UNIVERSITAT POLITÈCNICA
DE CATALUNYA

Ph.D. Thesis

Interference in Wireless Networks

Cancelation, Impact, Practical Management, and Complexity

Author: **Eduard Calvo Page**

Advisors: **Dr. Javier Rodríguez Fonollosa**

Dr. Josep Vidal Manzano

SPCOM Group
Department of Signal Theory and Communications
Universitat Politècnica de Catalunya

Barcelona, December 2008

Abstract

The layered organization of traditional wired network design has continuously regarded communication links as bit pipes delivering data at some fixed rate with a certain error probability. While this modeling of the underlying physical layer may result appropriate for wired networks, it is certainly naive in the wireless domain. Unlike the fixed wired network, where the channel is time-invariant, the propagation physics of wireless channels and potential user mobility render the wireless network conditions very dynamic and time varying. Much worse, the pipe model disregards multiuser interference: by giving shared access to the same limited pool of resources to many users, the transmission rates of the communication links get coupled and the decomposition of the network into a set of independent single-user links turns out to be meaningless.

While the role of multiuser interference is widely recognized in prospective systems design, its impact on network performance is diverse. In this respect, the aim of the present Ph.D. thesis is to adopt a broad approach in the study of multiuser interference in wireless networks, recognizing it as a phenomenon with many different facets out of which we concentrate on four of them: cancelation, impact, practical management, and complexity.

We start studying when and how to perform partial interference cancelation, a technique that requires full statistical knowledge of the interfering signals at the receivers. We find that coding and decoding complexity can be traded whenever interference is under the control of the same source. In other words, the need for partial interference cancelation at the receivers can be alleviated through the use of appropriate coding techniques exploiting signal correlation at the transmitters. Additionally, we propose a transmission strategy based on superposition coding and aided decoding that yields an achievable region at least as large as the best long-standing region for the interference channel.

Useful as it is, interference cancelation becomes infeasible in applications backed by decentralized wireless networks with uncoordinated nodes. That leaves each sender-destination pair armed only with point-to-point (single user) strategies. This motivates the study of the totally asynchronous interference channel with single-user receivers. Having a capacity region rather involved, the evaluation of achievable rates is tackled based on simpler single-letter inner and outer bounds. The study of these bounds reveals that the impact of interference on the achievable rates can be mitigated through statistical signal design. Besides, the performance losses associated to the lack of transmission synchronism and the use of single-user decoders in the low- and high-power and low- and high-interference regimes are also quantified.

Next, the focus is on how to manage interference in a practical scenario where the receivers are again interference unaware but now frame-synchronous. A practical transmission scheme that allows for the design of optimal allocation policies of the limited transmission resources of the network is proposed. Giving special attention to a cellular configuration under practical conditions, efficient allocation schemes achieving Pareto and sequential optimality, respectively, are proposed and compared. The emphasis at this point is on the performance-complexity and throughput-fairness tradeoffs.

While recognizing that multiuser interference and the availability of receiver information modifies the fundamental limits and the practical figures of merit of wireless networks, the thesis concludes by studying a related aspect: how multiuser interference impacts on the complexity required for the evaluation of the previous quantities. Efficient methods for the evaluation of the capacity region of multiuser channels are proposed and, unlike the single-user case, non-convexities in optimization problems need to be unavoidably faced.

Resumen

La estructura de capas que rige el diseño de red ha modelado tradicionalmente los enlaces de comunicación como tuberías que transportan datos a una tasa fija con una cierta probabilidad de error. Si bien este modelo implícito de capa física puede resultar apropiado para redes cableadas, ciertamente resulta demasiado simple en el dominio inalámbrico. El modelo de tuberías ignora la interferencia multiusuario: al permitir el acceso compartido al mismo conjunto limitado de recursos por parte de varios usuarios, las tasas de transmisión de los diferentes enlaces de comunicación se acoplan y la red ya no puede descomponerse en un conjunto independiente de enlaces punto a punto.

Mientras que el papel de la interferencia multiusuario es ampliamente reconocido en el diseño de sistemas futuros, su impacto en las prestaciones de red es diverso. Es por tanto la intención de la presente tesis doctoral el adoptar un amplio enfoque en el estudio de la interferencia multiusuario, reconociéndola como un fenómeno con múltiples facetas, de las cuales se tratan cuatro de ellas: cancelación, impacto, gestión práctica y complejidad.

Empezamos estudiando el cuándo y el cómo de la cancelación parcial de interferencia, una técnica que requiere conocimiento estadístico completo de las señales interferentes en los receptores. Se ha demostrado que se pueden intercambiar la complejidad de codificación y decodificación cuando la interferencia está bajo el control de la misma fuente. En otras palabras, la necesidad de cancelar interferencia en los receptores se puede relajar gracias al uso de técnicas de codificación que aprovechen la correlación entre señales en los transmisores. Adicionalmente, se ha propuesto una estrategia de transmisión basada en codificación superpuesta y decodificación ayudada que ofrece una región alcanzable al menos tan grande como la mejor región conocida para el canal de interferencia.

Siendo útil, la cancelación de interferencia es impracticable en aplicaciones respaldadas por redes inalámbricas descentralizadas con nodos no coordinados. Así pues, cada par fuente-destino se ve relegado a usar únicamente estrategias punto a punto (monousuario). Este hecho motiva el estudio del canal de interferencia totalmente asíncrono con receptores monousuario. Al tener una región de capacidad compleja, la evaluación de las tasas alcanzables se realiza apoyándose en cotas interiores y exteriores más simples. El estudio de estas cotas revela que el impacto de la interferencia se puede mitigar a través del diseño estadístico de las señales transmitidas.

Además, se han cuantificado las pérdidas de prestaciones asociadas a la pérdida de sincronismo y el uso de receptores monousuario tanto en condiciones de baja y alta potencia, como de baja y alta interferencia.

A continuación se presta atención a cómo gestionar la interferencia en un escenario práctico donde los receptores de nuevo ignoran la presencia de interferencia, pero ahora mantienen sincronismo de trama. Dando especial atención a una configuración de red celular, se han propuesto y comparado esquemas eficientes de asignación de recursos capaces de alcanzar optimalidad de Pareto y secuencial. El énfasis en el análisis ha residido en los compromisos prestaciones-complejidad y throughput-igualdad.

Reconociendo que es la interferencia multiusuario y la información que sobre ella se tiene en los receptores lo que condiciona tanto los límites fundamentales como las figuras de mérito prácticas en las redes inalámbricas, la tesis concluye con el estudio de un aspecto relacionado: el impacto de la interferencia en la complejidad requerida para evaluar estas cantidades. Se han propuesto métodos eficientes para la evaluación de la región de capacidad de canales multiusuario y, al contrario que en el caso multiusuario, los problemas de optimización involucrados presentan no convexidades inevitables.

A la Rosa Maria,
A mis padres,

Agradecimientos

La realización de un doctorado no es un empleo de nueve a siete. Es quedarse con el compañero a apurar el deadline hasta el alba. También es marcharse a casa a las cinco con la cabeza a punto de estallar de tanta ecuación. Es resolver problemas en la ducha. Pero también es darse cuenta de que la demostración tenía un fallo, y, a veces, poder resolverlo. No es un más de lo mismo. Es un incesable proceso de crecimiento personal que culmina con la alumbración de una tesis, el fruto de innumerables anhelos, esperanzas, sufrimientos e ilusiones. En el doctorado, la línea que separa trabajo y diversión se vuelve difusa y delgada y, sin pasión, de nada sirven los conocimientos. Querer es poder, como dice el refrán.

Tras cuatro años de intenso trabajo, llega el momento de poner fin a una etapa que me ha permitido vivir muchas experiencias. Escribir estas líneas está sin duda entre las más gratificantes. Porque implica que se ha conseguido el objetivo deseado y porque apetece mucho ser agradecido y justo con todas aquellas personas que han contribuído a hacer de este periodo una época a recordar con cariño.

Quiero empezar agradeciendo la labor de mis directores de tesis, Javier R. Fonollosa y Josep Vidal. Tener una dirección bicéfala ha sido una experiencia muy enriquecedora, tanto a nivel profesional como personal. Tengo que darles las gracias por muchas cosas, pero en aras de la brevedad y de la contención de la exaltación de la amistad, resaltaré su apoyo férreo a mi trabajo, tanto en los buenos momentos como en los menos buenos. Y quién diga que de los últimos no hay en una tesis, miente.

I am very grateful to Milica Stojanovic from the Massachusetts Institute of Technology/Northeastern University for giving me another opportunity to establish fruitful collaboration with her, meet each other, and enjoy the New England way of living. Zoran Zvonar is the perfect gentleman: I want to be like him when I grow up.

Por enseñarme lo que es la pasión por el trabajo y predicar con el ejemplo, gracias a Gregori Vázquez y a Daniel P. Palomar.

Por permitirme disfrutar enseñando, gracias a mis alumnos y compañeros de Epsilon.

Mis companẽros del D5 se merecen un templo. Por las discusiones técnicas, los cafés de maitines, los debates de caña y bravas, las noches en las que hemos quemado la ciudad y, en esencia, por crear un ambiente de trabajo productivo, alegre y sano. No entiendo cómo no se han dado cuenta todavía que aguantar mis chistes malos a cambio de aprender tanto de ellos es un muy mal trato. Todos merecen mis agradecimientos más sinceros, aunque el azar, la actividad profesional o la afinidad personal ha hecho que con algunos de ellos la relación haya desbordado el ámbito de lo estrictamente laboral. Así pues, me quedo con la decisión de Andreu Urruela, el amor por lo bien hecho de José Antonio López, la dedicación de Alejandro Ramírez, las ocurrencias de Xavi Artigas, la destreza argumental de Julio Rolón, la sensibilidad de Luis García y el carácter de Pau Closas. Os admiro y os cuento entre mis Amigos.

Els meus Amics JoanMa Izquierdo i Jordi Cerdà m'han acompanyat des de l'adolescència i també mereixen un sincer agraïment per seguir sempre al peu del canó.

Me siento muy afortunado de tener la familia que tengo. Por ser mi club de fans pero también mis primeros críticos. Y eso no se paga con dinero. Su apoyo y cariño incondicional han resultado fundamentales desde siempre. A mis padres Selesio y Carmen les doy mil gracias por haber hecho todo lo posible para que tuviera la mejor educación y pudiera estudiar: aunque en esta carrera la meta la voy a cruzar yo, vosotros me habéis llevado en limusina hasta la línea de salida. Estoy muy orgulloso de vosotros. Y de mi hermana Sonia y mi cuñado Santi (el mejor padrino de boda que conozco). I de la meva nova família per part de la Rosa Maria, que també han seguit amb il·lusió l'evolució d'aquesta tesi. Y aunque él no lo sepa todavía, mi sobrino Héctor me ha hecho muy feliz. Héctor, el tío ya no ha tenido argumentos para responder a la mama si no había dicho que acabaría la tesis cuando ella fuera madre...

Diuen que de res serveix pujar una escala molt alta si quan arribes dalt de tot t'adónes que està recolzada en la paret equivocada. Alta o baixa, què importa, la meva escala es recolza en tu, Rosa Maria. Tu fas que jo vulgui ser millor persona. Les bones persones no necessàriament fan bones tesis. Però si les fan amb amor i et tenen a tu al costat, són felices. I això és del que anava la pel·lícula, oi?

Eso es to, eso es to, eso es todo amigos.

Eduard Calvo Page
Diciembre de 2008

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia a través del programa de Formación de Profesorado Universitario (F.P.U.) mediante la beca AP-2004-3549.

Contents

List of Figures

List of Tables

Notation

Throughout this dissertation, boldface lower-case letters shall denote column vectors, with $\mathbf{0}_n$ and $\mathbf{1}_n$ standing for the all-zero and all-one column vectors of length n, respectively. We shall denote by x_i the i-th entry of vector $\mathbf{x}$, and use $\geq$ and $\leq$ for scalar and component-wise inequalities indistinctly. Similarly, $\mathbf{z} = \min\{\mathbf{x}, \mathbf{y}\}$ denotes the vertical stacking of the component-wise minimum of two vectors. Boldface upper-case letters are used for matrices, with $\mathbf{I}_n$ standing for the $n \times n$ identity matrix and $A_{i,j}$ denoting the entry of the i-th row and j-th column of matrix $\mathbf{A}$, whose transpose and Hermitian are $\mathbf{A}^T$ and $\mathbf{A}^\dagger$, respectively. The i-th ordered eigenvalue (singular value) of an square (arbitrary) matrix $\mathbf{A}$ is denoted by $\lambda_i(\mathbf{A})$, where $\lambda_i(\mathbf{A}) \geq \lambda_{i+1}(\mathbf{A})$. Whenever needed, the superscript $(\cdot)^\star$ shall denote the optimal value of a variable.

As for random variables, we shall denote by X_k^n an n-dimensional random vector taking on the value x_k^n over the finite set $\mathcal{X}_k^n$ with probability $P_{X_k^n}(x_k^n)$. The i-th component of X_k^n is denoted by $X_{k,i}$, whereas $X_k^i = [X_{k,1} \dots X_{k,i}]$ and $X_{k,i}^n = [X_{k,i} \dots X_{k,n}]$. If the context is appropriate (no ambiguity is possible), we shall use the equivalence $[a, b] \equiv \{a, a+1, \dots, b\}$, for arbitrary integers a, b such that $b > a$. Other specific notation is introduced as follows:

$(x)^+$	The projection of the real variable x onto the non-negative semiaxis, i.e., $(x)^+ = \max\{x, 0\}$.
$\propto$	Proportional to.
$\delta[x]$	Kronecker delta. That is, $\delta[x] = \begin{cases} 1, & x = 0 \\ 0, & \text{otherwise.} \end{cases}$
$\mathrm{Co}\{\mathcal{S}\}$	Convex hull of the set $\mathcal{S}$.
$\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$	Multivariate Normal (Gaussian) distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$.
$\mathbf{A} \odot \mathbf{B}$	Hadamard (i.e. element-wise) product of matrices. That is, $[\mathbf{A} \odot \mathbf{B}]_{i,j} = A_{i,j} B_{i,j}$.
$a \ll b$	a much smaller than b.
$a \gg b$	a much greater than b.

Acronyms

AB	Arimoto-Blahut
AEP	Asymptotic Equipartition Property
AWGN	Additive White Gaussian Noise
BA-MAC	Binary Adder MAC
BC	Broadcast Channel
BEC	Binary Erasure Channel
BS	Base Station
BSC	Binary Symmetric Channel
BS-MAC	Binary Switching MAC
CSI	Channel State Information
DC	Difference of Convex
dDMBC	Degraded DMBC
DL	Downlink
DMAC	Discrete memoryless MAC
DMBC	Discrete Memoryless BC
DMC	Discrete Memoryless Channel
DMIC	Discrete Memoryless IC
FDD	Frequency Division Duplexing
FFT	Fast Fourier Transform
FUSC	Full Usage of Sub-Channels
GIC	Gaussian IC
H&K	Han and Kobayashi
IC	Interference Channel
IFFT	Inverse FFT

KKT	Karush-Kuhn-Tucker
LOS	Line Of Sight
MAC	Multiple Access Channel
MIMO	Multiple-Input Multiple-Output
MS	Mobile Station
nBS-MAC	Noisy BS-MAC
NLOS	Non-Lone Of Sight
NU	Network Utility
OFDMA	Orthogonal Frequency Division Multiple Access
PUSC	Partial Usage of Sub-Channels
QoS	Quality Of Service
RC	Relay Channel
RS	Relay Station
SISO	Single-Input Single-Output
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
UL	Uplink

Chapter 1

Introduction

More than half a century ago, the birth of Information Theory settled the fundamental principles governing the feasibility of reliable communication in single-user channels. Subsuming elegantly the impact of both channel distortion and noise in the single figure of merit of channel capacity, this new research framework not only ended with the long-standing folk theorem that regarded transmission rate and accuracy as an unavoidable tradeoff. It also facilitated the advent of modern digital communications and digital data recording: now, reliability could be achieved at any rate below capacity. The rest is well known history.

In parallel, military interests triggered in the 1960s the creation of another apparently unrelated field of research: networking. However, what started with early studies on packet switching and followed with the beginning of Internet, could no longer be kept within the defense umbrella. The scalability, robustness to link failure, and business appeal of Internet turned it into a mass phenomenon, giving rise to new user needs and services that rendered indispensable what was unprecedented. From http to ftp, from file sharing communities to social networks, the eclosion of communication networks has changed every single aspect of our behavior, labor, and leisure time.

In studying network performance, the layered organization of traditional wired network design has continuously regarded communication links as bit pipes delivering data at some fixed rate with a certain error probability. While this modeling of the underlying physical layer may result appropriate for wired networks, it is certainly naive in the wireless domain. Unlike the fixed wired network, where the channel is time-invariant, the propagation physics of wireless channels and potential user mobility render wireless network conditions very dynamic and time varying. Much worse, the pipe model disregards *multiuser interference*: by giving shared access to the same limited pool of resources to many users (either information sources or sinks), the transmission rates of the communication links get coupled and the decomposition of the network into a set of independent single-user links turns out to be meaningless. It is at this point when Information Theory can come back to help.

Information Theory has not been unaware of the multiuser interference problem in multi-terminal scenarios, as the first studies of multiuser channels date back to as soon as 1961, only 13 years after its birth in 1948. Besides, interference was not the only inherent phenomenon of wireless networks that received attention from the information theoretic community: issues such as cooperation and feedback also played an important role. In an impressive synthetic effort, every possible degree of freedom of physical layer interaction between the terminals of a wireless network were condensed in five multiuser channel models: the two-way channel, the broadcast channel, the multiple access channel, the relay channel, and the interference channel. Any network under study can therefore be decomposed into some combination of these building blocks.

While the study of just one of the previous canonical multiuser channel models could well fit the entire scope of a Ph.D. dissertation, it is the aim of the present one to adopt a higher level approach and study multiuser interference in wireless networks as a phenomenon with many facets by itself. Thus, from pure network information theory to the organization of real network scenarios, from optimization theory to statistical analysis, the focus of this dissertation is diverse in topics, yet it hopes to be specific in conclusions.

1.1 Motivation and Objectives

Although the role of multiuser interference is widely recognized in prospective systems design, its impact on network performance is diverse. To serve as an example, wireless sensor networks, wireless multihop networks, and mobile cellular systems, three key scenarios that shall enable most of upcoming communication services, exhibit different characteristics that may cause the eventual transmission strategies of the network terminals to be significantly different. In this respect, we identify the information that each receiver has about the structure of the interference to be the principal bottleneck constraining both the performance and sophistication of the potential interference mitigation techniques that can be carried out. This has motivated the study of four different research lines, each one represented by a different word in the subtitle of the present dissertation. Namely,

- *Cancelation.* When the receivers of a network have complete knowledge of the statistical structure of the interference,
  - Is it beneficial to allow all the receivers of the network to perform partial interference cancelation simultaneously?
  - If so, what is the most advantageous partial interference cancelation scheme?
- *Impact.* If no knowledge at all is available at the receivers, what is the performance loss experimented? Can proper signal design mitigate the performance degradation?

- *Practical Management.* If no knowledge at all is available at the receivers, but some sort of centralized coordination is possible, is there a way of getting rid of the performance losses? If so, how can we get the most out of the available transmission resources?

- *Complexity.* In evaluating the fundamental limits of wireless networks, what is the complexity increase due to multiuser interference?

An individual chapter is dedicated to each of the above research lines and a brief summary of these chapters is presented next.

1.2 Thesis Outline

Chapter 2

This chapter focuses first on the benefits of allowing all the receivers of a network to perform partial interference cancelation simultaneously. While it is clear that such feature can never lead to performance losses, we find out that sometimes it does neither produce rate gains as compared to a situation where cancelation is performed alternatively at the receivers. Intuitively, when all the interference is under the control of the same source, correlated coding can alleviate all the users from performing partial interference cancelation simultaneously. This work has led to a layered generalization of the random binning technique devised by Slepian and Wolf in the context of coding of correlated sources for multiuser coding.

Besides, borrowing from the work of Marton for the broadcast channel, a novel transmission strategy for the interference channel based on superposition coding and aided decoding is found to yield an achievable region at least as large as the best known achievability result for this channel.

Chapter 3

While Chapter 2 implicitly assumed perfect knowledge of the codebooks of the interferent users, Chapter 3 focuses on the opposite situation of interference unaware receivers. That is often the case in decentralized wireless networks with uncoordinated nodes, and motivates the study of the totally asynchronous interference channel with single-user receivers. The capacity region of this channel is characterized within an Information Spectrum approach, although more amenable single-letter inner and outer bounds are also provided. As an interesting result, Gaussian-distributed codes are found not to be optimal for the Gaussian case, as other practical codes are shown to outperform them. An analytical characterization of the conditions for the existence of other input statistics superior to Gaussian reveals that, essentially, the channel needs to be interference-limited.

Chapter 4

As a result of the performance losses accounted for the lack of interference information in Chapter 3, the motivation of Chapter 4 is i) to mitigate them in a practical scenario where the receivers are also interference unaware but frame-synchronous, and ii) to propose a practical transmission scheme that allows for the design of optimal allocation policies of the limited transmission resources of the network. Special attention is given to a cellular configuration, and, under practical conditions, efficient allocation schemes achieving Pareto and sequential optimality, respectively, are proposed and compared. The emphasis in this Chapter is on the performance-complexity and throughput-fairness tradeoffs.

Chapter 5

While the focus of Chapters 2, 3, and 4 is on how multiuser interference and the availability of receiver information modifies the fundamental limits (achievable rates in Chapter 2 and capacity regions in 3) or the practical figures of merit (user throughput, Chapter 4) of wireless networks, a related and so far unexplored aspect is how the same issues impact on the evaluation of these quantities. Thus, Chapter 5 is devoted to propose efficient evaluation methods of capacity regions of multiuser channels. It will be shown that, unlike the single-user case, multiuser interference brings about non-negligible complexity in the computation of the performance limits. Particular attention is payed to the multiple access channel and the degraded broadcast channel.

1.3 Research Contributions

The work conducted within the present thesis resulted in the publication of several contributions in technical journals and international conferences. The details of the research contributions in each chapter are as follows.

Chapter 2

The achievability result for the interference channel based on superposition coding and aided decoding, lead to the following contribution:

- E. Calvo, Javier R. Fonollosa, and J. Vidal, “A simple achievable rate region for the interference channel”, *unpublished.*

Unfortunately, by the time we were to submit this journal paper (March 2006) we found that, in an independent work, another set of authors had already published the same exact result one month before [Cho06a].

Chapter 3

The main results of this chapter are currently under review in one journal paper:

- [Cal08a] E. Calvo, J. R. Fonollosa, and J. Vidal, "The totally asynchronous interference channel with single-user receivers", *submitted to IEEE Transactions on Information Theory*, September 2008.

Chapter 4

One journal paper currently under review and one conference paper summarize the research contributions of the chapter:

- [Cal08e] E. Calvo, J. Vidal, and J. R. Fonollosa, "Optimal resource allocation in relay-assisted cellular networks with partial CSI", *submitted to IEEE Transactions on Signal Processing*, May 2008.

- [Cal07e] E. Calvo, J. Vidal, and J. R. Fonollosa, "Resource allocation in multihop OFDMA broadcast networks", *Proc. IEEE Workshop on Signal Process. Advances for Wireless Commun. (SPAWC)*, Helsinki, Finland, June 2007.

Chapter 5

Regarding the evaluation of capacity regions of multiuser channels, the research contributions of this chapter are in the form of one journal paper currently under review and two conference papers:

- [Cal07d] E. Calvo, D. P. Palomar, J. R. Fonollosa, and J. Vidal, "On the computation of the capacity region of the discrete MAC", *submitted to IEEE Transactions on Information Theory*, May 2007.

- [Cal07c] E. Calvo, D. P. Palomar, J. R. Fonollosa, and J. Vidal, "The computation of the capacity region of the discrete MAC is a rank-one non-convex optimization problem", *Proc. Intl. Symposium on Information Theory (ISIT)*, pp. 2396-2400, Nice, France, June 2007.

- [Cal08c] E. Calvo, D. P. Palomar, J. R. Fonollosa, and J. Vidal, "The computation of the capacity region of the discrete degraded BC is a non-convex DC problem", *Proc. Intl. Symposium on Information Theory (ISIT)*, pp. 1721-1725, Toronto, Canada, July 2008.

Other contributions not directly related with this dissertation

Apart from the topics covered in this dissertation, other research areas have been addressed during the period of Ph.D. studies. Some of these topics were related with research projects for private industry and public administrations, and the most relevant publications are listed below.

⋆ **Research in multiuser detection for underwater communications:**

- [Cal08d] E. Calvo and M. Stojanovic, "Efficient channel estimation-based multi-user detection for underwater CDMA systems", *accepted for publication in IEEE Journal of Oceanic Engineering*, 2008.
- [Cal05] E. Calvo and M. Stojanovic, "A coordinate descent algorithm for multichannel multiuser detection in underwater acoustic DS-CDMA systems", *Proc. IEEE OCEANS Europe Conference*, Brest, France, June 2005.

⋆ **Research in resource allocation with perfect CSI:**

- [Cal07b] E. Calvo, J. R. Fonollosa, and J. Vidal, "Near-optimal joint power and rate allocation for OFDMA broadcast channels", *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, HI, April 2007.
- [Cal07a]E. Calvo and J. R. Fonollosa, "Efficient resource allocation for orthogonal transmission in broadcast channels", *Proc. IEEE Workshop on Signal Process. Advances for Wireless Commun. (SPAWC)*, Helsinki, Finland, June 2007.

⋆ **Additional research in resource allocation with imperfect CSI:**

- [Muñ07] O. Muñoz, J. Vidal, A. Agustín, E. Calvo, and A. Alcón, "Resource management for relaying-enhanced WiMAX: OFDM and OFDMA", *Workshop Trends in Radio Resource Management*, Barcelona, November 2007.

⋆ **Research in wireless sensor networks:**

- [Clo07] P. Closas, E. Calvo, J. Fernández, and A. Pagès, "Coupling noise effect in self-synchronizing wireless sensor networks", *Proc. IEEE Workshop on Signal Process. Advances for Wireless Commun. (SPAWC)*, Helsinki, Finland, June 2007.

⋆ **Research in 4G systems**

- [Cal08b] E. Calvo, I. Kovács, L. García, and J. R. Fonollosa, "A reconfigurable downlink air interface: design, simulation methodology, and performance evaluation", *Proc. ICT-Mobile Summit*, Stockholm, Sweden, June 2008.

Chapter 2

Partial Interference Cancelation: When and How

Wireless networks are made of several information sources and sinks that share the same transmission resources in their pursuit of reliable communication of multiple information flows. The fact that the transmission resources are always limited in practice together with the potential conflicts of interest that may arise between neighboring links couples individual performance and complicates analysis. A rigorous way of approaching analysis is through the study of multiuser channels borrowed from Network Information Theory [Cov06, Ch. 14]: the broadcast channel (BC) [Cov72, Ber73, Gal74, Cov75, Mar79, Cov81a, Cov98, Wei06], the multiple-access channel (MAC) [Ahl71, Lia72], the relay channel (RC) [Cov79], and the interference channel (IC) [Car75, Car78, Han81, Car83].

Simple as they are, many problems regarding the characterization of their fundamental communication limits (capacity regions) are still open and aging. We believe that the main difficulty driver in their analysis is *multiuser interference*. In non-degenerated channels we cannot assume that the transmissions of the sources take place over orthogonal channels: the channel output of each receiver depends on the codewords transmitted by all the senders. This fact creates interdependencies between achievable rates and renders impossible the decomposition of a network into a set of independent single-user links.

Given that the presence of multiuser interference is often unavoidable, the next question is what to do with it. The right answer depends on the degree of knowledge that each receiver has about it. If completely unaware about interference, each sender-destination pair shall use single-user codes and expect performance losses (see Chapter 3 for quantification and minimization of these losses). Oppositely, if the identities and the codebooks of the interferers are perfectly known at each destination, the solution is to perform partial interference cancelation. This latter situation is the focus of this chapter.

2.1 Introduction

In a strict sense, the only two channels that suffer from multiuser interference are the BC and the IC. While in the MAC all the signals received at the destination have to be reliably decoded and, hence, cannot be regarded as interference, in the RC there is no interference at all since the role of the relay is to enhance the communication of the principal sender-destination pair. Yet both the BC and the IC can be viewed as the two sides of the same tapestry: the BC can be interpreted as an IC where all the senders are physically located inside the same transmitter, or an IC where all senders can cooperate.

The study of the IC has direct applications in wireless network models where communication between different sender-receiver pairs can take place simultaneously. Some examples are wireless sensor networks, wireless multi-hop networks, and mobile cellular systems with small bandwidth reuse factor suffering from large inter-cell interference. On the other hand, the BC models the situation in which one sender wishes to transmit information to a number of receivers (the information for each receiver can be different or not). This description perfectly matches the downlink of a mobile cellular system, with the base station acting as the only sender of the scenario. In this chapter, we shall concentrate on these two channels with two clear objectives in mind regarding the potential of partial interference cancelation:

- The analysis of *when* it is useful from an achievable rate standpoint.
- The analysis of *how* to perform it efficiently.

We address the first item in Section 2.2, the second in Section 2.3, and conclude this chapter in Section 2.4.

2.1.1 When

We start focusing on the BC. The capacity region of the discrete memoryless BC (DMBC) is known when the channel is degraded, that is, when the different channel outputs form a Markov chain in some specific order. In this case, Bergmans [Ber73] provided an achievable rate region that was later found to be the capacity region thanks to the converse theorems of Wyner [Wyn73] (for the specific case of binary symmetric channels) and Gallager [Gal74] (general case). For general DMBCs, the first achievable rate region was obtained in Cover's seminal paper [Cov72], and was later extended by Van der Meulen [Meu75] and Cover [Cov75]. They considered the general situation in which a common message may be sent to all the receivers. Here, the approach will be that the information for each user is independent and there is no common message. For a review of the most important contributions in the context of the DMBC the reader is referred to [Meu77] and [Cov98].

The most advantageous approach so far to the DMBC without common information came along with the work of Marton [Mar79] (see El Gamal and Van der Meulen [Gam81] for a simpler proof based on typicality arguments). The most salient feature of Marton's contribution is that it allows for arbitrarily correlated auxiliary random variables, hence enlarging the region with respect to independent intermediate encoding. Marton's general achievable rate region for the 1×2 discrete memoryless broadcast channel is stated in Theorem 2.1.

Theorem 2.1 (Marton) *Consider the region $\mathcal{R}_{\mathsf{MT}}(P_{WU_1U_2X})$ consisting of the set of (R_1, R_2) rate pairs satisfying*

$$
\begin{aligned}
R_1 &\leq I(WU_1; Y_1) && (2.1)\\
R_2 &\leq I(WU_2; Y_2) && (2.2)\\
R_1 + R_2 &\leq \min\{I(W; Y_1), I(W; Y_2)\} + I(U_1; Y_1|W) + I(U_2; Y_2|W) - I(U_1; U_2|W) && (2.3)
\end{aligned}
$$

for some joint probability distribution $P_{WU_1U_2X}$ defined on $\mathcal{W} \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X}$. Denote by $\mathcal{P}$ the set of all such distributions. The region

$$
\mathcal{R}_{\mathsf{MT}} = \bigcup_{P_{WU_1U_2X} \in \mathcal{P}} \mathcal{R}_{\mathsf{MT}}(P_{WU_1U_2X}), \tag{2.4}
$$

is achievable for the 1×2 DMBC.

The region $\mathcal{R}_{\mathsf{MT}}$ is characterized by two features. Namely, i) the auxiliary random variables U_1 and U_2 bearing private information for each of the receivers respectively are correlated using the random binning technique borrowed from Slepian-Wolf coding of correlated sources [Sle73]; ii) the additional auxiliary random variable W bearing information for only one of the receivers is reliably decoded at the intended receiver and simultaneously benefits the decoding of the non-intended receiver. The non-intended receiver does not reliably decode W but only takes it into account to compute the typical sets for decoding its information-bearing signal. In this sense, the non-intended receiver partially mitigates the interference arising from the transmission of information to the other user by considering W. The full region is obtained time-sharing the situations where W is intended for the first or for the second receiver.

Whereas in Marton's setting the maximal points of the region are achieved when only one receiver performs partial interference cancelation on the signals intended for the other user, we study whether an increase on the achievable rates can be realized by allowing both receivers to cancel interference simultaneously. To that end, we consider in Section 2.2 an extension of Marton's coding/decoding schemes in which both receivers are allowed to simultaneously take into account information about the message of the undesired user. The extended coding scheme uses four random variables: W_1, W_2, U_1, U_2. As in Marton's setting, U_1 and U_2 bear private information to receiver 1 and receiver 2, respectively. The difference comes with the signal W_1, which is intended for receiver 1 but is taken into account in receiver 2; the role of W_2 is dual. The pair (U_1, U_2) is generated conditionally on (W_1, W_2) and both pairs ((W_1, W_2) and (U_1, U_2))

are correlated using a layered random binning scheme. Hence, the two-layered random binning coding scheme is suitable for all the arbitrary joint probability distributions $P_{W_1W_2U_1U_2X}$.

To illustrate whether or not this enhanced feature may lead to an increase on the achievable rates, we analyze the probability of error of the proposed coding and decoding schemes and derive a new achievable rate region, the two-layered random binning region. We prove in Section 2.2 that Marton's region and the two-layered random binning region are equal. That is, for the BC the achievable rates do not increase if only one user at a time or both of them perform partial interference cancelation of the undesired signal.

This result is indeed not valid for the IC [Car78], for which the best achievable regions [Han81, Cho06a, Cho06b] are obtained with coding/decoding strategies which allow for simultaneous partial interference cancelation. In general, these strategies can be shown to strictly outperform the regions obtained with non-simultaneous partial interference cancelation. The rationale behind this fact is that in the BC, since all the interference is originated by the same source, we can benefit from being able to correlate the signals intended for both users. In contrast, in the IC, there is room for rate increase as the codewords of the users are forced to be independent since no cooperation between sources is possible.

Section 2.2 is organized as follows: in Section 2.2.1 we address the necessary definitions that will be needed in Section 2.2.2, where the analysis of the two-layered random binning region is performed. The two-layered random binning region is shown to be equal to Marton's region in Section 2.2.3, hence showing that the generalization of Marton's scheme does not increase the achievable rates in the BC. Section 2.2.4, traces a parallelism with the IC, for which the result of Section 2.2.3 cannot be extrapolated (a counterexample is provided).

2.1.2 How

If the previous section discussed the convenience of allowing all the receivers to perform partial interference cancelation simultaneously, we now move on to study how to perform cancelation in a way that the achievable rates are the largest possible. While Section 2.2 unveils some desirable features of an interference cancelation strategy for the BC, it is their application to the IC that yields a real improvement.

The characterization of the capacity region of the general IC remains as an open problem, although it has been solved in several cases. Namely:

- When all the channel outputs are statistically equivalent [Car83].
- When the interference is very strong [Car75] or strong [Sat78, Han81, Sat81, Cos87].
- A class of discrete additive degraded interference channels [Ben79].
- A class of deterministic interference channels [Gam82].

Other references that obtained inner and outer bounds on the capacity region of the IC are [Car78, Han81, Car83] for the general case and [Car78, Han81, Car83, Cos87, Sas04, Kra04] for the Gaussian case.

Our work particularly draws on the previous research made by Carleial [Car78] and Han and Kobayashi (H&K) [Han81]. Both provided achievable rate regions for the discrete memoryless IC (DMIC) which were based on the application of the superposition coding technique envisioned by Cover for the broadcast channel [Cov72, Ber73]. While [Car78] considered *successive* superposition, [Han81] used *simultaneous* superposition. In particular, the most general achievable rate region derived in [Han81] is shown to include the achievable rate region of [Car78] and has remained as the best achievable rate region for the DMIC so far. Oppositely to what is discussed in [Han81], we do not think that the goodness of the region of [Han81] must be attributed to the superiority of simultaneous superposition coding over successive superposition coding but to the superiority of the decoding scheme that was used in [Han81] with respect to that of [Car78].

Following the terminology of [Cov72, Ber73, Meu75], Carleial's and H&K's regions are based on encapsulating "cloud centers" and additional high-resolution information ("satellites") inside the transmitters' codewords. The behavior of the decoding at each receiver is ruled by the information-bearing random variables of the desired transmitter and the cloud center of the interfering sender. While Carleial adopted a sequential decoding strategy, H&K applied joint decoding. Since the prior is a special case of the latter, H&K's decoding strategy results superior.

In Section 2.3, we extend H&K's decoding strategy by making explicit that each receiver is only interested in the information-bearing random variables of its desired user. For instance, receiver 1 may consider the ensemble of cloud centers of sender 2 to *aid* the decoding of its intended message: ambiguity on the actual message carried by the cloud center of sender 2 can be allowed as long as there is no ambiguity on its desired message. While H&K and Carleial forced receiver 1 to reliably decode the message born in the cloud center of sender 2, we just consider that cloud center as an aid for the decoding of the intended message. The concept of *aided* decoding, is also present in the work of Marton [Mar79] for the broadcast channel. As previously explained, in [Mar79, Thm. 2] an information-bearing random variable W is communicated from one sender to one receiver that reliably decodes it, while the non-intended receiver uses it just for aiding its decoding. In the context of the IC, the decoding of each receiver is *aided* by the cloud centers of the interfering sender.

We show in Section 2.3 that the gains associated to the use of aided decoding instead of joint decoding are in the form of fewer rate constraints arising from the expression of the probability of error. That leads to an achievable rate region at least as large as H&K's region. While potential rate gains could also be expected from this result, [Cho06b] showed that both regions are equal. This does not imply that simplification is the only plus we get from aided decoding: having less constraints, the resulting error exponents in the finite blocklength regime can be larger with aided decoding than with H&K's decoding strategy.

The coverage of this item in Section 2.3 is structured in the following manner. Section 2.3.1 addresses the definitions and some preliminaries regarding the 2×2 DMIC. The achievable rate regions with successive superposition and aided decoding are presented in Section 2.3.2, where it is compared with the regions by Carleial [Car78] and H&K [Han81], and Section 2.3.3 shows an example of a 2×2 DMIC for which the proposed achievable rate region outperforms H&K's region for a given choice of the input statistics.

2.1.3 Summary of contributions

This chapter focuses first on the convenience of allowing all the receivers of a network to perform partial interference cancelation simultaneously. The study of this issue has resulted in the following contributions:

- The coding/decoding strategy of Marton [Mar79] for the BC is extended to allow for a scenario of simultaneous interference cancelation by all the receivers, giving rise to the two-layered random binning region.

- The two layered random binning region and Marton's region are shown to be equal, thus concluding that simultaneous and alternate partial interference cancelation result in the same achievable rates in the BC.

- This fact has been compared to what happens in the IC, where simultaneous partial interference cancelation clearly outperforms alternate cancelation.

From the study of the coding scheme of Marton for the BC, some ideas have been extrapolated to the IC, where the target has been to extend the long-standing H&K's achievable region [Han81]. In this respect, the contributions are the following:

- A new decoding scheme, aided decoding, is introduced and shown to result in a achievable rate region which contains only a subset of the inequalities that described H&K's region.

- An example is provided where, for some fixed input statistics, the achievable rates with aided decoding strictly outperform those of H&K.

- Based upon [Cho06b], which shows that both H&K and the aided decoding region are equal, the advantages of the aided decoding region are precluded to simplicity (less rate constraints and less auxiliary random variables) and potentially better error exponents.

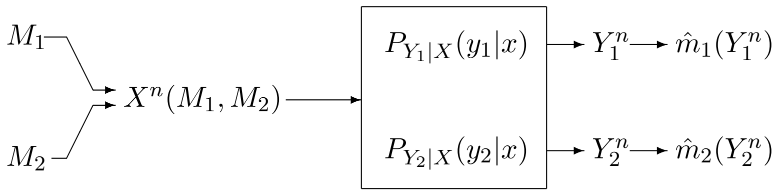

Figure 2.1: The 1×2 discrete memoryless broadcast channel without common information.

2.2 Simultaneous vs Alternate Partial Interference Cancelation in the BC

2.2.1 Definitions

A 1×2 DMBC without common information consists of one information source (sender) and 2 receivers. The sender wishes to communicate independent information to each of the receivers.

Definition 2.1 *A 1×2 discrete memoryless broadcast channel (DMBC) consists of one finite input alphabet set $\mathcal{X}$, two finite output alphabet sets $\mathcal{Y}_1$, $\mathcal{Y}_2$ and two marginal transition probability distributions $P_{Y_1|X}$ and $P_{Y_2|X}$ such that $\mathbb{P}\{y_k^n|x^n\} = \prod_{j=1}^n P_{Y_k|X}(y_{k,j}|x_j)$, $k = 1, 2$.*

Note that the DMBC is not defined in terms of its joint transition probability $P_{Y_1Y_2|X}$ since the capacity region depends only on the marginal distributions [Cov06, Thm. 14.6.1].

Definition 2.2 *A $\big((2^{nR_1}, 2^{nR_2}), n\big)$ code for the 1×2 DMBC without common information consists of two sets of integers $\big[1, 2^{nR_k}\big]$ ($k = 1, 2$) called the message sets, one encoding function,*

$$X^n \;:\; \big[1, 2^{nR_1}\big] \times \big[1, 2^{nR_2}\big] \longrightarrow \mathcal{X}^n \tag{2.5}$$

and 2 decoding functions

$$\hat{m}_k \;:\; \mathcal{Y}_k^n \longrightarrow \big[1, 2^{nR_k}\big] \;\;, \quad k = 1, 2. \tag{2.6}$$

The sender draws the indices (M_1, M_2) independently and uniformly from the message sets $\big[1, 2^{nR_k}\big]$ ($k = 1, 2$) and sends the corresponding codeword $X^n(M_1, M_2)$, of length n, over the channel. The k-th receiver uses its channel output, Y_k^n, to create an estimate of the message M_k, $\hat{m}_k(Y_k^n)$. This process is shown in Figure 2.1.

