality

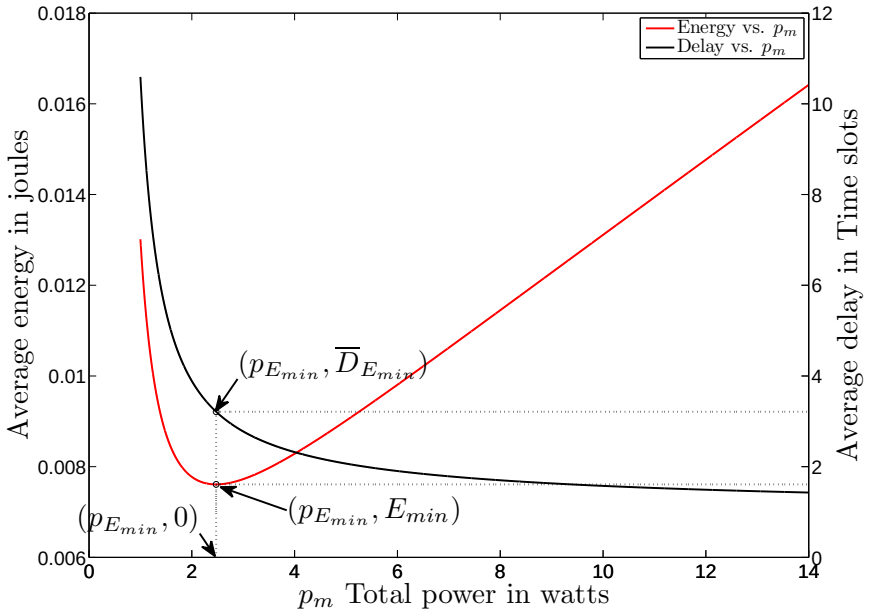
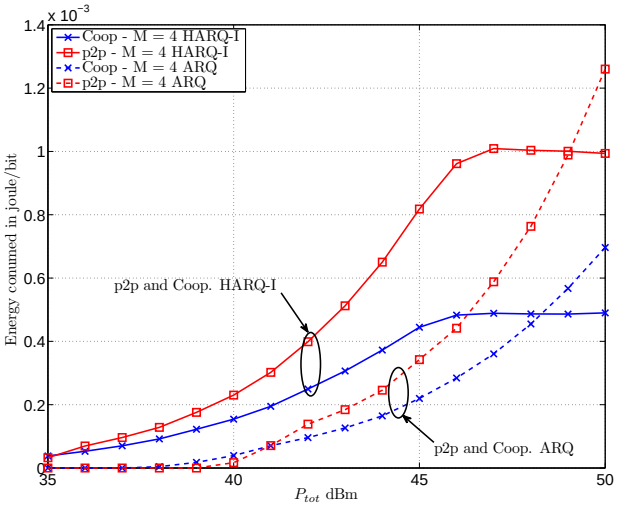


Figure 4.6: Energy-Delay versus power $p_m = p^c_{0,k,m} + p^c_{k,m}$

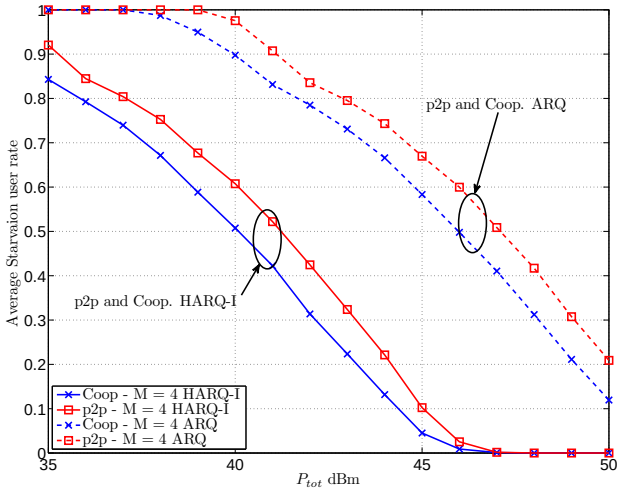
constraint in (c_2) is shifted to the left of the minimal point $(p_{E_{min}})$ for every user that have $\overline{\mathcal{D}}_m \geq \overline{D}_{E_{min}}$. Where $p_{E_{min}}$ is the total optimal power that leads to a minimum energy consumption regardless the delay constraint and $p_{E_{min}}$ corresponds to $\overline{D}_{E_{min}}$. Therefore, as μ increases, the system energy consumption increases (Theorem 2). Thus, if the *equality* in the delay constraint (c_2) is satisfied $\forall\ m$, and there is still no sufficient power that satisfy all users, then the problem is unfeasible. In this case, the resource scheduler rejects the user that has the greater energy-delay ratio R_m in watts (step 21) defined in (C.3) (Appendix C.3) and restarts allocation according to the steps defined above.

4.3.9 Numerical analysis

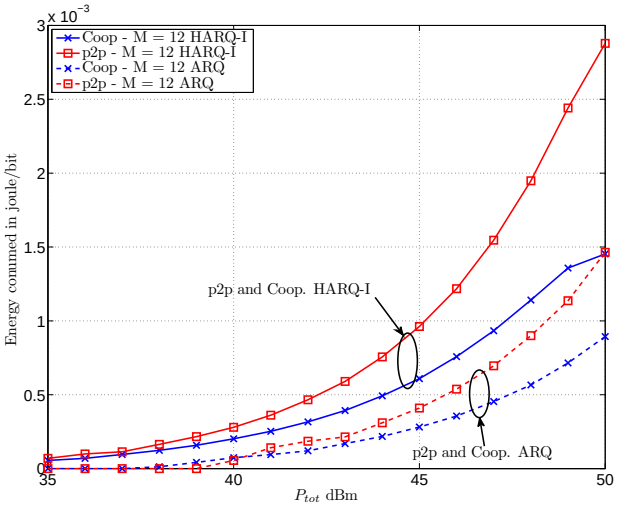
Consider a circular cell of radius R = 1 km consisting of 4 circularly distributed relays located at a distance R/2 from BS. We also consider that $M = 4$ and 12 users are randomly distributed in the network according to a uniform distribution with heterogeneous delay constraints $\overline{\mathcal{D}}_m = 1.5$ and 2.5 TSs. We consider a BPSK modulation, where each packet contains $L = 102$ information bits. The ARQ and HARQ-I retransmission protocols are used as well as an LDPC code with rate $R_0 = \frac{1}{2}$. The point-to-point transmission (p2p) is also presented for reference.



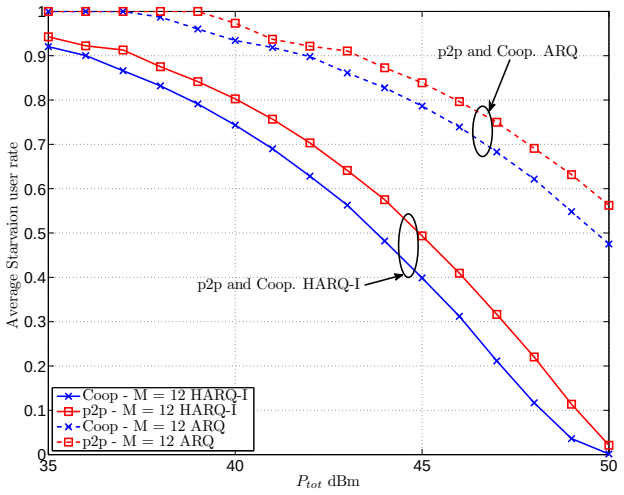
(a) Average energy consumed for $M = 4$



(b) Average Starvation user rate for $M = 4$



(c) Average energy consumed for $M = 12$



(d) Average Starvation user rate for $M = 12$

Figure 4.7: Average energy consumed in joules/bit and the average starvation user rate versus P_{tot} for $M = 4$ and 12. Delay constraints $\overline{\mathcal{D}}_m = 1.5$ and 2.5 TSs

4.3.10 Optimal energy consumed and average starvation rates

Using the resource allocation algorithm, figs. 4.7a and 4.7b plot the average energy consumed in joules/bit for $M = 4$ versus the available total power constraint P_{tot} in dBm. It can be shown that as P_{tot} increases, the energy consumption increases due to the decrease in the average starvation user rate. At nearly 46 dBm, the average starvation user rate becomes null for p2p and relay-assisted communication

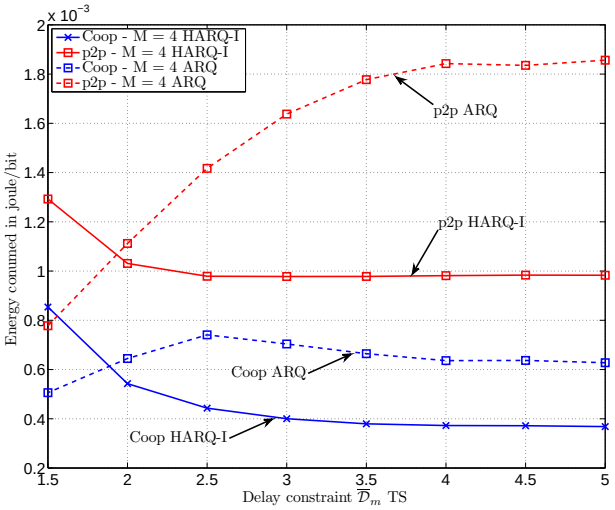
adopting HARQ-I retransmission protocol. However, the average energy consumed became stable and achieves 0.44×10^{-3} and 1×10^{-3} joules/bit for relay-assisted and p2p communications respectively. Thus, adopting relay-assisted network leads to an energy saving of more than 50 %. This significant gain is due to the path-loss gain offered by the relaying technique. However, using ARQ scheme and at $P_{tot} = 50$ dBm, the average energy consumption is about 0.7×10^{-3} and 1.25×10^{-3} joules/bit for cooperative and p2p schemes respectively with a corresponding starvation rate of about 0.2 and 0.12 respectively.
Now, referring to figs 4.7c and 4.7d and increasing M to 12 users, the average energy consumed increases significantly. It attains about 1.48×10^{-3} and 2.7×10^{-3} joules/bit for cooperative and p2p HARQ-I communications respectively with zero starvation rate. This increase in energy consumption is due to the increase of the number of users in the network. However, the total power needed is about $P_{tot} = 50$ dBm for both. Moreover, the p2p and cooperative ARQ schemes have a high average starvation user rates. This high starvation shows that there is no sufficient power to satisfy all users and hence the algorithm rejects users according to their energy-delay ratio R_m until the power constraint is satisfied.

Therefore, comparing ARQ to HARQ-I, we can state that a higher energy consumption is induced with a high starvation user rate. Moreover, a lower power constraint is needed to satisfy all users with HARQ-I scheme. This is due to the utilized channel coding that enhances the probability of correct detection, implying in turn a lower delay at the same SNR. In particular, relay-assisted HARQ-I based network can achieve a lower energy consumption as well as a lower starvation rate.

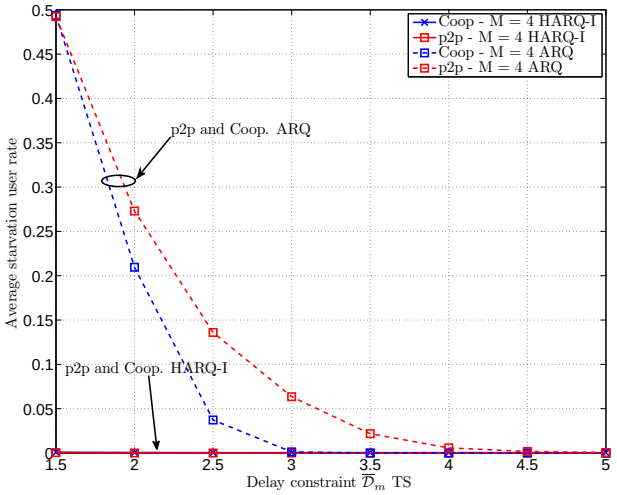
4.3.11 Energy consumption and starvation rate versus delay requirement

Le us now consider that all users have the same delay constraint $\overline{\mathcal{D}}_m$ and the total power constraint is $P_{tot} = 50$ dBm. For M = 4, figs. 4.8a and 4.8b depict respectively the average energy consumed in joule/bit and the average starvation user rate versus $\overline{\mathcal{D}}_m$. It can be shown that adopting HARQ-I, a zero starvation rate can be obtained for both p2p and cooperative communications. At $\overline{\mathcal{D}}_m = 1.5$, the average consumed energy is about 0.83×10^{-3} and 1.3×10^{-3} joules/bit for cooperative and p2p communications respectively. Thus adopting cooperation, an energy saving of about 36 % is enabled. Moreover, at high $\overline{\mathcal{D}}_m$, the average energy becomes stable with 0.38×10^{-3} and 0.98×10^{-3} joules/bit for p2p and cooperative networks respectively. Hence the energy saving increases to approximately 60 %. Therefore, cooperation enables a substantial energy saving as the delay constraint increases. This, energy saving is due to the pathloss gain exploited by cooperation. However, for $\overline{\mathcal{D}}_m < 2$, the energy gain offered by cooperative communication degrades, since some of the packets must be received at the first transmission instant in order to satisfy these stringent delay constraints. In other words, the link between the BS and every user u_m must be robust.