Since the message for each user is drawn independently, the average probability of error of a $\big((2^{nR_1}, 2^{nR_2}), n\big)$ code for the DMBC is defined as

$$P_e^{(n)} = 2^{-n(R_1+R_2)} \sum_{\substack{m_k \in [1, 2^{nR_k}] \\ k=1,2}} \mathbb{P}\left\{ \{\hat{m}_1(Y_1^n) \neq m_1\} \cup \{\hat{m}_2(Y_2^n) \neq m_2\} \middle| \, (m_1, m_2) \text{ sent} \right\}. \tag{2.7}$$

Definition 2.3 *A rate pair* (R_1, R_2) *is said to be achievable for the* 1×2 *DMBC if there exists a sequence of* $\big((2^{nR_1}, 2^{nR_2}), n\big)$ *codes with* $P_e^{(n)} \to 0$ *for arbitrarily large block length* n.

2.2.2 The two-layered random binning achievable rate region

Consider the following generalization of Marton's setting:

- Two auxiliary random variables W_1 and W_2 are correlated using the random binning technique.
- A pair of auxiliary random variables U_1 and U_2 are generated conditionally on W_1 and W_2 and are correlated also using the random binning technique.
- The signals (W_k, U_k) bear the information of M_k to the k-th receiver, $k = 1, 2$.
- The k-th receiver decodes (W_k, U_k) aided by W_j $(j \neq k)$ when computing the typical sets, $k = 1, 2$. That is, each user performs partial interference cancelation by considering the signal W_j.

The achievable rate region based on the above points is called the two-layered random binning region, and is stated in the following theorem.

Theorem 2.2 *Consider the region* $\mathcal{R}(P_{W_1W_2U_1U_2X})$ *consisting of the set of* (R_1, R_2) *rate pairs that can be expressed as* $R_1 = R_{W_1} + R_{U_1}$, $R_2 = R_{W_2} + R_{U_2}$ *satisfying*

$$R_{W_1} \le H(W_1) \quad ; \quad R_{U_1} \le \widetilde{R}_{U_1} \tag{2.8}$$
$$R_{W_2} \le H(W_2) \quad ; \quad R_{U_2} \le \widetilde{R}_{U_2} \tag{2.9}$$
$$R_{W_1} + R_{W_2} \le H(W_1W_2) \quad ; \quad R_{U_1} + R_{U_2} \le \widetilde{R}_{U_1} + \widetilde{R}_{U_2} - I(U_1; U_2|W_1W_2) \tag{2.10}$$
$$\widetilde{R}_{U_1} \le I(U_1; Y_1|W_1W_2) \quad ; \quad \widetilde{R}_{U_2} \le I(U_2; Y_2|W_1W_2) \tag{2.11}$$
$$R_{W_1} + \widetilde{R}_{U_1} \le I(W_1U_1; Y_1|W_2) \quad ; \quad R_{W_2} + \widetilde{R}_{U_2} \le I(W_2U_2; Y_2|W_1) \tag{2.12}$$
$$R_{W_2} + \widetilde{R}_{U_1} \le I(W_2U_1; Y_1|W_1) \quad ; \quad R_{W_1} + \widetilde{R}_{U_2} \le I(W_1U_2; Y_2|W_2) \tag{2.13}$$
$$R_{W_1} + R_{W_2} + \widetilde{R}_{U_1} \le I(W_1W_2U_1; Y_1) \quad ; \quad R_{W_1} + R_{W_2} + \widetilde{R}_{U_2} \le I(W_1W_2U_2; Y_2). \tag{2.14}$$

for some joint probability distribution $P_{W_1W_2U_1U_2X}$ *defined on* $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X}$ *and non-negative reals* $\widetilde{R}_{U_1}$, $\widetilde{R}_{U_2}$. *Denote by* $\mathcal{P}'$ *the set of all such distributions. The region*

$$\mathcal{R} = \mathrm{Co}\left(\bigcup_{P_{W_1W_2U_1U_2X} \in \mathcal{P}'} \mathcal{R}(P_{W_1W_2U_1U_2X}) \right), \tag{2.15}$$

called the two-layered random binning region, is achievable for the 1×2 *DMBC.*

Proof. See Appendix 2.A. □

2.2.3 Equality of Marton's region and the two-layered random binning region

For the purpose of comparing Marton's achievable rate region to the achievable rate region with simultaneous partial interference cancelation, the two-layered random binning region, we consider the following alternative rephrasing of $\mathcal{R}_{\mathsf{MT}}$.

Lemma 2.1 *Consider the region* $\mathcal{R}_{\mathsf{MT}}(P_{W_1W_2U_1U_2X})$ *consisting of the set of* (R_1, R_2) *rate pairs satisfying*

$$\begin{aligned} R_1 &\leq I(W_1W_2U_1;Y_1) &(2.16)\\ R_2 &\leq I(W_1W_2U_2;Y_2) &(2.17)\\ R_1+R_2 &\leq \min\{I(W_1W_2;Y_1), I(W_1W_2;Y_2)\} &(2.18)\\ &+ I(U_1;Y_1|W_1W_2)+I(U_2;Y_2|W_1W_2)-I(U_1;U_2|W_1W_2) &(2.19)\end{aligned}$$

for some joint probability distribution $P_{W_1W_2U_1U_2X}(w_1,w_2,u_1,u_2,x) \in \mathcal{P}'$. *Then, the following holds,*

$$\mathcal{R}_{\mathsf{MT}} = \bigcup_{P_{W_1W_2U_1U_2X}\in\mathcal{P}'} \mathcal{R}_{\mathsf{MT}}(P_{W_1W_2U_1U_2X}). \tag{2.20}$$

Proof. Note that

$$\mathcal{R}_{\mathsf{MT}} = \bigcup_{P_{WU_1U_2X}\in\mathcal{P}} \mathcal{R}_{\mathsf{MT}}(P_{WU_1U_2X}) \subseteq \bigcup_{P_{W_1W_2U_1U_2X}\in\mathcal{P}'} \mathcal{R}_{\mathsf{MT}}(P_{W_1W_2U_1U_2X}) \tag{2.21}$$

since we can always set $(W_1,W_2)=(W,\phi)$ or $(W_1,W_2)=(\phi,W)$ in $\mathcal{R}_{\mathsf{MT}}(P_{W_1W_2U_1U_2X})$, where ϕ is a constant. On the other hand, consider the Markov chain $(W_1,W_2)\to W\to(U_1,U_2)$ obtained when $W=f(W_1,W_2)$, a deterministic function of (W_1,W_2). Then, by the data processing theorem and the fact that $W=f(W_1,W_2)$ we obtain

$$\begin{aligned} I(WU_k;Y_k) &\geq I(W_1W_2U_k;Y_k), \qquad k=1,2 &(2.22)\\ I(W;Y_k) &\geq I(W_1W_2;Y_k), \qquad k=1,2 &(2.23)\\ I(U_k;Y_k|W)=I(U_k;Y_k|f(W_1,W_2)) &= I(U_k;Y_k|W_1W_2), \qquad k=1,2 &(2.24)\\ I(U_1;U_2|W)=I(U_1;U_2|f(W_1,W_2)) &= I(U_1;U_2|W_1W_2). &(2.25)\end{aligned}$$

The mutual information terms of the most right hand side terms of the above inequalities depend on $P_{W_1W_2U_1U_2X}$. Since we can cover all the distributions P_W by passing the distribution $P_{W_1W_2}$ through a deterministic function, the following holds

$$\mathcal{R}_{\mathsf{MT}} = \bigcup_{P_{WU_1U_2X}\in\mathcal{P}} \mathcal{R}_{\mathsf{MT}}(P_{WU_1U_2X}) \supseteq \bigcup_{P_{W_1W_2U_1U_2X}\in\mathcal{P}'} \mathcal{R}_{\mathsf{MT}}(P_{W_1W_2U_1U_2X}), \tag{2.26}$$

which concludes the proof. □

The two-layered random binning region $\mathcal{R}$ of Section 2.2.2 is the natural extension of Marton's achievable rate region to the case in which both users perform simultaneous partial interference cancelation on the signal intended for the other user. On the other hand, Lemma 2.1 allows us to compare it with Marton's region expressed in terms of the same ensemble of auxiliary random variables. The main result of this Section is stated in the following Theorem.

Theorem 2.3 *The achievable rates are not enlarged when both users perform simultaneous partial interference cancelation with respect to time-sharing between only one user canceling interference, i.e.,*

$$\mathcal{R} = \mathcal{R}_{\mathsf{MT}}. \tag{2.27}$$

Marton's achievable rate region and the two-layered random binning region are equal.

Proof. See Appendix 2.B. □

2.2.4 Comparison with the interference channel

For the general discrete memoryless IC, the best achievable rate region is that of Han and Kobayashi [Han81], which is based on embedding the message M_k into the codeword X_k ($k = 1, 2$) in two steps: first, some partial information about M_k is born by an auxiliary random variable; second, the auxiliary random variable together with the rest of information is embedded into X_k. While the auxiliary random variable of each user helps the decoding of the interfered user when computing the typical sets (like the signals W_1, W_2 of region $\mathcal{R}$), the information not embedded in that variable (analogous to the indexes j_1, j_2 of $\mathcal{R}$) remains private. The improvement on [Han81] discussed in Section 2.3 is that the partial information carried by the auxiliary random variable of the interfering transmitter (analogous to the indexes i_1, i_2 of $\mathcal{R}$) needs not to be reliably decoded at the interfered receiver.

Oppositely to what happened in the BC, the best achievable rates for the IC [Han81] are achieved by coding schemes that allow for simultaneous partial interference cancelation. These schemes can be shown to strictly outperform time-sharing among non-simultaneous interference cancelation setups, contrarily to what happens in the DMBC (Theorem 2.3), for which the achievable rates remain unchanged.

Consider a DMIC with very strong interference, that is, a DMIC for which the following holds

$$I(X_1; Y_1|X_2) \leq I(X_1; Y_2) \tag{2.28}$$

$$I(X_2; Y_2|X_1) \leq I(X_2; Y_1). \tag{2.29}$$

The capacity region of such a channel [Car75] is given by the convex hull of the set of (R_1, R_2) rate pairs satisfying

$$R_1 \leq I(X_1; Y_1|X_2) \tag{2.30}$$

$$R_2 \leq I(X_2; Y_2|X_1) \tag{2.31}$$

for some product probability distribution $P_{X_1}P_{X_2}$. The capacity region is achieved when both receivers perform successive decoding. This is equivalent to consider that both receivers aid their decoders by including the codeword of the undesired transmitter when computing the typical sets. Therefore, the capacity-achieving scheme performs simultaneous interference cancelation. If we consider the time-sharing of two non-simultaneous interference cancelation schemes, we are only able to achieve the region given by the convex hull of the set of (R_1, R_2) rate pairs satisfying

$$R_1 \leq I(X_1; Y_1|X_2) \tag{2.32}$$

$$R_2 \leq I(X_2; Y_2) \tag{2.33}$$

or

$$R_1 \leq I(X_1; Y_1) \tag{2.34}$$

$$R_2 \leq I(X_2; Y_2|X_1) \tag{2.35}$$

for some product probability distribution $P_{X_1}P_{X_2}$. Clearly, the last region is a subset of the capacity region. For some channels, it is an strict subset (consider for instance the Gaussian IC [Car78]). Hence, for DMICs with strong interference, simultaneous interference cancelation can strictly outperform non-simultaneous interference cancelation. The consideration of the interference channel with very strong interference shows that the result of Theorem 2.3 cannot be easily extended to other channels but that it describes an particular characteristic of the BC.

This fact can be explained in the following manner. A 1×2 DMBC can be viewed as a 2×2 DMIC in which both information sources are physically located inside a unique transmitter and hence there is only one input codeword. Whereas in the interference channel the codewords of each transmitter are independent, in the broadcast channel all the information is originated at the same logical entity and hence the codebooks of the auxiliary random variables can be arbitrarily correlated to improve the achievable rates. This advantage at the coding side of the channel provides the transmitter with additional degrees of freedom that can be exploited in the form of clever coding schemes for which non-simultaneous interference cancelation is capable of achieving the same rates as if simultaneous interference cancelation was performed with more complicated codes. Since in the interference channel there is no possibility to correlate the codewords of the transmitters, it is strictly beneficial for the achievable rates that both receivers simultaneously cancel out part of their interference.

2.3 Partial Interference Cancelation through Aided Decoding

2.3.1 Definitions

Following the terminology of [Han81], we shall distinguish between the 2×2 DMIC and the 2×2 *modified* DMIC.

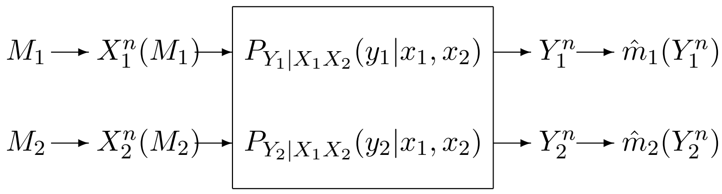

Figure 2.2: The 2×2 Discrete Memoryless IC (DMIC).

2.3.1.1 The 2×2 DMIC

Definition 2.4 *A 2×2 DMIC consists of 2 finite input alphabet sets, $\mathcal{X}_1$, $\mathcal{X}_2$, two finite output alphabet sets $\mathcal{Y}_1$, $\mathcal{Y}_2$ and two marginal transition probability distributions $P_{Y_1|X_1X_2}$ and $P_{Y_2|X_1X_2}$ such that $\mathbb{P}\{y_k^n|x_1^n,x_2^n\} = \prod_{j=1}^{n} p(y_{k,j}|x_{1,j},x_{2,j})$, $k = 1,2$.*

Definition 2.5 *A $\big((2^{nR_1},2^{nR_2}),n\big)$ code for the 2×2 DMIC consists of two sets of integers $\big[1,2^{nR_k}\big]$ $(k = 1,2)$, called the message sets, two encoding functions,*

$$X_k^n \;:\; \big[1,2^{nR_k}\big] \longrightarrow \mathcal{X}_k^n \;, \quad k = 1,2, \tag{2.36}$$

and two decoding functions

$$\hat{m}_k \;:\; \mathcal{Y}_k^n \longrightarrow \big[1,2^{nR_k}\big] \;, \quad k = 1,2. \tag{2.37}$$

Sender k draws an integer M_k uniformly from its message set $\big[1,2^{nR_k}\big]$ and sends the corresponding codeword $X_i^n(M_k)$, of length n, over the channel $(k = 1,2)$. Receiver k uses its channel output Y_k^n to create an estimate $\hat{m}_k(Y_k^n)$ of the transmitted message. This process is shown in Figure 2.2.

Since the message of each user is drawn independently, the average per-receiver probabilities of error of a $\big((2^{nR_1},2^{nR_2}),n\big)$ code for the DMIC can be expressed as

$$P_k^{(n)}(e) = 2^{-n(R_1+R_2)} \sum_{\substack{m_k \in [1,2^{nR_k}] \\ k=1,2}} \mathbb{P}\left\{\hat{m}_k(Y_k^n) \neq m_k | \,(m_1,m_2) \text{ sent}\right\}, \quad k = 1,2. \tag{2.38}$$

Definition 2.6 *A rate tuple (R_1,R_2) is said to be achievable for the 2×2 DMIC if there exists a sequence of $\big((2^{nR_1},2^{nR_2}),n\big)$ codes with $P_k^{(n)}(e) \to 0$, $k = 1,2$, for an arbitrarily large block length n.*

2.3.1.2 The 2×2 modified DMIC

In the modified DMIC, the way the information is coded and sent through the channel is different. Essentially, each user sends two independent flows of information instead of just one.

Definition 2.7 *A* $\left((2^{nR_{10}}, 2^{nR_{11}}, 2^{nR_{20}}, 2^{nR_{22}}), n\right)$ *code for the* 2×2 *modified DMIC consists of four sets of integers* $\left[1, 2^{nR_{k0}}\right], \left[1, 2^{nR_{kk}}\right]$ *(*$k = 1, 2$*), called the message sets, two encoding functions,*

$$X_k^n \; : \; \left[1, 2^{nR_{k0}}\right] \times \left[1, 2^{nR_{kk}}\right] \longrightarrow \mathcal{X}_k^n \; , \quad k = 1, 2, \tag{2.39}$$

and two decoding functions

$$(\hat{m}_{k0}, \hat{m}_{kk}) \; : \; \mathcal{Y}_k^n \longrightarrow \left[1, 2^{nR_{k0}}\right] \times \left[1, 2^{nR_{kk}}\right] \; , \quad k = 1, 2. \tag{2.40}$$

Sender k draws two integers (M_{k0}, M_{kk}) uniformly from its message sets $\left[1, 2^{nR_{k0}}\right] \times \left[1, 2^{nR_{kk}}\right]$ and sends the corresponding codeword $X_k^n(M_{k0}, M_{kk})$, of length n, over the channel ($k = 1, 2$). M_{k0} will be termed the low resolution message, while M_{kk} will be termed the high resolution message, $k = 1, 2$. Receiver k uses its channel output Y_k^n to create an estimate $(\hat{M}_{k0}(Y_k^n), \hat{M}_{kk}(Y_k^n))$ of the transmitted messages.

Again, the messages of each user are drawn independently, and the average per receiver probabilities of error of a $\left((2^{nR_{10}}, 2^{nR_{11}}, 2^{nR_{20}}, 2^{nR_{22}}), n\right)$ code for the 2×2 modified DMIC can be expressed as

$$\begin{aligned} P_k^{(n)}(e) &= 2^{-n(R_{10}+R_{11}+R_{20}+R_{22})} \times \\ &\sum_{\substack{m_{10}, m_{11} \\ m_{20}, m_{22}}} \mathbb{P}\left\{ (\hat{m}_{k0}(Y_k^n), \hat{m}_{kk}(Y_k^n)) \neq (m_{k0}, m_{kk}) | \, (m_{10}, m_{11}, m_{20}, m_{22}) \text{ sent} \right\}, \quad k = 1, 2. \end{aligned} \tag{2.41}$$

Definition 2.8 *A rate tuple* $(R_{10}, R_{11}, R_{20}, R_{22})$ *is said to be achievable for the* 2×2 *modified DMIC if there exists a sequence of* $\left((2^{nR_{10}}, 2^{nR_{11}}, 2^{nR_{20}}, 2^{nR_{22}}), n\right)$ *codes with* $P_k^{(n)}(e) \to 0$, $k = 1, 2$*, for an arbitrarily large block length* n*.*

There are two main reasons for considering the modified interference channel: **C**:

- Any achievability result in the modified channel can be extrapolated to the original one: if $(R_{10}, R_{11}, R_{20}, R_{22})$ is achievable for the prior, $(R_1, R_2) = (R_{10} + R_{11}, R_{20} + R_{22})$ is achievable for the latter [Han81, Cor. 2.1]. In fact, the achievable regions of Carleial, H&K, and the proposed one have been obtained as projections into the (R_1, R_2) plane of achievable regions in the modified setup. This will become explicit in Section 2.3.2.

- By splitting the information flow of one sender-receiver pair in one high- and another low-resolution message components, partial interference cancelation arises in a natural way by trying to cancel out the impact of the low-resolution signal of the interfering user.

2.3.2 The achievability result

Consider the region $\mathcal{R}(P_{QZ_1Z_2X_1X_2})$ consisting of the set of (R_1, R_2) rate pairs such that $R_1 = R_{10} + R_{11}$, $R_2 = R_{20} + R_{22}$ satisfying

$$R_{11} \le I(X_1;Y_1|Z_1Z_2Q) \quad ; \quad R_{22} \le I(X_2;Y_2|Z_1Z_2Q) \tag{2.42}$$
$$R_{10} + R_{11} \le I(X_1;Y_1|Z_2Q) \quad ; \quad R_{20} + R_{22} \le I(X_2;Y_2|Z_1Q) \tag{2.43}$$
$$R_{11} + R_{20} \le I(X_1Z_2;Y_1|Z_1Q) \quad ; \quad R_{22} + R_{10} \le I(X_2Z_1;Y_2|Z_2Q) \tag{2.44}$$
$$R_{10} + R_{11} + R_{20} \le I(X_1Z_2;Y_1|Q) \quad ; \quad R_{20} + R_{22} + R_{10} \le I(X_2Z_1;Y_2|Q) \tag{2.45}$$

for some joint probability distribution of the form $P_{QZ_1Z_2X_1X_2} = P_Q P_{Z_1|Q} P_{X_1|Z_1Q} P_{Z_2|Q} P_{X_2|Z_2Q}$ defined on $\mathcal{Q} \times \mathcal{Z}_1 \times \mathcal{Z}_2 \times \mathcal{X}_1 \times \mathcal{X}_2$. Denote by $\mathcal{P}$ the set of all such valid joint distributions. Q is a time-sharing random variable whose support set satisfies $|\mathcal{Q}| \le 3$ owing to Carathéodory's theorem.

Theorem 2.4 *The region*

$$\mathcal{R} = \bigcup_{P_{QZ_1Z_2X_1X_2} \in \mathcal{P}} \mathcal{R}(P_{QZ_1Z_2X_1X_2}) \tag{2.46}$$

is achievable for the 2×2 *interference channel.*

Proof. Please refer to Appendix 2.C. □

2.3.2.1 Carleial's achievable rate region

Consider the region $\mathcal{R}_{\mathsf{C}}(P_{QZ_1Z_2X_1X_2})$ consisting of the set of (R_1, R_2) rate pairs such that $R_1 = R_{10} + R_{11}$, $R_2 = R_{20} + R_{22}$ satisfying

$$\{R_{10} \le I(Z_1;Y_1|Q); R_{20} \le I(Z_2;Y_1|Z_1Q)\} \text{ or } \{R_{10} \le I(Z_1;Y_1|Z_2Q); R_{20} \le I(Z_2;Y_1|Q)\} \tag{2.47}$$
$$\{R_{10} \le I(Z_1;Y_2|Q); R_{20} \le I(Z_2;Y_2|Z_1Q)\} \text{ or } \{R_{10} \le I(Z_1;Y_2|Z_2Q); R_{20} \le I(Z_2;Y_2|Q)\} \tag{2.48}$$
$$R_{11} \le I(X_1;Y_1|Z_1Z_2Q) \quad ; \quad R_{22} \le I(X_2;Y_2|Z_1Z_2Q) \tag{2.49}$$

for some joint probability distribution of the form $P_{QZ_1Z_2X_1X_2} \in \mathcal{P}$ defined on $\mathcal{Q} \times \mathcal{Z}_1 \times \mathcal{Z}_2 \times \mathcal{X}_1 \times \mathcal{X}_2$.

Theorem 2.5 (Carleial) *The region*

$$\mathcal{R}_{\mathsf{C}} = \bigcup_{P_{QZ_1Z_2X_1X_2} \in \mathcal{P}} \mathcal{R}_{\mathsf{C}}(P_{QZ_1Z_2X_1X_2}) \tag{2.50}$$

is achievable for the 2×2 *interference channel.*

Althouth Carleial formulated $\mathcal{R}_{\mathsf{C}}$ using the convex-hull operation instead of using the time-sharing random variable Q, we have done so in order to establish a fair comparison with the proposed achievable rate region. Moreover, as pointed out in [Han81], the results obtained with the formulation in terms of Q are conjectured to be strictly superior than the ones obtained with the convex-hull operation for the general IC.

2.3.2.2 Han and Kobayashi's achievable rate region

Consider the region $\mathcal{R}_{\mathsf{HK}}(P_{QZ_1Z_2U_1U_2X_1X_2})$ consisting of the set of (R_1, R_2) rate pairs such that $R_1 = R_{10} + R_{11}$, $R_2 = R_{20} + R_{22}$ satisfying

$$R_{10} \leq I(Z_1; Y_1|U_1Z_2Q) \quad ; \quad R_{10} \leq I(Z_1; Y_2|U_2Z_2Q) \tag{2.51}$$
$$R_{20} \leq I(Z_2; Y_1|U_1Z_1Q) \quad ; \quad R_{20} \leq I(Z_2; Y_2|U_2Z_1Q) \tag{2.52}$$
$$R_{11} \leq I(U_1; Y_1|Z_1Z_2Q) \quad ; \quad R_{22} \leq I(U_2; Y_2|Z_1Z_2Q) \tag{2.53}$$
$$R_{10} + R_{20} \leq I(Z_1Z_2; Y_1|U_1Q) \quad ; \quad R_{10} + R_{20} \leq I(Z_1Z_2; Y_2|U_2Q) \tag{2.54}$$
$$R_{10} + R_{11} \leq I(Z_1U_1; Y_1|Z_2Q) \quad ; \quad R_{20} + R_{22} \leq I(Z_2U_2; Y_2|Z_1Q) \tag{2.55}$$
$$R_{11} + R_{20} \leq I(U_1Z_2; Y_1|Z_1Q) \quad ; \quad R_{22} + R_{10} \leq I(U_2Z_1; Y_2|Z_2Q) \tag{2.56}$$
$$R_{10} + R_{11} + R_{20} \leq I(Z_1U_1Z_2; Y_1|Q) \quad ; \quad R_{20} + R_{22} + R_{10} \leq I(Z_2U_2Z_1; Y_2|Q) \tag{2.57}$$

for some probability distribution $P_{QZ_1Z_2U_1U_2X_1X_2}$ defined on $\mathcal{Q} \times \mathcal{Z}_1 \times \mathcal{Z}_2 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X}_1 \times \mathcal{X}_2$ such that

- Z_1, Z_2, U_1, and U_2 are conditionally independent given Q.
- $X_1 = f_1(Z_1, U_1|Q)$, $X_2 = f_2(Z_2, U_2|Q)$.

Denote by $\mathcal{P}'$ the set of all such valid distributions.

Theorem 2.6 (Han and Kobayashi) *The region*

$$\mathcal{R}_{\mathsf{HK}} = \bigcup_{P_{QZ_1Z_2U_1U_2X_1X_2} \in \mathcal{P}'} \mathcal{R}_{\mathsf{HK}}(P_{QZ_1Z_2U_1U_2X_1X_2}) \tag{2.58}$$

is achievable for the 2×2 *interference channel.*

It is shown in [Han81] that the inclusion $\mathcal{R}_{\mathsf{C}} \subseteq \mathcal{R}_{\mathsf{HK}}$ holds. We show next that the inclusion $\mathcal{R}_{\mathsf{HK}} \subseteq \mathcal{R}$ also holds.

Lemma 2.2 *The achievable rate region* $\mathcal{R}_{\mathsf{HK}}$ *is a subset of* $\mathcal{R}$.

Proof. It is sufficient to prove that for each $P_{QZ_1Z_2U_1U_2X_1X_2} \in \mathcal{P}'$, there exists some $S_{QZ_1Z_2X_1X_2} \in \mathcal{P}$ such that $\mathcal{R}_{\mathsf{HK}}(P_{QZ_1Z_2U_1U_2X_1X_2}) \subseteq \mathcal{R}(S_{QZ_1Z_2X_1X_2})$.

Let $S_{QZ_1Z_2X_1X_2} \in \mathcal{P}$ be such that $S_{QZ_1Z_2X_1X_2} = S_Q S_{Z_1X_1|Q} S_{Z_2X_2|Q}$, where $S_Q = P_Q$ and

$$S_{Z_kX_k|Q}(z_k, x_k|q) = \sum_{u_k} P_{Z_kU_k|Q}(z_k, u_k|q) \cdot 1_{\{x_k = f_k(z_k, u_k|q)\}}, \quad k = 1, 2. \tag{2.59}$$

Note that this choice of $S_{QZ_1Z_2X_1X_2}$ satisfies $S_{QZ_1Z_2X_1X_2} = P_{QZ_1Z_2X_1X_2}$. To the end of showing that $\mathcal{R}_{\mathsf{HK}}(P_{QZ_1Z_2U_1U_2X_1X_2}) \subseteq \mathcal{R}(S_{QZ_1Z_2X_1X_2})$, first note that (2.51), (2.52), and (2.54) are constraints that limit the boundary of $\mathcal{R}_{\mathsf{HK}}(P_{QZ_1Z_2U_1U_2X_1X_2})$ that do not have counterparts in $\mathcal{R}(S_{QZ_1Z_2X_1X_2})$. On the other hand, the constraints (2.53), (2.55), (2.56), and (2.57) satisfy

$$I(U_k;Y_k|Z_1Z_2) = I(U_kZ_k;Y_k|Z_1Z_2) \leq I(X_k;Y_k|Z_1Z_2), \quad k = 1, 2 \tag{2.60}$$

$$I(Z_kU_k;Y_k|Z_j) \leq I(X_k;Y_k|Z_j), \quad k = 1, 2 \ \ j \neq k \tag{2.61}$$

$$I(U_kZ_j;Y_k|Z_k) = I(Z_kU_kZ_j;Y_k|Z_k) \leq I(X_kZ_j;Y_k|Z_k), \quad k = 1, 2 \ \ j \neq k \tag{2.62}$$

$$I(Z_kU_kZ_j;Y_k) \leq I(X_kZ_j;Y_k), \quad k = 1, 2 \ \ j \neq k. \tag{2.63}$$

These inequalities follow from the data processing theorem since $(Z_k, U_k) \to X_k \to (Y_1, Y_2)$ forms a Markov chain, $k = 1, 2$. Note that all the inequalities are satisfied with equality if $f_1(\cdot,\cdot)$, $f_2(\cdot,\cdot)$ are such that the following functions

$$Z_1 = h_{11}(X_1, U_1) \quad ; \quad Z_2 = h_{21}(X_2, U_2) \tag{2.64}$$

$$U_1 = h_{12}(Z_1, X_1) \quad ; \quad U_2 = h_{22}(Z_2, X_2) \tag{2.65}$$

exist [Gam82]. □

2.3.2.3 Comparison of the coding strategies of $\mathcal{R}_{\mathsf{C}}$, $\mathcal{R}_{\mathsf{HK}}$, and $\mathcal{R}$

In $\mathcal{R}_{\mathsf{C}}$, Carleial [Car78] considered superposition coding in which the cloud centers Z_k were encapsulated *probabilistically* into the input codewords X_k, while H&K [Han81] considered the simultaneous *deterministic* encapsulation (represented by the functions $f_1(\cdot,\cdot)$ and $f_2(\cdot,\cdot)$) of the cloud centers Z_k and the high-resolution information-bearing random variables U_k. For the achievability of the region $\mathcal{R}$ we adopted Carleial's approach based on the following two arguments: i) Lemma 2.2 shows that there is no information loss in using such a coding scheme but the possibility of increasing the mutual information; ii) the achievable rate region can be formulated in terms of a fewer number of random variables, obtaining simpler expressions.

We believe that the coding scheme by Carleial embodies a wider class of coding strategies than the ones considered by H&K. H&K require, for a given value of Q, the Z_k's and the U_k's to be independent. However, in the sequential superposition encoding approach, the codewords X_k may be formed as a deterministic function of two random variables Z_k and U_k, not necessarily meeting the independence constraint.

2.3.2.4 Comparison of the decoding strategies of $\mathcal{R}_{\mathsf{C}}$, $\mathcal{R}_{\mathsf{HK}}$, and $\mathcal{R}$

The region $\mathcal{R}_{\mathsf{C}}$ was obtained assuming a sequential behavior of the decoders. Receiver 1, for instance, tries first to reliably decode the cloud centers of both senders in a successive man-

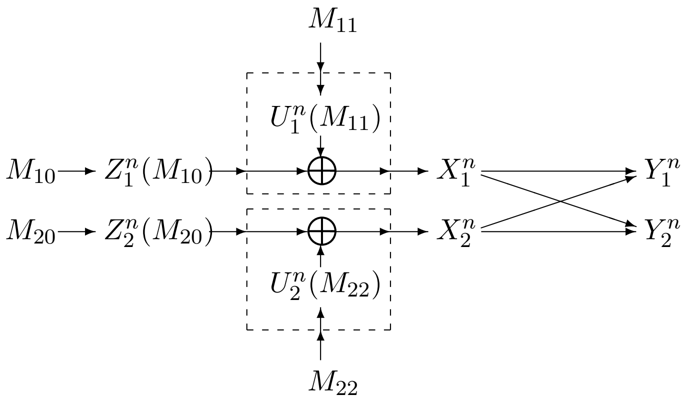

Figure 2.3: An example 2×2 modified interference channel.

ner (either Z_1 first and Z_2 second, or vice versa) and finally attempts to reliably decode the high-resolution desired message contained in X_1. In contrast, H&K improved the decoding performance by allowing the decoders to perform joint decoding and not requiring an specific decoding order. In this case, receiver 1 tries to reliably decode the information carried by (Z_1, U_1, Z_2) jointly.

We propose a decoding scheme that relaxes further the decoding strategy of the receivers. The key point is noticing that, for instance, receiver 1 does not require to reliably decode Z_2. Clearly, not considering Z_2 at receiver 1 would limit the achievable rates. However, receiver 1 can get benefit from taking Z_2 into account without decoding the low-resolution message of sender 2 with the *aided* decoding concept. The rationale behind the aided decoding concept is to include Z_2 to form the typical sets at receiver 1 but not to consider having ambiguity about the information carried by Z_2 as an error event. This way, the probability of error can be decomposed into fewer terms that yield fewer rate constraints, as seen in Section 2.3.2.2.

2.3.3 An example

In order to illustrate Lemma 2.2, we present an example for fixed $Q = \phi$ where, for a given probability distribution $P_{\phi Z_1 Z_2 U_1 U_2 X_1 X_2} \in \mathcal{P}'$, the region $\mathcal{R}(P_{\phi Z_1 Z_2 X_1 X_2})$ (where $P_{\phi Z_1 Z_2 X_1 X_2}$ is the appropriate marginalization of $P_{\phi Z_1 Z_2 U_1 U_2 X_1 X_2}$) strictly extends $\mathcal{R}_{\mathsf{HK}}(P_{\phi Z_1 Z_2 U_1 U_2 X_1 X_2})$.

Consider the following binary-inputs ($\mathcal{X}_k = \{0, 1\}$, $k = 1, 2$) binary-outputs ($\mathcal{Y}_k = \{0, 1\}$, $k = 1, 2$) modified interference channel of Figure 2.3.

Let

$$P_{Y_1|X_1X_2}(1|x_1, x_2) = [0.8\ 0.9\ 0.2\ 1], \quad P_{Y_2|X_1X_2}(1|x_1, x_2) = [0.8\ 0.1\ 0.8\ 0.2], \tag{2.66}$$

specify the transition probability matrices of the channels (the columns match the natural or-

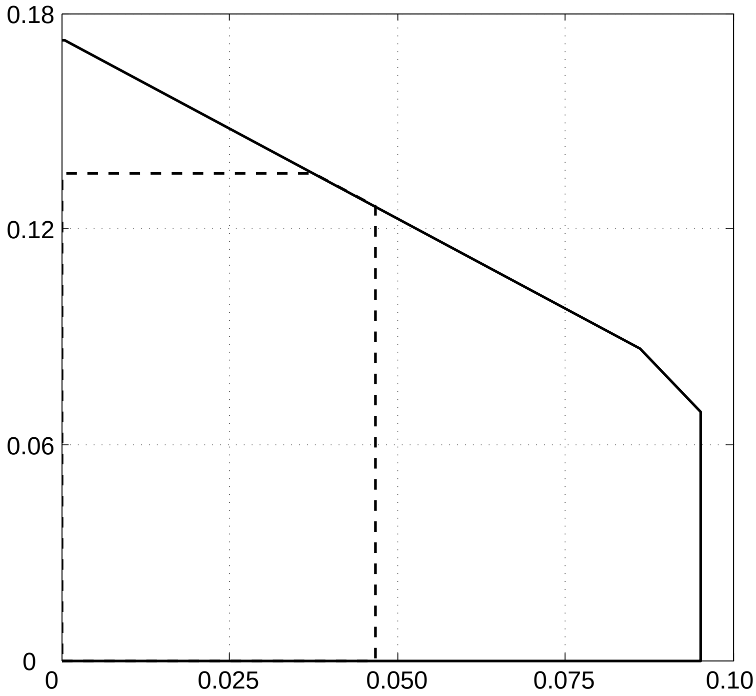

Figure 2.4: The regions $\mathcal{R}_{\mathsf{HK}}(P_{\phi Z_1 Z_2 U_1 U_2 X_1 X_2})$ (dashed) and $\mathcal{R}(P_{\phi Z_1 Z_2 X_1 X_2})$ (solid) computed for the binary interference channel of Figure 2.3. Units are [nat/ch. use].

dering of the inputs). Let $P_{\phi Z_1 Z_2 U_1 U_2 X_1 X_2} \in \mathcal{P}'$ be such that

$$P_{Z_1}(1) = 0.8,\ P_{U_1}(1) = 0.9\ \ ;\ \ P_{Z_2}(1) = 0.8,\ P_{U_2}(1) = 0.9, \tag{2.67}$$

and, as shown in Figure 2.3, $x_k = f_k(z_k, u_k) = z_k \oplus u_k$, $k = 1, 2$. The dashed boxes of Figure 2.3 represent how, for the sequential superposition coding used in $\mathcal{R}$, X_k is obtained in a transparent manner from Z_k in the sense that there is no explicit labeling of the high-resolution information-bearing random variable as done in $\mathcal{R}_{\mathsf{HK}}$ (the U_k signal). For the sake of fairness, we compute $\mathcal{R}_{\mathsf{HK}}(P_{\phi Z_1 Z_2 U_1 U_2 X_1 X_2})$ with the distribution above and $\mathcal{R}(P_{\phi Z_1 Z_2 X_1 X_2})$ with the same distributions of the cloud centers and the conditional distributions of the X_k's given the Z_k's obtained from (2.67) as well. Both regions are shown in Figure 2.4.

Although Figure 2.4 shows that $\mathcal{R}(P_{\phi Z_1 Z_2 X_1 X_2})$ strictly outperforms $\mathcal{R}_{\mathsf{HK}}(P_{\phi Z_1 Z_2 U_1 U_2 X_1 X_2})$, this does not imply that the full region $\mathcal{R}$ *strictly* outperforms $\mathcal{R}_{\mathsf{HK}}$. Lemma 2.2 only proves that $\mathcal{R}_{\mathsf{HK}} \subseteq \mathcal{R}$, but we haven't shown strict improvement of $\mathcal{R}$ over $\mathcal{R}_{\mathsf{HK}}$. To this respect, it was recently shown in [Cho06b] that both $\mathcal{R}$ and $\mathcal{R}_{\mathsf{HK}}$ are *equal*. That is, after performing the union over $\mathcal{P}$ and $\mathcal{P}'$, respectively, both regions coincide.

2.4 Conclusions

In scenarios where the presence of multiple neighboring sender-receiver pairs originates multiuser interference, the impact on the achievable rates is significant. One useful tool that can be used to mitigate the performance losses when the codebooks of the interferent users are known is partial interference cancelation. In this chapter we have studied when and how to perform it, giving special emphasis to the BC and the IC.

In the BC, the best known achievable rate region is obtained by time-sharing two situations in which only one of the receivers partially mitigates the interference that arises from the transmission of information to another user. If the coding strategy is generalized so as to comprise more general schemes in which both receivers perform partial interference cancelation simultaneously, the achievable rates remain unchanged. In the interference channel, oppositely, this result does not hold. Intuitively, when all the interference is under the control of the same source, correlated coding (as a way of precoding) can alleviate all the users from performing partial interference cancelation simultaneously (this is the case in the BC). On the other hand, if the interference is generated by non-cooperative sources that are physically disperse, the control of each information source over the resulting aggregate signal at the intended receiver is rather limited. This is the case in the IC, where partial interference cancelation significantly improves overall performance.

Motivated by the way interference cancelation is carried out in Marton's decoding strategy for the BC, we also worked towards improving achievability results in the IC through the use superposition coding at the sources and aided decoding at the receivers. By aided decoding we understand that the decoding at each receiver is aided by the inclusion of a low-resolution message component of the interfering user in the typical sets. This, in contrast to other approaches, does not imply to reliably decode part of the message of the interfering user. As a result, some rate constraints are relaxed. However, that only leads to an improvement on simplicity with respect to the best known achievability results since i) the new achievable region is described using a fewer number of rate inequalities, and ii) it requires a fewer number of auxiliary random variables. Although rate gains are precluded, the fewer number of rate inequalities of the proposed region may result in improved error exponents in the finite blocklength regime. Out of the scope of this dissertation, however, this line has not been explored.

2.A Appendix: Proof of Theorem 2.2

Let $m_1 \in [1, 2^{nR_1}]$, $m_2 \in [1, 2^{nR_2}]$ denote the messages intended for receivers 1 and 2, respectively. The messages are encoded into the codeword $X^n(m_1, m_2)$ through a two-stage coding process. To start with, we define the following isomorphisms:

$$\Phi_k : [1, 2^{R_k}] \longrightarrow [1, 2^{nR_{W_k}}] \times [1, 2^{nR_{U_k}}] \tag{2.68}$$
$$m_k \longrightarrow (i_k, j_k) = \left(\left\lfloor (m_k - 1)2^{-nR_{U_k}} \right\rfloor + 1, m_k - (i_k - 1)2^{nR_{U_k}} \right) \tag{2.69}$$

and their inverses

$$\Phi_k^{-1} : [1, 2^{nR_{W_k}}] \times [1, 2^{nR_{U_k}}] \longrightarrow [1, 2^{R_k}] \tag{2.70}$$
$$(i_k, j_k) \longrightarrow m_k = (i_k - 1)2^{nR_{U_k}} + j_k, \tag{2.71}$$

where $R_k = R_{W_k} + R_{U_k}$ and $k = 1, 2$. Note that, since Φ_k is an isomorphism, it is equivalent to transmit the message pair (m_1, m_2) than to transmit the message tuple $(i_1, j_1, i_2, j_2) = (\Phi_1(m_1), \Phi_2(m_2))$. The idea of the coding scheme is the following:

- (i_k, j_k) are integers to be communicated to the k-th receiver.
- The integers (i_1, i_2) are encoded simultaneously using the random binning technique with the random variables W_1 and W_2.
- For a given value of (i_1, i_2) the integers (j_1, j_2) are encoded simultaneously using the random binning technique with the random variables U_1 and U_2 conditioned on (W_1, W_2).
- Receiver 1 attempts to reliably decode (i_1, j_1) aided by W_2, i.e., it tries to mitigate the interference that arises from the transmission of m_2 by taking into account the random variable that encodes i_2 (without eventually decoding it). Receiver 2 operates similarly.