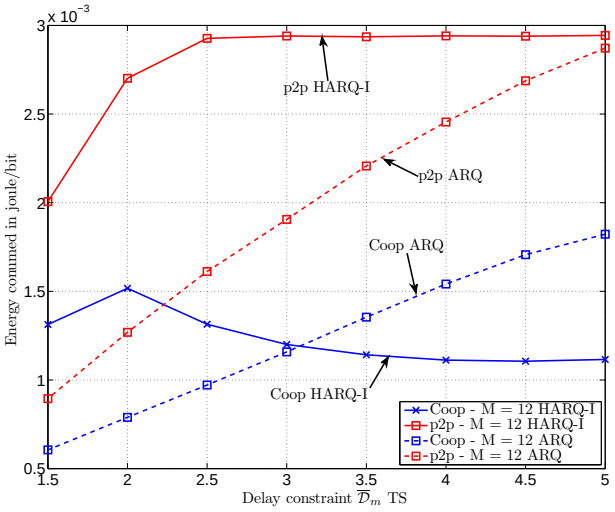
Now adopting ARQ protocol, a higher energy consumption and starvation rate are jointly observed. At $\overline{\mathcal{D}}_m = 1.5$, the average starvation rate is about 0.5 for both p2p and cooperative schemes, and hence a lower energy is consumed, because of the little number of satisfied users in the network. However, as $\overline{\mathcal{D}}_m$



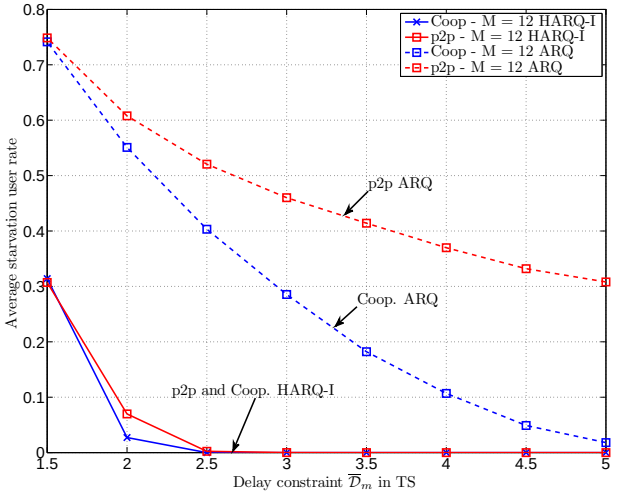
(a) Average energy consumed for $M = 4$



(b) Average Starvation user rate for $M = 4$



(c) Average energy consumed for $M = 12$



(d) Average Starvation user rate for $M = 12$

Figure 4.8: Average energy consumed in bits/joule and the average starvation user rate versus $\overline{\mathcal{D}}_m$ for $M = 4$ and 12. Power constraint is fixed at $P_{tot} = 50$ dBm

increases, the starvation rate decreases rapidly leading to an increase in the total energy consumption. A zero starvation rate can be observed at $\overline{\mathcal{D}}_m = 3$ and 4 for p2p and cooperative ARQ communications respectively. Their corresponding energies are respectively 1.8×10^{-3} and 0.65×10^{-3} joules/bit. Once again, we have an energy saving of about 50 % w.r.t. the ARQ schemes.

Increasing M to 12 users (see figs. 4.8c and 4.8d), the average energy consumption and the average starvation rate increase significantly. This is due to the lack of power to satisfy all users. Furthermore,

as the delay constraint increase, a low power constraint P_{tot} is needed and hence the average starvation rate decreases. In this case, the energy consumption starts increasing to achieve a stable state at high $\overline{\mathcal{D}}_m$. But the HARQ-I always offers the lower starvation w.r.t. ARQ protocol and at more than 2.5 TS a zero starvation rate for p2p and cooperation can be observed. In contrast, the ARQ schemes lead to approximately zero starvation for cooperative network and 0.4 for p2p network at $\overline{\mathcal{D}}_m = 5$. Therefore, as $\overline{\mathcal{D}}_m$ increases, the substantial energy saving is due to the increase of the number of retransmissions per packet that enables an increase in the temporal diversity in the system. Moreover, this allows to consume lower power and hence more users will access the radio resources.

In summary, the HARQ-I scheme is more energy efficient since it adopts the channel coding for every transmitted packet and the system reliability enhances. Moreover, HARQ-I cooperative network enables jointly lower energy consumption and starvation user rates, since it exploits the path-loss gain and also the channel coding leading to an energy efficient communication. At high delay constraint, it offers a significant energy gain up to 60 % compared to p2p communication with zero average starvation user rate.

4.4 Conclusion

In this chapter the energetic issues of cooperative cellular network in multi-user OFDMA scenarios are studied. The contribution of this chapter are summarized as follows:

1. A theoretical analysis of the energy efficiency (bits/joule) for HARQ-I and HARQ-CC types in point-to-point and relay-assisted networks is provided. Our proposed analysis takes into consideration finite packet length and a given MCS.
2. We further studied an energy minimization problem in relay-assisted networks considering an HARQ-I mechanism under delay-constrained users and total power consumption. This problem is solved efficiently by the Lagrange dual method. The convexity of this problem is proved and an energy minimization algorithm is proposed. The later jointly allocates the power at the BS and relay stations and performs the relay selection based on a minimal energy consumption criterion. Simulations have shown a clear energy saving as well as a better starvation rate compared to non cooperative transmissions in cellular networks. Moreover, the proposed solution of this problem is independent of the form of the delay function complexity and hence the proposed algorithm might be adapted to other HARQ schemes provided that the delay closed form expressions are derived and the convexity of the problem is guaranteed.

In a further work, other issues could be considered such as:

1. A subcarriers assignment strategy can be jointly considered with the other resources.
2. We can further suppose an MCS selection algorithm that chooses the best modulation and coding schemes leading to a minimal energy consumption.

Conclusions and Perspectives

Conclusion

Efficient resource allocation algorithms in multi-user cooperative cellular networks and the performance metrics analysis for cross-layer schemes are addressed in this thesis. The work provided herein accommodates the wireless standards demands such as 3GPP-LTE and 4G. We focused on two main issues: i) Satisfying the QoS requirements for every mobile user with ii) Optimal power consumption or minimum energy costs. We provided a theoretical analysis of the performance metrics of the HARQ schemes in *direct* and *relay-assisted* networks such as PER, throughput, end-to-end delay and the energy efficiency. In certain fading environments, such as slow fading channels, the average delay w.r.t. the fading realizations is irrelevant due to the non ergodicity of the process, hence the outage analysis is invoked. We theoretically analyzed the delay outage probability in relay-assisted slow fading channel. We further proposed two resource allocation algorithms in QoS aware cooperative cellular networks, addressing two issues i.e. a global rate maximization and an aggregate energy minimization.

Chapter 1 introduced the basic materials utilized in this dissertation. We started by presenting cooperative communication and pointing out the substantial gains it leads i.e. path-loss and diversity gains. We also gave a brief overview about the HARQ retransmission schemes which are incorporated in cross-layer protocols. We further provided the most prominent performance metrics for communication schemes with HARQ protocols e.g. the PER, throughput efficiency, average delay, energy efficiency and the outage probability. We then introduced the energy consumption model adopted in this work. Finally, some convex tools for optimization problems are presented.

In chapter 2, we focused on a resource optimization problem that aims to maximize the total throughput in multi-user relay-assisted OFDMA cellular network. This problem considers that each user has a throughput constraint to be satisfied and a maximum transmission power is available for the entire cell. The convexity of this optimization problem was proven and hence an optimal solution has been derived. The main novelty of this problem, is that a new resource scheduling policy has been suggested. It jointly optimizes the allocated power, assign subcarriers and selects relays to every user when the instantaneous CSI are known at the BS. Furthermore, our algorithm has demonstrated to outperform the existing

solutions in literature.

To accomplish the attributed task and to analyze the concepts of cross-layer protocols, the performances of HARQ-I and HARQ-II of type Chase combining in direct and relay-assisted scenarios has been theoretically analyzed in chapter 3. In more details, the mostly prominent metrics such as PER, end-to-end delay, throughput were tightly quantified in closed forms. These analyses take into account the practical issues of HARQ protocols, in which any finite packet length and modulation and coding schemes such as LDPC can be considered.

We also provided a theoretical analysis of the delay outage probability for relay-assisted HARQ-I system in slow fading channels which has never been done in literature. We provided hence a tight approximation for the delay outage probability. These theoretical expressions allow an important amount of simulation time to be saved. Hence, these analyses might be useful to precisely and rapidly evaluate the performance of cross-layer schemes in relay-assisted communication systems.

The contribution of chapter 4 focused on the analysis of the energetic issues of HARQ protocols in direct and relay-assisted networks. We firstly started by analyzing theoretically the energy efficiency for different HARQ schemes. Thereafter, an energy minimization problem in multi-user relay-assisted HARQ network has been investigated. This problem considers the static and dynamic energetic aspects in the cost function and the total power in the constraint function, provided that the QoS designated by the delay is guaranteed for every user. The main novelty of this optimization problem, is that it jointly considers the HARQ protocols and a given MCS when optimizing resources. Moreover, this algorithm is general enough to be applied for any HARQ scheme, such that the convexity of the cost and constraint functions are guaranteed. Therefore, based on the obtained optimal solution we proposed a new energy minimization algorithm that allocates the optimal power at the base station and relay stations. Moreover, during the resource allocation, a relay selection is performed and a decision rule to determine whether direct or cooperative transmission is preferred leading to the more energy-efficient solution.

Perspectives

There are still some open issues which could be addressed in future works:

1. The theoretical analysis of HARQ protocols started from the assumption that the ACK/NACK packets to be error-free with a negligible delay. However in reality it is possible to have a false ACK message. Therefore, it would be important to consider this point in future work.
2. The theoretical analysis of the performance metrics of incremental redundancy HARQ (HARQ-IR) scheme is not studied in this thesis. The study of the error probability at any transmission instant is still a challenging task in this case due to the high complexity of the relationship between the SNR and the error probability in HARQ-IR schemes. Hence, if a tractable expression of the

error probability could be found out, the study of the PER, average delay, throughput and energy efficiency could be addressed in future work.

3. In our energy minimization problem, we considered a simple linear energy consumption model. However, this linearity is no longer guaranteed if we take into account the consumption of the digital signal processing part. Therefore, a realistic energy consumption model to approach practical communication schemes is needed.
4. The proposed resource allocation algorithms should be implemented in an experimental platform in order to evaluate the interests of the theoretical analysis in real conditions.
5. In cellular networks, the rapid increase in the number of mobile users per cell and the relay deployments may cause interference with the adjacent cells. Therefore proposing efficient interference mitigation algorithms in HARQ schemes could be a future research topic.