Codebook generation. Generate $2^{n\widetilde{R}_{W_1}}$ w_1^n sequences of length n, with each component drawn i.i.d. $\sim P_{W_1}$. Similarly, generate $2^{n\widetilde{R}_{W_2}}$ w_2^n sequences of length n, with each component drawn i.i.d. $\sim P_{W_2}$. Distribute the $2^{n\widetilde{R}_{W_1}}$ w_1^n sequences uniformly into $2^{nR_{W_1}}$ bins, each one containing $2^{n(\widetilde{R}_{W_1} - R_{W_1})}$ sequences on average ($\widetilde{R}_{W_1} \geq R_{W_1}$). Proceed similarly with the w_2^n sequences and throw them into $2^{nR_{W_2}}$ bins ($\widetilde{R}_{W_2} \geq R_{W_2}$). Let the integers i_1, i_2 index the W_1 and W_2 bins. Let a $W_1 \times W_2$ product bin be the product space of a W_1 bin and a W_2 bin. The $2^{n(R_{W_1} + R_{W_2})}$ $W_1 \times W_2$ product bins remain conceptually in the first layer.

For each $W_1 \times W_2$ product bin, select arbitrarily a jointly typical pair (w_1^n, w_2^n) and label it as a product bin head. The head of the (i_1, i_2)-th $W_1 \times W_2$ product bin is denoted by $(w_1^n(i_1, i_2), w_2^n(i_1, i_2))$. For each product bin, discard all the pairs of sequences that are not product bin heads.

For each $W_1 \times W_2$ product bin, generate $2^{n\widetilde{R}_{U_1}}$ u_1^n sequences of length n, with each component drawn i.i.d. $\sim P_{U_1|W_1W_2}$ (conditioned by the $W_1 \times W_2$ product bin head). Similarly, generate

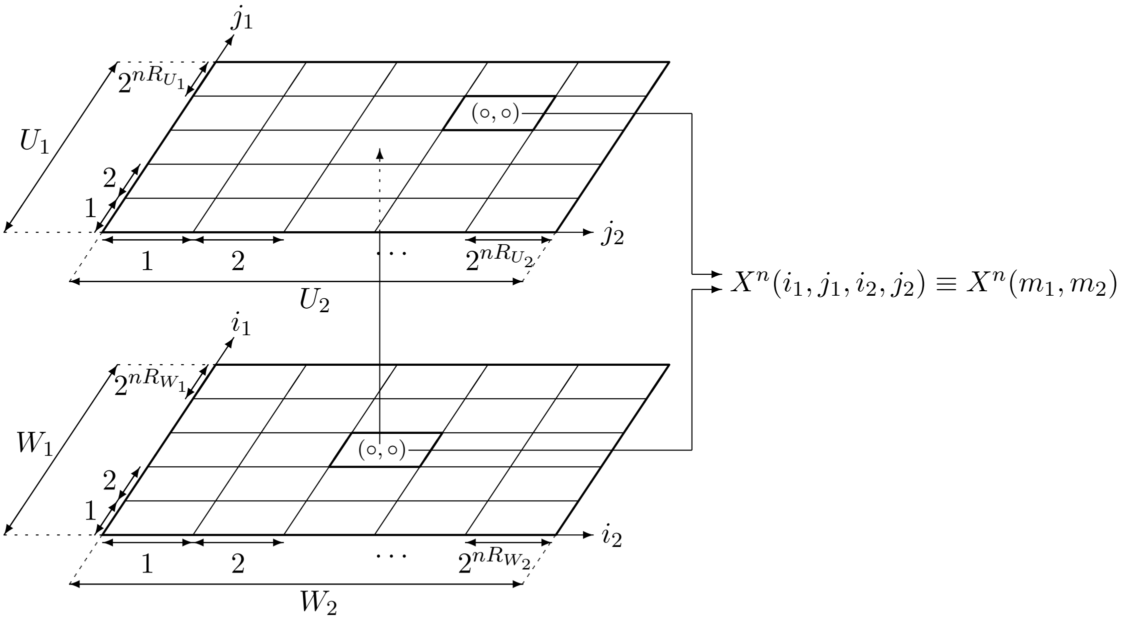

Figure 2.5: Diagram of the coding scheme: two layered random binning.

$2^{n\widetilde{R}_{U_2}}$ u_2^n sequences of length n, with each component drawn i.i.d. $\sim P_{U_2|W_1W_2}$ (conditioned by the $W_1 \times W_2$ product bin head). Distribute the $2^{n\widetilde{R}_{U_1}}$ u_1^n sequences uniformly into $2^{nR_{U_1}}$ bins, each one containing $2^{n(\widetilde{R}_{U_1}-R_{U_1})}$ sequences on average ($\widetilde{R}_{U_1} \geq R_{U_1}$). Proceed similarly with the u_2^n sequences and throw them into $2^{nR_{U_2}}$ bins ($\widetilde{R}_{U_2} \geq R_{U_2}$). Let the integers j_1, j_2 index the U_1 and U_2 bins. Let a $U_1 \times U_2$ product bin be the product space of a U_1 bin and a U_2 bin. For a given $W_1 \times W_2$ product bin, the $2^{n(R_{U_1}+R_{U_2})}$ $U_1 \times U_2$ product bins remain conceptually in a second layer.

For each $W_1 \times W_2$ product bin, select arbitrarily in all the associated $U_1 \times U_2$ product bins a jointly typical pair (u_1^n, u_2^n) and label it as a product bin head. The product bin head of the (j_1, j_2)-th $U_1 \times U_2$ product bin derived from the (i_1, i_2)-th $W_1 \times W_2$ product bin is denoted by $(u_1^n(i_1, i_2, j_1, j_2), u_2^n(i_1, i_2, j_1, j_2))$. For each $U_1 \times U_2$ product bin, discard all the pairs of sequences that are not product bin heads. The surviving u_k^n sequences are arbitrarily indexed as $u_k^n(i_1, i_2, l_k)$, where $l_k \leq 2^{n\min\{\widetilde{R}_{U_k}, R_{U_1}+R_{U_2}\}} \leq 2^{n\widetilde{R}_{U_k}}$, $k = 1, 2$. Denote by $\varphi_k(i_1, i_2, j_k)$ the set of $u_k^n(i_1, i_2, l_k)$ sequences belonging to the j_k-th U_k bin, whose cardinality is coarsely bounded by $|\varphi_k(i_1, i_2, j_k)| \leq 2^{n\widetilde{R}_{U_k}}$. In addition, $\sum_{j_k=1}^{2^{nR_{U_k}}} |\varphi_k(i_1, i_2, j_k)| \leq 2^{n\min\{\widetilde{R}_{U_k}, R_{U_1}+R_{U_2}\}} \leq 2^{n\widetilde{R}_{U_k}}$ also.

Encoding. If the source wants to transmit the message pair (m_1, m_2), it first computes $(i_1, j_1) = \Phi_1(m_1)$ and $(i_2, j_2) = \Phi_2(m_2)$. Then, the (i_1, i_2)-th product bin head of the W_1W_2 layer and the (j_1, j_2)-th product bin head of the U_1U_2 layer derived from $(w_1^n(i_1, i_2), w_2^n(i_1, i_2))$ are selected. An encoding error is declared if no product bin head could be defined in any of the layers. Finally, a codeword x^n of length n is drawn with each component i.i.d. $\sim P_{X|W_1W_2U_1U_2}$ (conditioned on $(w_1^n(i_1, i_2), w_2^n(i_1, i_2), u_1^n(i_1, i_2, j_1, j_2), u_2^n(i_1, i_2, j_1, j_2)))$ and sent through the channel. This process is shown in Figure 2.5.

Decoding. The decoding is symmetrical at both receivers.

- Receiver 1 computes the set $A_{\epsilon,1}^{(n)}$ of ϵ-typical $(w_1^n(i_1,i_2), w_2^n(i_1,i_2), u_1^n(i_1,i_2,l_1), y_1^n)$ sequences (for all l_1) and chooses $(\hat{i}_1,\hat{j}_1) \in [1, 2^{nR_{W_1}}] \times [1, 2^{nR_{U_1}}]$ as the unique pair of integers such that $(w_1^n(\hat{i}_1,i_2), w_2^n(\hat{i}_1,i_2), u_1^n(\hat{i}_1,i_2,l_1), y_1^n) \in A_{\epsilon,1}^{(n)}$ for some $i_2 \in [1, 2^{nR_{W_2}}]$ and some l_1 such that $u_1(\hat{i}_1,i_2,l_1) \in \varphi_1(\hat{i}_1,i_2,\hat{j}_1)$. If none or more than one such $(\hat{i}_1,\hat{j}_1)$ pair can be found, a decoding error is declared.

- Assuming that an estimate of both integers $(\hat{i}_1,\hat{j}_1)$ is available and a decoding error has not been declared, the estimate of the message m_1 is obtained as $\hat{m}_1 = \Phi_1^{-1}(\hat{i}_1,\hat{j}_1) = (\hat{i}_1 - 1)2^{nR_{U_1}} + \hat{j}_1$.

- Receiver 2 operates similarly by interchanging the roles of w_1^n and w_2^n, replacing u_1^n by u_2^n, and using Φ_2^{-1}.

Analysis of the probability of error. An error event will occur if any of the receivers declares a decoding error, or if the transmitter declares an encoding error in any of the layers. Without loss of generality, it can be assumed that the integers $(m_1, m_2) = (1,1)$ were sent. Equivalently, assume that $(i_1, i_2, j_1, j_2) = (1,1,1,1)$ were sent (since $\Phi_k(1) = (1,1)$, $k = 1,2$). Hence, by the union bound,

$$P_e^{(n)} \le P_{\mathsf{enc},1}^{(n)} + P_{\mathsf{enc},2}^{(n)} + P_{\mathsf{dec},1}^{(n)} + P_{\mathsf{dec},2}^{(n)}, \tag{2.72}$$

where :

- $P_{\mathsf{enc},1}^{(n)}$ is the probability of not finding a suitable product bin head in the $(1,1)$-th $W_1 \times W_2$ product bin.

- $P_{\mathsf{enc},2}^{(n)}$ is the probability of not finding a suitable product bin head in the $(1,1)$-th $U_1 \times U_2$ product bin belonging to the layer derived from the $(1,1)$-th $W_1 \times W_2$ product bin.

- $P_{\mathsf{dec},k}^{(n)}$ is the probability of decoding error of the k-th receiver, that is,

$$P_{\mathsf{dec},k}^{(n)} = \mathbb{P}\big\{(\hat{i}_k,\hat{j}_k) \neq (1,1) | (1,1,1,1) \text{ sent}\big\}. \tag{2.73}$$

The conditions on the rates that allow an arbitrary low probability of encoding error at both layers ($P_{\mathsf{enc},i}^{(n)} \to 0$, $i = 1,2$) can be derived and extrapolated from the work of Marton [Mar79] and the posterior alternative proof by El Gamal and Van der Meulen [Gam81],

$$R_{W_1} \le \widetilde{R}_{W_1} \ ; \ R_{U_1} \le \widetilde{R}_{U_1} \tag{2.74}$$

$$R_{W_2} \le \widetilde{R}_{W_2} \ ; \ R_{U_2} \le \widetilde{R}_{U_2} \tag{2.75}$$

$$R_{W_1} + R_{W_2} \le \widetilde{R}_{W_1} + \widetilde{R}_{W_2} - I(W_1; W_2) \ ; \ R_{U_1} + R_{U_2} \le \widetilde{R}_{U_1} + \widetilde{R}_{U_2} - I(U_1; U_2 | W_1 W_2). \tag{2.76}$$

Now consider the term $P^{(n)}_{\mathsf{dec},1}$. Let us define the following events

$$E(i_1,i_2,j_1) = \{(w_1^n(i_1,i_2), w_2^n(i_1,i_2), u_1^n, y_1^n) \in A^{(n)}_{\epsilon,1} \text{ for some } u_1^n \in \varphi_1(i_1,i_2,j_1) | (1,1,1,1) \text{ sent}\}, \tag{2.77}$$

which allow us to bound the probability of decoding error as

$$P^{(n)}_{\mathsf{dec},1} \leq \mathbb{P}\{\overline{E_0}\} + \sum_{j=1}^{6} \mathbb{P}\{E_j\}. \tag{2.78}$$

The definition of the error events E_j and their associated rate constraints are the following

- E0) $E_0 \triangleq E(1,1,1)$. By the Asymptotic Equipartition Property (AEP) [Cov06, Ch. 3], $\mathbb{P}\{E_0^c\} \leq \epsilon$.
- E1) $E_1 \triangleq \left\{ \bigcup_{i_1 \neq 1} E(i_1,1,1) \right\}$.

$$\begin{aligned}
\mathbb{P}\{E_1\} &= \mathbb{P}\left\{E(i_1,1,1) \text{ for some } i_1 \neq 1\right\} \overset{(a)}{\leq} 2^{nR_{W_1}} \mathbb{P}\left\{E(2,1,1)\right\} && (2.79)\\
&\overset{(b)}{\leq} 2^{nR_{W_1}} |\varphi_1(2,1,1)| \cdot \mathbb{P}\{(w_1^n(2,1), w_2^n(2,1), u_1^n(2,1,l_1), y_1^n) \in A^{(n)}_{\epsilon,1} | (1,1,1,1) \text{ sent}\} && (2.80)\\
&\overset{(c)}{\leq} 2^{n(R_{W_1}+\widetilde{R}_{U_1})} \mathbb{P}\{(w_1^n(2,1), w_2^n(2,1), u_1^n(2,1,l_1), y_1^n) \in A^{(n)}_{\epsilon,1} | (1,1,1,1) \text{ sent}\} && (2.81)\\
&\overset{(d)}{=} 2^{n(R_{W_1}+\widetilde{R}_{U_1})} \sum_{(w_1^n,w_2^n,u_1^n,y_1^n)\in A^{(n)}_{\epsilon,1}} P_{W_1W_2}(w_1,w_2) P_{U_1|W_1W_2}(u_1|w_1,w_2) P_{Y_1|W_2}(y_1|w_2) && (2.82)\\
&\leq 2^{n(R_{W_1}+\widetilde{R}_{U_1})} 2^{n(H(W_1,W_2,U_1,Y_1)+\epsilon)} 2^{-n(H(W_1W_2)-\epsilon)} 2^{-n(H(U_1|W_1W_2)-2\epsilon)} 2^{-n(H(Y_1|W_2)-2\epsilon)} && (2.83)\\
&\leq 2^{n(R_{W_1}+\widetilde{R}_{U_1})} 2^{-n(H(Y_1|W_2)-H(Y_1|W_1W_2U_1)-6\epsilon)} = 2^{n(R_{W_1}+\widetilde{R}_{U_1})} 2^{-n(I(W_1U_1;Y_1|W_2)-6\epsilon)} && (2.84)
\end{aligned}$$

The inequalities (a) and (b) and (c) follow from the union bound and the symmetry of the code. In (c), it has been used that $|\varphi_1(2,1,1)| \leq 2^{n\widetilde{R}_{U_1}}$ (in the worst case scenario, all the u_1^n sequences are concentrated in the same U_1 bin). The equality (d) takes into account that when $(1,1,1,1)$ is sent, y_1^n does not depend on $w_1^n(2,1)$, but may depend on $w_2^n(2,1)$ if the same w_2^n sequence happens to belong to the product bin head of the $(1,1)$ and the $(2,1)$ $W_1 \times W_2$ product bins. In addition, it has been considered that l_1 is such that $u_1^n(i_1,i_2,l_1) \in \varphi_1(2,1,1)$. Since ϵ can be made arbitrarily small, the probability of the error event E_1 is driven to zero with n if

$$R_{W_1} + \widetilde{R}_{U_1} \leq I(W_1U_1;Y_1|W_2). \tag{2.85}$$

A similar analysis is performed for the rest of error events. Note that, in the rest of cases, the mutual information constraining the appropriate sum of rates follows after subtracting $H(Y_1|W_1W_2U_1)$ to the entropy of Y_1 conditioned by an appropriate subset of random variables that follows from the dependence tree in each error event. We will use this fact to simplify the derivations of the remaining rate constraints.

- E2) $E_2 \triangleq \left\{ \bigcup_{i_1,i_2 \neq 1} E(i_1,i_2,1) \right\}$. Similarly and up to some epsilons,

$$R_{W_1} + R_{W_2} + \widetilde{R}_{U_1} \leq H(Y_1) - H(Y_1|W_1W_2U_1) = I(W_1W_2U_1;Y_1). \tag{2.86}$$

Note that in this situation y_1^n is independent of any of the other signals.

- E3) $E_3 \triangleq \Big\{ \bigcup_{i_1, j_1 \neq 1} E(i_1, 1, j_1) \Big\}$. As in E1), the channel output y_1^n may depend on $w_2^n(i_1, 1)$. Hence,
$$R_{W_1} + \widetilde{R}_{U_1} \leq H(Y_1|W_2) - H(Y_1|W_1W_2U_1) = I(W_1U_1; Y_1|W_2). \tag{2.87}$$
- E4) $E_4 \triangleq \Big\{ \bigcup_{j_1 \neq 1} E(1, 1, j_1) \Big\}$. Here, y_1^n shows dependence on $w_1^n(1,1)$ and $w_2^n(1,1)$. Therefore,
$$\widetilde{R}_{U_1} \leq H(Y_1|W_1W_2) - H(Y_1|W_1W_2U_1) = I(U_1; Y_1|W_1W_2). \tag{2.88}$$
- E5) $E_5 \triangleq \Big\{ \bigcup_{i_2, j_1 \neq 1} E(1, i_2, j_1) \Big\}$. Oppositely to what happened in E1), y_1^n does not depend on $w_2^n(1,2)$, but may depend on $w_1^n(1,2)$ if the same w_1^n sequence happens to belong to the product bin head of the $(1,1)$ and the $(1,i_2)$ $W_1 \times W_2$ product bins.
$$R_{W_2} + \widetilde{R}_{U_1} \leq H(Y_1|W_1) - H(Y_1|W_1W_2U_1) = I(W_2U_1; Y_1|W_1). \tag{2.89}$$
- E6) $E_6 \triangleq \Big\{ \bigcup_{i_1, i_2, j_1 \neq 1} E(i_1, i_2, j_1) \Big\}$. When there is mismatch in all the variables, y_1^n is independent of w_1^n, w_2^n and u_1^n hence yielding
$$R_{W_1} + R_{W_2} + \widetilde{R}_{U_1} \leq H(Y_1) - H(Y_1|W_1W_2U_1) = I(W_1W_2U_1; Y_1). \tag{2.90}$$

Notice that E1) and E2) are redundant. If we gather the useful rate constraints for receiver 1 together with the extrapolated rate constraints for receiver 2, we obtain the following system of constraints:

$$\widetilde{R}_{U_1} \leq I(U_1; Y_1|W_1W_2) \quad ; \quad \widetilde{R}_{U_2} \leq I(U_2; Y_2|W_1W_2) \tag{2.91}$$
$$R_{W_1} + \widetilde{R}_{U_1} \leq I(W_1U_1; Y_1|W_2) \quad ; \quad R_{W_2} + \widetilde{R}_{U_2} \leq I(W_2U_2; Y_2|W_1) \tag{2.92}$$
$$R_{W_2} + \widetilde{R}_{U_1} \leq I(W_2U_1; Y_1|W_1) \quad ; \quad R_{W_1} + \widetilde{R}_{U_2} \leq I(W_1U_2; Y_2|W_2) \tag{2.93}$$
$$R_{W_1} + R_{W_2} + \widetilde{R}_{U_1} \leq I(W_1W_2U_1; Y_1) \quad ; \quad R_{W_1} + R_{W_2} + \widetilde{R}_{U_2} \leq I(W_1W_2U_2; Y_2), \tag{2.94}$$

which ensure vanishing probabilities of decoding error at both receivers. Taking into account the rate constraints that arose from the necessity of driving to zero the probability of an encoding error at the source

$$R_{W_1} \leq \widetilde{R}_{W_1} \ ; \ R_{U_1} \leq \widetilde{R}_{U_1} \tag{2.95}$$
$$R_{W_2} \leq \widetilde{R}_{W_2} \ ; \ R_{U_2} \leq \widetilde{R}_{U_2} \tag{2.96}$$
$$R_{W_1} + R_{W_2} \leq \widetilde{R}_{W_1} + \widetilde{R}_{W_2} - I(W_1; W_2) \ ; \ R_{U_1} + R_{U_2} \leq \widetilde{R}_{U_1} + \widetilde{R}_{U_2} - I(U_1; U_2|W_1W_2), \tag{2.97}$$

it can be noticed that there is no constraint put on $\widetilde{R}_{W_1}$, $\widetilde{R}_{W_2}$. Hence, they can be increased up to their maximum value ($\widetilde{R}_{W_k} \leq H(W_k)$, $k = 1, 2$ due to the necessity of finding enough typical sequences for the codebook construction). This results in the achievability of $\mathcal{R}(P_{W_1W_2U_1U_2X})$. By timesharing arguments, the convex hull of the union over all the possible probability distributions in $\mathcal{P}'$, the region $\mathcal{R}$, is achievable. □

2.B Appendix: Proof of Theorem 2.3

We shall prove that $\mathcal{R}_{\mathsf{MT}} \subseteq \mathcal{R}$ and $\mathcal{R} \subseteq \mathcal{R}_{\mathsf{MT}}$ hold.

2.B.1 Proof of $\mathcal{R}_{\mathsf{MT}} \subseteq \mathcal{R}$

Consider the expression of $\mathcal{R}(P_{W_1W_2U_1U_2X})$ and set $W_1 = W$ and $W_2 = \phi$ (constant). After the following steps:

- set $R_{W_2} = 0$ ($W_2 = \phi$),
- eliminate the variables $\widetilde{R}_{U_1}$ and $\widetilde{R}_{U_2}$ using the Fourier-Motzkin elimination technique [Dan97],
- set the change of variables: $R_{U_1} = R_1 - R_{W_1}$, $R_{U_2} = R_2$,
- eliminate R_{W_1} using Fourier-Motzkin elimination,

the achievable rate region $\mathcal{R}(P_{W\phi U_1U_2X})$ is suitably projected into the (R_1, R_2) plane. Its expression becomes

$$\begin{aligned} R_1 &\leq I(WU_1; Y_1) &(2.98)\\ R_2 &\leq I(U_2; Y_2|W) &(2.99)\\ R_1 + R_2 &\leq \min\{I(W; Y_1), I(W; Y_2)\} + I(U_1; Y_1|W) + I(U_2; Y_2|W) - I(U_1; U_2|W), &(2.100)\end{aligned}$$

which depends only on $P_{WU_1U_2X} \in \mathcal{P}$. Similarly, when $W_1 = \phi$ and $W_2 = W$ we obtain that the projection of $\mathcal{R}(P_{\phi WU_1U_2X})$ into the (R_1, R_2) plane has the expression

$$\begin{aligned} R_1 &\leq I(U_1; Y_1|W) &(2.101)\\ R_2 &\leq I(WU_2; Y_2) &(2.102)\\ R_1 + R_2 &\leq \min\{I(W; Y_1), I(W; Y_2)\} + I(U_1; Y_1|W) + I(U_2; Y_2|W) - I(U_1; U_2|W), &(2.103)\end{aligned}$$

which again depends only on $P_{WU_1U_2X}$. Finally, note that

$$\begin{aligned} \mathcal{R}_{\mathsf{MT}} &= \bigcup_{P_{WU_1U_2X} \in \mathcal{P}} \mathcal{R}_{\mathsf{MT}}(P_{WU_1U_2X}) \overset{(a)}{=} \mathrm{Co}\left(\bigcup_{P_{WU_1U_2X} \in \mathcal{P}} \mathcal{R}_{\mathsf{MT}}(P_{WU_1U_2X}) \right) &(2.104)\\ &\overset{(b)}{=} \mathrm{Co}\left(\bigcup_{P_{WU_1U_2X} \in \mathcal{P}} \mathrm{Co}\left(\mathcal{R}(P_{W\phi U_1U_2X}) \cup \mathcal{R}(P_{\phi WU_1U_2X})\right) \right) &(2.105)\\ &= \mathrm{Co}\left(\bigcup_{\substack{P_{W_1W_2U_1U_2X} \in \mathcal{P}' \\ W_1=\phi \text{ or } W_2=\phi}} \mathcal{R}(P_{W_1W_2U_1U_2X}) \right) \subseteq \mathrm{Co}\left(\bigcup_{P_{W_1W_2U_1U_2X} \in \mathcal{P}'} \mathcal{R}(P_{W_1W_2U_1U_2X}) \right) &(2.106)\\ &= \mathcal{R}, &(2.107)\end{aligned}$$

where the equality (a) follows from the convexity of $\mathcal{R}_{\mathsf{MT}}$, and (b) follows from noticing that the two maximal points of $\mathcal{R}_{\mathsf{MT}}(P_{WU_1U_2X})$ are contained within $\mathcal{R}(P_{W\phi U_1U_2X}) \cup \mathcal{R}(P_{\phi WU_1U_2X})$. □

2.B.2 Proof of $\mathcal{R} \subseteq \mathcal{R}_{\mathsf{MT}}$

Consider the expression of $\mathcal{R}(P_{W_1W_2U_1U_2X})$ in Theorem 2.2. After the steps

- eliminate the variables $\widetilde{R}_{U_1}$ and $\widetilde{R}_{U_2}$ using the Fourier-Motzkin elimination technique,
- set the change of variables: $R_{U_1} = R_1 - R_{W_1}$, $R_{U_2} = R_2 - R_{W_2}$,

the projection of $\mathcal{R}(P_{W_1W_2U_1U_2X})$ into the $(R_1, R_2, R_{W_1}, R_{W_2})$ hyperplane is obtained. Among the rate inequalities that define the region in that hyperplane, we reproduce here the following subset:

$$\begin{aligned}
R_1 &\leq I(W_1U_1;Y_1|W_2) && (2.108)\\
R_2 &\leq I(W_2U_2;Y_2|W_1) && (2.109)\\
R_1+R_2 &\leq I(W_1W_2U_1;Y_1) + I(U_2;Y_2|W_1W_2) - I(U_1;U_2|W_1W_2) && (2.110)\\
R_1+R_2 &\leq I(W_2U_1;Y_1|W_1) + I(W_1U_2;Y_2|W_2) - I(U_1;U_2|W_1W_2) && (2.111)\\
R_1+R_2 &\leq I(W_1U_1;Y_1|W_2) + I(W_2U_2;Y_2|W_1) - I(U_1;U_2|W_1W_2) && (2.112)\\
R_1+R_2 &\leq I(U_1;Y_1|W_1W_2) + I(W_1W_2U_2;Y_2) - I(U_1;U_2|W_1W_2), && (2.113)
\end{aligned}$$

which defines a subset of $\mathcal{R}_{\mathsf{MT}}(P_{W_1W_2U_1U_2X})$ (Lemma 2.1). Therefore, there is no need to consider the rest of rate inequalities of $\mathcal{R}(P_{W_1W_2U_1U_2X})$ to show that

$$\mathcal{R}(P_{W_1W_2U_1U_2X}) \subseteq \mathcal{R}_{\mathsf{MT}}(P_{W_1W_2U_1U_2X}), \quad (2.114)$$

and, consequently,

$$\mathcal{R} = \mathrm{Co}\left(\bigcup_{P_{W_1W_2U_1U_2X}\in\mathcal{P}'} \mathcal{R}(P_{W_1W_2U_1U_2X})\right) \subseteq \mathrm{Co}\left(\bigcup_{P_{W_1W_2U_1U_2X}\in\mathcal{P}'} \mathcal{R}_{\mathsf{MT}}(P_{W_1W_2U_1U_2X})\right) \stackrel{(a)}{=} \mathcal{R}_{\mathsf{MT}}, \quad (2.115)$$

where equality (a) follows from the convexity of $\mathcal{R}_{\mathsf{MT}}$, which makes the convex hull unnecessary. □

2.C Appendix: Proof of Theorem 2.4

It is sufficient to show the achievability of (2.42)-(2.45) for the modified interference channel. Suppose we have a coding scheme like the one in Figure 2.6.

Codebook generation. Generate a random sequence Q^n of length n with each element i.i.d. $\sim P_Q$. For each sequence q^n, sender 1 generates $2^{nR_{10}}$ independent codewords of length n

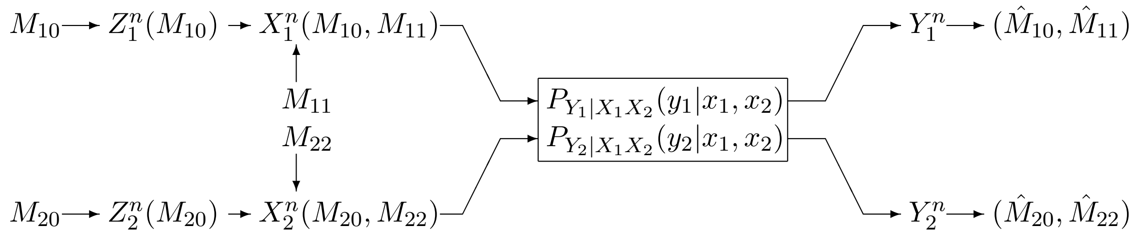

Figure 2.6: Coding scheme for the interference channel. The role of the time-sharing random variable Q has been omitted.

indexed as $Z_1^n(M_{10}|q^n)$ ($M_{10} \in [1, 2^{nR_{10}}]$) with each element i.i.d. $\sim P_{Z_1|Q}$. For each generated $(q^n, Z_1^n(M_{10}|q^n))$ pair, sender 1 generates $2^{nR_{11}}$ independent codewords of length n indexed as $X_1^n(M_{10}, M_{11}|q^n)$ ($M_{11} \in [1, 2^{nR_{11}}]$) with each element i.i.d. $\sim P_{X_1|Z_1Q}$. The rate of sender 1 when projected into the original interference channel is $R_1 = R_{10} + R_{11}$. Sender 2 proceeds similarly using the same sequence Q^n (its value is revealed to both senders).

Encoding. To send $(m_{k0}, m_{kk}) \in [1, 2^{nR_{k0}}] \times [1, 2^{nR_{kk}}]$, sender k transmits $X_k^n(m_{k0}, m_{kk}|q^n)$, $k = 1, 2$.

Decoding. Receiver 1 looks for the unique pair of integers $(\hat{m}_{10}, \hat{m}_{11}) \in [1, 2^{nR_{10}}] \times [1, 2^{nR_{11}}]$ satisfying

$$(q^n, z_1^n(m_{10}|q), x_1^n(m_{10}, m_{11}|q^n), z_2^n(m_{20}|q), y_1^n) \in A_\epsilon^{(n)}, \tag{2.116}$$

for some $m_{20} \in [1, 2^{nR_{20}}]$ (one or many of them), where $A_\epsilon^{(n)}$ is the set of ϵ-typical sequences. Note that the value of q^n used in the construction of the codewords is also told to both receivers. If none or more than such a pair exists, a decoding error is declared. Note that in order not to have ambiguity on its decision, receiver 1 requires a unique pair $(\hat{m}_{10}, \hat{m}_{11})$ to trigger the jointly typical set with some z_2^n sequence. It might have ambiguity about m_{20} if such z_2^n was not unique. However, this is not a problem for receiver 1 since it is only interested in decoding the information-bearing random variables of sender 1 (an ambiguity on m_{20} does not affect $P_1^{(n)}(e)$). Receiver 2 proceeds similarly.

Analysis of the probability of error. Let us focus on the probability of error at receiver 1. Thanks to the symmetry of the random code, we can express the average probability of error (2.41) at receiver 1 as

$$P_1^{(n)}(e) = \mathbb{P}\big\{(\hat{m}_{10}, \hat{m}_{11}) \neq (1, 1)|(1, 1, 1, 1) \text{ sent}\big\}, \tag{2.117}$$

where $(1, 1, 1, 1)$ indicates the value of the tuple $(m_{10}, m_{11}, m_{20}, m_{22})$. Next, by defining the events

$$E(m_{10}, m_{11}, m_{20}) \triangleq \big\{(q^n, z_1^n(m_{10}|q^n), x_1^n(m_{10}, m_{11}|q^n), z_2^n(m_{20}|q^n), y_1^n) \in A_\epsilon^{(n)}|(1, 1, 1, 1) \text{ sent}\big\} \tag{2.118}$$

we can set a complete list of the different situations in which a decoding error is declared at receiver 1 along with their associated rate constraints in Table 2.1.

Case	Error event at receiver 1	Associated rate constraint
E1	$\overline{E}(1,1,1)$	-
E2	$E(m_{10},1,1)\|E_1(1,1,1)$, $m_{10}\neq 1$	$R_{10} < I(X_1;Y_1\|Z_2Q)$
E3	$E(1,m_{11},1)\|E_1(1,1,1)$, $m_{11}\neq 1$	$R_{11} < I(X_1;Y_1\|Z_1Z_2Q)$
E4	$E(m_{10},m_{11},1)\|E_1(1,1,1)$, $m_{10},m_{11}\neq 1$	$R_{10}+R_{11} < I(X_1;Y_1\|Z_2Q)$
E5	$E(m_{10},1,m_{20})\|E_1(1,1,1)$, $m_{10},m_{20}\neq 1$	$R_{10}+R_{20} < I(X_1Z_2;Y_1\|Q)$
E6	$E(1,m_{11},m_{20})\|E_1(1,1,1)$, $m_{11},m_{20}\neq 1$	$R_{11}+R_{20} < I(X_1Z_2;Y_1\|Z_1Q)$
E7	$E(m_{10},m_{11},w_{20})\|E_1(1,1,1)$, $m_{10},m_{11},m_{20}\neq 1$	$R_{10}+R_{11}+R_{20} < I(X_1Z_2;Y_1\|Q)$

Table 2.1: Situations of decoding error and associated rate constraints.

Note that, oppositely to what happened with H&K's decoding scheme, we do not consider $E(1,1,m_{20})$, for some $m_{20}\neq 1$, an error event for receiver 1. This fact allows us to disregard two inequalities (one for each receiver). The rate constraints associated to each error event are obtained using standard results on joint typicality in a not difficult but rather long manner. For this reason, we will skip some of the parts of the proof which can be easily derived.

Consider first the error event E1. Thanks to the AEP, we can bound $\mathbb{P}\{\overline{E}(1,1,1))\}\leq\epsilon$.

With respect to the error event E2, we have:

$$\begin{aligned}
\mathbb{P}\{E2\} &= \mathbb{P}\{E_1(w_{10},1,1) \text{ for some } w_{10}\neq 1\} \overset{(a)}{\leq} 2^{nR_{10}}\mathbb{P}\{E_1(2,1,1)\} && (2.119)\\
&= 2^{nR_{10}}\mathbb{P}\{(q^n,z_1^n(2|q^n),x_1^n(2,1|q^n),z_2^n(1|q^n),y_1^n)\in A_\epsilon^{(n)}|(1,1,1,1)\text{ sent}\} && (2.120)\\
&\overset{(b)}{=} 2^{nR_{10}}\sum_{(q_1^n,z_1^n,x_1^n,z_2^n,y_1^n)\in A_\epsilon^{(n)}} P_{QZ_1X_1}(q^n,z_1^n,x_1^n)P_{Z_2|Q}(z_2^n|q^n)P_{Y_1|Z_2Q}(y_1^n|z_2^n,q^n) && (2.121)\\
&\overset{(c)}{\leq} 2^{nR_{10}}2^{n(H(Q,Z_1,X_1,Z_2,Y_1)+\epsilon)}2^{-n(H(Q,Z_1,X_1)-\epsilon)}2^{-n(H(Z_2|Q)-2\epsilon)}2^{-n(H(Y_1|Z_2Q)-2\epsilon)} && (2.122)\\
&= 2^{nR_{10}}2^{n(H(Y_1|Z_1,X_1,Z_2,Q)-H(Y_1|Z_2Q)-6\epsilon)}\\
&\overset{(d)}{=} 2^{nR_{10}}2^{-n(I(X_1;Y_1|Z_2Q)-6\epsilon)}, && (2.123)
\end{aligned}$$

where (a) follows from the symmetry of the code construction and the union bound, (b) follows from the dependence tree between $z_1^n(2|q^n), x_1^n(2,1|q^n), z_2^n(1|q^n)$, and y_1^n. Finally, (c) follows from applying [Cov06, Thm. 14.2.1] and (d) follows since $Z_1\to X_1\to Y_1$ forms a Markov chain given Q. Using similar upper-bounding techniques for the other cases of error event we obtain the rest of rate constraints of Table 2.1.

Note that the union bound can be used to upper bound $P_1^{(n)}(e)$ as the sum of the probabilities of cases E1 to E2. Hence, if all the rate constraints of Table 2.1 are satisfied, a total rate of $R_1 = R_{10}+R_{11}$ is reliably communicated from sender 1 to receiver 1. Dual constraints with interchanged subindexes arise to minimize $P_2^{(n)}(e)$. Table 2.2 summarizes all the rate constraints that must be fulfilled for the achievability of a $(R_1,R_2)=(R_{10}+R_{11},R_{20}+R_{22})$ rate pair.

	Rate constraint (receiver 1)		Rate constraint (receiver 2)
1	$R_{11} < I(X_1; Y_1 \vert Z_1 Z_2 Q)$	5	$R_{22} < I(X_2; Y_2 \vert Z_1 Z_2 Q)$
2	$R_{10} + R_{11} < I(X_1; Y_1 \vert Z_2 Q)$	6	$R_{20} + R_{22} < I(X_2; Y_2 \vert Z_1 Q)$
3	$R_{11} + R_{20} < I(X_1 Z_2; Y_1 \vert Z_1 Q)$	7	$R_{22} + R_{10} < I(X_2 Z_1; Y_2 \vert Z_2 Q)$
4	$R_{10} + R_{11} + R_{20} < I(X_1 Z_2; Y_1 \vert Q)$	8	$R_{20} + R_{22} + R_{10} < I(X_2 Z_1; Y_2 \vert Q)$

Table 2.2: Rate constraints for the achievability of a $(R_1, R_2) = (R_{10} + R_{11}, R_{20} + R_{22})$ rate pair.

Redundant constraints have been removed. □

Chapter 3

The Totally Asynchronous Interference Channel with Single-User Receivers

In Chapter 2 we discussed about the convenience of allowing the receivers to perform partial interference cancelation. On the one hand, it was found that this feature enhances the achievable rates as long as each transmitter is not capable of shaping all the multiuser interference (e.g., in the Interference Channel). On the other hand, when useful, each receiver does not need to reliably decode the underlying messages carried by the interfering signals: by including them in the definition of the typical sets, the chances of correct detection of the desired message are increased even when the interfering messages cannot be resolved.

It is implicitly assumed in the previous reasoning that the receivers have perfect knowledge of the codebooks of the interfering users. In other words, each receiver is able to form a finite list of potential interference realizations to aid the decoding of the desired message. Whereas this is a reasonable assumption in small-size centralized networks, there are many other practical applications where the interferers and/or the coupling channels are unknown that prevent us from blissfully adopting it. In particular, in this chapter we focus on the achievable rates in decentralized wireless networks with uncoordinated sender-destination pairs by considering an equivalent information theoretic model.

This motivates the study of the totally asynchronous interference channel with single-user receivers. Since this channel is not information stable, we characterize its capacity region resorting to information density, although more amenable single-letter inner and outer bounds are provided as well. The single-letter inner bound is obtained by constraining each sender to use stationary inputs with i.i.d. letters. Aiming at numerical evaluation of achievable rates, we subsequently concentrate on this achievable rate region.

When considering the Gaussian case, the Gram-Charlier expansion approximation of mutual information allows us to i) show that Gaussian-distributed codes are not always optimal, and ii) characterize sufficient conditions under which other input distributions may increase simultaneously the mutual information of both links beyond the Gaussian achievable rates. These conditions depend exclusively on the transmit power constraints of the senders and the coupling coefficients of the channel, and essentially require the receivers to be hampered by at least moderate interference. As a constructive verification of our results, some non-Gaussian-distributed codes are studied to show that their achievable rates outperform those of Gaussian-distributed codes when the conditions derived under the finite expansion approximation statistical analysis are fulfilled.

3.1 Introduction

The interference channel (IC) [Car78] is the network scenario that models the interactions between several disjoint non-cooperative (in the sense that no relays are used for cooperation) sender-receiver communication links sharing, generally in a non-orthogonal manner, the same physical medium. Interference is generated across links, and the achievable rates of each pair are hence coupled. The fact that each destination is interested in decoding only one among all the information-bearing codewords on which its channel output depends is what makes this channel difficult to analyze.

Finding a single-letter characterization of the capacity region of the IC remains an open problem which, however, has been solved in some particular cases: i) statistically equivalent channel outputs [Car83], ii) very strong [Car75,Car78] or strong interference [Sat78,Han81,Sat81, Cos87], iii) a class of discrete additive degraded ICs [Ben79], and iv) a class of deterministic ICs [Gam82]. As for results that hold in general, inner and outer bounds on the capacity region have been found for discrete memoryless [Car78, Han81, Car83, Cho06b] and Gaussian ICs [Car78, Han81, Car83, Kra04, Sas04, Sha07, Etk07].

The original definition of the IC assumes perfect frame synchronization between the senders and the receivers, i.e. that the codewords sent by the transmitters are received at unison at each destination. Instead, in the absence of frame synchronism, there exist unpredictable offsets between the epochs at which codewords are received at the destinations. In decentralized wireless networks consisting of several autonomous sender-receiver pairs this latter situation is likely to be predominant. Additionally, without a centralized entity broadcasting information about who and when is transmitting, the destinations may be unaware of the potential interference hampering the transmission of their intended user. Routing paths that change, fading dynamics, or node mobility may render useless any interference estimation attempt at the destinations by causing fast time-varying changes on the network topology and hence on the interference. The following are example scenarios and applications dealing with asynchronism and unknown interference.

- *Decentralized networks with simple receivers* [Scu08a, Scu08b, Pan08, Les06]. In an ad-hoc network consisting of a number of noncooperative wideband links sharing the same physical resources, signalling between transmitters (for synchronization purposes) and between transmitters and unintended receivers (to facilitate multi-user decoding) is hampered by interference, propagation delay, and the dynamics of the network topology. In the absence of a common clock reference, a practical choice is to employ OFDMA with independent coding across subcarriers and perform single-user decoding in each of them.