APPENDIX

Appendix A

A.1 Derivation of A_n

As defined in eq. (3.36), A_n can be written as follows:

$$A_n = \int_0^{\gamma_M} \int_0^{\gamma_M-\gamma_1} \cdots \int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n-1}} p_{\gamma_{sd}}(\gamma_1)p_{\gamma_{sd}}(\gamma_2)\cdots p_{\gamma_{sd}}(\gamma_n)d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{A.1}$$

Therefore, in order to obtain a closed form expression of the above integral, we make a change of variable by dividing $\gamma_1, \gamma_2, \cdots, \gamma_n$ by γ_M. Hence, we obtain the following integral

$$A_n = \left(\frac{\gamma_M}{\gamma_{sd}}\right)^n \int_0^1 \int_0^{1-\gamma_1} \cdots \int_0^{1-\gamma_1-\cdots\gamma_{n-1}} \exp\left(\frac{-\gamma_1\gamma_M}{\overline{\gamma}_{sd}}\right)\cdots\exp\left(\frac{-\gamma_n\gamma_M}{\overline{\gamma}_{sd}}\right)d\gamma_1\cdots d\gamma_n \tag{A.2}$$

Therefore, using the following multiple integral identity [80, eq. 2 pp. 614]:

$$\int_{\substack{x_1\geq 0 \\ \left(\frac{x_1}{q_1}\right)^{\alpha_1}+\cdots+\left(\frac{x_n}{q_n}\right)^{\alpha_n}\leq 1}} \cdots \int_{x_n\geq 0} f\left[\left(\frac{x_1}{q_1}\right)^{\alpha_1}+\cdots+\left(\frac{x_n}{q_n}\right)^{\alpha_n}\right] \times x_1^{p_1-1}\cdots x_n^{p_n-1}dx_1\cdots dx_n =$$

$$\frac{q_1^{p_1}\cdots q_n^{p_n}}{\alpha_1\cdots\alpha_n}\frac{\Gamma\left(\frac{p_1}{\alpha_1}\right)\cdots\Gamma\left(\frac{p_n}{\alpha_n}\right)}{\Gamma\left(\frac{p_1}{\alpha_1}+\cdots+\frac{p_n}{\alpha_n}\right)}\int_0^1 f(x)x^{\frac{p_1}{\alpha_1}+\cdots+\frac{p_n}{\alpha_n}-1}dx \tag{A.3}$$

under the assumptions that the one-dimensional integral on the right converges absolutely and p_i, q_i and α_i are positive $\forall\ i$.
A_n can be calculated in simple form as follows:

$$A_n = \left(\frac{\gamma_M}{\gamma_{sd}}\right)^n \frac{1}{\Gamma(n)}\int_0^1 \exp\left(-\frac{\gamma_M}{\overline{\gamma}_{sd}}x\right)x^{n-1}dx \tag{A.4}$$

Using integrals and series in [76, eq. (11) §2.33, pp. 108], A_n is computed as follows:

$$A_n = 1 - \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)\sum_{k=0}^{n-1}\frac{1}{k!}\left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^k \tag{A.5}$$

A.2 Derivation of B_n

As its defined in eq. (3.37), B_n can be written as:

$$B_{n,m} = \int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots-\gamma_{m-1}}\int_{\gamma_M-\gamma_1-\cdots-\gamma_m}^{\infty}\int_0^{\infty}\cdots\int_0^{\infty} a\exp\big(-g(\gamma_1+\cdots+\gamma_n)\big)p_{\gamma_{sd}}(\gamma_1)\cdots p_{\gamma_{sd}}(\gamma_n)\, d\gamma_1\cdots d\gamma_n \tag{A.6}$$

Therefore, integrating over the variables $\gamma_{m+1},\cdots,\gamma_n$, we have:

$$B_{n,m} = \left(\frac{1}{\bar{\gamma}_{sd}}\right)^n\left[\frac{1}{g+\frac{1}{\bar{\gamma}_{sd}}}\right]^{n-m}\exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots-\gamma_{m-1}} d\gamma_1\cdots d\gamma_m \tag{A.7}$$

Hence, dividing each variable by γ_M, we get the following:

$$B_{n,m} = \gamma_M^m\left(\frac{1}{\bar{\gamma}_{sd}}\right)^n\left[\frac{1}{g+\frac{1}{\bar{\gamma}_{sd}}}\right]^{n-m}\exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)\int_0^{1}\cdots\int_0^{1-\gamma_1-\cdots-\gamma_{m-1}} d\gamma_1\cdots d\gamma_m \tag{A.8}$$

and using [80, eq. 1 pp. 613], we have

$$\int_0^{1}\cdots\int_0^{1-\gamma_1-\cdots-\gamma_{m-1}} d\gamma_1 d\cdots d\gamma_m = \frac{1}{m!} \tag{A.9}$$

Finally, $B_{n,m}$ is given in closed form as follows:

$$B_{n,m} = \left(\frac{{\gamma_M}^m}{m!}\right)\left[\frac{1}{1+g\bar{\gamma}_{sd}}\right]^{n-m}\left(\frac{1}{\bar{\gamma}_{sd}}\right)^m\exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \tag{A.10}$$

A.3 Derivation of C_n

As defined in eq. C_n can be written as follows:

$$C_n = \int_{\gamma_M}^{\infty} \int_0^{\infty} \cdots \int_0^{\infty} a \exp\big(- g(\gamma_1 + \cdots + \gamma_n)\big) p_{\gamma_{sd}}(\gamma_1) p_{\gamma_{sd}}(\gamma_2) \cdots p_{\gamma_{sd}}(\gamma_n) d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{A.11}$$

It can be noticed that eq. (A.11) is an integral of separable function, hence it can be readily proved that C_n can be written as:

$$C_n = \left[\frac{1}{1 + g\bar{\gamma}_{sd}}\right]^n \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \tag{A.12}$$

A.4 Derivation of $A(n, n')$

$A(n, n')$ is written as:

$$A(n, n') = \int_0^{\gamma_M} \cdots \int_0^{\gamma_M - \gamma_1 \cdots \gamma_{n-1}} p_{\gamma_{sd}}(\gamma_1) \cdots p_{\gamma_{sd}}(\gamma_{n'}) p_{\gamma_{rd}}(\gamma_{n'+1}) \cdots p_{\gamma_{rd}}(\gamma_n) d\gamma_1 \cdots d\gamma_n \tag{A.13}$$

By dividing each variable in eq. (A.13) by γ_M, $A(n, n')$ can be rewritten as:

$$A(n, n') = \left(\frac{1}{\bar{\gamma}_{sd}}\right)^{n'} \left(\frac{1}{\bar{\gamma}_{rd}}\right)^{n-n'} \cdot {\gamma_M}^n \int_0^1 \cdots \int_0^{1-\gamma_1 \cdots - \gamma_{n'-1}} \int_0^{1-\gamma_1 \cdots - \gamma_{n'}} \ldots \int_0^{1-\gamma_1 - \cdots \gamma_{n-1}} \exp\left(-\frac{\gamma_1 \gamma_M}{\bar{\gamma}_s}\right) \cdots \exp\left(-\frac{\gamma_{n'} \gamma_M}{\bar{\gamma}_s}\right) \exp\left(-\frac{\gamma_{n'+1} \gamma_M}{\bar{\gamma}_r}\right) \cdots \exp\left(-\frac{\gamma_n \gamma_M}{\bar{\gamma}_r}\right) d\gamma_1 d\gamma_2 \ldots d\gamma_n \tag{A.14}$$

We firstly integrate w.r.t. $\gamma_{n'+1}, \cdots, \gamma_n$, i.e we integrate the function I_1 defined as:

$$I_1 = \int_0^{1-\gamma_1 \cdots - \gamma_{n'}} \ldots \int_0^{1-\gamma_1 - \cdots \gamma_{n-1}} \exp\left(-\frac{\gamma_{n'+1} \gamma_M}{\bar{\gamma}_{rd}}\right) \cdots \exp\left(-\frac{\gamma_n \gamma_M}{\bar{\gamma}_{rd}}\right) d\gamma_{n'+1} \ldots d\gamma_n \tag{A.15}$$

Dividing each variable in this integral by $1 - \gamma_1 \cdots - \gamma_{n'}$, I_1 can be written as:

$$I_1 = \int_0^1 \ldots \int_0^{1-\gamma_{n'+1} - \cdots \gamma_{n-1}} \exp\left(-\frac{\gamma_M (1 - \gamma_1 \cdots - \gamma_{n'})(\gamma_{n'+1} + \cdots + \gamma_n)}{\bar{\gamma}_{rd}}\right) (1 - \gamma_1 \cdots - \gamma_{n'})^{n-n'} d\gamma_{n'+1} \ldots d\gamma_n$$

As we notice, I_1 is an integral over a simplex domain [76] and using the identity in eq. (A.3). I_1 can be transformed into one dimensional integral form as follows:

$$I_1 = \frac{(1-\gamma_1 \cdots - \gamma_{n'})^{n-n'}}{\Gamma(n-n')} \int_0^1 x^{n-n'-1} \exp\left(-\frac{\gamma_M(1-\gamma_1 \cdots - \gamma_{n'})}{\bar{\gamma}_{rd}} \cdot x\right) dx \tag{A.16}$$

Therefore, I_1 is written in closed form as follows [76, eq. (11) §2.33, pp. 108]:

$$I_1 = \left(\frac{\overline{\gamma}_{rd}}{\gamma_M}\right)^{n-n'} \left(1 - \exp\left(-\frac{\gamma_M(1-\gamma_1 \cdots - \gamma_{n'})}{\bar{\gamma}_{rd}}\right) \sum_{k=0}^{n-n'-1} \frac{1}{k!} \left(\frac{\gamma_M(1-\gamma_1 \cdots - \gamma_{n'})}{\bar{\gamma}_{rd}}\right)^k\right) \tag{A.17}$$

Inserting I_1 in eq. (A.14) and after some simplifications, $A(n, n')$ can be written as:

$$A(n, n') = \varphi_1 - \varphi_2, \tag{A.18}$$

with φ_1 and φ_2 are respectively given by:

$$\varphi_1 = \left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^{n'} \int_0^1 \cdots \int_0^{1-\gamma_1 \cdots - \gamma_{n'-1}} \exp\left(-\frac{\gamma_1 \gamma_M}{\bar{\gamma}_s}\right) \cdots \exp\left(-\frac{\gamma_{n'} \gamma_M}{\bar{\gamma}_s}\right) d\gamma_1 \cdots d\gamma_{n'} \tag{A.19}$$

$$\varphi_2 = \left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^{n'} \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{rd}}\right) \cdot \sum_{k=0}^{n-n'-1} \left[\int_0^1 \int_0^{1-\gamma_1} \cdots \int_0^{1 \cdots - \gamma_{n'-1}} \exp\left(\left(\frac{\gamma_M}{\bar{\gamma}_r} - \frac{\gamma_M}{\bar{\gamma}_s}\right) \cdot (\gamma_1 \cdots + \gamma_{n'})\right) \frac{(1-\gamma_1 \cdots - \gamma_{n'})^k}{k!} d\gamma_1 d\gamma_2 ... d\gamma_{n'}\right] \tag{A.20}$$

Hence using (A.3) can be transformed into one dimensional integral. Then using [76, eq. (11) §2.33, pp. 108], φ_1 is given in closed form:

$$\varphi_1 = \left[1 - \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \sum_{k=0}^{n'-1} \frac{1}{k!} \left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^k\right] \tag{A.21}$$

Moreover, the integral φ_2 is a integral over a simplex domain. Therefore, using eq. (A.3) it can be transformed into one dimensional integral and is given by:

$$\varphi_2 = \frac{1}{\Gamma(n')}\left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^{n'} \exp\left(-\frac{\gamma_M}{\bar{\gamma}_{rd}}\right) \sum_{k=0}^{n-n'-1} \frac{1}{k!}\left(\frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^{k} Z_1(k) \tag{A.22}$$

where $Z_1(k)$ is written as follows:

$$Z_1(k) = \int_0^1 \exp\left(\left(\frac{\gamma_M}{\bar{\gamma}_{rd}} - \frac{\gamma_M}{\bar{\gamma}_{sd}}\right)\cdot x\right) x^{n'-1}(1-x)^k dx \tag{A.23}$$

Using the binomial theorem $(1-x)^k = \sum_{i=0}^{k} \binom{k}{i}(-1)^i x^i$, $Z_1(k)$ can be written as [76, eq. (11) §2.33, pp. 108]:

$$Z_1 = \sum_{i=0}^{k} \binom{k}{i} (-1)^i \frac{(n'+i-1)!}{\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^{(n'+i)}} \times \left[1 - \exp\left(-\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)\right) \sum_{j=0}^{n'+i-1} \frac{\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^j}{j!}\right] \tag{A.24}$$

A.5 Derivation of $B(n, n', m)$

We have $B(n, n', m)$ is given by,

$$B(n,n',m) = \underbrace{\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{m-1}}}_{1\to m} \underbrace{\int_{\gamma_M-\cdots\gamma_m}^{\infty}\int_0^{\infty}\cdots\int_0^{\infty}}_{m+1\to n} p_{\gamma_{sd}}(\gamma_1)\cdots p_{\gamma_{sd}}(\gamma_{n'}) p_{\gamma_{rd}}(\gamma_{n'+1})\cdots p_{\gamma_{rd}}(\gamma_n)$$
$$\times a \exp\left(-g\left(\gamma_1+\cdots+\gamma_n\right)\right) d\gamma_1 d\gamma_2 \cdots d\gamma_n \tag{A.25}$$

Hence we have two conditions according to n' values:
If $n' \geq m+1$ eq. (A.25) can be written as:

$$B(n,n',m) = \underbrace{\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{m-1}}\int_{\gamma_M-\cdots\gamma_m}^{\infty}}_{1\to m+1}\underbrace{\int_0^{\infty}\cdots\int_0^{\infty}}_{m+2\to n'}\underbrace{\int_0^{\infty}\cdots\int_0^{\infty}}_{n'+1\to n}\underbrace{p_{\gamma_{sd}}(\gamma_1)\cdots p_{\gamma_{sd}}(\gamma_{m+1})}_{1\to m+1}$$
$$\times\underbrace{p_{\gamma_{sd}}(\gamma_{m+2})\cdots p_{\gamma_{sd}}(\gamma_{n'})}_{m+2\to n'}\underbrace{p_{\gamma_{rd}}(\gamma_{n'+1})\cdots p_{\gamma_{rd}}(\gamma_n)}_{n'+1\to n}\times a\exp\left(-g\left(\gamma_1+\cdots+\gamma_{m+1}\right)\right)\exp\left(-g\left(\gamma_{m+2}+\cdots+\gamma_{n'}\right)\right)$$
$$\exp\left(\left(\gamma_{n'+1}+\cdots+\gamma_n\right)\right)d\gamma_1 d\gamma_2\cdots d\gamma_n \quad \text{(A.26)}$$

Theretofore, integrating the terms from $m+2$ to n, we get the following:

$$B(n,n',m) = a\left(\frac{1}{\bar{\gamma}_{sd}}\right)^n\left(\frac{1}{\bar{\gamma}_{rd}}\right)^{n-n'}\left[\frac{\bar{\gamma}_{sd}}{1+g\bar{\gamma}_{sd}}\right]^{n'-m-1}\left[\frac{\bar{\gamma}_{rd}}{1+g\bar{\gamma}_{rd}}\right]^{n-n'}\times\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{m-1}}\int_{\gamma_M-\cdots\gamma_m}^{\infty}$$
$$p_{\gamma_{sd}}(\gamma_1)\cdots p_{\gamma_{sd}}(\gamma_{m+1})\exp\left(-g\left(\gamma_1+\cdots+\gamma_{m+1}\right)\right)d\gamma_1\cdots d\gamma_{m+1}$$

Hence, integrating over the variable γ_{m+1}, we get the following:

$$B(n,n',m) = a\left(\frac{1}{\bar{\gamma}_{sd}}\right)^n\left(\frac{1}{\bar{\gamma}_{rd}}\right)^{n-n'}\left[\frac{\bar{\gamma}_{sd}}{1+g\bar{\gamma}_{sd}}\right]^{n'-m}\left[\frac{\bar{\gamma}_{rd}}{1+g\bar{\gamma}_{rd}}\right]^{n-n'}\exp\left(-\gamma_M\left(g+\frac{1}{\bar{\gamma}_{sd}}\right)\right)$$
$$\times\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{m-1}}d\gamma_1\cdots d\gamma_m \quad \text{(A.27)}$$

Then, using the integrals and series in [80, (eq. 2 pp. 614)]:

$$\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{m-1}}d\gamma_1\cdots d\gamma_m = \frac{{\gamma_M}^m}{m!} \quad \text{(A.28)}$$

Hence, finally $B(n,n',m)$ can be written in closed form as follows:

$$B(n,n',m) = \frac{1}{m!}\left(\frac{\gamma_M}{\bar{\gamma}_{sd}}\right)^m\left[\frac{1}{1+g\bar{\gamma}_{rd}}\right]^{n-n'}\left[\frac{1}{1+g\bar{\gamma}_{sd}}\right]^{n'-m}\times\exp\left(-\frac{\gamma_M}{\bar{\gamma}_{sd}}\right) \quad \text{(A.29)}$$

If $n' > m+1$ eq. (A.25) can be written as:

$$B(n,n',m) = \underbrace{\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n'-1}}}_{1\to n'}\underbrace{\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n'}}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{m-1}}\int_{\gamma_M-\cdots\gamma_m}^{\infty}}_{n'+1\to m+1}\underbrace{\int_0^{\infty}\cdots\int_0^{\infty}}_{m+2\to n}\underbrace{p_{\gamma_{sd}}(\gamma_1)\cdots p_{\gamma_{sd}}(\gamma_{n'})}_{1\to n'}$$
$$\underbrace{p_{\gamma_{rd}}(\gamma_{n'+1})\cdots p_{\gamma_{rd}}(\gamma_{m+1})}_{n'+1\to m+1}\underbrace{p_{\gamma_{rd}}(\gamma_{m+2})\cdots p_{\gamma_{rd}}(\gamma_n)}_{m+2\to n}\times a\exp\left(-g\left(\gamma_1+\cdots+\gamma_{n'}\right)\right)\exp\left(-g\left(\gamma_{n'+1}+\cdots+\gamma_{m+1}\right)\right)$$
$$\exp\left((\gamma_{m+2}+\cdots+\gamma_n)\right)d\gamma_1\cdots d\gamma_n \tag{A.30}$$

Therefore, integrating over the variables $\gamma_{m+1},\cdots,\gamma_n$, $B(n,n',m)$ can be written as:

$$B(n,n',m) = a\left(\frac{1}{\bar{\gamma}_{sd}}\right)^{n'}\left(\frac{1}{\bar{\gamma}_{rd}}\right)^{n-n'}\left[\frac{1}{g+\frac{1}{\bar{\gamma}_{rd}}}\right]^{n-(m+2)+2}\exp\left(-\gamma_M\left(g+\frac{1}{\bar{\gamma}_{rd}}\right)\right)$$
$$\underbrace{\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n'-1}}}_{1\to n'}\underbrace{\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n'}}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{m-1}}}_{n'+1\to m}\exp\left(\left(g+\frac{1}{\bar{\gamma}_{rd}}\right)\left(\gamma_1+\cdots+\gamma_m\right)\right)$$
$$\times\exp\left(-g(\gamma_{n'+1}+\cdots+\gamma_m)\right)\exp\left(-\frac{(\gamma_1+\cdots+\gamma_n')}{\bar{\gamma}_{sd}}\right)\times\exp\left(-\frac{(\gamma_{n'+1}+\cdots+\gamma_m)}{\bar{\gamma}_{rd}}\right)d\gamma_1\cdots d\gamma_m \tag{A.31}$$

After some simplification, $B(n,n',m)$ can be re-written as:

$$B(n,n',m) = \left(\frac{1}{\bar{\gamma}_{sd}}\right)^{n'}\left(\frac{1}{\bar{\gamma}_{rd}}\right)^{n-n'}\left[\frac{1}{g+\frac{1}{\bar{\gamma}_{rd}}}\right]^{n-(m+2)+2}\exp\left(-\frac{\gamma_M}{\bar{\gamma}_{rd}}\right)$$
$$\underbrace{\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n'-1}}}_{1\to n'}\underbrace{\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n'}}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{m-1}}}_{n'+1\to m}\exp\left(-\left(\frac{1}{\bar{\gamma}_{sd}}-\frac{1}{\bar{\gamma}_{rd}}\right)\left(\gamma_1+\cdots+\gamma_{n'}\right)\right)d\gamma_1\cdots d\gamma_m \tag{A.32}$$

Now, integrating over the variables, $\gamma_{n'+1},\cdots,\gamma_m$, we have the following [76, eq. (1) §4.632, pp. 613]:

$$\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n'}}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{m-1}}d\gamma_{n'+1}\cdots d\gamma_m = \frac{(\gamma_M-\gamma_1-\cdots\gamma_{n'})^{m-n'}}{(m-n')!} \tag{A.33}$$

Hence, inserting eq. (A.33) in (A.35), we get the following:

$$B(n,n',m) = \frac{1}{(m-n')!}\left(\frac{1}{\bar{\gamma}_{sd}}\right)^{n'}\left(\frac{1}{\bar{\gamma}_{rd}}\right)^{n-n'}\left[\frac{1}{g+\frac{1}{\bar{\gamma}_{rd}}}\right]^{n-(m+2)+2}\exp\left(-\frac{\gamma_M}{\bar{\gamma}_{rd}}\right)\times$$
$$\int_0^{\gamma_M}\cdots\int_0^{\gamma_M-\gamma_1-\cdots\gamma_{n'-1}}\exp\left(-\left(\frac{1}{\bar{\gamma}_{sd}}-\frac{1}{\bar{\gamma}_{rd}}\right)(\gamma_1+\cdots+\gamma_{n'})\right)(\gamma_M-\gamma_1-\cdots\gamma_{n'})^{m-n'}d\gamma_1\cdots d\gamma_m \quad \text{(A.34)}$$

Now, dividing each variable by γ_M, we have:

$$B(n,n',m) = \frac{\gamma_M^{m-n'}}{(m-n')!}\left(\frac{1}{\bar{\gamma}_{sd}}\right)^{n'}\left(\frac{1}{\bar{\gamma}_{rd}}\right)^{n-n'}\left[\frac{1}{g+\frac{1}{\bar{\gamma}_{rd}}}\right]^{n-(m+2)+2}\exp\left(-\frac{\gamma_M}{\bar{\gamma}_{rd}}\right)\times$$
$$\int_0^{1}\cdots\int_0^{1-\gamma_1-\cdots\gamma_{n'-1}}\exp\left(-\left(\frac{1}{\bar{\gamma}_{sd}}-\frac{1}{\bar{\gamma}_{rd}}\right)(\gamma_1+\cdots+\gamma_{n'})\,\gamma_M\right)(1-\gamma_1-\cdots\gamma_{n'})^{m-n'}d\gamma_1\cdots d\gamma_m \quad \text{(A.35)}$$

Then, using the identity in (A.3), $B(n,n',m)$ can be transformed into one dimensional integral as follows:

$$B(n,n',m) \quad = \quad \frac{\gamma_M^{m-n'}}{(m-n')!}\left(\frac{1}{\bar{\gamma}_{sd}}\right)^{n'}\left(\frac{1}{\bar{\gamma}_{rd}}\right)^{n-n'}\left[\frac{1}{g+\frac{1}{\bar{\gamma}_{rd}}}\right]^{n-(m+2)+2}\exp\left(-\frac{\gamma_M}{\bar{\gamma}_{rd}}\right)\times\frac{Z_2}{\Gamma(n')} \quad \text{(A.36)}$$

with

$$Z_2 = \int_0^1 \exp\left(-\gamma_M\left(\frac{1}{\bar{\gamma}_{sd}}-\frac{1}{\bar{\gamma}_{rd}}\right)x\right)(1-x)^{m-n'}x^{n'-1}dx \quad \text{(A.37)}$$

Applying binomial theorem, we can write

$$(1-x)^{m-n'} = \sum_{j=0}^{m-n'}\binom{m-n'}{j}(-1)^j(x)^j \quad \text{(A.38)}$$

Therefore, using eq. (A.36), Z_2 can be computed using the integral and series identity in [76, eq. (11) §2.33, pp. 108]:

$$Z_2 = \sum_{j=0}^{m-n'} \binom{m-n'}{j} (-1)^j \times \int_0^1 x^{j+n'-1} \exp\left(-\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right) x\right) dx \tag{A.39}$$

$$= \sum_{j=0}^{m-n'} \binom{m-n'}{j} (-1)^j \frac{(j+n'-1)!}{\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^{(j+n')}} \times \left[1 - \exp\left(-\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)\right) \sum_{i=0}^{j+n'-1} \frac{\left(\frac{\gamma_M}{\bar{\gamma}_{sd}} - \frac{\gamma_M}{\bar{\gamma}_{rd}}\right)^i}{i!}\right] \tag{A.40}$$