- *Networks with non-stationary interference* [Big07b]. Consider a cellular system with users logging in and out independently. The statistics of the user session duration result in an stochastic process counting the number of interferers from a neighboring cell either i) unable to be tracked, or ii) requiring neighbor discovery mechanisms and adaptive multiuser decoding techniques of unaffordable complexity. In this scenario, the computational burden of the receivers can be alleviated by treating interference as noise.
- *The throughput scaling law of multihop wireless networks* [Gup00, Xie04]. The throughput analysis of asymptotically large multihop wireless networks requires the use of mathematically amenable decoding strategies at the receivers. To that end, the assumption of totally asynchronous transmissions in conjunction with single-user decoding (of generally Gaussian-distributed codes) give rise to closed-form results.

Motivated by the operational and practical constraints of these scenarios, we focus on the 2×2 IC in the limited setup where the receivers are restricted to treat interference as noise (we shall refer to this strategy as single-user detection) and transmission is totally asynchronous (in the sense of [Hui85]). The lack of frame synchronism has been studied in the discrete multiple-access channel, and diverse results are obtained depending on the degree of asynchronism [Hui85, Cov81b] and the memory of the channel [Ver89]. To the best of our knowledge, its impact on the achievable rates of the interference channel has not been addressed previously. Similarly, the rate penalty for precluding multiuser decoding strategies may also depend on the nature of the channel as, surprisingly, for sufficiently small interference single-user detection achieves the sum capacity of the frame-synchronous Gaussian IC [Sha07].

We first consider the discrete alphabet case in Section 3.2, where we define the channel model and show that it is information unstable [Pin64]. Consequently, the capacity region does not admit a single-letter characterization, and, rather, is given using an Information Spectrum approach [Ver94]. This same approach was used in [SB06] to characterize the capacity region of the frame-synchronous discrete IC. Pursuing analytical results, we provide an achievable rate region and an outer bound to the capacity region using single-letter expressions. The single-letter inner bound is achieved by stationary inputs with i.i.d. letters and, given its simplicity, we subsequently focus on this achievable rate region throughout the rest of the chapter.

Next, we move on to the Gaussian IC (GIC). It is a common practice to use Gaussian-distributed codes in the frame-synchronous GIC, the main reasons being that i) they allow for closed-form characterizations of achievable rate regions, and ii) they achieve capacity in the extreme cases when interference is strong [Han81, Sat81] or absent. Whether the optimality of Gaussian-distributed codes can be extrapolated to all the ranges of interference intensity or to the scope of this work is not known. Interestingly, [Che93] showed that Gaussian inputs fall short of achieving the capacity region of the GIC when it is characterized using a limiting expression. However, that tells us little about the possible optimality of Gaussian inputs in single-letter characterizations of the capacity region (their gap to capacity is known to be at most 1 bit/ch. use [Etk07]), provided that they exist. Again, when it comes to frame asynchronism and single-user decoding, nothing is known.

In studying whether stationary i.i.d. Gaussian inputs maximize the achievable rates without frame synchronism and single-user decoders, Section 3.3 defines optimality of input distributions with respect to the achievable rates. Additionally, the Gram-Charlier expansion of mutual information is introduced. Making use of this tool, we show in Section 3.4 that Gaussian-distributed codes do not always achieve all the points of the achievable rate region. Additionally, simple sufficient conditions for non-optimality of Gaussian-distributed codes are derived that only depend on the coupling coefficients of the channel and the transmit power constraints. For the symmetric GIC (equal coupling coefficients, equal transmit power constraints) they reduce to exceeding a threshold transmit power. In other words, Gaussian codewords are not optimal when the channel is interference-limited.

This result is intuitive in the following sense: only when the distribution of the interference-plus-noise term is dominated by the distribution of interference the optimization of each input distribution may impact on mutual information and significant gains with respect to Gaussian-distributed codes can be expected. Some non-Gaussian-distributed codes are numerically shown to outperform the Gaussian achievable rates in accordance with the analytical results of Section 3.5, hence constructively proving them. Finally, Section 3.6 concludes the chapter.

3.1.1 Summary of contributions

The contributions of this chapter are the following:

- The capacity region of the totally asynchronous interference channel with single-user receivers is characterized using an Information Spectrum approach, but practical single-letter inner and outer bounds are also provided.
- A finite series approximation analysis of mutual information subsumes the impact of input statistics on the achievable rates through a quadratic form around the Gaussian achievable rates.
- The statistical analysis of mutual information reveals analytic conditions under which Gaussian-distributed codes are not optimal, and practical alternative codes are shown to outperform the Gaussian achievable rates in excellent accordance with the derived expressions.

3.2 The Capacity Region

Let us denote by $\mathcal{X}_1$ and $\mathcal{X}_2$ the input alphabets of the two senders, and by $\mathcal{Y}_1$ and $\mathcal{Y}_2$ the output alphabets of the two destinations. The 2×2 frame-asynchronous discrete memoryless IC (DMIC) consists of two conditional distributions $\{P_{Y_k|X_1X_2} : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}_k\}_{k=1,2}$ that describe the underlying channels, and two collections of distributions $\{P_{D_1,n}, P_{D_2,n}\}_{n=1}^{\infty}$ both defined on $\{0, 1, \dots, n-1\}_{n=1}^{\infty}$ that describe the degree of asynchronism. We say that the channel is totally asynchronous [Hui85] when $P_{D_1,n}, P_{D_2,n}$ are uniform for all n. In contrast, a frame-synchronous IC is characterized by $P_{D_1,n}(0) = P_{D_2,n}(0) = 1$ for all n.

A code for the 2×2 frame-asynchronous DMIC consists of two encoding functions

$$X_k^n \ : \ \{1, \dots, 2^{nR_k}\} \to \mathcal{X}_k^n, \quad k = 1, 2, \tag{3.1}$$

and two decoding functions

$$\hat{m}_k \ : \ \mathcal{Y}_k^n \to \{1, \dots, 2^{nR_k}\}, \quad k = 1, 2. \tag{3.2}$$

Sender 1 draws a message M_1 uniformly from $\{1, \dots, 2^{nR_1}\}$ and sends the corresponding codeword X_1^n, of length n, over the channel. We assume without loss of generality[1] that receiver 1 is frame-synchronized with sender 1, and thus Y_1^n is a sufficient statistic for the message M_1. Since receiver 1 is unaware of the presence of an interferer, the random delay D_1 (with distribution $P_{D_1,n}$) experienced by the codewords of sender 2 is unknown. Similar arguments apply for sender 2 and receiver 2.

Thus, by treating the transmission of the interferer as noise, the transmission of each transmitter faces a channel whose conditional distribution depends on the delay of the interferer. That is, the channel realization from transmitter 1 and receiver 1 is determined by the random variable D_1, which is chosen according to $P_{D_1,n}$ at the beginning of transmission and then held fixed, i.e.,

$$P_{Y_1^n|X_1^n,D_1}(y_1^n|x_1^n, d_1) = \sum_{x_2^n \in \mathcal{X}_2^n} P_{X^n_{2,n-d_1+1}}(x_{2,1}^{d_1}) P_{X_{2,1}^{n-d_1}}(x_{2,d_1+1}^n) \prod_{i=1}^{n} P_{Y_1|X_1X_2}(y_{1,i}|x_{1,i}x_{2,i}). \tag{3.3}$$

We shall use W_1^n to denote the channel from sender 1 to receiver 1 without particularizing on the delay instance,

$$P_{W_1^n}(y_1^n|x_1^n) = \sum_{d_1=0}^{n-1} P_{D_1,n}(d_1) P_{Y_1^n|X_1^n,D_1}(y_1^n|x_1^n, d_1) = \sum_{d_1=0}^{n-1} \frac{1}{n} P_{Y_1^n|X_1^n,D_1}(y_1^n|x_1^n, d_1). \tag{3.4}$$

Expression (3.3) reflects the fact that each received frame of receiver 1 depends on two codewords of user 2. Since each of them corresponds to an independent draw of the message M_2, they are independent as well. As nothing is imposed on the distribution of the interference, the channel (3.3) may have memory in general. A related model to (3.3) is the composite channel [Eff08], or

[1] Synchronization can be achieved via the use of periodic preamble sequences at negligible rate penalty [Hui85].

averaged channel [Ahl68], where each component channel (for every instance of D_1) is assumed stationary and ergodic. The composite channel is information unstable [Pin64], so hence is the frame-asynchronous DMIC with single-user decoders. In order to characterize the capacity of this channel, we can therefore treat each link as a general single-user channel and apply the capacity formula by Verdú and Han [Ver94].

Theorem 3.1 *The capacity region of the totally-asynchronous DMIC with single-user receivers is given by*

$$\mathcal{C} = \bigcup_{P_{\mathbf{X}_1}, P_{\mathbf{X}_2}} \big\{R_k \leq \sup\{\alpha_k : F_{\mathbf{X}_k}(\alpha_k) = 0\},\ k = 1, 2\big\}, \tag{3.5}$$

where $P_{\mathbf{X}_k} = \{P_{X_k^n}\}_{n=1}^{\infty}$ *is a sequence of finite-dimensional distributions , and* $F_{\mathbf{X}_k}$ *is the limit of the cumulative distribution function of the normalized information density of the* k*-th link, i.e.,*

$$F_{\mathbf{X}_k}(\alpha_k) = \lim_{n\to\infty} P_{X_k^n W_k^n} \left\{ \frac{1}{n} i_{X_k^n W_k^n}(X_k^n; Y_k^n) \leq \alpha_k \right\}, \tag{3.6}$$

where the information density amounts to

$$i_{X_k^n W_k^n}(x_k^n; y_k^n) = \log \frac{P_{W_k^n}(y_k^n|x_k^n)}{P_{Y_k^n}(y_k^n)} = \log \frac{\sum_{d_k=0}^{n-1} P_{D_k,n}(d_k) P_{Y_k^n|X_k^n,D_k}(y_k^n|x_k^n, d_k)}{\sum_{d_k=0}^{n-1} P_{D_k,n}(d_k) P_{Y_k^n|D_k}(y_k^n|d_k)} \tag{3.7}$$

$$= \log \frac{\sum_{d_k=0}^{n-1} P_{Y_k^n|X_k^n,D_k}(y_k^n|x_k^n, d_k)}{\sum_{d_k=0}^{n-1} P_{Y_k^n|D_k}(y_k^n|d_k)}. \tag{3.8}$$

Proof. The proof is a natural extension of the capacity formula proved in [Ver94] for any arbitrary single-user channel. □

Given the complex form of the expressions in Theorem 3.1, we aim at deriving simpler expressions upper and lower bounding the capacity region. To that end, we start by noticing that the sequences of distributions $P_{\mathbf{X}_1}$ and $P_{\mathbf{X}_2}$ in (3.5) can be interpreted as the distributions generating the codebooks of senders 1 and 2, respectively [Eff08]. As these distributions can be arbitrary in the union in (3.5), an inner bound is obtained when they are constrained to be i.i.d..

Lemma 3.1 *The achievable rate region*

$$\mathcal{R} = \bigcup_{P_{X_1}, P_{X_2}} \{R_1 \leq I(X_1; Y_1),\ R_2 \leq I(X_2; Y_2)\} \tag{3.9}$$

is an inner bound of the capacity region; $\mathcal{R} \subseteq \mathcal{C}$.

Proof. An inner bound is obtained when we restrict the union in (3.5) to input distributions of the form $P_{X_k^n}(x^n) = \prod_{i=1}^n P_{X_k}(x_{k,i})$. In this case, since the interference seen by receiver 1 is i.i.d. irrespective of the realization of the random variable D_1, the channel between sender 1 and receiver 1 becomes memoryless with conditional distribution

$$P_{Y_1|X_1}(y_1|x_1) = \sum_{x_2 \in \mathcal{X}_2} P_{X_2}(x_2) P_{Y_1|X_1X_2}(y_1|x_1, x_2), \tag{3.10}$$

and the normalized information density $i_{X_1^n W_1^n}(X_1^n; Y_1^n)/n$ converges to $I(X_1; Y_1)$ by the law of large numbers. Thus, $F_{\mathbf{X}_1}(\alpha_1) = u(\alpha_1 - I(X_1; Y_1))$, where $u(\cdot)$ is the unit step function, and the supremum in (3.5) is hence $I(X_1; Y_1)$. Lemma 3.1 follows by applying similar arguments to the link between sender 2 and receiver 2 and noticing that the achievable rates now depend on the input distributions through P_{X_1} and P_{X_2} only. □

Lemma 3.2 *The region*

$$\mathcal{R}_o = \bigcup_{P_{X_1}, P_{X_2}} \{R_1 \leq I(X_1; Y_1|X_2),\ R_2 \leq I(X_2; Y_2|X_1)\} \tag{3.11}$$

is an outer bound to the capacity region; $\mathcal{C} \subseteq \mathcal{R}_o$.

Proof. The proof follows by noticing that

$$\sup\{\alpha_1 : F_{\mathbf{X}_1}(\alpha_1) = 0\} \overset{(a)}{\leq} \lim \inf_{n\to\infty} \frac{1}{n} I(X_1^n; Y_1^n) \overset{(b)}{\leq} \lim \inf_{n\to\infty} \frac{1}{n} I(X_1^n; Y_1^n|X_2^n) \overset{(c)}{\leq} I(X_1; Y_1|X_2), \tag{3.12}$$

where (a) follows from [Ver94, Thm. 8.(h)], (b) is a consequence of the independence between X_1^n and X_2^n, and (c) holds for some pair of distributions P_{X_1} and P_{X_2} and follows from [Hui85, Thms. 3.3-3.4]. Similar arguments hold for the link between sender 2 and receiver 2, where [Hui85, Thms. 3.3-3.4] guarantees that $\sup\{\alpha_2 : F_{\mathbf{X}_2}(\alpha_2) = 0\} \leq I(X_2; Y_2|X_1)$ where $I(X_2; Y_2|X_1)$ is evaluated using the *same* distributions P_{X_1}, P_{X_2} as in $I(X_1; Y_1|X_2)$ (3.12). □

Although the outer bound on the capacity region is trivial, it is worth pointing out that both $\mathcal{R}$ and $\mathcal{R}_o$ are characterized in terms of the union of regions, without any convex hull operation. Intuitively, the lack of frame synchronism precludes time-sharing between distributions, as happens in the discrete-multiple access channel [Hui85]. Due to its simplicity and amenability for numerical computation, we subsequently focus on the achievable rate region $\mathcal{R}$ throughout the rest of the chapter.

3.3 The Gaussian IC

Consider the 2×2 standard-form GIC [Car78],

$$Y_1 = X_1 + c_{21} X_2 + Z_1 \tag{3.13}$$

$$Y_2 = X_2 + c_{12} X_1 + Z_2, \tag{3.14}$$

where $Z_k \sim \mathcal{N}(0,1)$ $k = 1,2$, and the codewords X_1 and X_2 are independent and independent of the noise samples Z_1 and Z_2. The input codewords satisfy the transmit power constraint $\mathbb{E}\{X_k^2\} \leq P_k$, $k = 1,2$. By fully symmetric GIC we denote a GIC where $c_{12} = c_{21} \triangleq c$ and $P_1 = P_2 \triangleq P$. In the totally asynchronous setup with single-user decoders, to maximize the achievable rate region $\mathcal{R}$, the senders must optimize the distribution of their codewords, described by the pdf's f_{X_1} and f_{X_2}.

3.3.1 Definition of optimality

Definition 3.1 *The input distributions f_{X_1} and f_{X_2} are α-optimal, $\alpha \in [0, \pi/2]$, if they achieve the rate pair of the boundary of $\mathcal{R}$ that intersects the line $R_2 = \tan(\alpha) R_1$. Denote such pair by $(R_1^\star(\alpha), R_2^\star(\alpha))$.*

It is not difficult to show that any pair of α-optimal distributions is a solution to the optimization problem

$$\underset{f_{X_1}, f_{X_2}}{\mathsf{maximize}} \quad \min\left\{\frac{I(X_1;Y_1)}{\cos(\alpha)}, \frac{I(X_2;Y_2)}{\sin(\alpha)}\right\} \tag{3.15}$$

$$\mathsf{subject\ to} \quad f_{X_k}(x) \geq 0 \ \forall x \in \mathbb{R},\ k = 1,2 \tag{3.16}$$

$$\int f_{X_k}(x) dx = 1,\ k = 1,2 \tag{3.17}$$

$$\int x^2 f_{X_k}(x) dx \leq P_k,\ k = 1,2. \tag{3.18}$$

The problem (3.15)-(3.18) is rather involved because of the intricate dependence of the mutual information on f_{X_1} and f_{X_2}. Consider for instance the mutual information of the first link,

$$I(X_1;Y_1) = I(X_1; X_1 + c_{21}X_2 + Z_1) = \iint f_{X_1Y_1}(x,y) \log \frac{f_{Y_1|X_1}(y|x)}{f_{Y_1}(y)} dxdy \tag{3.19}$$

$$= \iint f_{X_1}(x) g(y-x) \log \frac{g(y-x)}{\int f_{X_1}(z) g(y-z) dz} dxdy, \tag{3.20}$$

where $g \equiv f_{c_{21}X_2+Z_1}$ is

$$g(x) = f_{c_{21}X_2}(x) * f_{Z_1}(x) = \frac{1}{|c_{21}|} \int f_{X_2}(z/c_{21}) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-z)^2} dz. \tag{3.21}$$

A similar expression holds for the mutual information of the second link.

Instead of finding a pair of α-optimal distributions, we are interested in knowing whether the common practice of using Gaussian-distributed codewords for the GIC is always an optimal choice in our setup. The achievable rate region of Gaussian-distributed codes is given by[2]

$$\mathcal{R}^G = \bigcup_{\substack{0 \leq p_k \leq P_k \\ k=1,2}} \left\{(R_1, R_2) : 0 \leq R_1 \leq \frac{1}{2}\log\left(1 + \frac{p_1}{1 + c_{21}^2 p_2}\right), 0 \leq R_2 \leq \frac{1}{2}\log\left(1 + \frac{p_2}{1 + c_{12}^2 p_1}\right)\right\}. \tag{3.22}$$

[2]Unless the logarithm basis is indicated, it can be chosen arbitrarily as long as both sides of the equations have the same units.

Gaussian-distributed codes are optimal, i.e. $\mathcal{R}^G = \mathcal{R}$, if and only if a pair of Gaussian distributions f_{X_1}, f_{X_2} is α-optimal $\forall\alpha \in [0, \pi/2]$. Let us denote by $(p_1(\alpha), p_2(\alpha))$ the power allocation that achieves the rate pair of the boundary of $\mathcal{R}^G$ that intersects the line $R_2 = \tan(\alpha)R_1$. The region $\mathcal{R}^G$ can hence be rephrased as

$$\mathcal{R}^G = \bigcup_{0\leq\alpha\leq\pi/2} \left\{(R_1, R_2) : 0 \leq R_1 \leq \frac{1}{2}\log\Big(1 + \frac{p_1(\alpha)}{1 + c_{21}^2 p_2(\alpha)}\Big), 0 \leq R_2 \leq \frac{1}{2}\log\Big(1 + \frac{p_2(\alpha)}{1 + c_{12}^2 p_1(\alpha)}\Big)\right\}. \tag{3.23}$$

To further explore the optimality of Gaussian-distributed codes, consider the following result.

Lemma 3.3 *For any fixed $\alpha \in [0, \pi/2]$, if $X_1 \sim \mathcal{N}(0, p_1)$ with $0 \leq p_1 \leq P_1$, the optimal distribution of X_2 that maximizes the achievable rates along the line $R_2 = \tan(\alpha)R_1$ is $X_2 \sim \mathcal{N}(0, p_2)$ for some $0 \leq p_2 \leq P_2$. An analogous result follows if the user indexes are swapped.*

Proof. Once X_1 follows a Gaussian distribution, the problem at hand is to determine the distribution of X_2 solution to

$$\underset{X_2:\mathbb{E}\{X_2^2\}\leq P_2}{\text{maximize}} \ \min\left\{\frac{I(X_1;Y_1)}{\cos(\alpha)}, \frac{I(X_2;Y_2)}{\sin(\alpha)}\right\}. \tag{3.24}$$

While $I(X_2; Y_2)$, equivalent to the mutual information in an AWGN channel with noise power $(1 + c_{12}^2 p_1)$, is maximized by a Gaussian X_2, the mutual information of the first link requires more attention. It can be rephrased as

$$\begin{aligned} I(X_1; Y_1) &= h(c_{21}X_2 + X_1 + Z_1) - h(c_{21}X_2 + Z_1) && (3.25)\\ &= h(X_2 + (X_1 + Z_1)/c_{21}) - h(X_2 + Z_1/c_{21}) && (3.26)\\ &\equiv h(X_2 + N_1) - h(X_2 + N_2), && (3.27)\end{aligned}$$

where $N_1 \sim \mathcal{N}(0, (1 + p_1)/c_{21}^2)$, $N_2 \sim \mathcal{N}(0, 1/c_{21}^2)$, and $h(\cdot)$ stands for differential entropy. The maximization of (3.27) with respect to the distribution of X_2 subject to a power constraint P_2 falls within the class of problems addressed by [Liu07, Thm. 1], where it was shown that the optimal X_2 is Gaussian, too. Since a Gaussian-distributed X_2 maximizes simultaneously both mutual informations, it only remains to optimize its power p_2, $0 \leq p_2 \leq P_2$, so that (3.24) is maximized. □

According to Lemma 3.3, Gaussian-distributed codes behave as a (possibly local) extremum in the maximization of the achievable rates. They are also a greedy strategy: although this input distribution maximizes mutual information if interference is absent, it also gives rise to the worst additive interference [Iha78, Dig01]. In other words, Gaussian-distributed codes maximize $h(Y_k)$ and $h(Y_k|X_k)$ simultaneously for $k = 1, 2$, but this does not necessarily imply that they maximize $I(X_k; Y_k)$ as well. Since direct construction of α-optimal distributions (3.15)-(3.18) seems overwhelming, we shall adopt a completely different approach for showing non-optimality of Gaussian-distributed codes. It is based on the relation between mutual information and the shape of the pdf of the codewords, as described by their cumulants.

3.3.2 Finite expansion analysis of mutual information

Denote by Ω_X the support set of a zero-mean continuous random variable X whose pdf is f_X. Let $\varphi_X(\omega)$ denote its characteristic function,

$$\varphi_X(\omega) = \mathbb{E}\{e^{j\omega X}\}. \tag{3.28}$$

The cumulants [Pap84] $\{\kappa_i(X)\}_{i=1}^{+\infty}$ of X are defined as the coefficients of the McLaurin series of the natural logarithm of the characteristic function,

$$\log_e(\varphi_X(\omega)) = \sum_{i=1}^{+\infty} \kappa_i(X) \frac{(j\omega)^i}{i!}. \tag{3.29}$$

The cumulants of the zero-mean random variable X can be related to its (central) moments, and have some interesting properties concerning the shape of f_X, as listed below [Nik93].

- First-order cumulant: $\kappa_1(X) = \mathbb{E}\{X\} = 0$.
- Second-order cumulant (variance): $\kappa_2(X) = \mathbb{E}\{X^2\} \equiv \sigma_X^2$.
- Third-order cumulant (skewness): $\kappa_3(X) = \mathbb{E}\{X^3\}$.
- Fourth-order cumulant (kurtosis): $\kappa_4(X) = \mathbb{E}\{X^4\} - 3\mathbb{E}\{X^2\}^2$.
- Symmetry: $f_X(x) = f_X(-x) \Rightarrow \kappa_{2i-1}(X) = 0 \ \forall i \geq 1$.
- Independence: X_1, X_2 independent $\Rightarrow \kappa_i(X_1 + X_2) = \kappa_i(X_1) + \kappa_i(X_2) \ \forall i$.
- Scaling: $\kappa_i(aX) = a^i \kappa_i(X) \ \forall a \in \mathbb{R}$.
- Cumulants of the Gaussian distribution: $X \sim \mathcal{N}(0, P) \Rightarrow \kappa_2(X) = P, \ \kappa_i(X) = 0 \ \forall i \neq 2$.

Skewness measures the lack of symmetry of a distribution, whereas kurtosis can be considered a measure of the non-Gaussianity (or peakedness) of X. Kurtosis is zero for a Gaussian random variable, it is typically positive for distributions with heavy tails and a peak at zero, and negative for flatter-than-Gaussian densities with lighter tails. While explicit distributions with infinite positive kurtosis exist (e.g. a limiting version of Pearson's type VII distribution[3] [Pea16]), kurtosis is fundamentally lower bounded as

$$\kappa_4(X_k) = \mathbb{E}\{X_k^4\} - 3\sigma_X^4 \geq \mathbb{E}\{X_k^2\}^2 - 3\sigma_X^4 = -2\sigma_X^4, \tag{3.30}$$

thanks to Jensen's inequality and the fact that X is zero-mean[4].

[3] The limiting pdf of a Pearson's type VII random variable X with zero mean, variance σ^2, zero skewness, and infinite kurtosis is $f_X(x) = \frac{3}{\sigma}(2 + (x/\sigma)^2)^{-5/2}$.

[4] The lower bound (3.30) is achievable by the distribution $f_X(x) = \frac{1}{2}(\delta(x - \sigma_X) + \delta(x + \sigma_X))$.

Finally, the principal tool that will be used subsequently to analyze the optimality of Gaussian-distributed codes is Gram-Charlier expansion [Nik93], which allows us to approximate the differential entropy $h(X)$ of the near-Gaussian random variable X around the entropy of a Gaussian random variable with the same variance as X. If $\frac{\kappa_i^2(X)}{\sigma_X^{2i}} \ll 1$ for $i > 2$ then

$$h(X) \approx \frac{1}{2}\log(2\pi e \sigma_X^2) - \left(\frac{1}{12}\frac{\kappa_3^2(X)}{\sigma_X^6} + \frac{1}{48}\frac{\kappa_4^2(X)}{\sigma_X^8}\right)\log(e), \tag{3.31}$$

is a fourth-order entropy approximation for X [Hel95]. Expression (3.31) will bridge the gap between mutual information and cumulants in Section 3.4. Before that, let us adopt without loss of generality the zero-mean assumption on both X_1 and X_2.

3.4 On the Optimality of Gaussian-Distributed Codes

Based on the Gram-Charlier expansion of entropy, mutual information can be approximated in terms of the cumulants of the input distributions of order four and below. This will allow us to obtain (fourth-order) analytical results on the optimality of Gaussian-distributed codes.

Lemma 3.4 *Assume that both X_1 and X_2 are zero-mean and $\mathbb{E}\{X_k^2\} = p_k$, $k = 1, 2$. A fourth-order expansion approximation of mutual information is*

$$I(X_1;Y_1) \approx \frac{1}{2}\log\left(1 + \frac{p_1}{1 + c_{21}^2 p_2}\right) + \boldsymbol{\kappa}^T \mathbf{A} \boldsymbol{\kappa} \tag{3.32}$$

$$I(X_2;Y_2) \approx \frac{1}{2}\log\left(1 + \frac{p_2}{1 + c_{12}^2 p_1}\right) + \boldsymbol{\kappa}^T \mathbf{B} \boldsymbol{\kappa}, \tag{3.33}$$

where $\boldsymbol{\kappa} \triangleq [\kappa_3(X_1)\ \kappa_4(X_1)\ \kappa_3(X_2)\ \kappa_4(X_2)]^T$,

$$[\mathbf{A}]_{1,1} = -\frac{\log(e)}{12\sigma_{Y_1}^6} \qquad [\mathbf{B}]_{1,1} = \frac{\log(e)c_{12}^6}{12}(\sigma_{Y_2|X_2}^{-6} - \sigma_{Y_2}^{-6}) \tag{3.34}$$

$$[\mathbf{A}]_{1,3} = -\frac{\log(e)c_{21}^3}{6\sigma_{Y_1}^6} \qquad [\mathbf{B}]_{1,3} = -\frac{\log(e)c_{12}^3}{6\sigma_{Y_2}^6} \tag{3.35}$$

$$[\mathbf{A}]_{2,2} = -\frac{\log(e)}{48\sigma_{Y_1}^8} \qquad [\mathbf{B}]_{2,2} = \frac{\log(e)c_{12}^8}{48}(\sigma_{Y_2|X_2}^{-8} - \sigma_{Y_2}^{-8}) \tag{3.36}$$

$$[\mathbf{A}]_{2,4} = -\frac{\log(e)c_{21}^4}{24\sigma_{Y_1}^8} \qquad [\mathbf{B}]_{2,4} = -\frac{\log(e)c_{12}^4}{24\sigma_{Y_2}^8} \tag{3.37}$$

$$[\mathbf{A}]_{3,3} = \frac{\log(e)c_{21}^6}{12}(\sigma_{Y_1|X_1}^{-6} - \sigma_{Y_1}^{-6}) \qquad [\mathbf{B}]_{3,3} = -\frac{\log(e)}{12\sigma_{Y_2}^6} \tag{3.38}$$

$$[\mathbf{A}]_{4,4} = \frac{\log(e)c_{21}^8}{48}(\sigma_{Y_1|X_1}^{-8} - \sigma_{Y_1}^{-8}) \qquad [\mathbf{B}]_{4,4} = -\frac{\log(e)}{48\sigma_{Y_2}^8}, \tag{3.39}$$

the rest of entries of $\mathbf{A}$ *and* $\mathbf{B}$ *are zero, and*

$$\sigma_{Y_k}^2 = 1 + c_{jk}^2 p_j + p_k,\ \sigma_{Y_k|X_k}^2 = 1 + c_{jk}^2 p_j \tag{3.40}$$

for $k, j = 1, 2,\ j \neq k$.

Proof. Consider $I(X_1;Y_1)$ (the analysis of $I(X_2;Y_2)$ is analogous). By using

$$\sigma_{Y_1}^2 = 1 + c_{21}^2 P_2 + P_1 \tag{3.41}$$

$$\sigma_{Y_1|X_1}^2 = 1 + c_{21}^2 P_2 \tag{3.42}$$

the Gram-Charlier expansion (3.31) can be applied to each entropy term involved in mutual information,

$$\begin{aligned}
I(X_1;Y_1) &= h(Y_1) - h(Y_1|X_1) = h(X_1 + c_{21}X_2 + Z_1) - h(c_{21}X_2 + Z_1) && (3.43)\\
&\overset{(a)}{\approx} \frac{1}{2}\log(2\pi e \sigma_{Y_1}^2) - \frac{\log(e)}{12}\frac{(\kappa_3(X_1)+c_{21}^3\kappa_3(X_2))^2}{\sigma_{Y_1}^6} - \frac{\log(e)}{48}\frac{(\kappa_4(X_1)+c_{21}^4\kappa_4(X_2))^2}{\sigma_{Y_1}^8} && (3.44)\\
&- \frac{1}{2}\log(2\pi e \sigma_{Y_1|X_1}^2) + \frac{\log(e)}{12}\frac{c_{21}^6\kappa_3^2(X_2)}{\sigma_{Y_1|X_1}^6} + \frac{\log(e)}{48}\frac{c_{21}^8\kappa_4^2(X_2)}{\sigma_{Y_1|X_1}^8} && (3.45)\\
&= \frac{1}{2}\log\Big(1 + \frac{p_1}{1 + c_{21}^2 p_2}\Big) - \frac{\log(e)}{12\sigma_{Y_1}^6}\kappa_3(X_1)^2 - \frac{\log(e)}{48\sigma_{Y_1}^8}\kappa_4(X_1)^2 && (3.46)\\
&+ \frac{\log(e)c_{21}^6}{12}(\sigma_{Y_1|X_1}^{-6} - \sigma_{Y_1}^{-6})\kappa_3(X_2)^2 + \frac{\log(e)c_{21}^8}{48}(\sigma_{Y_1|X_1}^{-8} - \sigma_{Y_1}^{-8})\kappa_4(X_2)^2 && (3.47)\\
&- \frac{\log(e)c_{21}^3}{6\sigma_{Y_1}^6}\kappa_3(X_1)\kappa_3(X_2) - \frac{\log(e)c_{21}^4}{24\sigma_{Y_1}^8}\kappa_4(X_1)\kappa_4(X_2) && (3.48)\\
&\overset{(b)}{=} \frac{1}{2}\log\Big(1 + \frac{p_1}{1 + c_{21}^2 p_2}\Big) + \boldsymbol{\kappa}^T\mathbf{A}\boldsymbol{\kappa}. && (3.49)
\end{aligned}$$

Expression (a) is obtained when (3.31) is applied in conjunction with the independence property, the scaling property, and the fact that Z_1 is Gaussian, while (b) follows from the definition of $\mathbf{A}$ (3.34)-(3.39). □

The matrices $\mathbf{A}$ and $\mathbf{B}$ are upper-triangular, and hence their eigenvalues lie on their diagonal entries (3.34), (3.36), (3.38) and (3.39). Since $\mathbf{A}$ and $\mathbf{B}$ are not negative definite (two eigenvalues are negative and the other two are non-negative), the possibility of finding a vector of cumulants $\boldsymbol{\kappa}$ inducing a pair of distributions outperforming Gaussian-distributed codes is not precluded.

Lemma 3.5 *Gaussian-distributed codes are not α-optimal for some fixed $\alpha \in [0, \pi/2]$ if the problem*

$$\begin{aligned}
&\textsf{find} && \boldsymbol{\kappa} && (3.50)\\
&\textsf{subject to} && \min\{\boldsymbol{\kappa}^T\mathbf{A}(\alpha)\boldsymbol{\kappa}, \boldsymbol{\kappa}^T\mathbf{B}(\alpha)\boldsymbol{\kappa}\} > 0 && (3.51)\\
& && [\boldsymbol{\kappa}]_2 \geq -2p_1^2(\alpha) && (3.52)\\
& && [\boldsymbol{\kappa}]_4 \geq -2p_2^2(\alpha) && (3.53)
\end{aligned}$$

is feasible, where $\mathbf{A}(\alpha), \mathbf{B}(\alpha)$ are equivalent to $\mathbf{A}, \mathbf{B}$ in Lemma 3.4 but with $\mathbb{E}\{X_k^2\} = p_k(\alpha)$, $k = 1, 2$, in (3.40).

Proof. Gaussian distributed codes are not α-optimal if another pair of distributions achieve a rate pair outside $\mathcal{R}^G$ in the direction given by the line $R_2 = \tan(\alpha)R_1$. That amounts to finding an input distribution that yields an objective value (3.15)-(3.18) larger than $\min\left\{\frac{R_1^G(\alpha)}{\cos(\alpha)}, \frac{R_2^G(\alpha)}{\sin(\alpha)}\right\}$, where

$$(R_1^G(\alpha), R_2^G(\alpha)) \triangleq \left(\frac{1}{2}\log\Big(1+\frac{p_1(\alpha)}{1+c_{21}^2 p_2(\alpha)}\Big), \frac{1}{2}\log\Big(1+\frac{p_2(\alpha)}{1+c_{12}^2 p_1(\alpha)}\Big)\right). \tag{3.54}$$

Since

$$\min\left\{\frac{I(X_1;Y_1)}{\cos(\alpha)}, \frac{I(X_2;Y_2)}{\sin(\alpha)}\right\} \overset{(a)}{\approx} \min\left\{\frac{R_1^G(\alpha)+\boldsymbol{\kappa}^T\mathbf{A}\boldsymbol{\kappa}}{\cos(\alpha)}, \frac{R_2^G(\alpha)+\boldsymbol{\kappa}^T\mathbf{B}\boldsymbol{\kappa}}{\sin(\alpha)}\right\} \tag{3.55}$$

$$\geq \min\left\{\frac{R_1^G(\alpha)}{\cos(\alpha)}, \frac{R_2^G(\alpha)}{\sin(\alpha)}\right\} + \min\left\{\frac{\boldsymbol{\kappa}^T\mathbf{A}\boldsymbol{\kappa}}{\cos(\alpha)}, \frac{\boldsymbol{\kappa}^T\mathbf{B}\boldsymbol{\kappa}}{\sin(\alpha)}\right\}, \tag{3.56}$$

where (a) follows from Lemma 3.4, a sufficient condition is that (3.50)-(3.53) is feasible, where (3.52)-(3.53) account for the fundamental lower bound on the kurtosis (3.30). □

In general, it is difficult to find a vector of cumulants satisfying Lemma 3.5 for a given GIC and α due to i) the lack of general closed-form expressions for $p_1(\alpha)$ and $p_2(\alpha)$ (which are the solution to a non-convex problem), and ii) the fact that neither $\mathbf{A}(\alpha)$ nor $\mathbf{B}(\alpha)$ are positive/negative definite. However, in some particular cases the optimal power allocation $(p_1(\alpha), p_2(\alpha))$ can be found in closed form.

Lemma 3.6 *The power allocation $(p_1(\alpha), p_2(\alpha))$ that achieves the unique rate pair resulting from the intersection of the line $R_2 = \tan(\alpha)R_1$ and the boundary of $\mathcal{R}^G$ for any fixed $\alpha \in [0, \pi/2]$ is given by*

$$(p_1(\alpha), p_2(\alpha)) = \begin{cases} (\{p_1 \geq 0 : \varphi(p_1, P_2; \alpha) = 0\}, P_2) & \text{if } \varphi(P_1, P_2; \alpha) \geq 0 \\ (P_1, \{p_2 \geq 0 : \varphi(P_1, p_2; \alpha) = 0\}) & \text{otherwise,} \end{cases} \tag{3.57}$$

where

$$\varphi(x, y; \alpha) \triangleq \Big(1+\frac{x}{1+c_{21}^2 y}\Big)^{1/\cos(\alpha)} - \Big(1+\frac{y}{1+c_{12}^2 x}\Big)^{1/\sin(\alpha)}. \tag{3.58}$$

Proof. Consider that $\mathbb{E}\{X_1^2\} = p$ and $\mathbb{E}\{X_2^2\} = \beta p$, for some fixed $0 \leq p \leq P_1$ and $0 \leq \beta \leq P_2/p$. Then, since

$$R_1^G = \frac{1}{2}\log\Big(1+\frac{p}{1+c_{21}^2\beta p}\Big)\ ,\ R_2^G = \frac{1}{2}\log\Big(1+\frac{\beta p}{1+c_{12}^2 p}\Big) \tag{3.59}$$

are both increasing in p, which cannot exceed $\min\{P_1, P_2/\beta\}$ to satisfy the transmit power constraints, if follows that at least one of the senders must transmit at its maximum power. If

$$\frac{1}{2\cos(\alpha)}\log\Big(1+\frac{P_1}{1+c_{21}^2 P_2}\Big) \geq \frac{1}{2\sin(\alpha)}\log\Big(1+\frac{P_2}{1+c_{12}^2 P_1}\Big), \tag{3.60}$$

which is equivalent to $\varphi(P_1, P_2; \alpha) \geq 0$, sender two must transmit at power $p_2(\alpha) = P_2$ and sender one must cut-down its transmit power until both sides of (3.60) are equal, which is achieved by $p_1(\alpha) = \{p_1 \geq 0 : \varphi(p_1, P_2; \alpha) = 0\}$. If

$$\frac{1}{2\cos(\alpha)}\log\Big(1+\frac{P_1}{1+c_{21}^2 P_2}\Big) < \frac{1}{2\sin(\alpha)}\log\Big(1+\frac{P_2}{1+c_{12}^2 P_1}\Big), \tag{3.61}$$

a similar result holds. □

Corollary 3.1 *In case $\alpha = \pi/4$, (3.57) reduces to*

$$(p_1(\pi/4), p_2(\pi/4)) = \begin{cases} \Big(\frac{1}{2c_{12}^2}(\sqrt{1+4c_{12}^2(1+c_{21}^2P_2)P_2}-1), P_2\Big) & \text{if } \frac{P_1}{1+c_{21}^2P_2} \geq \frac{P_2}{1+c_{12}^2P_1} \\ \Big(P_1, \frac{1}{2c_{21}^2}(\sqrt{1+4c_{21}^2(1+c_{12}^2P_1)P_1}-1)\Big) & \text{otherwise} \end{cases}, \quad (3.62)$$

which is the power allocation that achieves the maximum symmetric achievable rate of Gaussian-distributed codes, defined as

$$R_{\text{sym}}^G = \max_{(R_1,R_2)\in\mathcal{R}^G} \min\{R_1, R_2\}. \quad (3.63)$$

Proof. It suffices to show that

$$\text{sign}(\varphi(P_1, P_2; \pi/4)) = \text{sign}\Big(\frac{P_1}{1+c_{21}^2P_2} - \frac{P_2}{1+c_{12}^2P_1}\Big), \quad (3.64)$$

and hence $\{p_1 \geq 0 : \varphi(p_1, P_2; \pi/4) = 0\}$ is the unique positive solution to the equation

$$c_{12}^2p_1^2 + p_1 - (1+c_{12}^2P_2)P_2 = 0. \quad (3.65)$$

□

Corollary 3.2 *For a fully symmetrical GIC with $c_{12} = c_{21} \triangleq c$ and $P_1 = P_2 \triangleq P$, (3.62) reduces to*

$$p_1(\pi/4) = p_2(\pi/4) = P. \quad (3.66)$$

Proof. It follows after straightforward manipulation of (3.62). □

Lemma 3.6, Corollary 3.1, and Corollary 3.2 allow us to state the following result.

Theorem 3.2 *Gaussian-distributed codes are not optimal for the totally asynchronous GIC with single-user receivers. For the fully symmetric GIC, R_{sym}^G is outperformed by other input distributions if*

$$P > P_{\text{th}}(c) \triangleq \frac{\sqrt{1+c^{-4}}-1}{1+c^2-\sqrt{1+c^4}}, \quad (3.67)$$

which implies that interference is at least moderate. The maximum symmetric achievable rate for zero-mean non-Gaussian X_1 and X_2 with symmetric pdf's, R_{sym}, is such that

$$R_{\text{sym}} - R_{\text{sym}}^G \propto \kappa_4^2, \quad (3.68)$$

where $\kappa_4 = \kappa_4(X_1) = \kappa_4(X_2)$ is their common kurtosis.