APPENDIX

Appendix B

B.1 Proof of Theorem 1

Eq. (3.68) can be split in two terms as $\mathbb{P}(O) = \mathbb{P}(O_1) + \mathbb{P}(O_2)$ with

$$\mathbb{P}(O_1) = \mathbb{P}\left(\overline{N}_s^c + \overline{N}_r^c > \theta_m; \overline{N}_s^d > \theta_m; \overline{N}_s^c > \theta_m\right) \tag{B.1}$$

$$\mathbb{P}(O_2) = \mathbb{P}\left(\overline{N}_s^c + \overline{N}_r^c > \theta_m; \overline{N}_s^d > \theta_m; \overline{N}_s^c \leq \theta_m\right) \tag{B.2}$$

$\mathbb{P}(O_1)$ corresponds to the case where outage occurs when the source is transmitting meaning the delays on the $s-d$ and $s-r$ links exceed θ_m. $\mathbb{P}(O_2)$ corresponds to the outage event when the delay on the link $s-r$ does not exceed θ_m, i.e. $\overline{N}_s^c \leq \theta_m$, but when the links $s-d$ and $s-r-d$ are in outage. Let us consider first the derivation of $\mathbb{P}(O_2)$. Using eq. (B.2) with eqs. (3.62), (4.11), (3.66) and (3.67):

$$\overline{N}_s^d > \theta_m \quad \Leftrightarrow \quad \alpha_1 < P(\gamma_{sd}) < 1 \tag{B.3}$$

$$\overline{N}_s^c \leq \theta_m \quad \Leftrightarrow \quad 0 < P(\gamma_{sr}) \leq \alpha_1 / P(\gamma_{sd}) \tag{B.4}$$

$$\overline{N}_s^c + \overline{N}_r^c > \theta_m \quad \Leftrightarrow \quad 0 < 1 - \frac{P(\gamma_{sd})(1 - P(\gamma_{sr}))}{\theta_m(1 - P(\gamma_{sd})P(\gamma_{sr})) - 1} < P(\gamma_{rd}). \tag{B.5}$$

with α_1 defined as in Theorem 1. Moreover, the first inequality in the right-hand side of eq. (B.5) is equivalent to $0 < P(\gamma_{sr}) < \frac{1}{P(\gamma_{sd})} - \frac{1}{\theta_m - 1}$ iff $\alpha_1 < P(\gamma_{sd}) < \alpha_2$ (which is always the case for $\theta_m \geq 2$) and α_2 defined as in Theorem 1. However for $\theta_m < 2$, we can have $\alpha_2 < P(\gamma_{sd}) < 1$ and hence the previous inequality on $P(\gamma_{sr})$ above is no longer valid. Hence, we have to verify that even in this case the events in eq. (B.2) are valid. In particular, we have to verify if $\alpha_2 < P(\gamma_{sd}) < 1$ implies $\overline{N}_s^c + \overline{N}_r^c > \theta_m$, the other terms remaining unaffected by this condition. The following lemma states about this property.

Lemma 1. *If $\theta_m < 2$ and $P(\gamma_{sd}) > \theta_m - 1$, then $\overline{N}_s^c + \overline{N}_r^c > \theta_m$*

Proof. *if $\theta_m < 2$, i.e. $\alpha_2 = \theta_m - 1$, and $\alpha_2 < P(\gamma_{sd}) < 1$, this implies that:*

$$\frac{1}{1 - P(\gamma_{sd})P(\gamma_{sr})} > \frac{1}{1 - \alpha_2 P(\gamma_{sr})}$$

Then using (4.11) *and* (3.66)

$$\begin{aligned}
\overline{N}_s^c + \overline{N}_r^c &> \frac{1}{1-\alpha_2 P(\gamma_{sr})}\left(1+\frac{P(\gamma_{sd})(1-P(\gamma_{sr}))}{1-P(\gamma_3)}\right) \\
&> \frac{1}{1-\alpha_2 P(\gamma_{sr})}\left(1+\frac{\alpha_2(1-P(\gamma_{sr}))}{1-P(\gamma_3)}\right) \\
&> \frac{1}{1-\alpha_2 P(\gamma_{sr})}\left(1+\alpha_2(1-P(\gamma_{sr}))\right) \\
&= 1+\frac{\alpha_2}{1-\alpha_2 P(\gamma_{sr})} \\
&> 1+\alpha_2 \\
&= \theta_m
\end{aligned}$$

■

Hence, $\mathbb{P}(O_2)$ can be split according to $P(\gamma_{sd})$ values as:

$$\mathbb{P}(O_2) = \mathbb{P}^{(1)}(O_2) + \mathbb{P}^{(2)}(O_2) \tag{B.6}$$

with $\mathbb{P}^{(1)}(O_2)$ is defined for $\alpha_1 < P(\gamma_{sd}) < \alpha_2$ and represents the outage probability when the link $s-r$ is not in outage, i.e. $\overline{N}_s^c \leq \theta_m$, however $s-d$ and the path $s-r-d$ are both in outage. It leads to the expression in eq. (B.10) at the top of the page [3]. $\mathbb{P}^{(2)}(O_2)$ accounts for the same physical situation but when $\alpha_2 < P(\gamma_{sd}) < 1$ and is given by:

$$\mathbb{P}^{(2)}(O_2) = \begin{cases} \mathbb{P}\left(\overline{N}_s^c < \theta_m; \alpha_2 < P(\gamma_{sd}) < 1\right) & \text{if } \theta_m < 2 \\ 0 & \text{otherwise} \end{cases}$$

which can be written as

$$\mathbb{P}^{(2)}(O_2) = \int_0^{P_{\gamma_{sd}}^{-1}(\alpha_2)} \int_{P_{\gamma_{sr}}^{-1}\left(\frac{\alpha_1}{P(\gamma_{sd})}\right)}^{\infty} p_{\gamma_{sd}}(\gamma_{sd})\, p_{\gamma_{sr}}(\gamma_{sr})\, d\gamma_{sd} d\gamma_{sr} \text{ iff } \theta_m < 2. \tag{B.7}$$

Let us now focus on $\mathbb{P}(O_1)$. Since, $\overline{N}_s^c > \theta_m \Rightarrow \overline{N}_s^c + \overline{N}_r^c > \theta_m$, $\mathbb{P}(O_1)$ can be written as:

3. obtained from eqs. (B.3) to (B.5).

$$\mathbb{P}^{(1)}(O_2) = \int_{P_{\gamma_{sd}}^{-1}(\alpha_2)}^{P_{\gamma_{sd}}^{-1}(\alpha_1)} \int_{P_{\gamma_{sr}}^{-1}\left(\frac{1}{P(\gamma_{sd})}-\frac{1}{\theta_m-1}\right)}^{\infty} \int_0^{P_{\gamma_3}^{-1}\left(1-\frac{P(\gamma_{sd})(1-P(\gamma_{sr}))}{\theta_m(1-P(\gamma_{sd})P(\gamma_{sr}))-1}\right)} p_{\gamma_{sd}}(\gamma_{sd}) p_{\gamma_{sr}}(\gamma_{sr}) p_{\gamma_3}(\gamma_3) d\gamma_{sd} d\gamma_{sr} d\gamma_3 \tag{B.10}$$

$$\mathbb{P}(O_1) = \mathbb{P}\left(\overline{N}_s^d > \theta_m; \overline{N}_s^c > \theta_m\right) \tag{B.8}$$

We have $\overline{N}_s^c > \theta_m \Leftrightarrow P(\gamma_{sr}) > \frac{\alpha_1}{P(\gamma_{sd})}$. Since $0 < P(\gamma_{sr}) < 1$, we have $P(\gamma_{sd}) > \alpha_1$. Hence, $\mathbb{P}(O_1)$ can be written as:

$$\begin{aligned} \mathbb{P}(O_1) &= \mathbb{P}\left(\alpha_1 < P(\gamma_{sd}) < 1; \frac{\alpha_1}{P(\gamma_{sd})} < P(\gamma_{sr}) < 1\right) \\ &= \int_0^{P_{\gamma_{sd}}^{-1}(\alpha_1)} \int_0^{P_{\gamma_{sr}}^{-1}\left(\frac{\alpha_1}{P(\gamma_{sd})}\right)} p_{\gamma_{sd}}(\gamma_{sd}) p_{\gamma_{sr}}(\gamma_{sr}) d\gamma_{sd} d\gamma_{sr} \end{aligned} \tag{B.9}$$

According to the value of θ_m, it can be noticed there exists overlapping domains in the calculation of $\mathbb{P}(O_1)$ and $\mathbb{P}^{(2)}(O_2)$. Indeed, the outer integral in (B.9) ranges from 0 to $P_{\gamma_{sd}}^{-1}(\alpha_1)$, i.e. $\alpha_1 < P(\gamma_{sd}) < 1$. On the other hand, the outer integral in (B.7) ranges from 0 to $P_{\gamma_{sd}}^{-1}(\alpha_2)$, i.e. $\alpha_2 < P(\gamma_{sd}) < 1$. Hence, if $\theta_m < 2$, the integration domain of (B.9) encompasses the one of (B.7), since $\alpha_1 < \alpha_2 < 1$. $\mathbb{P}(O_1)$ can hence be split into two terms, we get $\mathbb{P}(O_1) = \mathbb{P}^{(1)}(O_1) + \mathbb{P}^{(2)}(O_1)$ with:

$$\mathbb{P}^{(1)}(O_1) = \int_{P_{\gamma_{sd}}^{-1}(\alpha_2)}^{P_{\gamma_{sd}}^{-1}(\alpha_1)} \int_0^{P_{\gamma_{sr}}^{-1}\left(\frac{\alpha_1}{P(\gamma_{sd})}\right)} p_{\gamma_{sd}}(\gamma_{sd}) p_{\gamma_{sr}}(\gamma_{sr}) d\gamma_{sd} d\gamma_{sr}$$

$$\mathbb{P}^{(2)}(O_1) = \int_0^{P_{\gamma_{sd}}^{-1}(\alpha_2)} \int_0^{P_{\gamma_{sr}}^{-1}\left(\frac{\alpha_1}{P(\gamma_{sd})}\right)} p_{\gamma_{sd}}(\gamma_{sd}) p_{\gamma_{sr}}(\gamma_{sr}) d\gamma_{sd} d\gamma_{sr}$$

where $\mathbb{P}^{(1)}(O_1)$ is the outage probability when the links $s-d$ and $s-r$ are both in outage with $\theta_m < 2$, i.e. $\alpha_1 < P(\gamma_{sd}) < \alpha_2$ and $\mathbb{P}^{(2)}(O_1)$ corresponds to the same case but when $\alpha_2 < P(\gamma_{sd}) < 1$. Hence for $\theta_m < 2$, we have $\mathbb{P}^{(2)}(O_1) + \mathbb{P}^{(2)}(O_2) = 1 - \exp(-P_{\gamma_{sd}}^{-1}(\alpha_2)/\overline{\gamma}_{sd})$. Furthermore, for $\theta_m \geq 2$, $P_{\gamma_{sd}}^{-1}(\alpha_2) = 0$. Thus, $\mathbb{P}^{(2)}(O_1) = \mathbb{P}^{(2)}(O_2) = 0$, which proves the expression (3.71) in Theorem 1.