Proof. Non-optimality of Gaussian-distributed codes follows if (3.50)-(3.53) is shown to be feasible for some fixed α, P_1, P_2, c_{21}, and c_{12}. To that end, consider the fully symmetric GIC and $\alpha = \pi/4$. By Corollary 3.2 it follows that

$$\sigma_Y^2 \equiv \sigma_{Y_1}^2 = \sigma_{Y_2}^2 = 1 + (1+c^2)P \quad (3.69)$$

$$\sigma_{Y|X}^2 \equiv \sigma_{Y_1|X_1}^2 = \sigma_{Y_2|X_2}^2 = 1 + c^2P, \quad (3.70)$$

and, according to the symmetry of the channel and the choice of α, we impose that X_1 and X_2 have the same distribution (with skewness κ_3 and kurtosis κ_4) and that their pdf is symmetric around zero. The symmetry property forces $\kappa_3 = 0$ and (3.50)-(3.53) reduces to

$$\underset{\kappa_4 \geq -2P^2}{\text{find}} \quad \kappa_4 \tag{3.71}$$

$$\text{subject to} \quad \frac{\log(e)}{48}\Big(\frac{c^8}{\sigma^8_{Y|X}} - \frac{(1+c^4)^2}{\sigma^8_Y}\Big)\kappa_4^2 > 0, \tag{3.72}$$

which is feasible when

$$\frac{\sigma^2_Y}{\sigma^2_{Y|X}} > \sqrt{1+c^{-4}} \tag{3.73}$$

or, equivalently, $P > P_{\text{th}}(c)$ (3.67). Moderate interference[5] accounts for $\frac{\sqrt{1+2P}-1}{2P} < c^2$ or, equivalently for transmission with power above $P_{\text{mod}}(c) = \frac{1/2-c^2}{c^4}$. Since

$$P_{\text{th}}(c) - P_{\text{mod}}(c) = \frac{c^2 + \sqrt{1+c^4} - 1}{2(1+c^2-\sqrt{1+c^4})c^4} \geq 0, \tag{3.74}$$

a necessary condition for non-optimality of Gaussian codes is that the receivers experience interference of at least moderate strength. Finally, (3.68) follows from (3.49) noticing that $R_{\text{sym}} - R^G_{\text{sym}}$ corresponds to the left hand side of inequality (3.72). □

In essence, Theorem 3.2 shows that Gaussian-distributed codes are not optimal when interference is significant enough (at least moderate). Since $P_{\text{th}}(c)$ is decreasing in c, the more limited the achievable rates are because of interference, the easier to outperform Gaussian-distributed codes at lower transmit powers. As long as the maximum symmetric achievable rate is concerned, kurtosis plays a key role on the performance of non-Gaussian codes (while it also concerns certain practical implementation aspects [Bha06]). Consistently with the fact that Gaussian codes are capacity-achieving in the AWGN, $P_{\text{th}}(c) \to +\infty$ when $c \to 0$.

Interestingly, [Sha07] showed that Gaussian-distributed codes and single-user detection are sum-rate optimal provided that the interference is low enough (noisy interference, as in the terminology of [Sha07]). Our result is consistent with that of [Sha07], which holds under weaker interference than weak interference [Cos85], and rules out the optimality of Gaussian codes and single-user detection for frame-synchronous GICs with stronger interference than moderate interference.

3.5 Numerical Results

We shall present here some numerical results to illustrate Theorem 3.2. Particularly, we shall show the existence of non-Gaussian distributions outperforming Gaussian-distributed codes for the fully symmetric GIC to support the fourth-order analysis of Section 3.3. To that end, let us consider the following non-Gaussian distributions:

[5] A fully symmetric GIC is hampered by moderate interference [Cos85] when $c^2 < 1$ and time-sharing is better than Gaussian-distributed codes with single-user decoders.

- Uniformly-distributed codes: the codewords are drawn uniformly about zero, i.e., $X_k \sim \mathcal{U}(-\sqrt{3p_k}, \sqrt{3p_k})$ for some power $0 \leq p_k \leq P_k$, $k = 1, 2$.

- Ternary-distributed codes: the codewords are drawn from the discrete alphabet $\{-\gamma_k, 0, \gamma_k\}$ according to the probabilities

$$\mathbb{P}(X_k = -\gamma_k) = \mathbb{P}(X_k = \gamma_k) = \frac{p_k}{2\gamma_k^2} \quad (3.75)$$

$$\mathbb{P}(X_k = 0) = 1 - 2\mathbb{P}(X_k = \gamma_k), \quad (3.76)$$

where $0 \leq p_k \leq P_k$ is the transmission power of the k-th sender, $\gamma_k = (3p_k + \kappa_4(X_k)/p_k)^{1/2}$, and $\kappa_4(X_k) \geq -2p_k^2$ is the kurtosis, $k = 1, 2$. The rationale of this choice resides in that (3.75)-(3.76) is the simplest distribution having an arbitrarily large kurtosis, a fact that in view of Theorem 3.2 is desirable in order to increase the achievable symmetric rate whenever near-Gaussianity holds.

Given a pair of input probability distributions, the mutual information is numerically computed by discretizing the integrals in (3.19)-(3.21). Although these distributions are not near-Gaussian (in the sense $\frac{\kappa_i^2}{\sigma_X^{2i}} \ll 1$, $i > 2$), we shall see that the results using the Gram-Charlier expansion hold even in this situation.

Figure 3.1 can be viewed as a constructive proof of Theorem 3.2. The achievable rate regions of Gaussian-, uniformly-, and ternary-distributed codes are plotted for two different channels with $P = 15$. For each direction (described by a different value of $\alpha \in [0, \pi/2]$), the optimum transmit powers for Gaussian-distributed codes $(p_1(\alpha), p_2(\alpha))$ are computed using Lemma 3.6, and plug into the rest of distributions to obtain their achievable rates. For ternary-distributed codes, the kurtosis of each user's distribution is independently optimized, too. While in the first channel $c = 0.1$ (left), transmission is clearly below the threshold $P_{\text{th}}(0.1) \approx 9950$, and none of the proposed non-Gaussian distributions can beat $\mathcal{R}^G$, in the second $c = 1/\sqrt{2}$ (right), transmission power is above the threshold $P_{\text{th}}(1/\sqrt{2}) \approx 3.24$, and achievable rate gains are realized, showing that $\mathcal{R}^G$ falls short of achieving the capacity region.

To see how accurate is the threshold power (3.67) given in Theorem 3.2 using the fourth-order analysis of mutual information, Figure 3.2 (left) plots the achievable symmetric rate of Gaussian-, uniformly-, and ternary-distributed codes as a function of P for $c = \{0.9245, 0.5436\}$, which yield the theoretical values of $P_{\text{th}} = 1$ and $P_{\text{th}} = 10$, respectively. The curves show excellent agreement with (3.67), and Gaussian-distributed codes are outperformed when the conditions of Theorem 3.2 are satisfied. To study this in a wider range of values of c, Figure 3.2 (right) shows the comparison between the theoretical value of the threshold power (3.67) and the actual threshold power of uniformly-distributed codes, which is computed using a bisection method up to a precision of $\pm 2.5\%$. Agreement between both curves is also observed.

Finally, to evaluate the impact of the lack of synchronization and the use of single-user decoders on the capacity of the GIC, we have compared the achievable symmetric rates of

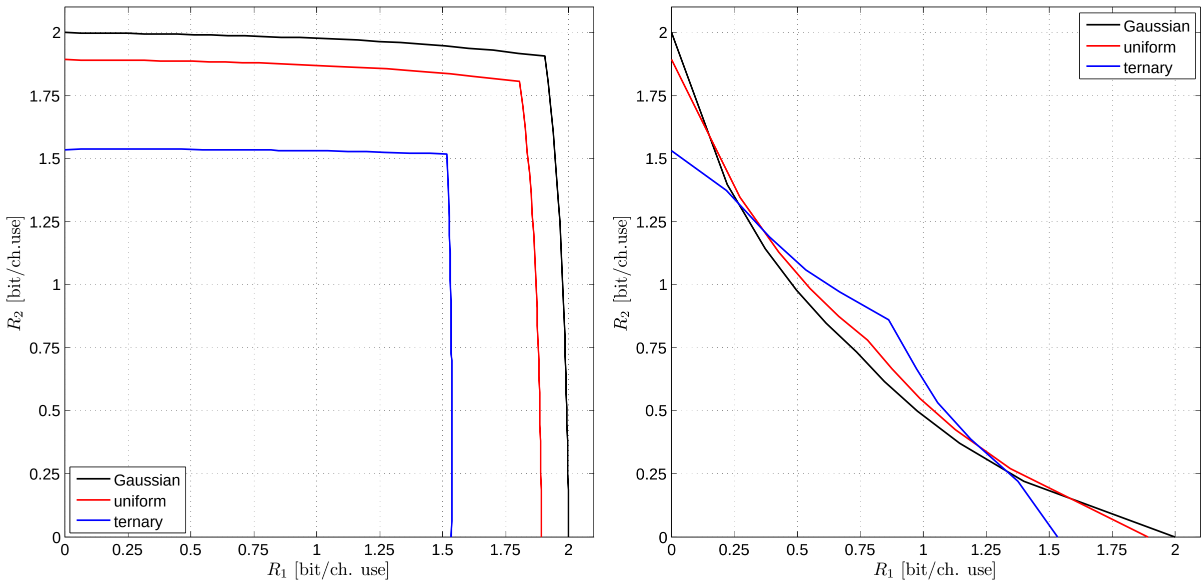

Figure 3.1: Achievable rate regions of Gaussian-, uniformly-, and ternary-distributed codes for a fully symmetric GIC with $P = 15$ and $c = 0.1$ (left) and $c = 1/\sqrt{2}$ (right), which correspond to a signal-to-interference ratio value of 20 dB and 3 dB, respectively.

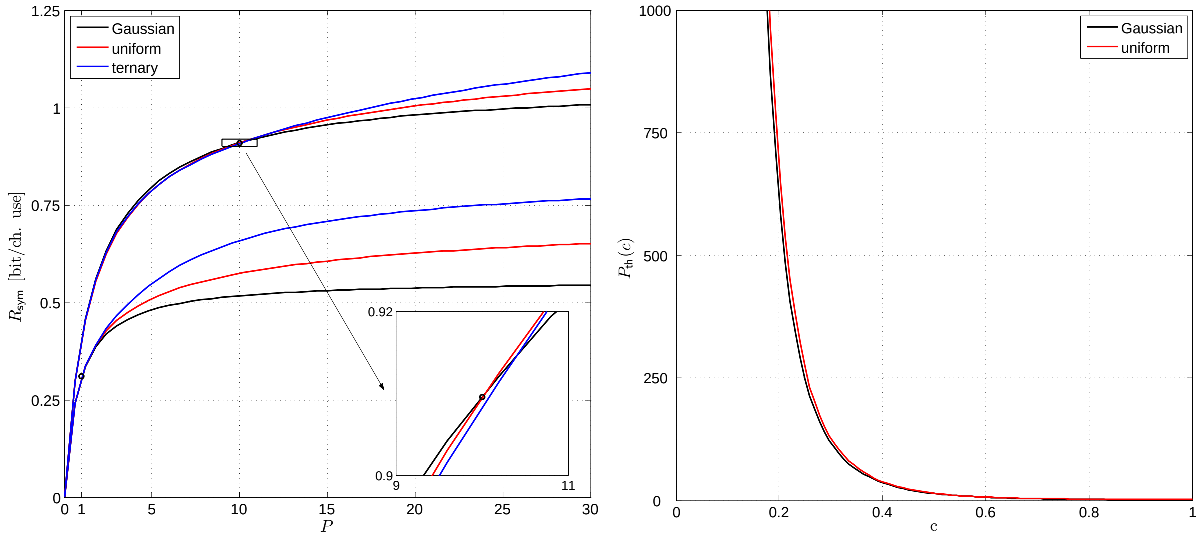

Figure 3.2: Achievable symmetric rate of Gaussian-, uniformly-, and ternary-distributed codes for two different values of c yielding theoretical threshold powers of 1 and 10 (left). Comparison between the theoretical value of $P_{\text{th}}(c)$ (3.63) and the threshold power of uniformly-distributed codes as a function of the coupling coefficient (right).

Gaussian-, uniformly-, and ternary-distributed codes with the following strategies (which clearly require synchronization and, some of them, the use of multi-user decoders):

- Time-sharing: the achievable symmetric rate is independent of c and equals $\frac{1}{4}\log(1+2P)$. Since we are focusing on the fully symmetric GIC, the symmetric rate is half of the sum-rate

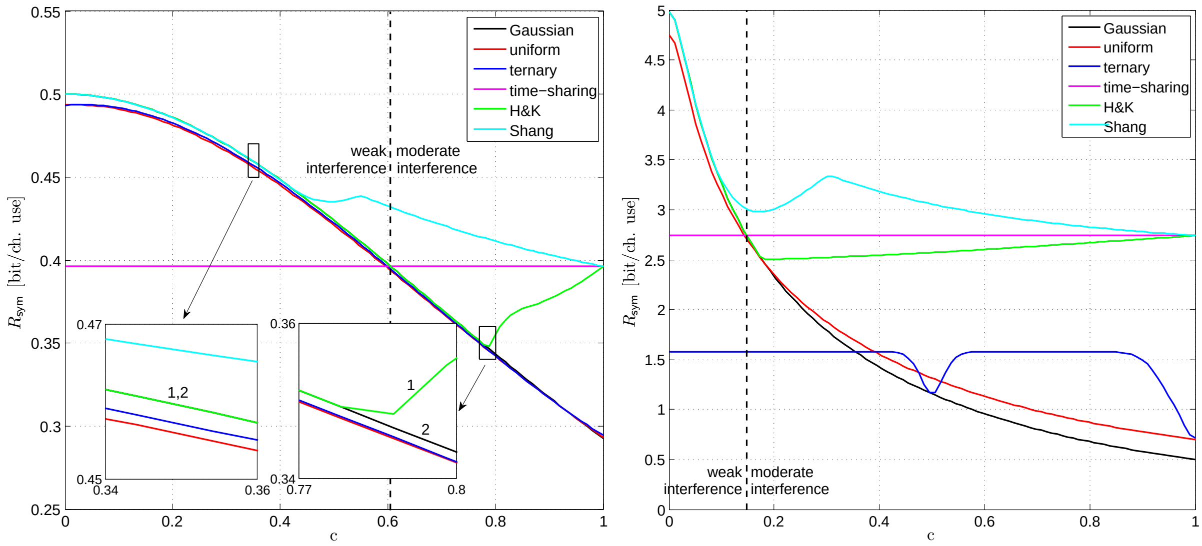

Figure 3.3: Achievable symmetric rates in the low-power regime, $P = 1$ (left), and high-power regime, $P = 1000$ (right), as a function of the coupling coefficient c.

and $\frac{1}{4}\log(1 + 2P)$ is also the symmetric rate of the achievable rate region of Sason [Sas04, Thm. 1].

- Han and Kobayashi (H&K) simplified region: being the most long-standing unbeaten achievable rate region for the GIC, the general region of Han and Kobayashi [Han81, Thm. 3.2] is widely accepted as the most successful approach towards the capacity region. However, the excessive number of random variables in which it is expressed and the cardinality of their alphabets prevents us from its direct computation. Instead, we will focus on the symmetric rate of its simpler subregion [Han81, Thm. 4.1] which tradeoffs computational complexity and achievable rate results.
- Shang's outer bound: one of the tightest known outer bounds of the capacity region of the GIC [Sha07], which improves upon [Kra04, Thm. 1].

Figure 3.3 compares the achievable symmetric rates of the transmission strategies described above as a function of the coupling coefficient c in both the low-power ($P = 1$, shown left) and high-power ($P = 1000$, shown right) regimes. When interference is weak, Gaussian-distributed codes perform undistinguishable to H&K, and are very close to Shang's upper bound. The rest of proposed non-Gaussian distributions, although very close to the Gaussian behavior, cannot improve upon it because for low interference the threshold power is very large (it grows as c^{-4} for small c). Nevertheless, the proposed non-Gaussian distributions, which are based on totally asynchronous transmission, are able to significantly outperform time-sharing (except for ternary-distributed codes in the high-power regime, where the discrete nature of their alphabet makes them reach the upper bound of $\log_2(3)$ bits/ch. use[6]). Therefore, when interference is

[6]Note that the symmetric rate is strictly below $\log_2(3)$ close to those values of c which make some of the ternary symbols of each user overlap at the receiver ($c = \{0.5, 1\}$).

weak the losses associated to the lack of transmission synchronism and the use of single-user decoders are moderate (as long as H&K's performance is taken as a benchmark).

When interference increases up to moderate, time-sharing appears to be the best strategy (as long as the symmetric rate is concerned). However, it is far from Shang's upper bound, which might not be tight in this regime. In this case, the use of the proposed non-Gaussian distributions is beneficial in the high-power regime (where transmission power is above the threshold (3.67)), where they can significantly reduce the capacity losses due to the lack of synchronism and the use of single-user decoders. For instance, in the high-power regime, for 3 dB signal-to-interference ratio the use of ternary-distributed codes can raise the achievable rates 99.8%.

3.6 Conclusions

Motivated by practical constraints arising in decentralized wireless networks with uncoordinated nodes, we studied the totally asynchronous interference channel with single-user receivers. In the discrete case, its capacity region was characterized using an Information Spectrum approach since the channel is not information stable. As single-letter inner and outer bounds of the capacity region were found, they allowed us to compute achievable rates for the Gaussian case tackling the rather complex expressions of the capacity region.

Gaussian-distributed codes are capacity-achieving whenever the capacity region of the 2×2 GIC is known and lead to closed-form characterizations of achievable rate regions. Despite their natural appeal, in the limited setup where transmission synchronism cannot be guaranteed and single-user decoders are required, they fall short of maximizing the achievable rates for all GIC instances. Sufficient conditions for the existence of other than Gaussian-distributed codes leading to higher achievable rates were found, that for the fully symmetric GIC reduced to ensuring that the transmission powers exceed a threshold that makes interference to be at least moderate.

These analytical results were supported by numerical experiments aiming at i) showing that explicit non-Gaussian distributed codes yielding higher achievable rates than Gaussian-distributed codes exist (Figure 3.1); ii) checking the agreement between the sufficient conditions and the performance of the explicitly proposed non-Gaussian codes (Figure 3.2); and iii) quantifying the losses associated to the lack of transmission synchronism and the use of single-user decoders and the losses/gains from using non-Gaussian distributed codes in the low- and high-power and low- and high-interference regimes (Figure 3.3).

Chapter 4

Optimal Resource Allocation in Cellular Networks with Partial CSI

Previous chapters addressed several aspects of multiuser interference. Namely, the study of how and when to partially cancel it (Chapter 2), and its impact on the achievable rates whenever that is not possible (Chapter 3). Focused on canonic small-size scenarios and the infinite blocklength regime, they disregarded part of the essential characteristics found in most of today practical multiuser systems, e.g. cellular systems. In this chapter we try to bridge this gap by targeting a cellular network with an arbitrary number of nodes using non-ideal codes.

Emerging cellular networks are likely to handle users with heterogeneous quality of service requirements attending to the nature of their underlying service application, the quality of their wireless equipment, or even their contract terms. While sharing the same physical resources (power, bandwidth, transmission time), the utility they get from using them may be very different and arbitrage is needed to optimize the global operation of the network. In this respect, we investigate resource allocation strategies maximizing network utility under practical constraints.

When it comes to considering a scenario with many users, the metric of interest can no longer be the capacity region: its complexity becomes as formidable as the chances of extracting some practical insights out of it for potential extrapolation to prospective system designs. It is therefore sensible to adopt a particular transmission strategy and try to get the most out of the stringently limited available resources. In particular, we focus on a cellular network with half-duplex, MIMO terminals and relaying infrastructure in the form of fixed and dedicated relay stations. Whereas orthogonal frequency division multiple access is assumed, it is seen as a frequency diversity enabler since path loss is the only channel state information (CSI) known at the transmitters, which is refreshed periodically.

The lack of complete centralized CSI of all the links leads us to nullify intracell interference by assigning resources disjointly. Despite the theoretical suboptimality of this choice from an information theory standpoint, we see a three-fold interest in avoiding interference in this scenario:

1. Not knowing the channel states between each destination and its potential undesired transmitters, the network performance is exposed to uncontrolled losses if resource allocation is performed ignoring interference.

2. Interference couples the performance of the different links possibly in a non-analytic manner which prevents closed form expressions (see (3.19)-(3.20) in Chapter 3).

3. Whenever analytic characterizations of achievable rate degradations due to multiuser interference exist, they often lead to nonconvex expressions which preclude the use of efficient global optimization approaches.

With this setup, the performance of a state-of-the art relay-assisted transmission protocol is characterized in terms of the ergodic achievable rates, for which novel concave lower bounds are developed. The use of these bounds allows us to derive two efficient algorithms computing resource allocations in polynomial time, which address the optimization of the uplink and downlink directions jointly. First, a global optimization algorithm providing one Pareto optimal solution maximizing network utility during all the validity of one CSI is studied, which acts as a performance upper bound. Second, a sequential optimization algorithm maximizing network utility frame by frame is considered as a simpler alternative. The performance of both schemes has been compared in practical scenarios, giving special attention to the performance-complexity and throughput-fairness tradeoffs.

4.1 Introduction

4.1.1 Motivation

The deployment of cellular networks has been traditionally associated to the provision of voice (and low-rate data) service to mobile users. The exclusivity of this purpose, however, is in conflict with the ubiquitous availability of wireless equipment and the steadily increasing traffic demands arising from new interactive, multimedia services, which have opened the door to a plethora of new potential network scenarios. From interactive gaming to wireless broadband access, different services with heterogeneous quality of service (QoS) requirements shall converge to the same service network. Regarding this paradigm, we identify three central issues in prospective network design which motivate the work of this chapter:

- How to characterize the user experience of the different services of the network using homogeneous performance measures?

- How to dynamically arbitrate on the shared use of the limited transmission resources of the network by competing flows which are of different nature?

- How to extract the largest possible system spectral efficiency from the physical layer?

With this in mind, the optimization of the operation of the network is hence a matter of allocating resources (power, bandwidth, rate, transmission time) efficiently for uplink (UL) and downlink (DL) scheduled transmissions among the serving users such that some network-wide cost function involving their service experience is maximized along time.

4.1.2 Adopted network setup

In this chapter, we tackle the network design problem adopting a cell-by-cell approach. Hence, we focus on a cell consisting of one base station (BS) serving M mobile stations (MS's, or users). To enable the realization of high spectral efficiencies and boost network performance, we assume that R relay stations (RS's) are deployed within the cell coverage area to enhance the communication between the users and the BS [Cov79, Cov06, Nab04, Wan05, Och06]. Interpreting the presence of relays as an extension of the network infrastructure enabling relay-assisted transmission, their locations are assumed to remain fixed, although they can indeed be optimized beforehand. All the terminals are assumed half-duplex for practical reasons.

Since the capacity of the relay channel is still an open problem (so is determining the optimal relaying strategy) we shall adopt here the cooperation protocol of [HM05, Prop. 2], based on the decode-and-forward strategy [Yu05], which comprises essentially some of the protocols in [Nab04, Och06, Val03, Doh04b] as particular cases and is able to work with partial knowledge of the channel state. To make our approach more general, we let the BS, the RS's, and the MS's be equipped with an arbitrary number of antennas, denoted by n_{BS}, n_{RS}, and $n_{\mathrm{MS},m}$, respectively ($n_{\mathrm{MS},m}$ is the number of antennas of the m-th MS, $1 \leq m \leq \mathsf{M}$) such that extra performance gains arising from MIMO [Big04, Tse05, Big07a] can be also captured.

Pursuing the application of our results to practical scenarios we are led to two important choices, the first one being the adoption of orthogonal frequency division multiple-access (OFDMA). OFDMA can be efficiently implemented via FFT/IFFT, and it is able to combat the inherent frequency selectivity of wireless channels while at the same time allowing a modular tone-based multiplexing of users. Additionally, it improves upon TDMA with respect to achievable rates and data latency, and allows for finer granularity in resource allocation [YJC07], a must in wideband systems. For these and other reasons, it results appealing for upcoming wireless networking standards [Gro06, Gro07, 20005].

The second choice is related to the availability and quality of channel state information (CSI) at each network location (BS, RS's, and MS's). In relayless OFDMA networks, centralized perfect CSI of all the links (in the form of per-tone fading state knowledge) can be used to allocate resources adaptively. Hence, bandwidth, power, and rate can be optimally assigned to align with the instantaneous network conditions [Yu06], yielding enormous performance gains. However, such perfect CSI is likely not to be available in all the ($\mathsf{R}+\mathsf{M}+\mathsf{RM}$) links of our scenario.

On the one hand, the amount of processing required to take advantage of perfect CSI can be formidable (the complexity has been shown to be NP-hard even for a relayless network [Won99]) and possibly non-affordable. On the other hand, for sufficiently fast time-varying channels, the necessary CSI refresh interval can happen to exceed the capacity of the limited-rate feedback channels of the network. Even worse, propagation and processing delays on the feedback channels may result in outdated, useless CSI at the beginning of a resource allocation phase. Thus, unlike other works [Yu06, Won99, Bae06, Lee06, Ng07] we shall study the network scenario of all transmitters having perfect knowledge of the *path loss* of each of the channels, a slowly varying scalar parameter, but being ignorant of each per-tone fading state. Although explicit path loss estimation techniques are out of the scope of this thesis, its accurate estimation seems reasonable provided that some pilot tones are placed within the transmission bandwidth, which is a common practice in OFDM-based standards such as the IEEE 802.16 suite [Gro06, Gro07, 20005], the 3GPP LTE [36.], and WiMAX [Teo07] for synchronization purposes.

With this setup, we aim at optimizing the network operation for maximizing network utility [Ng07, Kel98, Lin06, Pal07, Xue06] in a cell-by-cell approach. Centralized optimization is hence performed at each BS which, upon collection of CSI, takes scheduling decisions and implements resource allocation strategies shaping the instantaneous rates of all the users involved in its cell. One nice feature of our network operation design framework is that the network resources (time, frequency, power, and rate) devoted to UL and DL transmissions are optimized *jointly*, instead of allocating a given portion of total resources to each direction in each transmission frame and optimizing them separately.

4.1.3 Summary of contributions

We propose a centralized optimization framework for the maximization of the cell performance based on the user experience of each serving MS. Under the setup of Section 4.1.2, the CSI of all the links (path loss) is collected at the BS which, together with the QoS requirements of each UL and DL flow[1] and its current degree of fulfillment, decides the resource allocation strategy to be followed during some period of time. This strategy is based on the maximization of network utility, a cell-wide performance measure which combines the service satisfaction of all the users, and has given rise to the following contributions to the problems raised in Section 4.1.1:

- User satisfaction is measured using utility functions. Thus, the same network infrastructure can flexibly reconfigure to optimally serve a variety of scenarios by properly choosing the user utility function of each service under operation such that their different profiles are conveniently reflected.

[1]We consider here that each user requires to send *and* receive information, hence generating one UL flow and another DL flow. The generalization to the setup where users may require more than one flow per direction (e.g. when accessing different services simultaneously using the same equipment) is straightforward as each pair of UL and DL flows can be treated as a different virtual user.

- An algorithm to efficiently compute a global optimal resource allocation strategy in polynomial time (by solving a series of convex optimization problems) is proposed. It is benchmarked against other simpler, suboptimal strategies able to retain a large fraction of performance with significant complexity savings.
- The optimal operation of the network that maximizes network utility is essentially cross-layer, as the joint optimization of user scheduling, resource allocation, and relay-assisted transmission is involved for UL and DL directions.
- In characterizing the performance of the adopted relay-assisted transmission protocol, tight concave lower bounds to the ergodic capacity of MIMO and distributed MIMO channels are obtained which may find applications outside the problems addressed in this thesis.

The rest of the chapter is structured as follows. Section 4.2 describes the adopted transmission strategy for OFDMA with partial CSI and some preliminaries regarding key system parameters. Next, Section 4.3 addresses the transmission protocol for relay-assisted communication. Its cell-wide short term performance is analytically characterized in Section 4.4 in terms of instantaneous achievable rate regions. Then, Section 4.5 builds upon this to i) introduce user utility functions as a useful tool to characterize user satisfaction with services of different nature, ii) pose optimal network strategy as the solution to an optimization problem which aims at maximizing network utility, and iii) propose an iterative algorithm to compute a global optimal solution to this problem in polynomial time. Additionally, a reduced-complexity algorithm computing a suboptimal network strategy is also proposed and benchmarked against the global optimal in Section 4.6, where simulation results of practical scenarios are provided. Finally, Section 4.7 concludes the chapter summarizing results and sketching lines for future work.

4.2 System Model and Preliminaries

Consider the network setup described in Section 4.1.2, where the BS, the RS's, and the MS's are power constrained to $p_{\mathrm{BS}}^{\max}$, $p_{\mathrm{RS}}^{\max}$, and $p_{\mathrm{MS}}^{\max}$, respectively. In every transmission frame interval, denoted by T, the same network bandwidth B is used in the UL and DL phases, of adjustable duration via TDD[2]. In each of them, the communication of each BS-MS pair is assisted by one RS. Let us denote the RS attached to the m-th MS by $\mathrm{RS}(m) \in \{1, \ldots, \mathsf{R}\}$. The RS assignment of the network is hence described by the connectivity matrix $\mathbf{L} \in \{0,1\}^{\mathsf{R}\times\mathsf{M}}$, where $L_{r,m} = \delta[r - \mathrm{RS}(m)]$ and $\delta[\cdot]$ is the Kronecker delta. Note that each BS-MS pair is assisted by one RS, but the same RS can serve more than one BS-MS pair. In fact, the number of BS-MS's pairs assisted by the r-th RS equals the number of non zero entries of the r-th row of the connectivity matrix $\mathbf{L}$.

[2]Although the proposed optimization framework can be extended to the FDD mode, we have ruled it out because it poses more restrictive complexity requirements on the RS's, which should be able to receive and transmit simultaneously on different frequency bands.

We shall use the vectors $\boldsymbol{\ell}_1(t) \in \mathbb{R}_+^{\mathsf{M}\times 1}$, $\boldsymbol{\ell}_2(t) \in \mathbb{R}_+^{\mathsf{R}\times 1}$, and $\boldsymbol{\ell}_3(t) \in \mathbb{R}_+^{\mathsf{M}\times 1}$ to denote the CSI collected at the beginning of the t-th frame. While $\ell_{1,m}(t)$ stands for the path loss between the BS and the m-th MS, $\ell_{2,r}(t)$ is the path loss between the BS and the r-th RS, and $\ell_{3,m}(t)$ is the path loss between the m-th MS and its associated RS (which is the RS(m)-th). All of them are assumed to satisfy reciprocity.

When OFDM is employed with the only knowledge of the link path loss at each transmitter, one practical strategy is to perform uniform power allocation among groups of tones sufficiently far apart such that their individual fading states are uncorrelated and frequency diversity is enabled. This is the case in the IEEE 802.16e - PUSC and FUSC standards [20005]. With this approach, coding across a sufficiently large number of tones makes the instantaneous achievable rate, denoted by $r(t)$, be upper bounded by the ergodic (or average) mutual information thanks to the law of large numbers[3]. By ergodic capacity we understand the instantaneous capacity given some fading state in the frequency domain averaged over all possible fading realizations in this domain. Therefore, no matter how short the transmission interval is nor how fast the channel response varies, the ergodic capacity will exclusively depend on the transmission bandwidth and the link signal-to-noise ratio, snr. The snr suffices to characterize the quality of a link since interference between neighboring transmitters is prevented by allocating bandwidth among the different BS-RS-MS triplets in a *disjoint* manner[4]: each BS-MS-RS triplet is assigned a fraction of the total bandwidth in exclusivity. This fraction may vary from UL to DL phases and also within each of them, depending on whether the RS is active (relay-transmit subphase) or not (relay-receive subphase). Whichever subphase we focus on, the snr of any link is given by

$$\mathsf{snr} = \frac{\ell G p}{N_0 F b}, \tag{4.1}$$

where ℓ is pathloss, G is antenna gain, p is transmit power, N_0 is the AWGN one-sided power spectral density, F is the noise factor, and b is bandwidth. Whereas the specific values of p and b are subject to optimization by the BS and N_0 and ℓ are given, we distinguish between F_{BS} (G_{BS}), F_{RS} (G_{RS}), and $F_{\text{MS},m}$ ($G_{\text{MS},m}$) to consider the general case of nodes equipped with RF front-ends of diverse quality.

[3]Consider a SISO point-to-point link for simplicity. When the transmit power is uniformly allocated over N tones spanning some total bandwidth B, the per-tone snr is constant. If $\{h_i\}_{i=1}^N$ denote the fading states of each tone (assumed i.i.d and unknown), the achievable rate satisfies

$$r(t) \leq \sum_{i=1}^{N} \frac{B}{N} \log_2(1+\mathsf{snr} h_i) = \frac{1}{N}\sum_{i=1}^{N} B\log_2(1+\mathsf{snr} h_i) \xrightarrow{(a)} \mathbb{E}\{B\log_2(1+\mathsf{snr} h)\},$$

where (a) follows for large N from the law of large numbers and convergence is in probability. For finite moderate values of N, outage events are not precluded. Its impact on system design, however, is beyond the scope of this thesis.

[4]Inter-cell interference due to frequency reuse in neighboring cells is not considered in this work.

4.3 Relay-Assisted Transmission

4.3.1 Maximum instantaneous achievable rates

The use of RS's in our network setup has the advantage of realizing performance gains arising from relay-assisted transmission. As the bandwidth is assigned orthogonally (disjointly) to each BS-RS-MS triplet, intra-cell interference is completely nulled and it suffices to study one single triplet to describe the overall behavior of the cell.

Considering that every RS operates in the half duplex mode, then for a given time duration the relay is in the receive mode (we call this period the relay-receive subphase), and in the transmit mode for the rest (we call this period the relay-transmit subphase)[5]. To illustrate the cooperation protocol, which is that of [HM05, Prop. 2], consider the specific BS-RS-MS triplet shown in Figure 4.1, where the DL phase is described and the MS and RS index are omitted for simplicity. The matrices $\mathbf{H}_1 \in \mathbb{C}^{n_{\mathrm{MS}} \times n_{\mathrm{BS}}}$, $\mathbf{H}_2 \in \mathbb{C}^{n_{\mathrm{RS}} \times n_{\mathrm{BS}}}$, and $\mathbf{H}_3 \in \mathbb{C}^{n_{\mathrm{MS}} \times n_{\mathrm{RS}}}$ represent the instantaneous fading states of each of the links at a given tone[6].

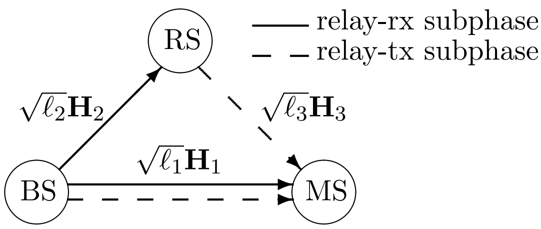

Figure 4.1: DL cooperation protocol: the DL phase is split into two subphases attending to the half duplex nature of the RS.

Although the details of the coding scheme can be found in [HM05, App. A], we provide here a brief sketch of it for the sake of clarity. The BS splits its message into two independent components: one which is transmitted directly to the MS without the help of the RS, and another which is transmitted through the RS to the MS. During the relay-receive subphase, of duration τ_1, the BS transmits one codeword related to the latter message component using some power p_1 while the RS and the MS listen. At the end of this subphase, the RS attempts to decode this message component. If successful, a relay-transmit subphase of duration τ_2 starts where both the RS (which re-encodes the decoded message component) and the BS (which now transmits a codeword associated to its other message component) transmit using powers p_3 and p_2, respectively. Otherwise the RS remains silent during this subphase. The receiver performs successive decoding: it first attempts to decode the relayed message component from the signal of both subphases and, if successful, it subtracts the signal transmitted by the RS in the second subphase prior to decoding the unrelayed message component. Assuming that

[5]We shall also refer to the protocol subphases as subphase 1 (relay-receive) and subphase 2 (relay-transmit).

[6]When UL cooperative transmission is considered, the instantaneous fading states can be described by using the transposed matrices $\{\mathbf{H}_j^T\}_{j=1}^3$

communication takes place over bandwidths B_1 (relay-receive subphase) and B_2 (relay-transmit subphase) and that uniform power allocation across antennas is performed (see [Lia07] for relay-assisted communication protocols where power allocation is performed assuming perfect CSI), the achievable rate r_{DL} in [bit/s] satisfies [HM05]

$$r_{\mathrm{DL}} \leq \min\{r_{\mathrm{DL}}^{(1)}, r_{\mathrm{DL}}^{(2)}\}, \tag{4.2}$$

where the min function models whether it is the source-relay or the source-destination who act as information bottlenecks for the relayed message component, and

$$r_{\mathrm{DL}}^{(1)} = \tau_1 B_1 \mathbb{E}\Big\{\log_2 \det\Big(\mathbf{I}_{n_{\mathrm{RS}}} + \frac{\mathsf{snr}_2}{n_{\mathrm{BS}}}\mathbf{H}_2\mathbf{H}_2^\dagger\Big)\Big\} + \tau_2 B_2 \mathbb{E}\Big\{\log_2 \det\Big(\mathbf{I}_{n_{\mathrm{MS}}} + \frac{\mathsf{snr}_{1,2}}{n_{\mathrm{BS}}}\mathbf{H}_1\mathbf{H}_1^\dagger\Big)\Big\} \tag{4.3}$$

$$r_{\mathrm{DL}}^{(2)} = \tau_1 B_1 \mathbb{E}\Big\{\log_2 \det\Big(\mathbf{I}_{n_{\mathrm{MS}}} + \frac{\mathsf{snr}_{1,1}}{n_{\mathrm{BS}}}\mathbf{H}_1\mathbf{H}_1^\dagger\Big)\Big\} \tag{4.4}$$

$$+ \tau_2 B_2 \mathbb{E}\Big\{\log_2 \det\Big(\mathbf{I}_{n_{\mathrm{MS}}} + \frac{\mathsf{snr}_{1,2}}{n_{\mathrm{BS}}}\mathbf{H}_1\mathbf{H}_1^\dagger + \frac{\mathsf{snr}_3}{n_{\mathrm{RS}}}\mathbf{H}_3\mathbf{H}_3^\dagger\Big)\Big\}, \tag{4.5}$$

where the snr's amount to

$$\mathsf{snr}_{1,j} = \frac{\ell_1 G_{\mathrm{BS}} p_j}{N_0 F_{\mathrm{MS}} B_j}, \quad \mathsf{snr}_2 = \frac{\ell_2 G_{\mathrm{BS}} p_1}{N_0 F_{\mathrm{RS}} B_1}, \quad \mathsf{snr}_3 = \frac{\ell_3 G_{\mathrm{RS}} p_3}{N_0 F_{\mathrm{MS}} B_2}, \tag{4.6}$$

where $j = 1, 2$.

While the success of decoding the relayed message component at the relay indeed impacts on the success of decoding at the destination, the behavior of the destination is independent of whether the relay was able to decode or not. The destination will attempt to decode first the relayed component, perform successive interference cancellation, and go for the direct component afterwards, no matter what happened to the relay. This makes the relay a transparent network feature as seen by the MS, as no signalling between them is required whatsoever. In fact, as we rely on the ergodic capacities to characterize performance (see Section 4.2), it can be assumed that all the transmissions are reliable as long as their information rates lie below capacity. Consequently, the performance (4.2) of the strategy [HM05] for the one-way relay channel with half-duplex relay is such that the transmission rate of the relayed message component always results in successful decoding at the relay.

It is important to remark that the upper bound (4.2) is only tight for Gaussian codes of infinite blocklength. When practical discrete alphabet codes of finite blocklength are used instead, decoding errors at the RS and the MS cannot be disregarded at rates below the corresponding ergodic capacities. However, expression (4.2) can still be used by introducing a penalizing gap Γ such that $\mathsf{snr}_{\mathrm{practical}} = \mathsf{snr}/\Gamma$ in (4.3)-(4.5)[7]. We will hence use the gap from now on and omit the subscript 'practical' in snr for simplicity. When UL transmission is considered, an analogous expression to (4.2) of the form $r_{\mathrm{UL}} \leq \min\{r_{\mathrm{UL}}^{(1)}, r_{\mathrm{UL}}^{(2)}\}$ readily follows by exchanging the roles of the MS and the BS and transposing the matrices $\{\mathbf{H}_j\}_{j=1}^3$.

[7] The gap can be further increased to model the impact of inter-cell interference on final performance via snr degradation.

Oppositely to [Ng07], where relays explicitly switched between amplify-and-forward and decode-and-forward depending on the achievable rates, the adopted cooperation protocol has the advantage of comprising other well-known cooperation strategies as particular cases such that the best one is implicitly selected when the resource allocation is optimized. While it mimics the philosophy of protocol I of [Nab04] and transmit diversity [Och06], it can also accommodate the following:

- *Protocol III* [Nab04], *simplified transmit diversity* [Och06] - Set p_1 to be too small to enable direct BS-MS reliable communication in the relay-receive subphase.
- *Protocol II* [Nab04], *receive diversity* [Nab04] - Set $p_2 = 0$.
- *Multihop relaying* [Och06, Doh04a, Cal07e] - Set $p_2 = 0$ and p_1 to be too small to enable direct BS-MS reliable communication in the relay-receive subphase.
- *Direct transmission* - Set $p_3 = 0$ and/or $\tau_2 = 0$ and/or $B_2 = 0$.

4.3.2 Universal concave lower bounds on the achievable rates

Transmission over multiple tones with uncorrelated fading makes the ergodic (or average) rates show up in (4.3)-(4.5). They involve computing three MIMO channel ergodic capacities and one distributed MIMO channel ergodic capacity (the τ_2 term in (4.5)). After averaging over the fading distribution, i.e., the distribution of the matrices $\{\mathbf{H}_j\}_{j=1}^3$, the resulting expectations depend only on the link snr's and the $\tau_j B_j$ products, and admit closed form expressions for both the MIMO [Shi03, Big04] and the distributed MIMO [Kie05] channel in case of Rayleigh fading. However, analytical expressions cannot be derived for other fading distributions like Ricean, that are common in the BS-RS link and include line-of-sight components (LOS). On top of that, equations (4.3)-(4.5) are not concave functions of the duration of the subphases, the allocated bandwidths, and the transmit powers. This prevents efficient methods to be applied for rate allocation in global optimization approaches.