Integrating $\mathbb{P}^{(1)}(O_1)$ w.r.t. γ_{sr} gives:

$$\mathbb{P}^{(1)}(O_1) = G\left(\frac{1}{\overline{\gamma}_{sd}}\right) - \exp\left(-\frac{\gamma_{th}}{\overline{\gamma}_2}\right)(\alpha_1)^{\frac{1}{a\overline{\gamma}_2}}\, I_0 \tag{B.11}$$

where I_0 is given by:

$$I_0 = \int_{P^{-1}_{\gamma_{sd}}(\alpha_2)}^{P^{-1}_{\gamma_{sd}}(\alpha_1)} \underbrace{(P(\gamma_{sd}) - \alpha_1)^{\frac{-1}{a\overline{\gamma}_2}}}_{A} p_{\gamma_{sd}}(\gamma_{sd}) d\gamma_{sd} \tag{B.12}$$

$P(\gamma_{sd})$ is a function of the form $1/(1+u)$. The domain of interest of the PER in HARQ systems is the range from high to medium PER (low PER makes HARQ mechanism useless). Hence, I_0 can be evaluated owing to the first order Taylor expansion of the function A near to $u_0 = \exp(-a\gamma_{th})$:

$$A \approx (\lambda_1)^{\frac{-1}{a\overline{\gamma}_{sr}}} \left(1 - \frac{\lambda_2}{\overline{\gamma}_{sr}} + \frac{\lambda_3 u}{\overline{\gamma}_{sr}}\right). \tag{B.13}$$

Thus, substituting u by $\exp(a(\gamma_{sd} - \gamma_{th}))$ and integrating w.r.t. γ_{sd}, the integral I_0 can be approximated as in eq. (3.73a) where λ_1, λ_2 and λ_3 are defined in table 3.5. Furthermore, integrating $\mathbb{P}^{(1)}(O_2)$ in eq. (B.10) w.r.t. γ_3 gives $\mathbb{P}^{(1)}(O_2) = F_1 - F_2$ with F_1 and F_2 respectively defined as:

$$F_1 = \int_{P^{-1}_{\gamma_{sd}}(\alpha_2)}^{P^{-1}_{\gamma_{sd}}(\alpha_1)} \int_{P^{-1}_{\gamma_{sr}}\left(\frac{1}{P(\gamma_{sd})} - \frac{1}{\theta_m - 1}\right)}^{\infty} p_{\gamma_{sd}}(\gamma_{sd}) p_{\gamma_{sr}}(\gamma_{sr}) d\gamma_{sd} d\gamma_{sr}$$

$$F_2 = \int_{P^{-1}_{\gamma_{sd}}(\alpha_2)}^{P^{-1}_{\gamma_{sd}}(\alpha_1)} \int_{P^{-1}_{\gamma_{sr}}\left(\frac{1}{P(\gamma_{sd})} - \frac{1}{\theta_m - 1}\right)}^{\infty} \exp\left(-\frac{\gamma_{th}}{\overline{\gamma}_{rd}}\right)(\theta_m - 1)^{\frac{1}{a\overline{\gamma}_{rd}}} \times \left(\frac{\frac{1}{P(\gamma_{sd})} - \frac{1}{\theta_m - 1} - P(\gamma_{sr})}{1 - P(\gamma_{sr})}\right)^{\left(\frac{1}{a\overline{\gamma}_{rd}}\right)} p_{\gamma_{sd}}(\gamma_{sd})$$
$$\times p_{\gamma_{sr}}(\gamma_{sr}) d\gamma_{sd} d\gamma_{sr}$$

Hence by integrating w.r.t. γ_{sr}, $F_1 = e^{-\gamma_{th}/\overline{\gamma}_{sr}} I_1$ with:

$$I_1 = \int_{P^{-1}_{\gamma_{sd}}(\alpha_2)}^{P^{-1}_{\gamma_{sd}}(\alpha_1)} \underbrace{\left(\frac{1}{\frac{1}{P(\gamma_{sd})} - \frac{1}{\theta_m - 1}} - 1\right)^{\frac{-1}{a\overline{\gamma}_{sr}}}}_{B} p_{\gamma_{sd}}(\gamma_{sd}) d\gamma_{sd} \tag{B.15}$$

The first order Taylor expansion of B near to the point $u = \exp(\gamma_m - \gamma_{th})$ gives [4]:

$$B \approx (\lambda_1)^{\frac{-1}{a\overline{\gamma}_{sr}}} \left(1 - \lambda_2 + \frac{\lambda_3}{\overline{\gamma}_{sr}} \exp(-a\gamma_{th})u\right). \tag{B.16}$$

Replacing u by $\exp(a(\gamma_{sd} - \gamma_{th}))$ and integrating w.r.t. γ_{sd}, I_1 can be written as in (3.73b) with λ_1, λ_2, λ_3 given in table 3.5. We let $g(\gamma_{sd}) = \frac{1}{P(\gamma_{sd})} - \frac{1}{\theta_m - 1}$, $u = \exp(-a(\gamma_{sr} - \gamma_{th}))$, F_2 can be re-written as follows:

$$F_2 = \frac{(\theta_m - 1)^{\frac{1}{a\overline{\gamma}_{rd}}}}{a\overline{\gamma}_{sr}} \exp\left(-\frac{\gamma_{th}}{\overline{\gamma}_{sr}} - \frac{\gamma_{th}}{\overline{\gamma}_{rd}}\right) \times \int_{P^{-1}_{\gamma_{sd}}(\alpha_2)}^{P^{-1}_{\gamma_{sd}}(\alpha_1)} (1 - g(\gamma_{sd}))^{\frac{1}{a\overline{\gamma}_{rd}}} \times I_3 d\gamma_{sd} \tag{B.17}$$

with:

$$I_3 = \int_0^{\frac{g(\gamma_{sd})}{1-g(\gamma_{sd})}} u^{(\frac{1}{a\overline{\gamma}_{sr}} - 1)} \left(\frac{g(\gamma_{sd})}{1 - g(\gamma_{sd})} - u\right)^{\frac{1}{a\gamma_3}} du. \tag{B.18}$$

I_3 can be written in closed form [76, p. 315], leading to:

$$F_2 = \left(\frac{(\theta_m - 1)^{\frac{1}{\overline{\gamma}_{rd}}}}{a\overline{\gamma}_{sr}}\right) \exp\left(-\left(\frac{\gamma_{th}}{\overline{\gamma}_{sr}} + \frac{\gamma_{th}}{\overline{\gamma}_{rd}}\right)\right) \beta\left(\frac{1}{a\overline{\gamma}_{rd}} + 1, \frac{1}{a\overline{\gamma}_{sr}}\right) I_2 \tag{B.19}$$

with $\beta(\cdot, \cdot)$ being the beta function and:

$$I_2 = \int_{P^{-1}_{\gamma_{sd}}(\alpha_2)}^{P^{-1}_{\gamma_{sd}}(\alpha_1)} (1 - g(\gamma_{sd}))^{\frac{1}{a\overline{\gamma}_{rd}}} \left(\frac{g(\gamma_{sd})}{1 - g(\gamma_{sd})}\right)^{\frac{1}{a\overline{\gamma}_{sr}} + \frac{1}{a\overline{\gamma}_{rd}}} p_{\gamma_{sd}}(\gamma_{sd}) d\gamma_{sd} \tag{B.20}$$

Therefore, using the same approximation method as in (B.16), eq. (3.73c) is obtained which concludes the proof.

4. where $\gamma_m = \frac{P^{-1}_{\gamma_{sd}}(\alpha_1) + P^{-1}_{\gamma_{sd}}(\alpha_2)}{2}$

APPENDIX

Appendix C

C.1 Proof of proposition 1

The problem is generally not a convex optimization problem, since constraints (c_4) and (c_5) are not convex. Thus, we try to search if the problem solutions are optimal when relay-assisted communications can be used, i.e. for $g^c_{0,m} > g^c_{0,k,m}$.
Consider the two functions $h(p^c_{0,k,m})$ and $u(p^c_{0,k,m})$ defined over $p^c_{0,k,m} \in \,]\,0, +\infty\,[$. The first derivative of the composite function $f = h \circ u$ is given by:

$$\frac{\partial f}{\partial p^c_{0,k,m}} = \left(\frac{\partial h(u)}{\partial u}\right)\left(\frac{\partial u}{\partial p^c_{0,k,m}}\right) \tag{C.1}$$

let $h(u) = \frac{P^c_{0,m}(u^{-1})(1-P^c_{0,k,m}(u^{-1}))}{(1-P^c_{0,m}(u^{-1})P^c_{0,k,m}(u^{-1}))}$ and $u(p^c_{0,k,m}) = \frac{1}{p^c_{0,k,m}}$. Then, $\frac{\partial u}{\partial p^c_{0,k,m}} = \frac{-1}{{p^c_{0,m}}^2} < 0$ for $p^c_{0,k,m} > 0$ and $\frac{\partial h(u)}{\partial u} = \frac{\left(1-P^c_{0,m}(u^{-1})\right)\left(1-P^c_{0,k,m}(u^{-1})\right)}{\left(1-P^c_{0,m}(u^{-1})P^c_{0,k,m}(u^{-1})\right)^2}\left(g^c_{0,m} - g^c_{0,k,m}P^c_{0,m}(u^{-1})\right)$. Therefore, $\frac{\partial h(u)}{\partial u} > 0$, since $g^c_{0,m} > g^c_{0,k,m}$ and $P^c_{0,m} \in]0,1[$. Therefore, $f(p^c_{0,k,m})$ is a strictly decreasing function. Hence, for any p_1 and p'_1 such that $p_1 < p'_1$ and $t \in \,]\,0,1\,[$ we have $f(tp_1+(1-t)p'_1) < \max\{f(p_1), f(p'_1)\}$. Therefore $f(p^c_{0,k,m})$ is a quasiconvex function [65]. Since, $f(p^c_{0,k,m})$ is a strictly decreasing function, then $\max\{f(p_1), f(p'_1)\} = f(p_1)$. Also consider the strictly decreasing function $l(p^c_{k,m}) = \frac{1}{(1-P^c_{k,m}(p^c_{k,m}))}$. Hence, for any different p_2 and p'_2 such that $p_2 < p'_2$ and $t \in \,]\,0,1\,[$ the following condition is satisfied

$$l(tp_2 + (1-t)p'_2) < \max\left\{l(p_2), l(p'_2)\right\} = l(p_2) \tag{C.2}$$

Therefore $l(p^c_{k,m})$ is a quasiconvex function. Hence, for any (p_1, p_2) and (p'_1, p'_2) such that $p_1 < p'_1$ and $p_2 < p'_2$ we have $f(tp_1+(1-t)p'_1)l(tp_2+(1-t)p'_2) \leq f(p_1)l(p_2) = \max\{f(p_1)l(p_2), f(p'_1)l(p'_2)\}$. Therefore, $f(p^c_{0,k,m})l(p^c_{k,m})$ is a quasiconvex function. Moreover $-\overline{N}^c_{k,m}$ is an affine decreasing function. Hence, for any different positive N_1 and N'_1 such that $N_1 < N'_1$ and $t \in \,]\,0,1\,[$ we have $-(tN_1 + (1-t)N'_1) < -N_1$. Finally, for any triple (p_1, p_2, N_1) and (p'_1, p'_2, N'_1) such that $p_1 < p'_1$, $p_2 < p'_2$, $N_1 < N'_1$ and $t \in \,]\,0,1\,[$ we have $(h(tp_1+(1-t)p'_1)l(tp_2+(1-t)p'_2) - tN_1 - (1-t)N'_1) < (h(p_1)l(p_2) - N_1) = \max\{(h(p_1)l(p_2) - N_1),$ $(h(p'_1)l(p'_2) - N'_1)\}$. Thus constraint (c_5) is a quasiconvex function. By the same analogy, it can be easily shown that constraints (c_3) and (c_4) are also quasiconvex functions. Moreover, the objective function and all other constraint functions are affine. Thus it can be deduced from the quasiconvexity properties

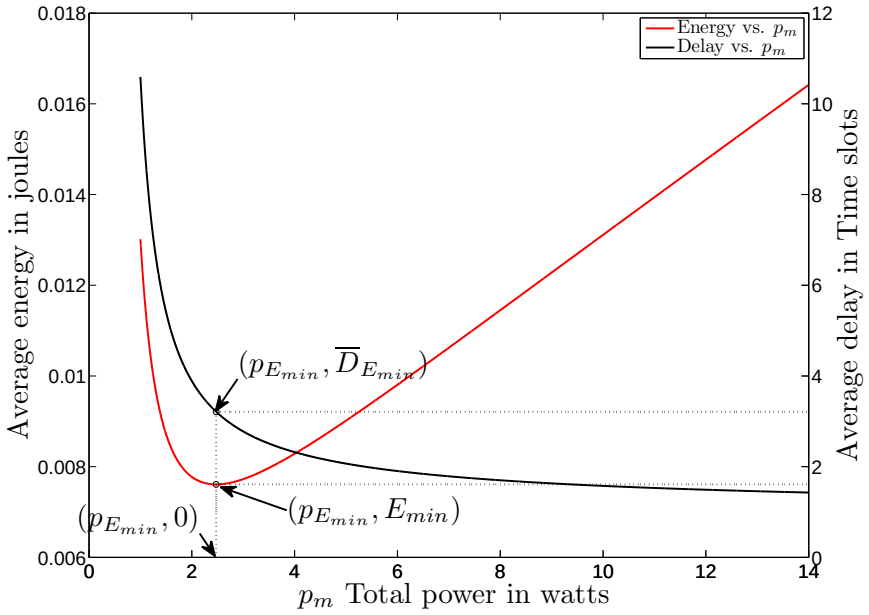


Figure 9: Energy-Delay versus power $p_m = p^c_{0,k,m} + p^c_{k,m}$

of (c_3), (c_4) and (c_5) and the linearity of the objective function and all other constraints that any local optimal solution is a global one [81].