Alternatively, we develop universal, simpler concave lower bounds of $r_{\mathrm{DL}}^{(1)}$ and $r_{\mathrm{DL}}^{(2)}$ that ease prospective optimization methods and allow for an easy concavity test. Here, by universal we mean that parametric lower bounds with the same *structure* can be applied to *any* fading distribution by changing the parameter values and not that the same expression holds for all of them. Other parametric approaches have been taken to approximate MIMO ergodic capacities [Doh05], but oppositely to our needs, concavity with respect to durations, bandwidths, and powers was not guaranteed, parameter values were not systematically found (i.e., curve fitting was performed), and the distributed MIMO case was not tackled. To start with, consider the following results upon which our lower bounds are based. Their concavity analysis will be left to the next section.

Lemma 4.1 *A lower bound to the ergodic capacity of an $n_t \times n_r$ MIMO channel is*

$$\mathbb{E}\Big\{\log_2 \det\Big(\mathbf{I}_{n_r} + \frac{\mathsf{snr}}{n_t}\mathbf{H}\mathbf{H}^\dagger\Big)\Big\} \geq \sum_{i=1}^{n_r} \log_2(1 + \rho_i(f_{\mathbf{H}})\mathsf{snr}/n_t), \tag{4.7}$$

where $\rho_i(f_{\mathbf{H}}) \triangleq \exp\big(\mathbb{E}\{\log \lambda_i\left(\mathbf{H}\mathbf{H}^\dagger\right)\}\big)$ *and* $f_{\mathbf{H}}(\cdot)$ *denotes the pdf of the channel matrix* $\mathbf{H}$[8].

Proof. Proceeding as in [Big04, App. E.1], we start from the expression that relates the ergodic capacity with the ordered eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$ to obtain

$$\begin{aligned}\mathbb{E}\Big\{\log_2 \det\Big(\mathbf{I}_{n_r} + \frac{\mathsf{snr}}{n_t}\mathbf{H}\mathbf{H}^\dagger\Big)\Big\} &= \sum_{i=1}^{n_r} \mathbb{E}\{\log_2(1 + \lambda_i(\mathbf{H}\mathbf{H}^\dagger)\mathsf{snr}/n_t)\} && (4.8)\\ &= \sum_{i=1}^{n_r} \mathbb{E}\{\log_2(1 + \exp(\log \lambda_i(\mathbf{H}\mathbf{H}^\dagger))\mathsf{snr}/n_t)\} && (4.9)\\ &\geq \sum_{i=1}^{n_r} \log_2(1 + \exp(\mathbb{E}\{\log \lambda_i(\mathbf{H}\mathbf{H}^\dagger)\})\mathsf{snr}/n_t), && (4.10)\end{aligned}$$

where (4.10) follows from Jensen's inequality and the convexity of the function $I(x) = \log_2(a + b\exp(x))$ for all $a, b \geq 0$. □

Lemma 4.2 *A lower bound to the ergodic capacity of an $n_{t,1} \times n_r$ and $n_{t,2} \times n_r$ distributed MIMO channel is*

$$\mathbb{E}\Big\{\log_2 \det\Big(\mathbf{I}_{n_r} + \frac{\mathsf{snr}_1}{n_{t,1}}\mathbf{H}_1\mathbf{H}_1^\dagger + \frac{\mathsf{snr}_2}{n_{t,2}}\mathbf{H}_2\mathbf{H}_2^\dagger\Big)\Big\} \geq \sum_{i=1}^{n_r} \log_2(1 + \rho_i(f_{\mathbf{H}_1})\mathsf{snr}_1/n_{t,1} + \rho_i(f_{\mathbf{H}_2})\mathsf{snr}_2/n_{t,2}). \tag{4.11}$$

Proof. Since $1 + \lambda_i(\mathbf{H}_1\mathbf{H}_1^\dagger)\mathsf{snr}_1/n_{t,1} + \lambda_i(\mathbf{H}_2\mathbf{H}_2^\dagger)\mathsf{snr}_2/n_{t,2} \geq 0$ for $1 \leq i \leq n_r$ and both $\mathbf{H}_1\mathbf{H}_1^\dagger$ and $\mathbf{H}_2\mathbf{H}_2^\dagger$ are Hermitian matrices, it follows from [Fie71] that

$$\mathbb{E}\Big\{\log_2 \det\Big(\mathbf{I}_{n_r} + \frac{\mathsf{snr}_1}{n_{t,1}}\mathbf{H}_1\mathbf{H}_1^\dagger + \frac{\mathsf{snr}_2}{n_{t,2}}\mathbf{H}_2\mathbf{H}_2^\dagger\Big)\Big\} \geq \sum_{i=1}^{n_r} \mathbb{E}\{\log_2(1 + \lambda_i(\mathbf{H}_1\mathbf{H}_1^\dagger)\mathsf{snr}_1/n_{t,1} + \lambda_i(\mathbf{H}_2\mathbf{H}_2^\dagger)\mathsf{snr}_2/n_{t,2})\}. \tag{4.12}$$

The lemma follows by similarly applying Jensen's inequality resorting twice to the function $I(x)$. □

Lemmas 4.1 and 4.2 lower bound the MIMO channel capacities with expressions that mimic equivalent transmissions through virtual parallel AWGN channels of gains $\rho_i(\cdot)$, which depend on the antenna configuration and the fading distribution, and whose tightness is analyzed in figures 4.2 and 4.3. As for the MIMO channel, Lemma 4.1 lower bound is extremely tight. The tightness of Lemma 4.2 lower bound with respect to the distributed MIMO channel capacity, however, depends on the snr.

[8]Note that since $\text{rank}\{\mathbf{H}\mathbf{H}^\dagger\} \leq \min\{n_t, n_r\}$, $\rho_i(f_{\mathbf{H}}) = 0$ for $\min\{n_t, n_r\} < i \leq n_r$.

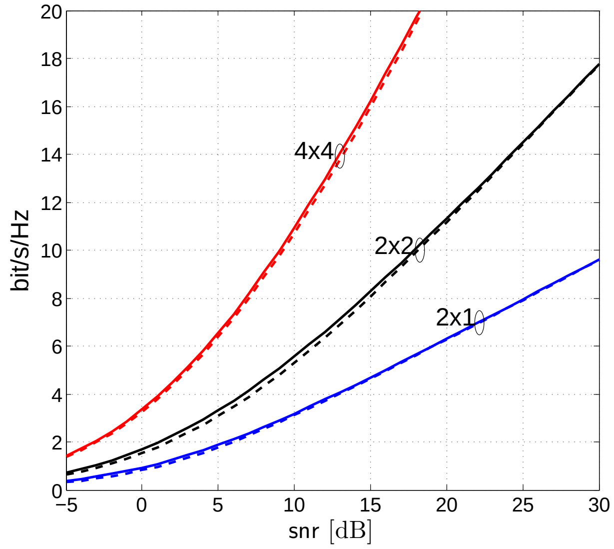

Figure 4.2: Exact ergodic capacity (solid lines) and Lemma 1 lower bound (dashed lines) vs snr for different antenna configurations and Rayleigh fading.

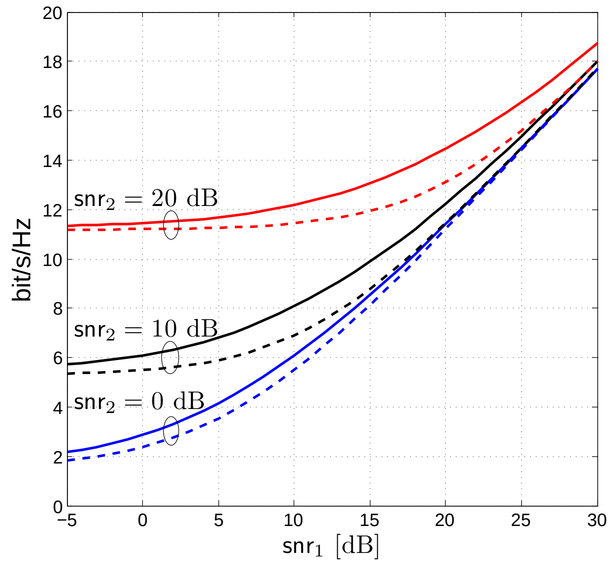

Figure 4.3: Exact ergodic capacity (solid lines) and Lemma 2 lower bound (dashed lines) vs snr_1 for different values of snr_2 and Rayleigh fading. The antenna configuration is $n_r = n_{t,1} = n_{t,2} = 2$.

The computation of the channel-dependent coefficients $\rho_i(\cdot)$ can be accurately performed offline by using Monte Carlo methods. However, as for Rayleigh fading and an $n \times 1$ or $1 \times n$ antenna configuration, results on the expectation of the logarithm of a Chi-square random variable [Lap03] can be applied to show that

$$\rho_i(n) = e^{-\Psi + \sum_{j=1}^{n-1} \frac{1}{j}} \delta[i-1], \tag{4.13}$$

where $\Psi \approx 0.577$ is the Euler-Mascheroni constant [Gra00, 4.352-1]. In any case, once the channel-dependent coefficients are computed, lemmas 1 and 2 allow us to state the main result of this section.

Corollary 4.1 *A lower bound on the maximum DL achievable rates of the adopted relay-assisted transmission protocol is*

$$r_{\mathrm{DL}} \leq \min\{\overline{r}_{\mathrm{DL}}^{(1)}, \overline{r}_{\mathrm{DL}}^{(2)}\}, \tag{4.14}$$

where

$$\overline{r}_{\mathrm{DL}}^{(1)} = \tau_1 B_1 \sum_{i=1}^{n_{\mathrm{RS}}} \log_2\Big(1+\rho_i(f_{\mathbf{H}_2})\frac{\mathsf{snr}_2}{n_{\mathrm{BS}}}\Big) + \tau_2 B_2 \sum_{i=1}^{n_{\mathrm{MS}}} \log_2\Big(1+\rho_i(f_{\mathbf{H}_1})\frac{\mathsf{snr}_{1,2}}{n_{\mathrm{BS}}}\Big) \leq r_{\mathrm{DL}}^{(1)} \tag{4.15}$$

$$\overline{r}_{\mathrm{DL}}^{(2)} = \tau_1 B_1 \sum_{i=1}^{n_{\mathrm{MS}}} \log_2\Big(1+\rho_i(f_{\mathbf{H}_1})\frac{\mathsf{snr}_{1,1}}{n_{\mathrm{BS}}}\Big) + \tau_2 B_2 \sum_{i=1}^{n_{\mathrm{MS}}} \log_2\Big(1+\rho_i(f_{\mathbf{H}_1})\frac{\mathsf{snr}_{1,2}}{n_{\mathrm{BS}}} + \rho_i(f_{\mathbf{H}_3})\frac{\mathsf{snr}_3}{n_{\mathrm{RS}}}\Big) \tag{4.16}$$

$$\leq r_{\mathrm{DL}}^{(2)}. \tag{4.17}$$

A similar lower bound on the maximum UL achievable rates holds by exchanging the roles of the BS and the MS and transposing $\{\mathbf{H}_j\}_{j=1}^{3}$.

4.4 Achievable Instantaneous Rates

Given the CSI at the beginning of the t-th frame, $\{\boldsymbol{\ell}_j(t)\}_{j=1}^{3}$, the instantaneous performance of the network is given by the DL and UL achievable rate regions, i.e., the set of all rate vectors $\mathbf{r}_{\mathrm{DL}}(t), \mathbf{r}_{\mathrm{UL}}(t) \in \mathbb{R}_{+}^{\mathsf{M}\times 1}$ that can be sustained during one frame duration. The achievable rates depend upon the frame format as described by the vector of fractional durations $\boldsymbol{\tau} \in \mathbb{R}_{+}^{4}$, $\mathbf{1}_4^T\boldsymbol{\tau} = 1$, whose components account for the DL subphase 1 (τ_1), DL subphase 2 (τ_2), UL subphase 1 (τ_3), and UL subphase 2 (τ_4). Each subphase duration shapes and couples the instantaneous achievable rate regions, denoted by $\mathcal{R}_{\mathrm{DL}}(t;\boldsymbol{\tau})$ (DL) and $\mathcal{R}_{\mathrm{UL}}(t;\boldsymbol{\tau})$ (UL), and will be subject to optimization later on, when rate allocation policies come into play in the next section. In this section, however, we shall focus on the dependence of the achievable rates on the disjoint allocations of power among transmitters and bandwidth among BS-RS-MS triplets.

4.4.1 DL instantaneous achievable rate region

Assuming that the duration of the DL subphases is fixed to $\tau_1 T$ and $\tau_2 T$, the instantaneous achievable rates depend upon the allocation of bandwidth and transmit power among the M competing flows. Let us describe the DL resource allocation by using the vectors $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}_{+}^{\mathsf{M}\times 1}$, which represent the fractional BS power allocation in subphases 1 and 2, the fractional RS's power allocation in subphase 2 ($p_{3,m}$ is the fraction of power transmitted by the RS(m)-th RS in assisting the m-th MS), and the fractional bandwidth allocation in subphases 1 and 2, respectively. By imposing non-negativity on each fraction and constraining the sum of resources it follows that

$$\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{b}_1, \mathbf{b}_2 \geq \mathbf{0}_{\mathsf{M}}, \tag{4.18}$$

$$\mathbf{1}_{\mathsf{M}}^T \mathbf{p}_j \le 1, \quad \mathbf{L}\mathbf{p}_3 \le \mathbf{1}_{\mathsf{R}}, \quad \mathbf{1}_{\mathsf{M}}^T \mathbf{b}_j \le 1, \tag{4.19}$$

where $j = 1, 2$ and $\mathbf{L}$ is the connectivity matrix defined in Section 4.2. Thus, applying Corollary 4.1 the DL achievable rate in [bit/s] of the m-th user, $r_{\mathrm{DL},m}(t)$, satisfies

$$r_{\mathrm{DL},m}(t) \le \min\{\overline{r}^{(1)}_{\mathrm{DL},m}(t), \overline{r}^{(2)}_{\mathrm{DL},m}(t)\}, \tag{4.20}$$

where

$$\overline{r}^{(1)}_{\mathrm{DL},m}(t) = B\tau_1 b_{1,m} \sum_{i=1}^{n_{\mathrm{RS}}} \log_2\Big(1 + c^i_{2,m}(t)\frac{p_{1,m}}{b_{1,m}}\Big) + B\tau_2 b_{2,m} \sum_{i=1}^{n_{\mathrm{MS},m}} \log_2\Big(1 + c^i_{1,m}(t)\frac{p_{2,m}}{b_{2,m}}\Big) \tag{4.21}$$

$$\overline{r}^{(2)}_{\mathrm{DL},m}(t) = B\tau_1 b_{1,m} \sum_{i=1}^{n_{\mathrm{MS},m}} \log_2\Big(1 + c^i_{1,m}(t)\frac{p_{1,m}}{b_{1,m}}\Big) \tag{4.22}$$

$$+ \; B\tau_2 b_{2,m} \sum_{i=1}^{n_{\mathrm{MS},m}} \log_2\Big(1 + \frac{c^i_{1,m}(t)p_{2,m} + c^i_{3,m}(t)p_{3,m}}{b_{2,m}}\Big) \tag{4.23}$$

condense CSI into the equivalent channel gains

$$c^i_{1,m}(t) = \frac{\rho_i(f_{\mathbf{H}_{1,m}})\ell_{1,m}(t)G_{\mathrm{BS}}p^{\max}_{\mathrm{BS}}}{n_{\mathrm{BS}}\Gamma N_0 F_{\mathrm{MS}} B} \tag{4.24}$$

$$c^i_{2,m}(t) = \frac{\rho_i(f_{\mathbf{H}_{2,\mathrm{RS}(m)}})\ell_{2,\mathrm{RS}(m)}(t)G_{\mathrm{BS}}p^{\max}_{\mathrm{BS}}}{n_{\mathrm{BS}}\Gamma N_0 F_{\mathrm{RS}} B} \tag{4.25}$$

$$c^i_{3,m}(t) = \frac{\rho_i(f_{\mathbf{H}_{3,m}})\ell_{3,m}(t)G_{\mathrm{RS}}p^{\max}_{\mathrm{RS}}}{n_{\mathrm{RS}}\Gamma N_0 F_{\mathrm{MS}} B}. \tag{4.26}$$

Following the notation of Section 4.3, we have used $f_{\mathbf{H}_{1,m}}$ to denote the DL fading distribution between the BS and the m-th MS, $f_{\mathbf{H}_{2,\mathrm{RS}(m)}}$ for the DL fading distribution between the BS and the serving RS of the m-th MS, and $f_{\mathbf{H}_{3,m}}$ for the DL fading distribution between the m-th MS and its serving RS. The DL achievable rate region is hence given by

$$\mathcal{R}_{\mathrm{DL}}(t;\boldsymbol{\tau}) = \bigcup \{\mathbf{0}_{\mathsf{M}} \le \mathbf{r}_{\mathrm{DL}}(t) \le \min\{\overline{\mathbf{r}}^{(1)}_{\mathrm{DL}}(t), \overline{\mathbf{r}}^{(2)}_{\mathrm{DL}}(t)\}\}, \tag{4.27}$$

where the union is taken over the allocations satisfying (4.18)-(4.19).

Lemma 4.3 *The DL instantaneous achievable rate region $\mathcal{R}_{\mathrm{DL}}(t;\boldsymbol{\tau})$ is convex.*

Proof. For fixed τ_1, τ_2, some properties of convex functions [Boy04] can be used to show that the right hand side of (4.20) is concave: the minimum of concave functions is concave, and the concavity of (4.21)-(4.23) can be shown resorting to the function $G(x, y) = ax\log(1 + by/x)$, which is concave in $x, y \ge 0$ $\forall a, b \ge 0$. This implies convexity of $\mathcal{R}_{\mathrm{DL}}(t;\boldsymbol{\tau})$ for fixed $\boldsymbol{\tau}$ [Boy04], which will turn out to be useful in ensuring global optimality in rate allocation problems. This desirable property, which follows from the use of the universal lower bounds derived Section 4.3.2, vanishes when the frame format (here in the form of the relative durations τ_1, τ_2) is subject to optimization too. However, this can be circumvented with the following variable change

$$\mathbf{q}_j \triangleq \tau_j \mathbf{p}_j, \quad \mathbf{q}_3 \triangleq \tau_2 \mathbf{p}_3, \quad \mathbf{w}_j \triangleq \tau_j \mathbf{b}_j, \tag{4.28}$$

where $j = 1, 2$. This variable change gives rise to a new set of allocation variables, in terms of which (4.21)-(4.23) become

$$\overline{r}^{(1)}_{\mathrm{DL},m}(t) = B\left[w_{1,m}\sum_{i=1}^{n_{\mathrm{RS}}}\log_2\left(1+c^i_{2,m}(t)\frac{q_{1,m}}{w_{1,m}}\right)+w_{2,m}\sum_{i=1}^{n_{\mathrm{MS},m}}\log_2\left(1+c^i_{1,m}(t)\frac{q_{2,m}}{w_{2,m}}\right)\right] \tag{4.29}$$

$$\overline{r}^{(2)}_{\mathrm{DL},m}(t) = B\left[w_{1,m}\sum_{i=1}^{n_{\mathrm{MS},m}}\log_2\left(1+c^i_{1,m}(t)\frac{q_{1,m}}{w_{1,m}}\right)+w_{2,m}\sum_{i=1}^{n_{\mathrm{MS},m}}\log_2\left(1+\frac{c^i_{1,m}(t)q_{2,m}+c^i_{3,m}(t)q_{3,m}}{w_{2,m}}\right)\right], \tag{4.30}$$

both of them concave functions regardless of τ_1, τ_2. The new set of feasible allocations transforms accordingly into

$$\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{w}_1, \mathbf{w}_2 \geq \mathbf{0}_{\mathsf{M}}, \tag{4.31}$$

$$\mathbf{1}^T_{\mathsf{M}}\mathbf{q}_j \leq \tau_j, \quad \mathbf{L}\mathbf{q}_3 \leq \tau_2\mathbf{1}_{\mathsf{R}}, \quad \mathbf{1}^T_{\mathsf{M}}\mathbf{w}_j \leq \tau_j, \tag{4.32}$$

where $j = 1, 2$ again. Formulated in terms of the new allocation variables, the region $\mathcal{R}_{\mathrm{DL}}(t;\boldsymbol{\tau})$ can be equivalently obtained by taking the union in (4.27) over (4.31)-(4.32), where (4.29)-(4.30) are used instead of (4.21)-(4.23). This way, the convexity of $\mathcal{R}_{\mathrm{DL}}(t;\boldsymbol{\tau})$ with respect to $\boldsymbol{\tau}$ is unveiled: (4.29)-(4.30) are concave and the feasible set (4.31)-(4.32) is the intersection of halfspaces and hence convex, something that was hidden with the original allocation variables. Needless to say, the variable change (4.28) can be straightforwardly reversed to obtain the allocated fractions of bandwidth and power. □

4.4.2 UL instantaneous achievable rate region

Proceeding similarly, if the relative duration of the subphases is fixed to $\tau_3 T$ and $\tau_4 T$, the UL resource allocation can be characterized in terms of the vectors $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^{\mathsf{M}}_+$. While the meaning of $\mathbf{b}_1, \mathbf{b}_2$, and $\mathbf{p}_3$ is identical, $\mathbf{p}_1$ and $\mathbf{p}_2$ refer now to the fractional MS's transmit power in subphases 1 and 2. Thus, the feasible set of UL resource allocations is

$$\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{b}_1, \mathbf{b}_2 \geq \mathbf{0}_{\mathsf{M}} \tag{4.33}$$

$$\mathbf{p}_j \leq \mathbf{1}_{\mathsf{M}}, \quad \mathbf{L}\mathbf{p}_3 \leq \mathbf{1}_{\mathsf{R}}, \quad \mathbf{1}^T_{\mathsf{M}}\mathbf{b}_j \leq 1, \tag{4.34}$$

where $j = 1, 2$. The application of Corollary 1 to the UL achievable rate in [bit/s] of the m-th user implies that $r_{\mathrm{UL},m}(t)$ satisfies

$$r_{\mathrm{UL},m}(t) \leq \min\{\overline{r}^{(1)}_{\mathrm{UL},m}(t), \overline{r}^{(2)}_{\mathrm{UL},m}(t)\}, \tag{4.35}$$

where

$$\overline{r}^{(1)}_{\mathrm{UL},m}(t) = B\tau_1 b_{1,m}\sum_{i=1}^{n_{\mathrm{RS}}}\log_2\left(1 + d^i_{3,m}(t)\frac{p_{1,m}}{b_{1,m}}\right) + B\tau_2 b_{2,m}\sum_{i=1}^{n_{\mathrm{BS},m}}\log_2\left(1 + d^i_{1,m}(t)\frac{p_{2,m}}{b_{2,m}}\right) \tag{4.36}$$

$$\overline{r}^{(2)}_{\mathrm{UL},m}(t) = B\tau_1 b_{1,m}\sum_{i=1}^{n_{\mathrm{BS},m}}\log_2\left(1 + d^i_{1,m}(t)\frac{p_{1,m}}{b_{1,m}}\right) \tag{4.37}$$

$$+ B\tau_2 b_{2,m}\sum_{i=1}^{n_{\mathrm{BS},m}}\log_2\left(1 + \frac{d^i_{1,m}(t)p_{2,m} + d^i_{2,m}(t)p_{3,m}}{b_{2,m}}\right) \tag{4.38}$$

use the equivalent channel gains

$$d^i_{1,m}(t) = \frac{\rho_i(f_{\mathbf{H}^T_{1,m}})\ell_{1,m}(t)G_{\mathrm{MS}}p^{\max}_{\mathrm{MS}}}{n_{\mathrm{MS}}\Gamma N_0 F_{\mathrm{BS}} B} \tag{4.39}$$

$$d^i_{2,m}(t) = \frac{\rho_i(f_{\mathbf{H}^T_{2,\mathrm{RS}(m)}})\ell_{2,\mathrm{RS}(m)}(t)G_{\mathrm{RS}}p^{\max}_{\mathrm{RS}}}{n_{\mathrm{RS}}\Gamma N_0 F_{\mathrm{BS}} B} \tag{4.40}$$

$$d^i_{3,m}(t) = \frac{\rho_i(f_{\mathbf{H}^T_{3,m}})\ell_{3,m}(t)G_{\mathrm{MS}}p^{\max}_{\mathrm{MS}}}{n_{\mathrm{MS}}\Gamma N_0 F_{\mathrm{RS}} B}. \tag{4.41}$$

Note that by transposing the DL fading state matrices, the fading distributions account for UL transmission. The UL achievable rate region can finally be expressed as

$$\mathcal{R}_{\mathrm{UL}}(t;\boldsymbol{\tau}) = \bigcup\{\mathbf{0}_{\mathsf{M}} \leq \mathbf{r}_{\mathrm{UL}}(t) \leq \min\{\bar{\mathbf{r}}^{(1)}_{\mathrm{UL}}(t), \bar{\mathbf{r}}^{(2)}_{\mathrm{UL}}(t)\}\}, \tag{4.42}$$

where the union is over (4.33)-(4.34).

Lemma 4.4 *The UL instantaneous achievable rate region $\mathcal{R}_{\mathrm{UL}}(t;\boldsymbol{\tau})$ is convex.*

Proof. As happened in the DL, the achievable rate region $\mathcal{R}_{\mathrm{UL}}(t;\boldsymbol{\tau})$ is convex when the frame format $\boldsymbol{\tau}$ is fixed, but not when it becomes an optimization variable. To avoid this handicap, the same variable change as in (4.28) is proposed. This leads to the concave expressions

$$\bar{r}^{(1)}_{\mathrm{UL},m}(t) = B\left[w_{1,m}\sum_{i=1}^{n_{\mathrm{RS}}}\log_2\left(1+d^i_{3,m}(t)\frac{q_{1,m}}{w_{1,m}}\right)+w_{2,m}\sum_{i=1}^{n_{\mathrm{BS},m}}\log_2\left(1+d^i_{1,m}(t)\frac{q_{2,m}}{w_{2,m}}\right)\right] \tag{4.43}$$

$$\bar{r}^{(2)}_{\mathrm{UL},m}(t) = B\left[w_{1,m}\sum_{i=1}^{n_{\mathrm{BS},m}}\log_2\left(1+d^i_{1,m}(t)\frac{q_{1,m}}{w_{1,m}}\right)+w_{2,m}\sum_{i=1}^{n_{\mathrm{BS},m}}\log_2\left(1+\frac{d^i_{1,m}(t)q_{2,m}+d^i_{2,m}(t)q_{3,m}}{w_{2,m}}\right)\right] \tag{4.44}$$

and the new feasible set

$$\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{w}_1, \mathbf{w}_2 \geq \mathbf{0}_{\mathsf{M}} \tag{4.45}$$

$$\mathbf{q}_j \leq \tau_j \mathbf{1}_{\mathsf{M}}, \quad \mathbf{L}\mathbf{q}_3 \leq \tau_2 \mathbf{1}_{\mathsf{R}}, \quad \mathbf{1}^T_{\mathsf{M}}\mathbf{w}_j \leq 1, \tag{4.46}$$

where $j = 1, 2$. Thus, a convex representation of $\mathcal{R}_{\mathrm{UL}}(t;\boldsymbol{\tau})$ follows by replacing (4.36)-(4.38) by (4.43)-(4.44) in (4.42) and performing the union over the allocations satisfying (4.45)-(4.46). □

4.5 Maximum Network Utility Rate Allocation Policies

In a cellular network handling users of services with different QoS requirements, some users might exhibit large sensitivity to transmission delays while others might only pay attention to their experienced long-term throughput. And still other key performance indicators may play a role, such as transmit buffer overflow probability, energy consumption... etc; even users of the same service, attending to their contract terms, can ask for differentiated QoS requirements.

This poses a challenging problem: upon collection of CSI at the beginning of the t-th frame, the BS must decide by whom, when, and at which rate any information will be transmitted/received until the next CSI refresh arrives. Assuming that CSI updates are received periodically every D frames, user scheduling and resource allocation needs to be jointly optimized for UL and DL transmission during the frames $\{t, t+1, \ldots, t+D-1\}$. On the one hand, flows of different nature may require completely different management policies; on the other, the network operation should seamlessly reconfigure as the scenario (users, services, QoS requirements...) varies with time.

We address this problem by using utility functions [Kel98, Ng07] that evaluate each user's satisfaction given the achieved throughput as compared to its requirements. By properly characterizing QoS requirements with the dependency of the utility on the throughput for each service under operation, the different nature of the serving flows is incorporated into the BS arbitrage. An arbitrage that will use the utility function of each user to allocate resources and perform scheduling decisions.

4.5.1 User utility functions

Utility functions were first used in [Kel98] to introduce the proportional fair criterion in resource allocation problems, and allow us to describe the satisfaction of one user given its served throughput. Although many schemes implicitly assume utilities proportional to throughputs [Yu06, Bae06, Cal07e], we shall adopt here a more general approach as in [Lee06, Kel98, Lin06, Pal07, Xue06, Ng07]. We define a user utility function $U(R)$ as a concave function of the (long-term) throughput R, which is computed using the exponentially weighted smoothing

$$R(t+1) = \lambda R(t) + (1-\lambda) r(t), \tag{4.47}$$

where $R(t+1)$ stands for the throughput as seen at the beginning of the $(t+1)$-frame, λ represents the smoothing memory, and $r(t)$ is the instantaneous rate achieved in the t-th frame. If for any reason the user satisfaction profile of a service cannot be described using a concave function (e.g, an S-type curve), the finding of efficient methods for network utility maximization is compromised. Fortunately, a plethora of common services comply with the concavity constraint.

4.5.1.1 QoS-oriented utility functions

We say a utility function is QoS-oriented whenever the QoS requirements of the service appear explicitly in its expression. Although we are not constrained to it for operational reasons, we focus on utility functions upper-bounded by 1. This way, we set the same maximum user satisfaction level as a reference for all the services under operation in the network. Otherwise, the use of unbounded utilities for the different services might cause the BS to bias its attention towards services with favored utility scales. The following examples show that this is not a major impairment to describe the satisfaction profile of services of different nature.

- *Example 1: Best-effort data service*

The user satisfaction of a best-effort data service (e.g., ftp, http) without any data latency or other QoS constraints than achieving the largest possible throughput can be modeled with the utility function

$$U(R) = 1 - e^{\log(1-U_0)\frac{R}{R_0}} = 1 - (1 - U_0)^{\frac{R}{R_0}}. \tag{4.48}$$

This utility is parameterized by the satisfaction level $0 < U_0 < 1$ achieved when the throughput is R_0.

- *Example 2: Delay-sensitive service*

The user of a delay-sensitive service is interested in achieving some target throughput under the constraint that data latency remains below some critical threshold: in practical applications (e.g. voice service, video streaming), bits exceeding the maximum allowed delay are dropped. Since such applications are usually of constant bit rate, allocation of rates larger than the target throughput renders suboptimal. In other words, an overusing of resources makes no real improvement for this user but compromises the QoS provision to the rest. This can be alternatively viewed in terms of imposing the instantaneous rate to be as constant as possible, thus avoiding bursty transmissions yielding the same throughput at the expense of larger idle periods (and hence delays). One suitable utility function is

$$U(R) = 1 - \left(\frac{R - R_0}{\sigma}\right)^2, \tag{4.49}$$

where R_0 is the target throughput and σ depends on the maximum allowable delay W_0 (in number of frames). To select σ, we choose to set the utility of one user that was initially served R_0 but is laid aside W_0 frames idle equal to $U_0 \ll 1$. This forces the satisfaction index of this user to move from the peak of (4.49), where utility was 1, to some unacceptable value U_0. This way, each frame one such user is idle it is able to warn the BS about its urgency for being scheduled by decreasing its utility. With this criterion, the appropriate σ is

$$\sigma = \frac{(1 - \lambda^{W_0})R_0}{\sqrt{1 - U_0}}. \tag{4.50}$$

In case users of several delay-sensitive services with different QoS requirements (as specified by R_0 and W_0) are present in the network, one has only to adjust σ according to (4.50) and use the resulting utility function (4.49).

4.5.1.2 Best-effort utility functions

In situations where the utility function of a service does not depend on the QoS requirements, we say that utility is best-effort. As we are not able to quantify how far we are from the user expectations, we rather use utility as a qualitative satisfaction index. Additionally, if there is only one service under operation in the network, there is no reason to focus on functions upper-bounded by 1. This can allow us to consider a wider class of utility functions. A useful example of best-effort utility function is the family [Mo00]

$$U(R) = \begin{cases} \log(R) & \text{if } \alpha = 1 \\ \frac{R^{1-\alpha}}{1-\alpha} & \text{if } \alpha \neq 1 \end{cases}, \tag{4.51}$$

where the choice of the parameter α governs the way resources are shared among users, and its role shall be discussed later at the end of Section 4.6.

4.5.2 Network utility maximization

To account for services with asymmetric requirements, consider different DL and UL utility functions, denoted by $U_{\mathrm{DL},m}$ and $U_{\mathrm{UL},m}$ respectively for the m-th user. Let $\mathbf{R}_{\mathrm{DL}}(t), \mathbf{R}_{\mathrm{UL}}(t) \in \mathbb{R}_{+}^{\mathsf{M}\times 1}$ denote the vectors corresponding to the vertical stacking of DL and UL per-user throughputs at the beginning of the t-th frame, respectively. As user throughput varies with time, so does user utility. A global snapshot is given by the vectors $\mathbf{U}_{\mathrm{DL}}(t), \mathbf{U}_{\mathrm{UL}}(t) \in \mathbb{R}_{+}^{\mathsf{M}\times 1}$, where

$$[\mathbf{U}_{\mathrm{DL}}(t)]_m = U_{\mathrm{DL},m}(R_{\mathrm{DL},m}(t)) \tag{4.52}$$

and a similar expression holds for UL utilities. Using (4.52), we define network utility as any concave non-decreasing function of the user utilities $\mathrm{NU}(t) \equiv \mathrm{NU}(\mathbf{U}_{\mathrm{UL}}(t), \mathbf{U}_{\mathrm{DL}}(t))$. It provides a cell-wide aggregate indicator rating the goodness of the scheduling and resource allocation strategy carried out at the BS as far as satisfaction of all the users of the cell is concerned. For instance, we could take a maxmin approach and set network utility as the minimum among all the users' satisfaction in either UL or DL directions, i.e.,

$$\mathrm{NU}(t) = \min_{1\leq m \leq \mathsf{M}} \Big\{ \min\{[\mathbf{U}_{\mathrm{UL}}(t)]_m, [\mathbf{U}_{\mathrm{DL}}(t)]_m\} \Big\}. \tag{4.53}$$

Thanks to the concavity of each user utility on the throughput and the fact that NU is a concave non-decreasing function of the utilities, it follows from the convexity properties of composite functions [Boy04] that (4.53) is a concave function of $(\mathbf{R}_{\mathrm{UL}}(t), \mathbf{R}_{\mathrm{DL}}(t))$. This is an important property since concavity of network utility is necessary for obtaining global optimal allocation strategies in polynomial time. Alternatively, if there is no pressure to focus on the utility achieved by the worst user, we can take a simpler choice and set network utility as the sum of all the user's utilities

$$\mathrm{NU}(t) = \mathbf{1}_{\mathsf{M}}^{T}(\mathbf{U}_{\mathrm{UL}}(t) + \mathbf{U}_{\mathrm{DL}}(t)). \tag{4.54}$$

When (4.54) is used in conjunction with (4.50), the parameter α is said to enable α-fairness [Mo00]. Fairness is a wide concept which refers to the fact of not penalizing some users arbitrarily, and by tuning α from 0 to $+\infty$ the network planner is given a tool to easily switch between popular fair schemes. While $\alpha = 0$ yields utilities equal to throughputs and therefore the objective becomes maximizing the cell throughput, $\alpha = 1$ yields proportional fairness [Kel98], and as $\alpha \to +\infty$ the network operation tends to apply the maxmin criterion to the user throughputs.

4.5.2.1 Optimal strategy

For a given CSI, the task of the BS is to maximize network utility until the CSI becomes outdated and a new one is received (this period spans D frames). Afterwards, the following CSI update triggers another network utility maximization procedure for the subsequent D frames, and so on. Expressed succinctly, the optimal strategy for a given CSI is the solution to the following optimization problem[9]

$$\underset{\substack{\{\boldsymbol{\tau}_i, \mathbf{r}_{\mathrm{UL}}(t+i), \mathbf{r}_{\mathrm{DL}}(t+i)\}_{i=0}^{D-1} \\ \{\mathbf{U}_{\mathrm{UL}}(t+i), \mathbf{U}_{\mathrm{DL}}(t+i)\}_{i=1}^{D}}}{\text{maximize}} \quad \{\mathrm{NU}(t+1), \mathrm{NU}(t+2), \ldots, \mathrm{NU}(t+D)\} \tag{4.55}$$

$$\text{subject to} \quad [\mathbf{U}_{\mathrm{DL}}(t+i)]_m \leq U_{\mathrm{DL},m}\Big(\lambda^i R_{\mathrm{DL},m}(t) + (1-\lambda)\sum_{j=0}^{i-1} \lambda^{i-1-j} r_{\mathrm{DL},m}(t+j)\Big) \tag{4.56}$$

$$[\mathbf{U}_{\mathrm{UL}}(t+i)]_m \leq U_{\mathrm{UL},m}\Big(\lambda^i R_{\mathrm{UL},m}(t) + (1-\lambda)\sum_{j=0}^{i-1} \lambda^{i-1-j} r_{\mathrm{UL},m}(t+j)\Big) \tag{4.57}$$

$$\mathbf{r}_{\mathrm{UL}}(t+i) \in \mathcal{R}_{\mathrm{UL}}(t; \boldsymbol{\tau}_i) \tag{4.58}$$

$$\mathbf{r}_{\mathrm{DL}}(t+i) \in \mathcal{R}_{\mathrm{DL}}(t; \boldsymbol{\tau}_i) \tag{4.59}$$

$$\mathbf{1}_4^T \boldsymbol{\tau}_i = 1, \boldsymbol{\tau}_i \geq \mathbf{0}_4, \tag{4.60}$$

where (4.56)-(4.57) apply for $1 \leq m \leq \mathsf{M}$ and $1 \leq i \leq D$, and (4.58)-(4.60) for $0 \leq i \leq D-1$. Note that in (4.55)-(4.60) we have made implicit the resource allocation optimization with the use of the instantaneous achievable rate regions $\mathcal{R}_{\mathrm{UL}}(t; \boldsymbol{\tau}_i)$ and $\mathcal{R}_{\mathrm{DL}}(t; \boldsymbol{\tau}_i)$.

Determining the best rate allocation for the D frames under consideration amounts to solving the multiobjective optimization problem (4.55)-(4.60). Multiobjective problems do not usually have unique optimal solutions, and one usually selects one solution from the set of Pareto optimal solutions[10] according to some priorization of the objectives in conflict in the problem. Since network utility represents *cell-wide* quality of service, our approach will be to provide the largest possible network utility in each of the frames under optimization indistinctly. Hence, we will first aim at maximizing the minimum network utility during D frames, then maximize the second smallest network utility with no penalty to the previous one, and so on. Under this criterion, one *global optimal* solution[11] can be iteratively computed using Algorithm 4.1.

Proposition 4.1 *The solution computed by Algorithm 4.1 is Pareto optimal.*

Proof. See Appendix 4.A. □

[9] Note that we have omitted the dependence of network utility on each of the user utilities in (4.55) for the sake of simplicity.

[10] Some resource allocation achieving $\{\mathrm{NU}(t+1), \mathrm{NU}(t+2), \ldots, \mathrm{NU}(t+D)\}$ belongs to the Pareto optimal set if for any other allocation achieving $\{\mathrm{NU}'(t+1), \mathrm{NU}'(t+2), \ldots, \mathrm{NU}'(t+D)\}$ it will never happen that $\mathrm{NU}'(t+i) \geq \mathrm{NU}(t+i)$ for *all* $1 \leq i \leq D$ and $\mathrm{NU}'(t+i) > \mathrm{NU}(t+i)$ for *some* $1 \leq i \leq D$.

[11] In case more than one global optimal resource allocation solution exists, their achieved network utility values are permuted versions of some reference $\{\mathrm{NU}^\star(t+1), \mathrm{NU}^\star(t+2), \ldots, \mathrm{NU}^\star(t+D)\}$ (see the proof of Proposition 4.1 in Appendix 4.A).

Algorithm 4.1 Global maximization of network utility

1: Initializations: $\mathcal{S}=\emptyset$, $\mathrm{NU}_{\min}(t+i)=-\infty$ for $1\leq i\leq D$.

2: **while** $|\mathcal{S}|<D$ **do**

3: Solve

$$\underset{\substack{\{\boldsymbol{\tau}_i,\mathbf{r}_{\mathrm{UL}}(t+i),\mathbf{r}_{\mathrm{DL}}(t+i)\}_{i=0}^{D-1}\\ \{\mathbf{U}_{\mathrm{UL}}(t+i),\mathbf{U}_{\mathrm{DL}}(t+i)\}_{i=1}^{D}}}{\text{maximize}} \quad \min_{i\in\{1,2,\ldots,D\}\setminus\mathcal{S}}\{\mathrm{NU}(t+i)\} \tag{4.61}$$

$$\text{subject to} \quad \text{constraints (4.56)-(4.60)} \tag{4.62}$$

$$\mathrm{NU}(t+i)\geq \mathrm{NU}_{\min}(t+i) \quad \forall i\in\mathcal{S}. \tag{4.63}$$

4: Compute $i_{\min}=\arg\min_{i\in\{1,2,\ldots,D\}\setminus\mathcal{S}}\{\mathrm{NU}^{\star}(t+i)\}$ and update $\mathcal{S}=\mathcal{S}\cup i_{\min}$, $\mathrm{NU}_{\min}(t+i_{\min})=\mathrm{NU}^{\star}(t+i_{\min})$.

5: **end while**

6: Use the optimal resource allocation to compute the UL and DL exact achievable rates (4.2): $\{\mathbf{r}^{\star}_{\mathrm{UL}}(t+i),\mathbf{r}^{\star}_{\mathrm{DL}}(t+i)\}_{i=0}^{D-1}$.

7: Update throughputs for $1\leq m\leq \mathsf{M}$, $1\leq i\leq D$:

$$R_{\mathrm{UL},m}(t+i) = \lambda^{D}R_{\mathrm{UL},m}(t)+(1-\lambda)\sum_{j=0}^{i-1}\lambda^{i-1-j}r^{\star}_{\mathrm{UL},m}(t+j) \tag{4.64}$$

$$R_{\mathrm{DL},m}(t+i) = \lambda^{D}R_{\mathrm{DL},m}(t)+(1-\lambda)\sum_{j=0}^{i-1}\lambda^{i-1-j}r^{\star}_{\mathrm{DL},m}(t+j). \tag{4.65}$$

Remark 4.1 Algorithm 4.1 is able to compute one global optimal solution in *polynomial* time. To see this, it is sufficient to show that each of the subproblems (4.61)-(4.63) are convex. We first require that the objective (4.61) is concave, which follows from the concavity of network utility and the fact that the minimum of concave functions is concave. Then, the left hand side of each of the inequality constraints in (4.62)-(4.63), when rephrased as a function of some optimization variables less than or equal to zero, should be convex. This follows from the concavity of the user utility functions with respect to throughput, the fact that the throughput relates linearly to the instantaneous rates, and the convexity of the UL and DL achievable rate regions (see Section 4.4).

Remark 4.2 In each of the problems (4.61)-(4.63) a three-fold optimization in each of the frames under consideration is performed: first, the frame formats (relative durations of each relay-assisted transmission subphase for UL and DL) as described by the corresponding four-dimensional vectors $\{\boldsymbol{\tau}_i\}_{i=0}^{D-1}$; second, the allocated instantaneous rates, which implicitly account for user scheduling (note that $r_{\mathrm{DL},m}(t+i)=0$ implies that the m-th user shall not receive any DL data in the $(t+i)$-th frame); and third, the allocated resources (bandwidth and transmit power), which are implicit in the definition of the DL and UL achievable rate regions (4.58)-(4.59) as defined in (4.27) and (4.42).

Remark 4.3 In order to pose the maximization of network utility as a series of *convex* optimization problems, we have resorted to the concave lower bounds on the achievable rates derived in Section 4.3.2. However, the throughput updates are performed evaluating the *exact* ergodic rates (4.2) achieved by the optimal resource allocation, and not their lower bounds (4.14). As for Rayleigh fading, we resort to [Kie05] for their computation.

4.5.2.2 Reduced-complexity suboptimal strategies

Although Algorithm 4.1 provides the best network strategy from a network utility point of view, its computational load may become prohibitive in systems either serving a large number of users per cell (large M) or refreshing the CSI slowly with respect to the frame duration (large D). To see this, consider the fact that D convex problems of $(14\mathsf{M}+4)D$ variables each (where we have considered utilities, rates, frame formats, power allocations, and bandwidth allocations) are involved in each optimization instance. Hence, two directions may be taken to cut down complexity: either reduce the optimization window D or the number of users M to be scheduled.

How to deal with the first one is immediate: simply replace D by D' in Algorithm 4.1 such that D' divides D (this is required to avoid optimization windows requiring CSI not received yet). The second direction requires the implementation of a time-domain scheduler on top of Algorithm 4.1 such that only a subset of the M MS's are selected for network utility maximization. We choose to select the $\mathsf{M}' < \mathsf{M}$ users that would have the smallest utilities at the end of the optimization window if not scheduled. Despite this may have an impact on final performance since scheduling decisions are no longer optimal, time-domain pre-scheduling renders crucial in scenarios with a large number of users. On top of the aforementioned complexity issues, in practice, the frame structure needs to be signalled to the MS's, and this represents and additional overhead. If all users are allowed to transmit and/or receive in the same frame, the average quantity of the allocated resource per user and frame may go down below practical operational values while this signalling overhead may increase significantly and thus hamper network utility.

We benchmark the global optimal solution of Algorithm 4.1 against the suboptimal strategy which takes $D' = 1$ and attempts to maximize network utility *sequentially* in a frame-by-frame basis. Thus, Algorithm 4.2 is run at the beginning of each frame, which allocates resources among the subset of M' users selected by a time-domain pre-scheduler. This way, we are able to quantify the performance loss of sequential optimization versus global optimization.

Algorithm 4.2 Sequential maximization of network utility

1: Solve

$$\underset{\substack{\boldsymbol{\tau},\mathbf{r}_{\mathrm{UL}}(t),\mathbf{r}_{\mathrm{DL}}(t),\\ \mathbf{U}_{\mathrm{UL}}(t+1),\mathbf{U}_{\mathrm{DL}}(t+1)}}{\text{maximize}} \quad \mathrm{NU}(t+1) \tag{4.66}$$

$$\text{subject to} \quad [\mathbf{U}_{\mathrm{DL}}(t+1)]_m \leq U_{\mathrm{DL},m}\Big(\lambda R_{\mathrm{DL},m}(t) + (1-\lambda) r_{\mathrm{DL},m}(t)\Big) \tag{4.67}$$

$$[\mathbf{U}_{\mathrm{UL}}(t+1)]_m \leq U_{\mathrm{UL},m}\Big(\lambda R_{\mathrm{UL},m}(t) + (1-\lambda) r_{\mathrm{UL},m}(t)\Big) \tag{4.68}$$

$$\mathbf{r}_{\mathrm{UL}}(t) \in \mathcal{R}_{\mathrm{UL}}(t;\boldsymbol{\tau}) \tag{4.69}$$

$$\mathbf{r}_{\mathrm{DL}}(t) \in \mathcal{R}_{\mathrm{DL}}(t;\boldsymbol{\tau}) \tag{4.70}$$

$$\mathbf{1}_4^T \boldsymbol{\tau} = 1, \boldsymbol{\tau} \geq \mathbf{0}_4, \tag{4.71}$$

2: Use the optimal resource allocation to compute the UL and DL exact achievable rates (4.2): $\mathbf{r}^\star_{\mathrm{UL}}(t), \mathbf{r}^\star_{\mathrm{DL}}(t)$.

3: Update throughputs for $1 \leq m \leq \mathsf{M}$:

$$R_{\mathrm{UL},m}(t+1) = \lambda R_{\mathrm{UL},m}(t) + (1-\lambda) r^\star_{\mathrm{UL},m}(t) \tag{4.72}$$

$$R_{\mathrm{DL},m}(t+1) = \lambda R_{\mathrm{DL},m}(t) + (1-\lambda) r^\star_{\mathrm{DL},m}(t). \tag{4.73}$$

4.6 Simulation Results

We focus on two different scenarios sharing the same target cell spectral efficiency but having different user population sizes. In either case, we simulate a circular cell of 500 m radius, with $\mathsf{R} = 5$ relays uniformly spaced along a circle at 375 m from the BS, which is located in the cell center. We assume that the MS-RS links are in line-of-sight and have path loss exponent 2.6, while the rest of links (BS-MS and MS-RS) are in non line-of-sight with path loss exponent 4.05. All links are hampered by Rayleigh fading. See Table 4.1 for a complete list of values of the rest of physical parameters involved.

All the users of the cell are mobile. If $\mathbf{x}_m(t) \in \mathbb{R}^2_+$ denotes the position of the m-th MS at the beginning of the t-th frame in Cartesian coordinates, then

$$\mathbf{x}_m(t+1) = \mathbf{x}_m(t) + v_m T \begin{bmatrix} \cos(\varphi(t)) \\ \sin(\varphi(t)) \end{bmatrix}, \tag{4.74}$$

where v_m is its speed (assumed constant) and $\varphi(t)$ is an AR(1) stochastic process describing its direction,

$$\varphi(t+1) = 0.9\varphi(t) + 0.02\psi_t, \tag{4.75}$$

where $\{\psi_t\}$ are i.i.d. uniform random variables on $-[\pi, \pi)$. Whenever a MS happens to exceed the limits of the cell, it is forced to bounce on the cell edge by changing its instantaneous

B	10 MHz
T	25 ms
$p_{\mathrm{BS}}^{\max}$, $p_{\mathrm{RS}}^{\max}$, and $p_{\mathrm{BS}}^{\max}$	33, 30, and 24 dBm
G_{BS}, G_{RS}, and G_{BS}	10.6, 5, and −1 dB
F_{BS}, F_{RS}, and F_{BS}	4, 4, and 7 dB
n_{BS}, n_{RS}, and n_{MS}	2, 2, and 1
BS, RS, and MS heights	15, 5, and 0 m
N_0	−114 dBm/MHz
Γ	4 dB
λ	0.95

Table 4.1: Physical layer setup of the simulated scenario

direction in order to keep constant the total number of users. As a simplifying assumption we set the same speed of $v = 3$ kmph for all the MS's, and consider a feedback update rate of 100 ms, i.e., $D = 4$. Each time CSI is refreshed, each MS is attached to the RS towards which the path loss is the smallest. The connectivity matrix $\mathbf{L}$ is updated accordingly.

With this setup, we first simulate an scenario consisting of $\mathsf{M} = 6$ best-effort users (4.48) of gold, silver, and bronze QoS classes. Gold users experience 0.9 utility when they are given 30 Mbps (DL) and 6 Mbps (UL) throughput. Silver users have the same utility level when served 20 Mbps (DL) and 4 Mbps (UL) throughput. Finally, bronze users require 10 Mbps (DL) and 2 Mbps (UL) throughput for 0.9 utility. There are two users of each QoS class and, under a maxmin choice (4.53), if a *network* utility of 0.9 was realized, the cell spectral efficiency would amount to 14.4 bps/Hz.

Figure 4.4 shows the deployment layout and compares the network utility achieved by global and sequential optimization. To quantify separately the performance gains provided by the presence of relays from those achieved by the optimization approach itself, we benchmark the global and the sequential optimization algorithms against their counterparts without RS's. That is, we also simulate resource allocation where the relay-transmit phase is always forced to have zero duration. To make this comparison fair, we increase the transmit power constraint at the BS and the MS's such that, frame by frame, the total UL and DL power is equal with and without RS's in both optimization strategies. Since the number of users is relatively small, there is no need for a time-domain pre-scheduler.

As a general trend, global optimization dominates sequential optimization in the long term, although at the beginning the opposite holds. This is because in the first frames, the resource allocation of sequential optimization benefits from evaluating the actual ergodic rates often (frame by frame) as compared to global optimization, where this evaluation is carried out every

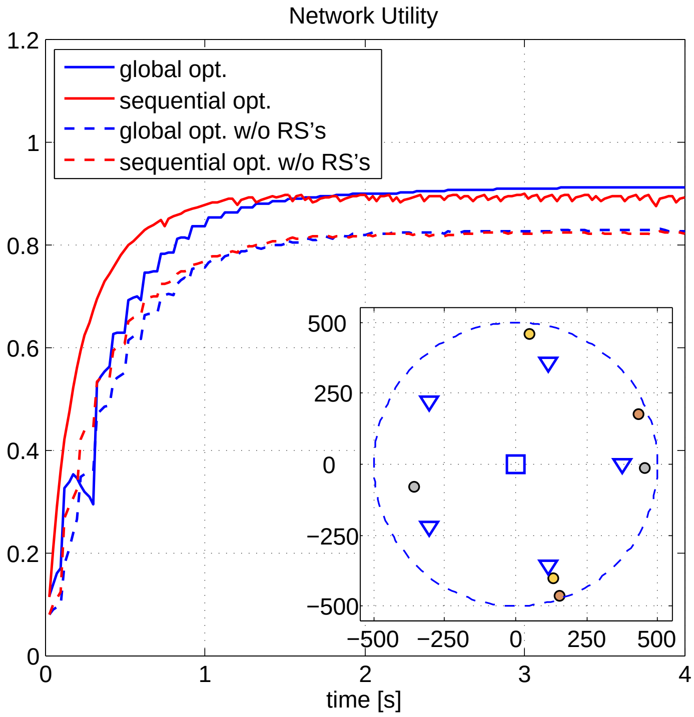

Figure 4.4: Network utility achieved by global (blue) and sequential (red) optimization, with (solid) and without (dashed) relaying infrastructure, and deployment layout.

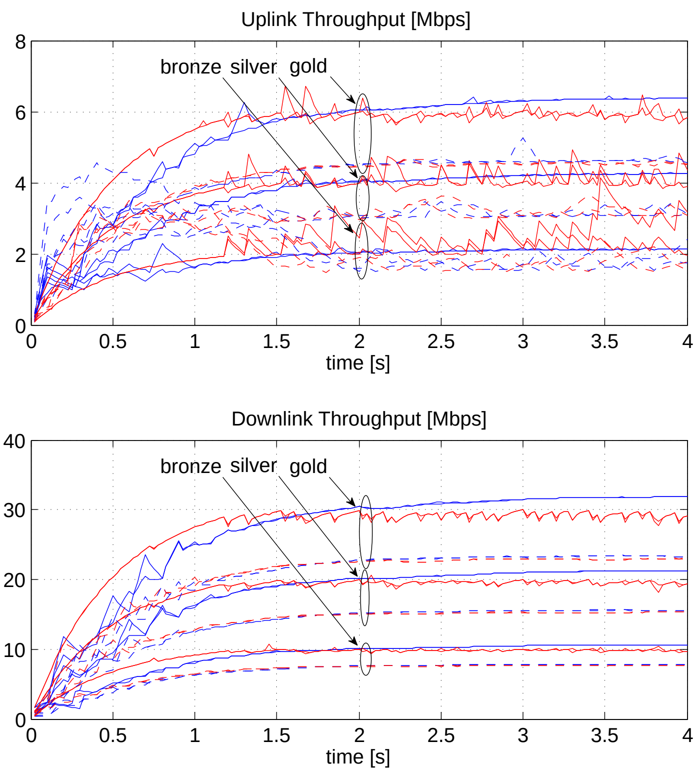

Figure 4.5: Per-user served throughput of global (blue) and sequential (red) optimization, with (solid) and without (dashed) relaying infrastructure.

Seq. Opt. w/o RS's	Global Opt. w/o RS's	Seq. Opt. w/ RS's	Global Opt. w/ RS's
1	1	2.6	33

Table 4.2: Relative execution times per transmission frame of the MATLAB implementations of all the resource allocation strategies.

group of D frames. This has a positive impact on user throughput, although this advantage vanishes quickly as confirmed also by Figure 4.5, where the per-user throughput achieved by global optimization becomes the largest in less than 2 s of network operation (80 frames).

Interestingly, sequential and global optimization perform almost undistinguishably without relays: as the ergodic capacity lower bounds are very tight in this setup (see Figure 4.2), the previous effect does not apply. In any case, the target spectral efficiency of 14.4 bps/Hz is achieved only with relaying infrastructure, as the steady-state performance without RS's falls roughly 25% below QoS targets. Indeed, as showed in Table 4.2 the price to pay is an increase in computation complexity.

Next, we focus on a practical scenario with the same target spectral efficiency at 0.9 network utility as before, adopting the maxmin utility criterion (4.53) again, but now serving 10 times more users with rate requirements reduced to 10% of the previous. Thus, $\mathsf{M} = 60$ best-effort MS's are present in the cell which split into 20 users per QoS class. Now, at 0.9 user utility gold users require 3/0.6 Mbps (DL/UL) throughput, silver users 2/0.4 Mbps (DL/UL) throughput, and bronze users 1/0.2 Mbps (DL/UL) throughput. With this system size, global optimization renders unfeasible and we restrict our attention to sequential optimization with time-domain pre-scheduling. In particular, we study the performance degradation as a function of M', the maximum number of users per frame. In this respect, Figure 4.6 shows the deployment layout and the network utility achieved with $\mathsf{M}' = 12, 8$, and 4. Clearly, the larger M', the larger network utility but also the algorithm complexity and signalling overhead. Assuming negligible this last effect for the range of values of M' studied[12], the steady-state network utility loss between $\mathsf{M}' = 12$ and $\mathsf{M}' = 4$ is on the order of 0.1.

In Figure 4.7 we show the per-user throughput averaged per QoS class, to show that, although the general trend is similar for each value of M', the performance degradation when $\mathsf{M}' = 4$ is due to the fact that gold users are served far below their 0.9 target while bronze users are satiated more than necessary. As M' decreases, the average number of idle frames between transmissions for a given user increases, but the instantaneous rate per user in an active frame increases.

[12] For each direction (UL and DL), each user should be signalled about the transmission rate, the fractional bandwidth allocation, the fractional power allocation, and the duration of the protocol subphases. Assuming that each of these parameters is quantized to n_b bits, the throughput penalty per user and direction is $n_b(6 + 2/\mathsf{M}')$ which decreases with M' since the subphase durations are common. Hence, although the global signalling overhead increases in M', the per-user penalty decreases. In particular for a practical value of $n_b = 5$ bits, this penalty is on the order of 1 kbps and, hence, negligible.

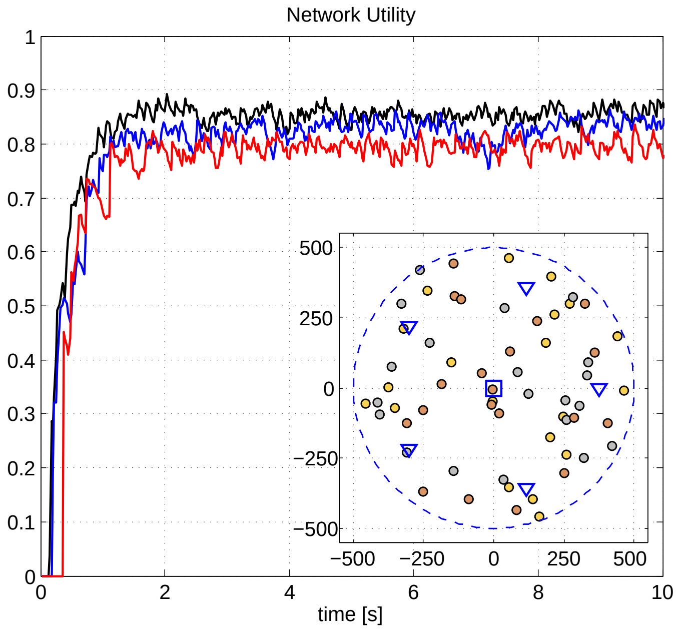

Figure 4.6: Network utility achieved by sequential optimization with time-domain pre-scheduling allowing a maximum of 12 (black), 8 (blue), and 4 (red) users per frame, and deployment layout.

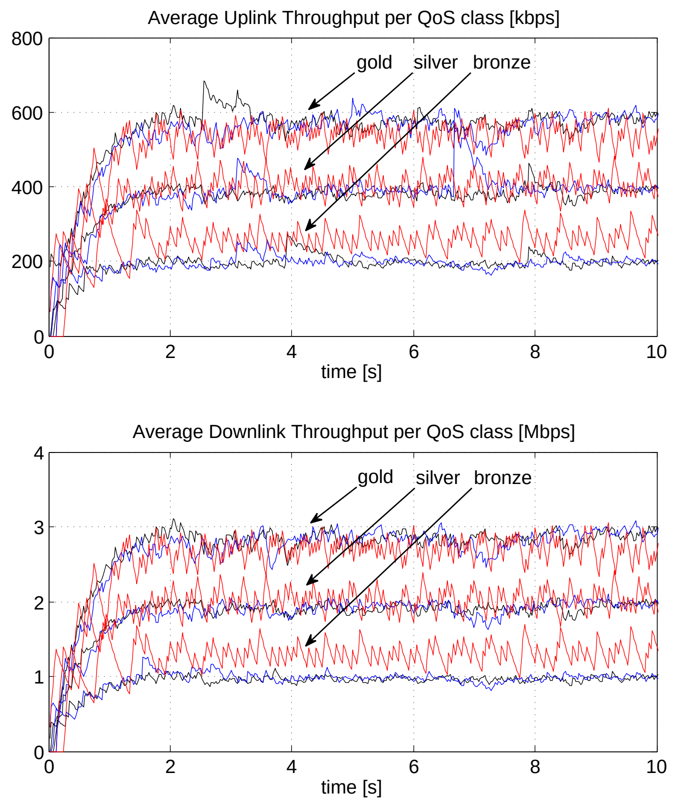

Figure 4.7: Average per-user served throughput per QoS class achieved by sequential optimization with time-domain pre-scheduling allowing a maximum of 12 (black), 8 (blue), and 4 (red) users per frame.

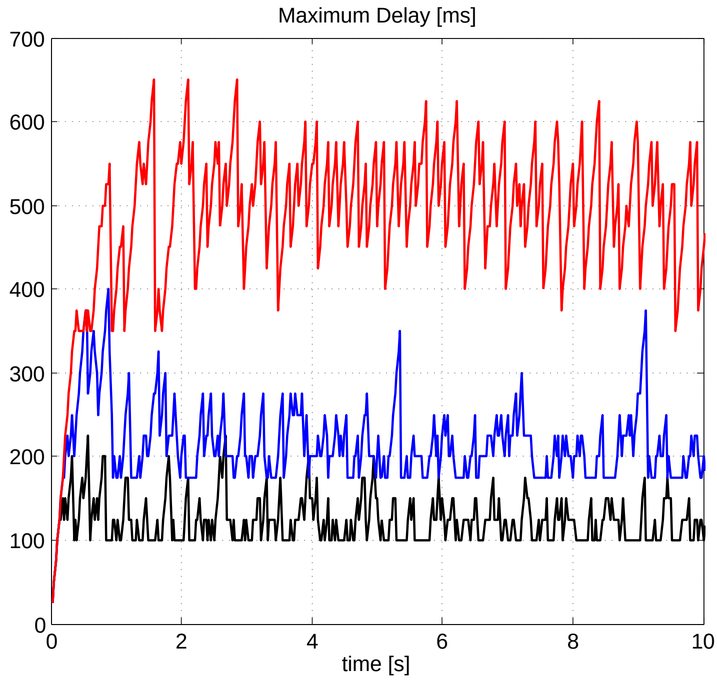

Figure 4.8: Maximum user delay (number of frames idle) achieved by sequential optimization with time-domain pre-scheduling allowing a maximum of 12 (black), 8 (blue), and 4 (red) users per frame.

The conclusion from Figure 4.7 is that if M' is too low the global effect is negative. Finally, Figure 4.8 addresses the maximum per-user delay in each setup, which is shown to be roughly proportional to $1/\mathsf{M}'$ in average.

So far, we have only explored maxmin network utility and QoS-oriented utility functions. To conclude this section, however, we shall modify the previous system with a total of $\mathsf{M} = 60$ users and $\mathsf{M}' = 4$ users at most per frame and explore the inherent tradeoff between fairness and throughput when other network utility functions are used. In particular, we now consider the situation where the UL and DL utility functions of every user are given by (4.51) and network utility is sum utility (4.54). Thus, by changing the parameter α we have a way of trading fairness and throughput. We capture this relation in Figure 4.9, where we focus on the average steady-state per-user and link direction throughput (UL and DL directions are averaged together as their utilities are the same). It is plotted against Jain's fairness index [Jai84] in each direction for a given cell deployment. This index rates the degree of fairness incurred in serving n competing flows with a real number between $1/n$ (worst case: one user gets it all) and 1 (best case: resources are equally shared). Figure 4.9 confirms that a simultaneous increase in both fairness and throughput cannot be achieved and quantifies the explicit tradeoff. Notice that the time-domain pre-scheduler prevents the fairness index to fall below acceptable levels.

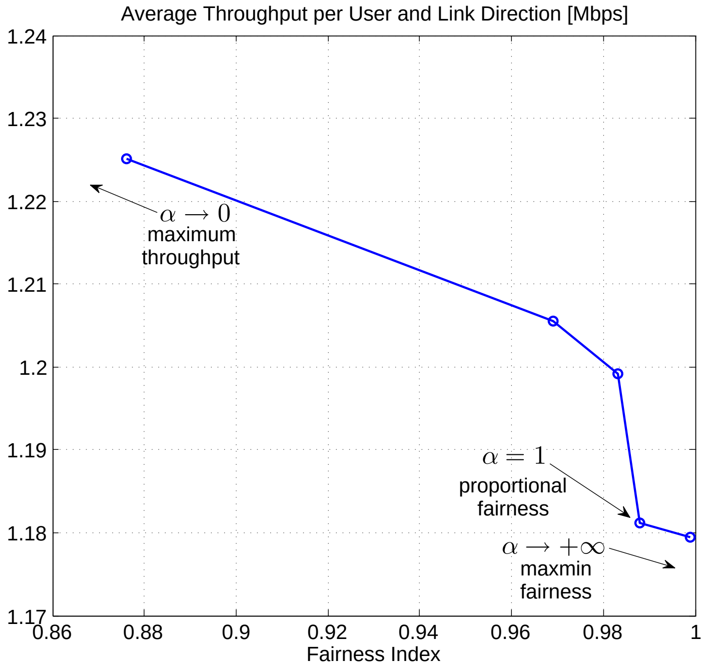

Figure 4.9: Average steady-state per-user and link-direction throughput versus fairness index. The corresponding values of α are 0.25, 0.50, 0.75, 1, and 5.

4.7 Conclusions

The work of this chapter concerns the performance characterization of a relay-assisted network deployment under practical constraints. In particular, terminals are half-duplex MIMO and path loss is the only CSI assumed to be known at the transmitters. Network performance is evaluated in terms of ergodic achievable rates by the development of novel lower bounds. These bounds are employed to derive two efficient algorithms for resource allocation optimization under heterogeneous QoS requirements. The first one provides one Pareto optimal solution whereas the second one performs a simpler frame by frame optimization by means of a sequential algorithm. The performance of both schemes has been evaluated showing that, whenever global optimization can be afforded, significant performance gains can be achieved with respect to sequential optimization. When systems dimensions are large, however, the complexity of sequential network optimization can be tuned by using time-domain pre-scheduling.

The proposed network resource optimization schemes could be generalized in at least two ways. First, by incorporating outage events into system design in those scenarios where the number of tones is limited and moderate. Second, by considering a multi-cell configuration allowing the incorporation of a prescribed maximum inter-cell interference as an additional constraint.

4.A Appendix: Proof of Proposition 4.1

Let $\{\mathrm{NU}^\star(t+1), \mathrm{NU}^\star(t+2), \ldots, \mathrm{NU}^\star(t+D)\}$ denote the network utility achieved by the solution of Algorithm 4.1, and define $i_1, i_2, \ldots, i_D$ such that $\mathrm{NU}^\star(t+i_1) \leq \mathrm{NU}^\star(t+i_2) \leq \ldots \leq \mathrm{NU}^\star(t+i_D)$. Assume it is *not* Pareto optimal. Hence there exists at least another allocation achieving $\{\mathrm{NU}(t+1), \mathrm{NU}(t+2), \ldots, \mathrm{NU}(t+D)\}$ and integers $j_1, j_2, \ldots, j_D$ such that $\mathrm{NU}(t+j_1) \leq \mathrm{NU}(t+j_2) \leq \ldots \leq \mathrm{NU}(t+j_D)$,

$$\begin{aligned} \mathrm{NU}(t+i) &\geq \mathrm{NU}^\star(t+i), \quad 1 \leq i \leq D, & (4.76)\\ \mathrm{NU}(t+i) &> \mathrm{NU}^\star(t+i), \quad \text{for some } i. & (4.77) \end{aligned}$$

Then,

$$\mathrm{NU}^\star(t+i_1) \overset{(a)}{\geq} \mathrm{NU}(t+j_1) \overset{(b)}{\geq} \mathrm{NU}^\star(t+j_1) \overset{(c)}{\geq} \mathrm{NU}^\star(t+i_1) \tag{4.78}$$

where (a) follows by construction of the solution of Algorithm 4.1, (b) from (4.76), and (c) from the definition of $\{i_n\}$. Equation (4.78) implies that $\mathrm{NU}^\star(t+i_1) = \mathrm{NU}(t+j_1)$. Two cases arise now:

- Case $i_1 \neq j_1$:

$$\mathrm{NU}^\star(t+j_1) \overset{(a)}{\geq} \mathrm{NU}^\star(t+i_1) \overset{(b)}{=} \mathrm{NU}(t+j_1) \overset{(c)}{\geq} \mathrm{NU}^\star(t+j_1), \tag{4.79}$$

  where (a) follows from the definition of $\{i_n\}$, (b) is a consequence of (4.78), and (c) follows from (4.76). Then, $\mathrm{NU}^\star(t+j_1) = \mathrm{NU}(t+j_1)$. Without loss of generality we can set $i_2 = j_1$, $j_2 = i_1$ and obtain $\mathrm{NU}^\star(t+i_2) = \mathrm{NU}(t+j_2)$.

- Case $i_1 = j_1$:

$$\mathrm{NU}^\star(t+i_2) \overset{(a)}{\geq} \mathrm{NU}(t+j_2) \overset{(b)}{\geq} \mathrm{NU}^\star(t+j_2) \overset{(c)}{\geq} \mathrm{NU}^\star(t+i_2) \tag{4.80}$$

  where (a) follows by construction of the solution of Algorithm 4.1, (b) is derived from (4.76), and (c) follows from the fact that $j_2 \neq j_1 = i_1$ and the definition of $\{i_n\}$. Expression (4.80) also implies $\mathrm{NU}^\star(t+i_2) = \mathrm{NU}(t+j_2)$.

Proceeding similarly, we can iteratively show that

$$\mathrm{NU}^\star(t+i_n) = \mathrm{NU}(t+j_n), \quad 1 \leq n \leq D, \tag{4.81}$$

which implies that the network utility values of both strategies are either equal or related by an arbitrary permutation. In either case, this contradicts (4.76)-(4.77), hence implying that the solution to Algorithm 4.1 is Pareto optimal. □

Chapter 5

Multiuser Interference and Evaluation of Capacity Regions

The computation of the channel capacity of a discrete memoryless channel is a convex problem that can be efficiently solved using alternate maximization methods. The presence of multiuser interference, however, complicates the extension of these methods to the evaluation of the capacity region of multiterminal networks. To start with, the capacity region of a general network is unknown, and we are left with only the cut-set outer bound [Cov06, Sec. 15.10]. Although many major breakthroughs in the field have been achieved (see [Cov06, Ch. 15] and [Wei06] and references therein), there are many open problems on single-letter characterizations of capacity regions. Second, not in all the particular setups where the capacity region is known it admits a computable characterization. By computable characterization we mean that the number of degrees of freedom that need to be explored in the evaluation of the achievable rates are finite (see Chapter 3, Section 3.2, for an example non-computable capacity region). Let alone the fact that, when computable expressions of capacity regions are known, multiuser interference makes their evaluation depend on the solution to non-convex problems.

Problems that are initially regarded as non-convex need not be necessarily difficult to be solved, as some hidden convexity enabling the use of efficient optimization methods can sometimes be found. This chapter is hence devoted to study the nature of the non-convexities that arise in the problems associated to the evaluation of the performance limits (capacity regions) of multiuser channels. The interest in doing so is three-fold:

1. By characterizing the class of non-convexities involved in the capacity region of a network setup, we can accurately quantify the impact of multiuser interference on the computational complexity of the evaluation of the achievable rates.

2. The study of these optimization problems often gives back some insight on the structure of the transmission strategies that achieve boundary points of capacity regions.

3. In some cases, upon unveiling the structure of the optimization problem at hand, efficient methods to compute inner and outer bounds on the capacity regions can be found.

Constrained by the limited amount of results on single-letter characterizations of capacity regions, our approach will not be general and we shall rather restrict our focus to two specific network configurations: the discrete memoryless multiple access channel (DMAC) and the degraded discrete memoryless broadcast channel (dDMBC). Neither the general broadcast nor interference channels shall be tackled due to the lack of capacity results.

5.1 Introduction

The evaluation of the capacity of an arbitrary single-user memoryless channel is a problem that admits a single-letter representation in the form of a maximization of a concave function over a convex set, e.g., a probability simplex for the Discrete Memoryless Channel (DMC). This is a convex problem that can be numerically evaluated efficiently in practice (i.e., with polynomial time worst-case complexity) [Boy04]. For the continuous Gaussian channel, for example, the solution admits a simple closed-form characterization as already obtained by Shannon in 1948 [Sha48]. For the DMC there is no closed-form expression but the popular practical Arimoto-Blahut (AB) algorithm [Ari72, Bla72], dating back to the early 1970s, can be efficiently used. Things are quite different in the multiuser case.

5.1.1 The DMAC

Fortunately, for the multiple-access channel (MAC) we have a single-letter representation of the capacity region [Ahl71, Lia72]. However, the characterization is not generally in the form of a convex optimization problem as happened in the single-user case. Again, for the continuous Gaussian channel, the characterization simplifies to a convex problem which can always be numerically evaluated in an efficient way [Yu04]. For the DMAC, however, the characterization is not in the form of a convex problem and, as a consequence, there is no efficient algorithm to compute the capacity region in practice. In this context, many authors have recently contributed toward the computation of the sum-capacity (or total capacity) of an arbitrary DMAC [Wat96, Wat02], and an algorithm for its exact computation has been found [Rez04].

It was shown in [Wat96] that any two-user DMAC can be decomposed into a finite number of elementary (two-user binary-input and binary-output) DMACs for which their total capacity can be computed by applying a necessary and sufficient condition for optimality of the input probability distributions. In addition, [Wat96] showed that for any 2-user non-binary inputs and binary output DMAC, the total capacity can be determined by computing the total capacities of the elementary DMACs of its decomposition, and provided an iterative algorithm. In a later work, [Wat02] extended the decomposition result for the K-user DMAC (with arbitrary input and output alphabets), which allowed [Rez04] to propose an algorithm for the computation of the total capacity based on a generalization of the AB algorithm. Other applications of generalizations of the AB algorithm can be found in the context of the computation of channel capacity with side information [Dup04].

Regarding the computation of the whole capacity region of the DMAC, not much work has been done due to the intractability of the problem because of its non-convexity. As a consequence of the non-convexity, brute-force algorithms or random search methods seem to be the only alternative to compute inner bounds on the capacity region with no quantification on the suboptimality.

In this chapter, we shall show that the key difficulty in computing the capacity region of an arbitrary DMAC can be identified as a rank-one constraint (a non-convex constraint) in an otherwise convex optimization problem. Optimization problems with this kind of constraint arise in areas such as control theory [Apk99, Hen01, Ors04] and signal processing [Jal03, Ma02], and cannot be solved optimally in polynomial time with state of the art knowledge. Hence, alternative suboptimal methods must be used to obtain good approximations of the capacity region. One approach that has reported near optimal performance when dealing with rank-one constraints in maximum-likelihood single-user [Jal03, Wie05] and multiuser [Ma02] detection is the use of relaxation methods. Relaxation methods are based on i) replacing the rank-one constrained problem by an approximate (not equivalent) tractable convex problem and ii) generating a potential solution to the original problem from the solution to the relaxed problem. This way, the use of computationally demanding algorithms is avoided since efficient interior point methods can be used to solve the convex approximation of the original problem in polynomial time.

We propose two efficient methods for the computation of both an inner and an outer bound of the capacity region of any arbitrary DMAC. The outer bound follows from removing the rank-one constraint and corresponds to the achievable rates in the situation of full transmit cooperation, since user codewords can thus be arbitrarily correlated. To generate potential solutions to the original problem, we first focus on randomization, an approach that has shown near optimal numerical performance in the previously mentioned areas. In essence, several rank-one input probability distributions are generated close (in the mean sense) to the optimal solution of the relaxed problem and the one yielding the largest achievable rates is kept. This can be viewed as a random search algorithm with guidance on the correlation matrix of the potential solutions.

Pursuing a simpler method allowing for performance analysis, we then study a deterministic alternative in which the solution to the relaxed problem is projected onto the feasible set via a minimum divergence criterion. This criterion yields a candidate solution which turns out to be the marginalization of the relaxed solution, a very simple operation scalable with the number of users. Regarding analytical results, there exists a class of channels for which this algorithm is able to compute exactly the capacity region. It comprises the subclass of channels with identical inner and outer bounds and the subclass of channels with strict outer bound and tight inner bound. Given a channel, we derive necessary and sufficient conditions for checking whether it belongs to the first subclass and sufficient conditions for verifying whether it belongs to the second subclass. These conditions are used to show that for all the two-user binary-input deterministic DMACs as well as for some non-deterministic channels simple marginalization of the full cooperation solution achieves capacity. Although we have not been able to fully characterize analytically the class of channels for which marginalization is optimal, numerical simulations for various channels show that, in practice, both randomization and marginalization perform indistinguishably to the optimal solution obtained with a computationally intensive brute-force full search.

5.1.2 The dDMBC

The two-user discrete memoryless broadcast channel (DMBC) [Cov72] models the situation in which one sender wishes to send information to two receivers. Although a single-letter characterization of the capacity region of this channel is not known in general, some achievable rate regions and outer bounds have been derived by Cover and van der Meulen [Meu75, Cov75], Marton [Mar79], and Nair and El Gamal [Nai07].

Achievable rate regions and outer bounds for this channel are usually expressed in terms of auxiliary random variables. We say a region is computable whenever the cardinalities of the support set of all the random variables involved are bounded. In this respect, Nair's outer bound and Cover-van der Meulen's achievable region in case of binary inputs [Haj79] are computable. However, apart from showing the feasibility of the evaluation of the regions, little work has been done towards finding efficient methods for their computation. We shall try to bridge this gap in the specific setup where the DMBC is degraded and there is no common information to be transmitted simultaneously to both receivers. Assuming these constraints, the capacity region of the channel is known [Ber73, Wyn73, Gal74] and computable.

By posing the computation of the capacity region of the two-user degraded DMBC (dDMBC) as a constrained optimization problem we show that it can be characterized as a difference of convex (DC) optimization problem [Hor99]. Since this kind of problems are not convex in general, there is no efficient algorithm available to compute the capacity region in practice. Additionally, the lack of convexity of this problem makes the Karush-Kuhn-Tucker (KKT) conditions [Boy04] not sufficient, not even necessary (some regularity conditions are required), for claiming optimality of an input distribution. Regarding this, we are able to show that in this problem they are necessary optimality conditions and, therefore, by choosing the best among those input distributions satisfying the KKT conditions the capacity region can be optimally computed. Moreover, since the dimensionality of the set of distributions satisfying the KKT conditions is in general substantially lower than the original feasible set, the complexity involved in computing the capacity region is greatly reduced. These results are applied to evaluate the capacity region of the BEC-BSC dDMBC.

5.1.3 Summary of contributions

The study of the efficient evaluation of multiuser capacity regions has led to the following contributions:

- The problem of the computation of the capacity region of the MAC has been rephrased in an alternative way that condensed all the non-convexities of the problem in a single rank-one constraint.
- This alternative formulation, which applies to any number of users, has paved the way to two efficient relaxation methods that provide inner and outer bounds on the capacity region: marginalization and randomization.
- The mathematical amenability of marginalization has allowed to determine analytical conditions under which its bounds are tight. These conditions have been applied to rederive existing capacity results but also to find the capacity region of new channels.
- The problem of the computation of the capacity region of the BC has been shown to be a DC problem.
- Necessary conditions for the optimality of an input distributions have led to a way of reducing the degrees of freedom that need to be explored to compute the capacity region, as shown for the BEC-BSC dDMBC.

This chapter consists of three sections. First, Section 5.2 is devoted to the DMAC. In particular, Section 5.2.1 introduces the problem of the computation of the capacity region of the DMAC and reformulates it as a rank-one constrained optimization problem. Section 5.2.2 describes the proposed relaxation-based methods for the computation of inner and outer bounds on the capacity region: randomization and marginalization. Analytical optimality conditions that determine when the marginalization bounds are tight are provided in Section 5.2.3. Then, the performance of the proposed algorithms among various channels is numerically compared to that of a random search method in Section 5.2.4.

Second, Section 5.3 concerns the dDMBC. We start again by introducing the problem of the computation of its capacity region in Section 5.3.1, where we also reformulate it as a DC optimization problem. Next, Section 5.3.2 describes the necessary optimality conditions and how they can be applied to obtain capacity-achieving distributions. Then, Section 5.3.3 provides the capacity region of the BEC-BSC dDMBC by using the results of Section 5.3.2.

Finally, Section 5.4 concludes the chapter by sketching its main contributions.

5.2 The DMAC

5.2.1 The capacity region as a rank-one constrained optimization problem

The computation of the capacity region of an arbitrary DMAC (a convex set) is a non-convex problem. It can be formulated in a matrix form that reveals all the non-convexity of the problem as a single rank-one constraint.

5.2.1.1 The problem of the capacity region for two users

The capacity region $\mathcal{C}$ of the two-user DMAC is the convex hull of the set[1] of rate pairs (R_1, R_2) satisfying

$$0 \leq R_1 \leq I(X_1; Y|X_2) \tag{5.1}$$

$$0 \leq R_2 \leq I(X_2; Y|X_1) \tag{5.2}$$

$$R_1 + R_2 \leq I(X_1 X_2; Y) \tag{5.3}$$

for a distribution of the form $P_{X_1X_2Y} = P_{X_1}P_{X_2}P_{Y|X_1X_2}$ on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$, where the input alphabets can be characterized as

$$\mathcal{X}_k = \{x_k^{(1)}, \ldots, x_k^{(|\mathcal{X}_k|)}\}, \quad k = 1, 2. \tag{5.4}$$

P_{X_k} is the input probability distribution of the k-th user ($k = 1, 2$), and $P_{Y|X_1X_2}$ is the given conditional distribution that characterizes the channel. It is well known that $\mathcal{C}$ is a convex set [Cov06, Thm. 14.3.2] and hence, by applying the supporting hyperplane theorem [Boy04, Sec. 2.5.2], the computation of the capacity region can be parameterized for $\theta \in [0, 1]$ as[2]

$$\underset{R_1,R_2,P_{X_1},P_{X_2}}{\text{maximize}} \quad \theta R_1 + (1-\theta)R_2 \tag{5.5}$$

$$\text{subject to} \quad 0 \leq R_1 \leq \sum_{x_1,x_2,y} P_{X_1}(x_1)P_{X_2}(x_2)P_{Y|X_1X_2}(y|x_1x_2)\log\frac{P_{Y|X_1X_2}(y|x_1x_2)}{\sum_{x_1'} P_{X_1}(x_1')P_{Y|X_1X_2}(y|x_1'x_2)} \tag{5.6}$$

$$0 \leq R_2 \leq \sum_{x_1,x_2,y} P_{X_1}(x_1)P_{X_2}(x_2)P_{Y|X_1X_2}(y|x_1x_2)\log\frac{P_{Y|X_1X_2}(y|x_1x_2)}{\sum_{x_2'} P_{X_2}(x_2')P_{Y|X_1X_2}(y|x_1x_2')} \tag{5.7}$$

$$R_1+R_2 \leq \sum_{x_1,x_2,y} P_{X_1}(x_1)P_{X_2}(x_2)P_{Y|X_1X_2}(y|x_1x_2)\log\frac{P_{Y|X_1X_2}(y|x_1x_2)}{\sum_{x_1',x_2'} P_{X_1}(x_1')P_{X_2}(x_2')P_{Y|X_1X_2}(y|x_1'x_2')} \tag{5.8}$$

$$\sum_{x_k} P_{X_k}(x_k) = 1 \ , \ P_{X_k}(x_k) \geq 0 \ \forall x_k \in \mathcal{X}_k \quad k = 1, 2, \tag{5.9}$$

[1] The convex hull is strictly necessary for convexification of $\mathcal{C}$ since otherwise it may not be convex in general [Bie79].