C.2 Energy-Delay versus power

Suppose that there exists only one user m to be served and a relay k is selected to help this user. In order to analyze how the resource scheduler allocates powers for a user m according to its delay and power constraint P_{tot}, let us assume that $p_m = p^c_{0,k,m} + p^c_{k,m}$. This assumption on the power variables is just to be able to analyze the energy and delay versus p_m and give some insight about our proposed algorithm. Fig. 9 shows the average energy and average delay versus the power p_m for user m. Moreover, the energy curve has a minimum E_{min} at $p_m = p_{E_{min}}$ which corresponds to a delay $\overline{D}_{E_{min}}$. Therefore, the average energy function is strictly decreasing for $p_m < p_{E_{min}}$ and is increasing for $p_m > p_{E_{min}}$. Hence, there are several conditions to be studied and stated as follows:

- If user m has a delay constraint $\overline{\mathcal{D}}_m < \overline{D}_{E_{min}}$, then the required power $p_m > p_{E_{min}}$ (since the delay versus p_m is strictly decreasing) and the power p_m that satisfy the equality $\overline{N}^c_{0,k,m} + \overline{N}^c_{k,m} = \overline{\mathcal{D}}_m$ leads to the optimal minimum energy (Since for $p_m > p_{E_{min}}$, the energy function is increasing). Hence, if $p_m \leq P_{tot}$, user m is served. Otherwise, it goes in starvation.
- If $\overline{\mathcal{D}}_m \geq \overline{D}_{E_{min}}$, then the required power $p_m \leq p_{E_{min}}$. Therefore, if $p_{E_{min}} \leq P_{tot}$, then the optimal

power p_m that leads to a minimum energy, is the solution of the equality $\overline{N}^c_{0,k,m} + \overline{N}^c_{k,m} = \overline{D}_{E_{min}}$, i.e. $p_m = p_{E_{min}}$ and the consumed energy is E_{min}. Moreover, it is equivalent to find the minimum of the energy function without considering the delay constraint ($\gamma_m = 0$).

- If $\overline{\mathcal{D}}_m \geq \overline{D}_{E_{min}}$ and $p_{E_{min}} > P_{tot}$, the resource scheduler has to allocate a power in $p_m \in]0, p_{E_{min}}[$ and thus the energy function starts increasing as p_m decreases (decreasing p_m means increasing the delay and thus increasing the energy consumption). Hence, the resource scheduler starts updating μ until it finds $p_m \leq P_{tot}$ and it satisfies the inequality $\overline{N}^c_{0,k,m} + \overline{N}^c_{k,m} \leq \overline{\mathcal{D}}_m$ and then it served. Otherwise, user m goes in starvation.

Hence, these analysis prove that the overall system can have a minimum total energy consumption if there is a sufficient power constraint, i.e. for high P_{tot}, the resource scheduler allocates power $p_m \geq p_{E_{min}}$ for all $m \in \mathcal{M}$. Hence, this concludes the proof of theorem 2.

C.3 Energy-delay ratio

The energy-delay ratio R_m is defined as the ratio of the average energy consumed and the average delay associated to user m and given as follow:

$$R_m = \frac{B_w \times \left(s_{k,m} \overline{E}^c_{k,m} + (1 - s_{k,m}) \overline{E}^d_{0,m} \right)}{\left(1 - \sum_{k=1}^{K} s_{k,m}\right) \overline{N}^d_{0,m} + \sum_{k=1}^{K} s_{k,m} \left(\overline{N}^c_{0,k,m} + \overline{N}^c_{k,m} \right)} \quad \text{Watts} \tag{C.3}$$

where B_w is the bandwidth. If the problem is unfeasible, i.e. there is no sufficient power that can satisfy all users, the user with greater R_m is rejected.

Bibliography

[1] M. Dohler, "Virtual Antenna Arrays," Ph.D. dissertation, King's College London, 2003.

[2] J. N. Laneman, "Cooperative diversity in wireless networks: Algorithms and architectures," Dissertation, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2002.

[3] D. Chase, "Code Combining–A Maximum-Likelihood Decoding Approach for Combining an Arbitrary Number of Noisy Packets," *Communications, IEEE Transactions on*, vol. 33, no. 5, pp. 385–393, 1985.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, 1948.

[5] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999. [Online]. Available: http://dx.doi.org/10.1002/ett.4460100604

[6] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[7] S. Alamouti, "A simple transmit diversity technique for wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1451–1458, 1998.

[8] Tarokh, Vahid and Jafarkhani, Hamid and Calderbank, A.R., "Space-time block codes from orthogonal designs," *Information Theory, IEEE Transactions on*, vol. 45, no. 5, pp. 1456–1467, 1999.

[9] E. van der Meulen, *Transmission of Information in a T-terminal Discrete Memoryless Channel*. University of California, 1968.

[10] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *Information Theory, IEEE Transactions on*, vol. 25, no. 5, pp. 572–584, 1979.

[11] A. Sendonaris, E. Erkip, and B. Aazhang, "Increasing uplink capacity via user cooperation diversity," in *Information Theory, 1998. Proceedings. 1998 IEEE International Symposium on*, 1998, pp. 156–.

[12] ——, "User cooperation diversity. Part I. System description," *Communications, IEEE Transactions on*, vol. 51, no. 11, pp. 1927–1938, 2003.

[13] ——, "User cooperation diversity. Part II. Implementation aspects and performance analysis," *Communications, IEEE Transactions on*, vol. 51, no. 11, pp. 1939–1948, 2003.

[14] D. Hui, V. Lau, and W. H. Lam, "Cross-Layer Design for OFDMA Wireless Systems With Heterogeneous Delay Requirements," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2872 –2880, august 2007.

[15] H. Li, H. Luo, X. Wang, and C. Li, "Throughput Maximization for OFDMA Cooperative Relaying Networks with Fair Subchannel Allocation," in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, april 2009, pp. 1 –6.

[16] W. Nam, W. Chang, S.-Y. Chung, and Y. Lee, "Transmit Optimization for Relay-Based Cellular OFDMA Systems," in *Communications, 2007. ICC '07. IEEE International Conference on*, june 2007, pp. 5714 –5719.

[17] Y. Teng, M. Song, F. Niu, G. Chen, and J. Song, "Cooperative OFDMA resource allocation for multi-QoS guarantee: A cross-layer utility scheduling approach," in *Pervasive Computing (JCPC), 2009 Joint Conferences on*, dec. 2009, pp. 267 –272.

[18] D. Zhang, Y. Wang, and J. Lu, "QoS aware relay selection and subcarrier allocation in cooperative OFDMA systems," *IEEE Communications Letters*, vol. 14, no. 4, pp. 294 –296, april 2010.

[19] C.-N. Hsu, H.-J. Su, and P.-H. Lin, "Maximizing sum rate of OFDM transmission with decode-and-forward relaying," in *Green Circuits and Systems (ICGCS), 2010 International Conference on*, june 2010, pp. 263 –268.

[20] K. Vardhe, D. Reynolds, and B. Woerner, "Joint power allocation and relay selection for multiuser cooperative communication," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1255 –1260, april 2010.

[21] B. e. e. Gergely, "Economic and Ecological Impact of ICT," EARTH project deliverable D2.1, Tech. Rep., 2012.

[22] M. A. Imran *et al.*, "Most suitable efficiency metrics and utility functions," INFSO-ICT-247733 EARTH, Tech. Rep., 2012.

[23] A. e. e. Gunther, "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," EARTH project deliverable D2.3, Tech. Rep., 2012.

[24] Arnold, O. and Richter, F. and Fettweis, G. and Blume, O., "Power consumption modeling of different base station types in heterogeneous cellular networks," in *Future Network and Mobile Summit, 2010*, 2010, pp. 1–8.

[25] A. Chatzipapas, S. Alouf, and V. Mancuso, "On the minimization of power consumption in base stations using on/off power amplifiers," in *Online Conference on Green Communications (GreenCom), 2011 IEEE*, Sept. 2011, pp. 18 –23.

[26] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *Wireless Communications, IEEE*, vol. 18, no. 5, pp. 40 –49, Oct. 2011.

[27] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. Gonzalez, H. Klessig, I. Godor, M. Olsson, M. Imran, A. Ambrosy, and O. Blume, "Flexible power modeling of LTE base stations," in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, April 2012, pp. 2858 –2862.

[28] S. McLaughlin, P. Grant, J. Thompson, H. Haas, D. Laurenson, C. Khirallah, Y. Hou, and R. Wang, "Techniques for improving cellular radio base station energy efficiency," *Wireless Communications, IEEE*, vol. 18, no. 5, pp. 10 –17, Oct. 2011.

[29] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks," *Communications Magazine, IEEE*, vol. 48, no. 11, pp. 74 –79, Nov. 2010.

[30] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *Vehicular Technology Magazine, IEEE*, vol. 6, no. 1, pp. 37 –43, March 2011.

[31] Z. Niu, S. Zhou, Y. Hua, Q. Zhang, and D. Cao, "Energy-aware network planning for wireless cellular system with inter-cell cooperation," *IEEE Transactions on Wireless Communications,*, vol. 11, no. 4, pp. 1412 –1423, april 2012.

[32] D. Cao, S. Zhou, C. Zhang, and Z. Niu, "Energy saving performance comparison of coordinated multi-point transmission and wireless relaying," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, Dec. 2010, pp. 1 –5.

[33] S. Han, C. Yang, G. Wang, and M. Lei, "On the energy efficiency of base station sleeping with multicell cooperative transmission," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, sept. 2011, pp. 1536 –1540.

[34] X. Wang, P. Krishnamurthy, and D. Tipper, "Cell sleeping for energy efficiency in cellular networks: Is it viable?" in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, April 2012, pp. 2509 –2514.

[35] H. Holtkamp, G. Auer, and H. Haas, "On Minimizing Base Station Power Consumption," in *Vehicular Technology Conference (VTC Fall), 2011 IEEE*, Sept. 2011, pp. 1 –5.

[36] M. A. Imran *et al.*, "Most suitable efficiency metrics and utility functions," INFSO-ICT-247733 EARTH, Tech. Rep., Dec. 2011.

[37] O. Blume *et al.*, "Most promising tracks of green network technologies," INFSO-ICT-247733 EARTH, Tech. Rep., Dec. 2011.

[38] G. Miao, N. Himayat, G. Li, and S. Talwar, "Distributed interference-aware energy-efficient power optimization," *IEEE Trans. Wireless Commun.,*, vol. 10, no. 4, pp. 1323 –1333, April 2011.

[39] E. V. Belmega, S. Lasaulce, and M. Debbah, "A survey on energy-efficient communications," in *IEEE 22nd Int. Symp. on Personal Indoor and Mobile Radio Commun. (PIMRC), 2011*, Sept. 2011, pp. 1–5.

[40] A. El Gamal and S. Zahedi, "Minimum energy communication over a relay channel," in *Proc. on IEEE Int. Symp. Inf. Theo.*, June-July 2003.

[41] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for awgn relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, April 2006.

[42] C. Shuguang, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of mimo and cooperative mimo techniques in sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1089–1098, aug. 2004.

[43] S. K. Jayaweera, "Virtual mimo-based cooperative communication for energy-constrained wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 5, pp. 984–989, may 2006.

[44] R. Zhang, J.-M. Gorce, and K. Jaffres-Runser, "Low bound of energy-latency trade-off of opportunistic routing in multi-hop networks," in *ICC '09. IEEE Int. Conf. on Comm., 2009.*, June 2009, pp. 1–6.

[45] R. Zhang, "Analysis of Energy-Delay Performance In Multi-hop Wireless Sensor Networks," Ph.D. dissertation, INSA de Lyon, France, 2009.

[46] D. Chiarotto, O. Simeone, and M. Zorzi, "Throughput and energy efficiency of opportunistic routing with type-i harq in linear multihop networks," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, dec. 2010, pp. 1 –6.

[47] I. Stanojev, O. Simeone, Y. Bar-Ness, and K. Dong Ho, "Energy efficiency of non-collaborative and collaborative Hybrid-ARQ protocols," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 326–335, Jan. 2009.

[48] Y. Qi, R. Hoshyar, M. Imran, and R. Tafazolli, "H2-ARQ-relaying: Spectrum and energy efficiency perspectives," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1547 –1558, Sept. 2011.

[49] C. Le Martret, A. Le Duc, S. Marcille, and P. Ciblat, "Analytical performance derivation of hybrid arq schemes at ip layer," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1305 –1314, May 2012.

[50] M. Dohler and Y. Li, *Cooperative Communications: Hardware, Channel and PHY*. Wiley, 2010.

[51] 3GPP, "Evolved Universal Terrestrial Radio Access: Relay architectures for E-UTRA (LTE-Advanced) ," 3rd Generation Partnership Project, Tech. Rep., 2010.

[52] D. N. C. Tse, P. Viswanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1859–1874, 2004.

[53] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 99th ed. Wiley-Interscience, Aug. 1991.

[54] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, dec. 2004.

[55] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *Communications, IEEE Transactions on*, vol. 36, no. 4, pp. 389–400, 1988.

[56] S. Kallel, "Analysis of a type II hybrid ARQ scheme with code combining," *Communications, IEEE Transactions on*, vol. 38, no. 8, pp. 1133–1137, 1990.

[57] ——, "Complementary punctured convolutional (CPC) codes and their applications," *Communications, IEEE Transactions on*, vol. 43, no. 6, pp. 2005–2009, 1995.

[58] K. C. Beh, A. Doufexi, and S. Armour, "Performance Evaluation of Hybrid ARQ Schemes of 3GPP LTE OFDMA System," in *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, 2007, pp. 1–5.

[59] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *Communications, IEEE Transactions on*, vol. 52, no. 8, pp. 1311–1321, 2004.

[60] M. Levorato and M. Zorzi, "Performance analysis of type II hybrid ARQ with low-density parity-check codes," in *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, 2008, pp. 804–809.

[61] M. Yazdani and A. Banihashemi, "On construction of rate-compatible low-density Parity-check codes," *Communications Letters, IEEE*, vol. 8, no. 3, pp. 159–161, 2004.

[62] A. L. DUC, "Performance Closed-form Derivations and Analysis of Hybrid ARQ Retransmission Schemes in a Cross-layer Context," *PhD desertion, Telecom ParisTech*, 2010.

[63] M. Zorzi and R. Rao, "On the use of renewal theory in the analysis of ARQ protocols," *IEEE Transactions on Communications*, vol. 44, no. 9, pp. 1077–1081, Sep 1996.

[64] H. Karl and A. Willig, *Protocols and Architectures for Wireless Sensor Networks.* John Wiley & Sons, 2005.

[65] S. Boyd and L. Vandenberghe, *Convex Optimization.* New York, NY, USA: Cambridge University Press, 2004.

[66] J. N. Laneman and G. W. Wornell, "Distributed Space-Time Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[67] M. Dohler and Y. Li, *Cooperative Communications: Hardware, Channel & PHY.* John Wiley & sons, Ltd, 2010.

[68] A. Ribeiro, X. Cai, and G. B. Giannakis, "Symbol Error Probabilities for General Cooperative Links," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1264–1273, May 2005.

[69] Y. H. Chen and K. L. Hsieh, "A Dual Least-Square Approach of Tuning Optimal Propagation Model for Existing 3G Radio Network," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, vol. 6, may 2006, pp. 2942–2946.

[70] B. Maham, A. Behnad, and M. Debbah, "Analysis of Outage Probability and Throughput for Half-Duplex Hybrid-ARQ Relay Channels," *,IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3061–3070, 2012.

[71] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *Communications, IEEE Transactions on*, vol. 52, no. 8, pp. 1311–1321, 2004.

[72] ——, "Incremental redundancy hybrid arq schemes based on low-density parity-check codes," *Communications, IEEE Transactions on*, vol. 52, no. 8, pp. 1311–1321, 2004.

[73] A. Perez-Neira and M. Campalans, *Cross-Layer Resource Allocation in Wireless Communications: Techniques and Models from PHY and MAC Layer Interaction.* Elsevier Science, 2010.

[74] Q. Liu, S. Zhou, and G. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated arq over wireless links," *IEEE Transactions on Wireless Communications,*, vol. 3, no. 5, pp. 1746 – 1755, sept. 2004.

[75] X. Lagrange, "Throughput of harq protocols on a block fading channel," *Communications Letters, IEEE*, vol. 14, no. 3, pp. 257 –259, march 2010.

[76] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed., A. Jeffrey and D. Zwillinger, Eds. Elsevier Academic Press, 2007.

[77] M. Maaz, P. Mary, and M. Hélard, "Energy efficiency analysis in relay assisted hybrid-ARQ communications," in *PIMRC*. IEEE, 2012, pp. 2263–2268.

[78] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, pp. 30–37, June 2011.

[79] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 8, pp. 1439–1451, Aug.

[80] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed., A. Jeffrey and D. Zwillinger, Eds. Academic Press, Mar. 2007.

[81] J. Papandriopoulos, S. Dey, and J. Evans, "Optimal and distributed protocols for cross-layer design of physical and transport layers in manets," *Networking, IEEE/ACM Transactions on*, vol. 16, no. 6, pp. 1392 –1405, dec. 2008.

INSA de RENNES
Service des Formations

AVIS DU JURY SUR LA REPRODUCTION DE LA THESE SOUTENUE

Titre de la thèse:
Allocation de ressource et analyse des critères de performance dans les réseaux cellulaires coopératifs

Nom Prénom de l'auteur : MAAZ MOHAMAD

Membres du jury :
- Madame LANGLAIS Charlotte
- Monsieur LAGRANGE Xavier
- Monsieur POULLIAT Charly
- Madame HELARD Maryline
- Monsieur MARY Philippe
- Monsieur ROVIRAS Daniel

Président du jury : M. Xavier Lagrange

Date de la soutenance : 03 Décembre 2013

Reproduction de la these soutenue

☒ Thèse pouvant être reproduite en l'état
☐ Thèse pouvant être reproduite après corrections suggérées

Fait à Rennes, le 03 Décembre 2013

Signature du président de jury

Le Directeur,

M'hamed DRISSI

Résumé

Dans les systèmes de communications sans fil, la transmission de grandes quantités d'information et à faible coût énergétique sont les deux principales questions qui n'ont jamais cessé d'attirer l'attention de la communauté scientifique au cours de la dernière décennie.

Récemment, il a été démontré que la communication coopérative est une technique intéressante notamment parce qu'elle permet d'exploiter la diversité spatiale dans le canal sans fil. Cette technique assure une communication robuste et fiable, une meilleure qualité de service (QoS) et rend le concept de coopération prometteur pour les futurs générations de systèmes cellulaires. Typiquement, les QoS sont le taux d'erreurs paquet, le débit et le délai. Ces métriques sont impactées par le délai, induit par les mécanismes de retransmission Hybrid-Automatic Repeat-Request (HARQ) inhérents à la réception d'un paquet erroné et qui a un retard sur la QoS demandée. En revanche, les mécanismes HARQ créent une diversité temporelle. Par conséquent, l'adoption conjointe de la communication coopérative et des protocoles HARQ pourrait s'avérer avantageux pour la conception de schémas cross-layer.

Nous proposons tout d'abord une stratégie de maximisation de débit total dans un réseau cellulaire hétérogène. Nous introduisons un algorithme qui alloue la puissance optimale à la station de base (BS) et aux relais, qui à chaque utilisateur attribue de manière optimale les sous-porteuses et les relais. Nous calculons le débit maximal atteignable ainsi que le taux d'utilisateurs sans ressources dans le réseau lorsque le nombre d'utilisateurs actifs varie. Nous comparons les performances de notre algorithme à ceux de la littérature existante, et montrons qu'un gain significatif est atteint sur la capacité globale.

Dans un second temps, nous analysons théoriquement le taux d'erreurs paquet, le délai ainsi que l'efficacité de débit des réseaux HARQ coopératifs, dans le canal à évanouissements par blocs. Dans le cas des canaux à évanouissement lents, le délai moyen du mécanisme HARQ n'est pas pertinent à cause de la non-ergodicité du processus. Ainsi, nous nous intéressons plutôt à la probabilité de coupure de délai en présence d'évanouissements lents. La probabilité de coupure de délai est de première importance pour les applications sensibles au délai. Nous proposons une forme analytique de la probabilité de coupure permettant de se passer de longues simulations.

Dans la suite de notre travail, nous analysons théoriquement l'efficacité énergétique (bits/joule) dans les réseaux HARQ coopératifs. Nous résolvons ensuite un problème de minimisation de l'énergie dans les réseaux coopératifs en liaison descendante. Dans ce problème, chaque utilisateur possède une contrainte de délai moyen à satisfaire de telle sorte que la contrainte sur la puissance totale du système soit respectée. L'algorithme de minimisation permet d'attribuer à chaque utilisateur la station-relai optimale et sa puissance ainsi que la puissance optimale de la BS afin de satisfaire les contraintes de délai. Les simulations montrent qu'en termes de consommation d'énergie, les techniques assistées par relais prédominent nettement les transmissions directes, dans tout système limité en délai. En conclusion, les travaux proposés dans cette thèse peuvent promettre d'établir des règles fiables pour l'ingénierie et la conception des futures générations de systèmes cellulaires énergétiquement efficaces.

Abstract

In wireless systems, transmitting large amounts of information with low energetic cost are two main issues that have never stopped drawing the attention of the scientific community during the past decade.

Later, it has been shown that cooperative communication is an appealing technique that exploits spatial diversity in wireless channel. Therefore, this technique certainly promises a robust and reliable communications, higher quality-of-service (QoS) and makes the cooperation concept attractive for future cellular systems. Typically, the QoS requirements are the packet error rate, throughput and delay. These metrics are affected by the delay, where each erroneous packet is retransmitted several times according to Hybrid-Automatic Repeat-Request (HARQ) mechanism inducing a delay on the demanded QoS but a temporal diversity is created. Therefore, adopting jointly cooperative communications and HARQ mechanisms could be beneficial for designing cross-layer schemes.

First, a new rate maximization strategy, under heterogeneous data rate constraints among users is proposed. We propose an algorithm that allocates the optimal power at the base station (BS) and relays, assigns subcarriers and selects relays. The achievable data rate is investigated as well as the average starvation rate in the network when the load, i.e. the number of active users in the network, is increasing. It showed a significant gain in terms of global capacity compared to literature.

Second, in block fading channel, theoretical analyses of the packet error rate, delay and throughput efficiency in relay-assisted HARQ networks are provided. In slow fading channels, the average delay of HARQ mechanisms w.r.t. the fading states is not relevant due to the non-ergodic process of the fading channel. The delay outage is hence invoked to deal with the slow fading channel and is defined as the probability that the average delay w.r.t. AWGN channel exceeds a predefined threshold. This criterion has never been studied in literature, although being of importance for delay sensitive applications in slow fading channels. Then, an analytical form of the delay outage probability is proposed which might be useful to avoid lengthy simulations. These analyses consider a finite packet length and a given modulation and coding scheme (MCS) which leads to study the performance of practical systems.

Third, a theoretical analysis of the energy efficiency (bits/joule) in relay-assisted HARQ networks is provided. Based on this analysis, an energy minimization problem in multiuser relay-assisted downlink cellular networks is investigated. Each user has an average delay constraint to be satisfied such that a total power constraint in the system is respected. The BS is assumed to have only knowledge about the average channel statistics but no instantaneous channel state information (CSI). Finally, an algorithm that jointly allocates the optimal power at BS, the relay stations and selects the optimal relay in order to satisfy the delay constrains of users is proposed. The simulations show the improvement in terms of energy consumption of relay-assisted techniques compared to non-aided transmission in delay-constrained systems. Hence, the work proposed in this thesis can give useful insights for engineering rules in the design of the next generation energy-efficient cellular systems.

N° d'ordre : **13ISAR 30 / D13 - 30**

Institut National des Sciences Appliquées de Rennes
20, Avenue des Buttes de Coëmes • CS 70839 • F-35708 Rennes Cedex
Tel : 02 23 23 82 00 - Fax : 02 23 23 83 96