[2] Unless the logarithm basis is indicated, it can be chosen arbitrarily as long as both sides of the equation have the same units.

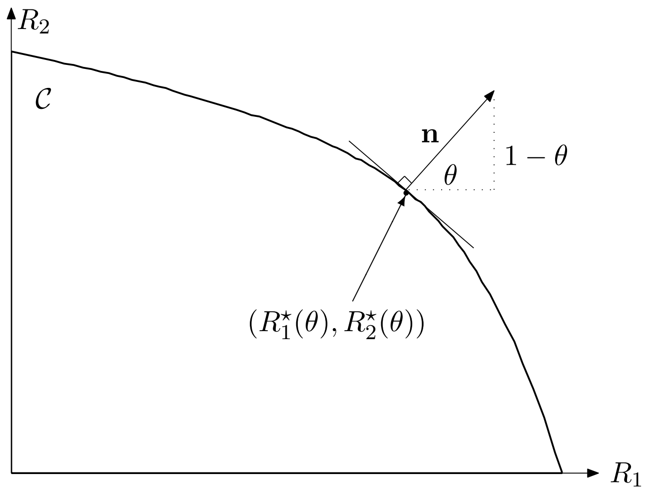

Figure 5.1: The boundary of $\mathcal{C}$ is obtained solving (5.5)-(5.9) for each $\theta \in [0, 1]$.

where the expressions (5.6)-(5.8) correspond to (5.1)-(5.3), respectively, instantiated for the DMAC. Note that the solutions $P^\star_{X_1}(\theta)$ and $P^\star_{X_2}(\theta)$ generally depend on θ. For each θ, the problem (5.5)-(5.9) computes the intersection between the contour of the capacity region and a tangent hyperplane with normal vector $\mathbf{n} = [\theta,\ 1-\theta]^T$, as illustrated in Figure 5.1. Hence, the capacity region is computed when (5.5)-(5.9) is solved for all $\theta \in [0, 1]$ and the convex hull of $\{R_1^\star(\theta), R_2^\star(\theta)\}_{\forall\theta\in[0,1]}$ is taken; in other words, the solutions $(R_1^\star(\theta), R_2^\star(\theta))$ are samples of the boundary of $\mathcal{C}$.

5.2.1.2 A rank-one constrained optimization problem

The problem (5.5)-(5.9) of the computation of the capacity region is non-convex because the constraints (5.6)-(5.8) are not jointly convex in P_{X_1} and P_{X_2}. For instance, the right hand side of the constraint in (5.8) is not concave (note that it should be concave for the problem to be convex). To see this, observe that even though $x\log(1/x)$ is concave, the composition with a linear combination of terms of the form xy is not[3]. Similar reasonings may be applied to the constraints (5.6) and (5.7) to obtain again that the lack of convexity follows from the presence of the product terms $P_{X_1}(x_1)P_{X_2}(x_2)$.

Although the problem (5.5)-(5.9) is not jointly convex in (P_{X_1}, P_{X_2}), it is separately convex in each of the input probability distributions. This would allow us to perform an alternate optimization procedure: $P^{(0)}_{X_1} \to P^{(0)}_{X_2} \to P^{(1)}_{X_1} \to P^{(1)}_{X_2} \to \ldots$, where $P^{(n)}_{X_k}$ denotes the optimal solution P_{X_k} at the n-th iteration. However, alternate optimization procedures applied to non-convex problems do not generally converge to global maxima of the cost function, and for this particular problem they do not yield acceptable results (as can be verified by numerical simulations).

[3]It is sufficient to note that the Hessian of $f(x, y) = xy\log(xy)$ has one positive and one negative eigenvalue at $(x, y) = (1/\sqrt{2}, 1/\sqrt{2})$.

Interestingly, if we allow the variables X_1 and X_2 to be dependent with distribution $P_{X_1X_2}$, then problem (5.5)-(5.9) becomes convex (recall that $x\log(1/x)$ is a concave function).

We will now reformulate the problem with a matrix-vector notation. Each of the input probability distributions P_{X_k} admits a vector representation of the form $\mathbf{p}_k$, where $[\mathbf{p}_k]_i = P_{X_k}(x_k^{(i)})$, $1 \le i \le |\mathcal{X}_k|$, $k = 1, 2$, while the joint distribution admits a matrix representation of the form $\mathbf{P}$, where $[\mathbf{P}]_{i,j} = P_{X_1X_2}(x_1^{(i)}, x_2^{(j)})$, $1 \le i \le |\mathcal{X}_1|$, $1 \le j \le |\mathcal{X}_2|$. Then, we define $\mathcal{P}_{\mathsf{prod}}$ as the subset containing all the *product* distributions (P_{X_1}, P_{X_2}) of X_1 and X_2,

$$\mathcal{P}_{\mathsf{prod}}=\left\{\mathbf{P}\in\mathbb{R}^{|\mathcal{X}_1|\times|\mathcal{X}_2|}\middle|[\mathbf{P}]_{i,j}=P_{X_1}(x_1^{(i)})P_{X_2}(x_2^{(j)})\text{ for some feasible }(P_{X_1},P_{X_2})\text{ on }\mathcal{X}_1\times\mathcal{X}_2\right\}. \tag{5.10}$$

For any joint probability matrix $\mathbf{P} \in \mathbb{R}^{|\mathcal{X}_1|\times|\mathcal{X}_2|}$, $\mathbf{P} \in \mathcal{P}_{\mathsf{prod}}$ is equivalent to $\operatorname{rank}(\mathbf{P}) = 1$, and hence the following simpler equivalent description of $\mathcal{P}_{\mathsf{prod}}$ can be given

$$\mathcal{P}_{\mathsf{prod}} = \left\{\mathbf{P}\in\mathbb{R}^{|\mathcal{X}_1|\times|\mathcal{X}_2|} \mid \operatorname{rank}(\mathbf{P}) = 1\ ,\ \mathbf{P}\ge\mathbf{0}\ ,\ \mathbf{1}^T\mathbf{P}\mathbf{1} = 1\right\}, \tag{5.11}$$

where $\ge$ denotes component-wise as well as scalar inequality indistinctly and $\mathbf{1}$ is an all-one column vector of appropriate length. The original problem (5.5)-(5.9) can now be expressed in an equivalent matrix form formulated in terms of $\mathbf{P} \in \mathcal{P}_{\mathsf{prod}}$, the joint distribution of X_1 and X_2, and its marginals $\mathbf{p}_1$ and $\mathbf{p}_2$, making use of expression (5.11) and the fact that

$$\mathbf{P}\mathbf{1} = \mathbf{p}_1, \quad \mathbf{P}^T\mathbf{1} = \mathbf{p}_2, \tag{5.12}$$

i.e., that P_{X_1} and P_{X_2} are the marginal distributions of $P_{X_1X_2}$. The following reformulation of the problem is the key point of the identification of (5.5)-(5.9) as a rank-one non-convex optimization problem.

Proposition 5.1 *The problem* (5.5)-(5.9) *of the computation of the capacity region of an arbitrary two-user DMAC is equivalent to the following rank-one non-convex optimization problem*

$$\begin{aligned}
\underset{R_1,R_2,\mathbf{P},\mathbf{p}_1,\mathbf{p}_2}{\mathsf{maximize}}\quad & \theta R_1 + (1-\theta)R_2 && (5.13)\\
\mathsf{subject\ to}\quad & 0 \le R_1 \le f_1(\mathbf{P},\mathbf{p}_2) && (5.14)\\
& 0 \le R_2 \le f_2(\mathbf{P},\mathbf{p}_1) && (5.15)\\
& R_1 + R_2 \le f_{12}(\mathbf{P}) && (5.16)\\
& \mathbf{P}\mathbf{1} = \mathbf{p}_1\ ,\ \mathbf{P}^T\mathbf{1} = \mathbf{p}_2 && (5.17)\\
& \mathbf{P}\ge\mathbf{0}\ ,\ \mathbf{1}^T\mathbf{P}\mathbf{1} = 1 && (5.18)\\
& \operatorname{rank}(\mathbf{P}) = 1, && (5.19)
\end{aligned}$$

where

$$f_1(\mathbf{P}, \mathbf{p}_2) \triangleq \sum_{i,j,y} [\mathbf{P}]_{i,j} P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \log \frac{P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})[\mathbf{p}_2]_j}{\sum_{i'}[\mathbf{P}]_{i',j} P_{Y|X_1X_2}(y|x_1^{(i')}x_2^{(j)})} \tag{5.20}$$

$$f_2(\mathbf{P}, \mathbf{p}_1) \triangleq \sum_{i,j,y} [\mathbf{P}]_{i,j} P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \log \frac{P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})[\mathbf{p}_1]_i}{\sum_{j'}[\mathbf{P}]_{i,j'} P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j')})} \tag{5.21}$$

$$f_{12}(\mathbf{P}) \triangleq \sum_{i,j,y} [\mathbf{P}]_{i,j} P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \log \frac{P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})}{\sum_{i',j'}[\mathbf{P}]_{i',j'} P_{Y|X_1X_2}(y|x_1^{(i')}x_2^{(j')})} \tag{5.22}$$

are concave in $(\mathbf{P}, \mathbf{p}_2)$*,* $(\mathbf{P}, \mathbf{p}_1)$*, and* $\mathbf{P}$*, respectively.*

Proof. See Appendix 5.A. □

Observe that if (5.19) were removed, the resulting problem would be convex.

While (5.18) ensures that $\mathbf{P}$ is a feasible probability matrix, (5.19) constrains it to $\mathcal{P}_{\mathsf{prod}}$, and (5.17) relates $\mathbf{P}$ with its marginals. The next results shows that, surprisingly, there is no need to include (5.17) as long as $(\mathbf{p}_1, \mathbf{p}_2)$ are feasible probability vectors on $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively.

Lemma 5.1 *The optimal solution to the optimization problem* (5.13)-(5.19) *does not change if* (5.17) *is replaced by*

$$\mathbf{p}_k \geq \mathbf{0}\ ,\ \mathbf{1}^T\mathbf{p}_k = 1 \quad k = 1, 2, \tag{5.23}$$

where $(\mathbf{p}_1, \mathbf{p}_2)$ *are probability vectors associated to the distributions of* X_1 *and* X_2 *that need not be the marginals induced by* $\mathbf{P}$*.*

Proof. We use $\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2$ to denote the marginals induced by $\mathbf{P}$. Since $(\mathbf{p}_1, \mathbf{p}_2)$ only affect the functions f_1 and f_2, it suffices to show that $\tilde{\mathbf{p}}_2$ is the vector distribution that maximizes f_1.

$$f_1(\mathbf{P}, \mathbf{p}_2) = f_1(\mathbf{P}, \tilde{\mathbf{p}}_2) - D(\tilde{\mathbf{p}}_2||\mathbf{p}_2) \leq f_1(\mathbf{P}, \tilde{\mathbf{p}}_2), \tag{5.24}$$

where $D(\mathbf{a}||\mathbf{b})$ is the Kullback-Leibler divergence between two probability distributions denoted by $\mathbf{a}$ and $\mathbf{b}$. By symmetry we can induce that $\tilde{\mathbf{p}}_1$ is the maximizing vector distribution of f_2. □

Remark 5.1 *Note that* (5.13)-(5.19) *can be extended to include cost constraints. Let* E_1 *and* E_2 *be the average expense requirements of user 1 and 2, respectively. By specifying the vectors* $\mathbf{e}_k \in \mathbb{R}_+^{|\mathcal{X}_k|}$*, where* $[\mathbf{e}_k]_i$ *is the expense of using* $x_k^{(i)}$*, the i-th letter of user k, the capacity region at expenses* (E_1, E_2)*,* $\mathcal{C}(E_1, E_2)$*, can be computed by introducing the constraints*

$$\mathbf{e}_1^T\mathbf{p}_1 \leq E_1, \quad \mathbf{e}_2^T\mathbf{p}_2 \leq E_2. \tag{5.25}$$

5.2.1.3 Extension to K users

The formulation of the computation of the capacity region as a rank-one constrained problem introduced in Section 5.2.1.2 for two users can be extended to the K-user case. Using similar equivalences but now involving tensors [RT05] of dimensionality up to K, the rank-one constraint applies to any number of users.

The capacity region $\mathcal{C}$ of the K-user DMAC is the convex hull of the set of rate tuples $(R_1, \ldots, R_K)$ satisfying

$$0 \le R_{\mathcal{S}} < I(X_{(\mathcal{S})}; Y | X_{(\mathcal{S}^c)}), \quad \forall \mathcal{S} \subseteq \{1, 2, \ldots, K\} \tag{5.26}$$

for a distribution of the form $P_{X_1 \ldots X_K Y} = P_{X_1} \cdot \ldots \cdot P_{X_K} P_{Y|X_1 \ldots X_K}$ on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_K \times \mathcal{Y}$. We denote by $\mathcal{S}^c = \{1, 2, \ldots, K\} \setminus \mathcal{S}$ the complement set of $\mathcal{S}$, $X_{(\mathcal{S})} = \{X_k : k \in \mathcal{S}\}$, and $R_{\mathcal{S}} = \sum_{k \in \mathcal{S}} R_k$. By defining $\mathcal{N} \triangleq \{1, 2, \ldots, K\}$, the computation of the capacity region can be parameterized as

$$\underset{\{R_k\}, \{P_{X_{(\mathcal{S})}}\}_{\forall \mathcal{S} \subseteq \mathcal{N}}}{\text{maximize}} \sum_{k=1}^{K} \theta_k R_k \tag{5.27}$$

$$\text{subject to} \quad 0 \le R_{\mathcal{S}} \le \sum_{x_{(\mathcal{N})}, y} P_{X_{(\mathcal{N})}}(x_{(\mathcal{N})}) P_{Y|X_{(\mathcal{N})}}(y|x_{(\mathcal{N})}) \log \frac{P_{Y|X_{(\mathcal{N})}}(y|x_{(\mathcal{N})})}{\sum_{x'_{(\mathcal{S})}} P_{X_{(\mathcal{S})}}(x'_{(\mathcal{S})}) P_{Y|X_{(\mathcal{N})}}(y|x'_{(\mathcal{S})}, x_{(\mathcal{S}^c)})} \tag{5.28}$$

$$P_{X_{(\mathcal{S})}}(x_{(\mathcal{S})}) = \prod_{k \in \mathcal{S}} P_{X_k}(x_k), \ \forall \mathcal{S} \subseteq \mathcal{N} \tag{5.29}$$

$$P_{X_k}(x_k) \ge 0 \ \forall x_k \in \mathcal{X}_k, \ \sum_{x_k} P_{X_k}(x_k) = 1, \quad \forall k \in \mathcal{N} \tag{5.30}$$

The solution to (5.27)-(5.30) for any given $\boldsymbol{\theta} = [\theta_1 \ldots \theta_K]^T$ such that $\boldsymbol{\theta} \ge \mathbf{0}$ and $\mathbf{1}^T \boldsymbol{\theta} = 1$ is a point $(R_1^\star(\boldsymbol{\theta}), \ldots, R_K^\star(\boldsymbol{\theta}))$ of the boundary of the capacity region $\mathcal{C}$.

Similarly to what happened in the two-user case, the problem (5.27)-(5.30) is not convex because the constraints in (5.29) are not jointly convex in $P_{X_1} \ldots P_{X_K}$. However, (5.27)-(5.30) can be also reformulated as a rank-one non-convex optimization problem if we allow the variables X_1, ..., X_K to be dependent with distribution $P_{X_{(\mathcal{N})}}$. In doing so, it is useful to extend the matrix-vector notation of Section 5.2.1.2 by using the tensor $\mathbf{P}_{(\mathcal{S})}$ to denote $P_{X_{(\mathcal{S})}}$, where $[\mathbf{P}_{(\mathcal{S})}]_{i_1, i_2, \ldots, i_{|\mathcal{S}|}} = P_{X_{(\mathcal{S})}}\left(x_{k_1}^{(i_1)}, x_{k_2}^{(i_2)}, \ldots, x_{k_{|\mathcal{S}|}}^{(i_{|\mathcal{S}|})}\right)$ $\forall 1 \le i_j \le |\mathcal{X}_{k_j}|$, $1 \le j \le |\mathcal{S}|$, $\forall \mathcal{S} = \{k_1, k_2, \ldots, k_{|\mathcal{S}|}\} \subseteq \mathcal{N}$. $P_{X_{(\mathcal{S})}}$ is the marginalization of $P_{X_{(\mathcal{N})}}$ into the set of user codewords $X_{(\mathcal{S})}$, i.e.,

$$P_{X_{(\mathcal{S})}}(x_{(\mathcal{S})}) = \sum_{x_{(\mathcal{S}^c)}} P_{X_{(\mathcal{N})}}(x_{(\mathcal{N})}) \tag{5.31}$$

or, equivalently in tensor notation,

$$\sum_{i_{(\mathcal{S}^c)}} \mathbf{P}_{(\mathcal{N})} = \mathbf{P}_{(\mathcal{S})}. \tag{5.32}$$

Proposition 5.2 *The problem* (5.27)-(5.30) *of the computation of the capacity region of an ar-*

bitrary K-user DMAC is equivalent to the following rank-one[4] *non-convex optimization problem:*

$$\underset{\{R_k\},\{\mathbf{P}_{(\mathcal{S})}\}_{\forall \mathcal{S}\subseteq\mathcal{N}}}{\text{maximize}} \quad \sum_{k=1}^{K} \theta_k R_k \tag{5.34}$$

$$\text{subject to} \quad 0 \leq R_{\mathcal{S}} \leq f_{\mathcal{S}}(\mathbf{P}_{(\mathcal{N})}, \mathbf{P}_{(\mathcal{S}^c)}), \quad \forall \mathcal{S} \subseteq \mathcal{N} \tag{5.35}$$

$$\sum_{i_{(\mathcal{S}^c)}} \mathbf{P}_{(\mathcal{N})} = \mathbf{P}_{(\mathcal{S})} \quad \forall \mathcal{S} \subseteq \mathcal{N} \tag{5.36}$$

$$\mathbf{P}_{(\mathcal{N})} \succeq 0 \ , \ \sum_{i_{(\mathcal{N})}} \mathbf{P}_{(\mathcal{N})} = 1 \tag{5.37}$$

$$\operatorname{rank}(\mathbf{P}_{(\mathcal{N})}) = 1, \tag{5.38}$$

where the functions

$$f_{\mathcal{S}}(\mathbf{P}_{(\mathcal{N})}, \mathbf{P}_{(\mathcal{S}^c)}) = \sum_{i_{(\mathcal{N})},y} [\mathbf{P}_{(\mathcal{N})}]_{i_{(\mathcal{N})}} P_{Y|X_{(\mathcal{N})}}\big(y|x_{(\mathcal{N})}^{i_{(\mathcal{N})}}\big) \log \frac{P_{Y|X_{(\mathcal{N})}}\big(y|x_{(\mathcal{N})}^{i_{(\mathcal{N})}}\big)[\mathbf{P}_{(\mathcal{S}^c)}]_{i_{(\mathcal{S}^c)}}}{\sum_{i'_{(\mathcal{S})}} [\mathbf{P}_{(\mathcal{N})}]_{i'_{(\mathcal{S})},i_{(\mathcal{S}^c)}} P_{Y|X_{(\mathcal{N})}}\big(y|x_{(\mathcal{S})}^{i'_{(\mathcal{S})}}, x_{(\mathcal{S}^c)}^{i_{(\mathcal{S}^c)}}\big)} \tag{5.39}$$

are concave in $(\mathbf{P}_{(\mathcal{N})}, \mathbf{P}_{(\mathcal{S}^c)})$.

Proof. This follows from extending the definition of $\mathcal{P}_{\mathsf{prod}}$ (5.11) to suit the K-th dimensional tensor $\mathbf{P}_{(\mathcal{N})}$ and noticing that the functions $f_{\mathcal{S}}$ are the generalizations of f_1 (5.20), f_2 (5.21), and f_{12} (5.22) to the K-user case. Hence, concavity of $f_{\mathcal{S}}$ can be showed by the same arguments than in the proof of Proposition 5.1. □

5.2.2 Relaxation methods

Proposition 5.1 identified all the non-convexity of the problem of the computation of the capacity region of an arbitrary DMAC as the rank-one constraint $\operatorname{rank}(\mathbf{P}) = 1$ in (5.19). Rank-one constrained optimization problems, even with linear matrix inequalities, are non-convex problems that cannot be solved optimally in polynomial time. Thus, we first choose to relax (5.13)-(5.19) by removing the rank-one constraint (5.19) to obtain a tractable, convex problem equivalent to solving the capacity region of a DMAC with arbitrarily dependent codewords (full transmitter cooperation). We will denote by $\mathcal{R}^{\circ}$ the outer bound on the true capacity region obtained with the relaxed problem (5.13)-(5.18).

If the optimal solution to the relaxed problem happens to be rank one, then it will also be an optimal solution to the original problem. Otherwise, this solution has to be projected onto $\mathcal{P}_{\mathsf{prod}}$ to obtain a candidate solution (not necessarily optimal). Many different approaches can

[4] The K-th dimensional tensor $\mathbf{P}_{(\mathcal{N})} \in \mathbb{R}_+^{|\mathcal{X}_1|\times\ldots|\mathcal{X}_K|}$ has rank one if and only if it can be written as

$$\mathbf{P}_{(\mathcal{N})} = \mathbf{p}_1 \otimes \mathbf{p}_2 \otimes \ldots \otimes \mathbf{p}_K, \tag{5.33}$$

where $\otimes$ denotes outer product and the vectors $\mathbf{p}_k \in \mathbb{R}_+^{|\mathcal{X}_k|}$ admit a similar equivalence to that of Section 5.2.1.2. Again, $\operatorname{rank}(\mathbf{P}_{(\mathcal{N})}) = 1$ is equivalent to imposing $P_{X_1\ldots X_K} = P_{X_1} \cdot \ldots \cdot P_{X_K}$.

be applied to approximate an arbitrary joint distribution $\mathbf{P}$ by a reduced rank distribution of the form $\mathbf{q}_1\mathbf{q}_2^T$; here we shall explore two of them: randomization and marginalization.

5.2.2.1 Randomization

A randomization approach generates random samples of candidate probability vectors $(\mathbf{q}_1, \mathbf{q}_2)$ such that $\mathbb{E}\{\mathbf{q}_1\mathbf{q}_2^T\} = \mathbf{P}$, thus approximating the solution to the relaxed problem in the mean sense. For each of the generated pairs, the achievable rates (5.5)-(5.9) are evaluated and the pair of distributions yielding the largest objective value is kept as the solution. This is equivalent to performing a random search on the original problem with guidance on the correlation matrix of the distributions under test taken from the relaxed problem.

The nature of the random pairs $(\mathbf{q}_1, \mathbf{q}_2)$, which are input probability distributions, prevents us from easily finding statistics for generating them yielding $\mathbb{E}\{\mathbf{q}_1\mathbf{q}_2^T\} = \mathbf{P}$ directly. Instead, we first choose to approximate $\mathbf{P}$ by the convex combination of n rank-one distributions under a minimum divergence criterion, i.e.,

$$\mathbf{P} \approx \sum_{i=1}^{n} \lambda_i \mathbf{a}_i \mathbf{b}_i^T, \tag{5.40}$$

where n is fixed, and

$$\begin{aligned} \{\lambda_i, \mathbf{a}_i, \mathbf{b}_i\} = \quad & \underset{\{\lambda_i, \mathbf{a}_i, \mathbf{b}_i\}}{\arg\min} \quad D(\mathbf{P} || \sum_{i=1}^{n} \lambda_i \mathbf{a}_i \mathbf{b}_i^T) & (5.41) \\ & \text{subject to} \quad \sum_{i=1}^{n} \lambda_i = 1, \; \lambda_i \geq 0 \quad 1 \leq i \leq n & (5.42) \\ & \qquad \mathbf{1}^T \mathbf{a}_i = 1, \; \mathbf{a}_i \geq \mathbf{0} \quad 1 \leq i \leq n & (5.43) \\ & \qquad \mathbf{1}^T \mathbf{b}_i = 1, \; \mathbf{b}_i \geq \mathbf{0} \quad 1 \leq i \leq n. & (5.44) \end{aligned}$$

The problem (5.41)-(5.44) is not jointly convex in $\{\lambda_i, \mathbf{a}_i, \mathbf{b}_i\}$ but is separately convex in $\{\lambda_i\}$, $\{\mathbf{a}_i\}$, and $\{\mathbf{b}_i\}$. Thus, a practical approximation as in (5.40) can be obtained through an alternate optimization $\{\lambda_i^{(0)}\} \rightarrow \{\mathbf{a}_i^{(0)}\} \rightarrow \{\mathbf{b}_i^{(0)}\} \rightarrow \{\lambda_i^{(1)}\} \rightarrow \dots$ until convergence is achieved $((.)^{(r)}$ denotes value at the r-th iteration).

Second, given that N random samples $(\mathbf{q}_1, \mathbf{q}_2)$ must be generated, the approximation in the mean sense $\mathbb{E}\{\mathbf{q}_1\mathbf{q}_2^T\} = \mathbf{P}$ can be achieved using (5.40) by drawing, $i = 1 \dots n$, $N\lambda_i$ pairs from a pair of independent distributions such that $\mathbb{E}\{\mathbf{q}_1\} = \mathbf{a}_i$, $\mathbb{E}\{\mathbf{q}_2\} = \mathbf{b}_i$ (because of statistical independence, this will imply $\mathbb{E}\{\mathbf{q}_1\mathbf{q}_2^T\} = \mathbf{a}_i\mathbf{b}_i^T$). To generate a random vector $\mathbf{q}$ with prescribed average $\mathbb{E}\{\mathbf{q}\} = \overline{\mathbf{q}}$ we use a distribution whose support $\Omega_{\mathbf{q}}$ is the largest sphere centered at $\overline{\mathbf{q}}$ such that all its boundary lies within the probability simplex, as illustrated in Figure 5.2 for the 3-dimensional case. While the radius r of such sphere can be analytically determined resorting to the point-line and point-plane distance formulas, its expression is omitted here for the sake of brevity. As for the distribution of $\mathbf{q}$, we choose to project a random vector $\mathbf{x}$ drawn uniformly

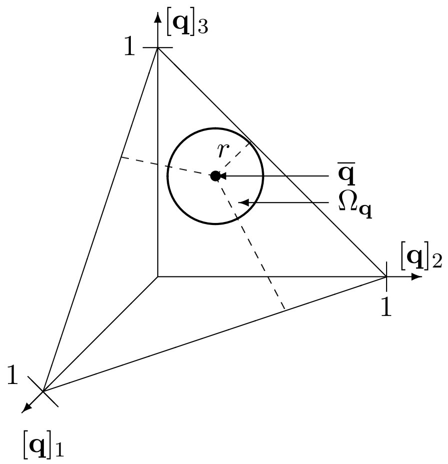

Figure 5.2: The support of the randomly generated probability distributions $\mathbf{q}$ is the largest circle centered at $\mathbb{E}\{\mathbf{q}\} = \overline{\mathbf{q}}$ that fits within the probability simplex.

over $\{\mathbf{x} \in \mathbb{R}^{|\mathcal{X}|\times 1} \; : \; \|\mathbf{x} - \overline{\mathbf{q}}\|_2^2 \leq r^2\}$[5] onto the probability simplex[6], i.e.,

$$\mathbf{q} = \mathbf{x} + \frac{1 - \mathbf{1}^T\mathbf{x}}{|\mathcal{X}|}\mathbf{1}, \tag{5.45}$$

which results in a circularly symmetric distribution with average $\overline{\mathbf{q}}$.

The evaluation of the achievable rates (5.5)-(5.9) yields the randomization inner bound, $\mathcal{R}^{\mathsf{i}}_{\mathsf{rand}}$ (see Algorithm 5.1 for a pseudocode description).

5.2.2.2 Marginalization

While the numerical performance of randomization is good, it is at the price of generating many potential solutions at random. Therefore, it is desirable to explore other simpler (deterministic) methods retaining most of its accuracy but also allowing for performance analysis.

To this end, we adopt a projection criterion based on the Kullback-Leibler divergence $D(\cdot\|\cdot)$. It has been used in [Cho68] and [Ku69] as the criterion for approximating joint discrete probability distributions given a dependence tree, although here the purpose is different. The use of the information divergence as the measure that quantifies the quality of the approximation offers several advantages, the most useful one being that, for some fixed $\mathbf{P}$ (with marginals $\tilde{\mathbf{p}}_1$ and $\tilde{\mathbf{p}}_2$), the pair $(\mathbf{p}_1, \mathbf{p}_2)$ that minimizes $D(\mathbf{P}||\mathbf{p}_1\mathbf{p}_2^T)$ follows easily from

$$D(\mathbf{P}||\mathbf{p}_1\mathbf{p}_2^T) = D(\mathbf{P}||\tilde{\mathbf{p}}_1\tilde{\mathbf{p}}_2^T) + D(\tilde{\mathbf{p}}_1||\mathbf{p}_1) + D(\tilde{\mathbf{p}}_2||\mathbf{p}_2) \geq D(\mathbf{P}||\tilde{\mathbf{p}}_1\tilde{\mathbf{p}}_2^T), \tag{5.46}$$

[5] To generate a vector $\mathbf{x}$ with uniform distribution on the sphere, we generate random vectors drawn uniformly on the hypercube $\left[[\overline{\mathbf{q}}]_1 - r, [\overline{\mathbf{q}}]_1 + r\right] \times \ldots \times \left[[\overline{\mathbf{q}}]_{|\mathcal{X}|} - r, [\overline{\mathbf{q}}]_{|\mathcal{X}|} + r\right]$ until the condition $\|\mathbf{x} - \overline{\mathbf{q}}\|_2^2 \leq r^2$ is satisfied.

[6] The projection onto a simplex usually has a water-filling form [Pal05], but, since by construction $\mathbf{x}$ belongs to the non-negative orthant, it reduces to (5.45), where the water level has been analytically found.

Algorithm 5.1 Randomization

```
for each value of θ ∈ [0,1] do
    Solve (5.13)-(5.18): (R°_1(θ), R°_2(θ)) = (R*_1, R*_2) and P(θ) = P*.
    Approximate P(θ) using (5.40) for some specified n: {λ_i, a_i, b_i}_{i=1}^n.
    for i = 1...n do
        for j = 1...Nλ_i do
            Generate a random pair (q_1, q_2) according to (5.45) such that E{q_1} = a_i, E{q_2} = b_i.
            Evaluate (5.5)-(5.9) using (q_1, q_2): (R_1^(i,j), R_2^(i,j)).
        end for
    end for
    Choose the best pair: (R^i_rand,1(θ), R^i_rand,2(θ)) = max_{i,j} θR_1^(i,j) + (1−θ)R_2^(i,j)
end for
Randomization inner bound: R^i_rand = Co({(R^i_rand,1(θ), R^i_rand,2(θ)), ∀θ}).
Outer bound: R° = Co({(R°_1(θ), R°_2(θ)), ∀θ}).
```

which shows that $(\mathbf{p}_1^\star, \mathbf{p}_2^\star) = (\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2)$. Therefore, marginalization is the solution to the minimum divergence criterion[8]. In order to obtain an approximation of $\mathcal{C}$, it is sufficient to solve (5.13)-(5.18) (or its equivalent form given in Lemma 5.1), take the marginal distributions of the solution, plug them into (5.5)-(5.9), and evaluate (R_1, R_2) (the problem (5.5)-(5.9) is convex for fixed input probability distributions). The solution to (5.5)-(5.9) in terms of (R_1, R_2) defines the marginalization inner bound $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}}$ (see Algorithm 5.2).

Algorithm 5.2 Marginalization

```
for each value of θ ∈ [0,1] do
    Solve (5.13)-(5.18): (R°_1(θ), R°_2(θ)) = (R*_1, R*_2) and (p_1(θ), p_2(θ)) = (p*_1, p*_2).
    Evaluate (R_1, R_2) (5.5)-(5.9) for fixed distributions (p_1(θ), p_2(θ)):
    (R^i_marg,1(θ), R^i_marg,2(θ)) = (R*_1, R*_2).
end for
Marginalization inner bound: R^i_marg = Co({(R^i_marg,1(θ), R^i_marg,2(θ)), ∀θ}).
Outer bound: R° = Co({(R°_1(θ), R°_2(θ)), ∀θ}).
```

5.2.2.3 Remarks

Remark 5.2 *The outer bound $\mathcal{R}^\circ$ can be further tightened by applying the algorithm of* [Rez04] *for the exact computation of the sum-capacity, denoted by C^{sum}. In particular, a tighter outer*

[8] Another deterministic strategy is to use the singular value decomposition (SVD) of $\mathbf{P}$ to choose $\mathbf{q}_1$ and $\mathbf{q}_2$ as the suitably normalized left and right singular vectors associated to the largest singular value. However, numerical simulations of this method performed over various channels have shown that it is outperformed by marginalization.

bound is $\mathcal{R}^{\mathsf{o}} \cap \mathcal{C}^{\mathsf{sum}}$*, where*

$$\mathcal{C}^{\mathsf{sum}} = \{(R_1, R_2) \in \mathbb{R}^2 \mid R_1 + R_2 \leq C^{\mathsf{sum}}\}. \tag{5.47}$$

Remark 5.3 *Both methods apply to the cost constrained case by adding the constraints in* (5.25) *to* (5.13)-(5.18) *when solving the relaxed problem. As for randomization, cost constraints have to be included also in* (5.41)-(5.44) *and the sphere radius* r *has to be chosen so as to belong to the feasible set.*

5.2.3 Performance analysis of marginalization

5.2.3.1 Analytical results

There exists a class of channels for which marginalization computes optimally the capacity region. Although we have not been able to fully characterize this class analytically, we have been able to show that some specific channels belong to it. For some channels, this can be proved by showing $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{R}^{\mathsf{o}} = \mathcal{C}$, while for others $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C} \subset \mathcal{R}^{\mathsf{o}}$. We restrict our attention to the two-user case for the sake of simplicity of the expressions.

It is worth point out that $\mathcal{R}^{\mathsf{o}}$ is also an outer bound of the capacity region of the two-user DMAC with feedback (see [Oza84, Sec. IV]), and it follows that the class of channels for which $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{R}^{\mathsf{o}}$ is a subset of the class of DMACs for which feedback does not increase the capacity region. Although the capacity region of the discrete memoryless MAC with feedback is not known in general (it is known in the continuous memoryless Gaussian (scalar) case [Oza84]), there are some achievability results [Bro05, Cov81a] and a class of DMACs for which the achievable region of [Cov81a] is tight (see [Wil82]). Let us start first with some optimality conditions that will be key for obtaining subsequent results.

Lemma 5.2 *A sufficient condition for optimality of the joint probability distribution* $P_{X_1X_2}$*, defined on* $\mathcal{X}_1 \times \mathcal{X}_2$*, with respect to the relaxed problem (5.13)-(5.18) for any fixed* $(\theta_1, \theta_2) = (\theta, 1-\theta)$*, assuming* $\theta_2 \geq \theta_1$*, is*[9]

$$\theta_1 D(P_{Y|X_1=x_1,X_2=x_2}||P_Y) + (\theta_2 - \theta_1) D(P_{Y|X_1=x_1,X_2=x_2}||P_{Y|X_1=x_1}) \begin{cases} = L_{\mathsf{o}}(\theta_1,\theta_2) \text{ if } P_{X_1X_2}(x_1,x_2) > 0 \\ \leq L_{\mathsf{o}}(\theta_1,\theta_2) \text{ if } P_{X_1X_2}(x_1,x_2) = 0 \end{cases} \tag{5.48}$$

for all $(x_1, x_2) \in (\mathcal{X}_1, \mathcal{X}_2)$ *and some* $L_{\mathsf{o}}(\theta_1, \theta_2) \geq 0$*, and*

$$I(X_1; Y|X_2) + I(X_2; Y|X_1) \geq I(X_1X_2, Y). \tag{5.49}$$

[9]If $\theta_2 \leq \theta_1$ the user indexes of θ and X in (5.48) or (5.51) must be swapped.

If (5.48) is satisfied by some joint distribution then it becomes also a necessary optimality condition with respect to the relaxed problem (5.13)-(5.18) and, since $L_{\mathsf{o}}(\theta_1, \theta_2)$ is precisely its objective value, it follows

$$L_{\mathsf{o}}(\theta_1, \theta_2) = R^\star_{\mathsf{o}}(\theta_1, \theta_2) = \max_{(R_1,R_2)\in\mathcal{R}^{\mathsf{o}}} \theta_1 R_1 + \theta_2 R_2. \tag{5.50}$$

Proof. See Appendix 5.B. □

Lemma 5.3 *A sufficient condition for optimality of the input probability distributions P_{X_1} and P_{X_2}, defined on $\mathcal{X}_1$ and $\mathcal{X}_2$, with respect to the capacity region for any fixed $(\theta_1, \theta_2) = (\theta, 1-\theta)$, assuming $\theta_2 \geq \theta_1$, is*[9]

$$\theta_1 D(P_{Y|X_1=x_1,X_2=x_2}||P_Y) + (\theta_2 - \theta_1) D(P_{Y|X_1=x_1,X_2=x_2}||P_{Y|X_1=x_1}) \begin{cases} = L(\theta_1,\theta_2) \text{ if } P_{X_1}(x_1)P_{X_2}(x_2) > 0 \\ \leq L(\theta_1,\theta_2) \text{ if } P_{X_1}(x_1)P_{X_2}(x_2) = 0 \end{cases}, \tag{5.51}$$

for all $x_k \in \mathcal{X}_k$, $k = 1, 2$. If (5.51) is satisfied for some pair of input probability distributions then it becomes also a necessary and sufficient condition for optimality with respect to the capacity region and to the relaxed problem and, since $L(\theta_1, \theta_2)$ is precisely its objective value, it follows

$$L(\theta_1, \theta_2) = C^\star(\theta_1, \theta_2) = \max_{(R_1,R_2)\in\mathcal{C}} \theta_1 R_1 + \theta_2 R_2. \tag{5.52}$$

Proof. Lemma 5.3 follows from the particularization of Lemma 5.2 to a product distribution of the form $P_{X_1X_2} = P_{X_1}P_{X_2}$, which always satisfies (5.49). Since such product distribution is a solution to the relaxed problem (5.13)-(5.18), it is capacity-achieving and hence optimal. □

Corollary 5.1 *$\mathcal{R}^{\mathsf{o}} = \mathcal{C}$ if and only if for each (θ_1, θ_2) there exists at least one pair of distributions satisfying the conditions in Lemma 5.3.*

Proof. The 'if' part is proved by noticing that existence of input distributions satisfying Lemma 5.3 for all (θ_1, θ_2) is equivalent to $R^\star_{\mathsf{o}}(\theta_1, \theta_2) = C^\star(\theta_1, \theta_2)$ $\forall(\theta_1, \theta_2)$, which implies $\mathcal{R}^{\mathsf{o}} = \mathcal{C}$, since both $\mathcal{C}$ and $\mathcal{R}^{\mathsf{o}}$ are convex sets. The 'only if' part follows from the fact that $\mathcal{R}^{\mathsf{o}} = \mathcal{C}$ implies that for each (θ_1, θ_2) there must exist at least one product distribution which is optimal with respect to the relaxed problem, and hence satisfies Lemma 5.3. □

Corollary 5.2 *$\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C}$ if for each (θ_1, θ_2) there exist one or more joint distributions satisfying Lemma 5.2, all of them with product distributions induced by their marginals satisfying Lemma 5.3.*

Proof. Taking into account the fact that the relaxed problem (5.13)-(5.18) may have multiple solutions not all of them product distributions, Corollary 5.2 implies that the inner bound $\mathcal{R}^{\mathsf{i}}$ of the capacity region provided by the relaxation method in Algorithm 5.1 is obtained with capacity achieving product distributions for all (θ_1, θ_2). □

Corollary 5.3 $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C} = \mathcal{R}^{\mathsf{o}}$ *if and only if for each* (θ_1, θ_2) *there exists at least one pair of distributions satisfying the conditions in Lemma 5.3 and all the joint (non-product) distributions satisfying Lemma 5.2 (if any such distributions exist) have product distributions induced by their marginals satisfying Lemma 5.3.*

Proof. This follows from corollaries 5.1 and 5.2. □

5.2.3.2 Applications

The previous results describe the conditions under which the bounds provided by the relaxation method in Algorithm 5.1 are tight[10], and provide us with a test for quantifying optimality of the marginalization approach. This test can be performed either analytically through any of the corollaries, or numerically (by checking if the product distribution that achieves $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}}$ satisfies the conditions in Lemma 5.3, for example). In addition, they can also be used as a tool for deriving new capacity results or for re-deriving existing ones, as will be illustrated with several examples.

Example 1: The binary switching MAC The binary switching MAC (BS-MAC) is a binary-input ternary-output ($\mathcal{Y} = \{0, 1, \infty\}$) deterministic multiple-access channel whose input-output relationship is

$$Y = X_2/X_1 \triangleq \begin{cases} 0 & \text{if } (X_1, X_2) = (1, 0) \\ 1 & \text{if } (X_1, X_2) = (1, 1) \\ \infty & \text{if } X_1 = 0 \end{cases} . \tag{5.53}$$

Note that, given the channel output Y, the value of X_1 is always decoded without ambiguity ($H(X_1|Y) = 0$). On the other hand, the information carried by X_2 can only be conveyed when $X_1 = 1$, since otherwise $Y = \infty$ independently of X_2. Therefore, a hierarchy is established among senders: sender 1 can always convey information to Y through the control of a switch which is also responsible for allowing sender 2 for effectively transmitting information to Y.

Proposition 5.3 (Vanroose [Van86], Vinck [Vin85a]) *The capacity region* $\mathcal{C}$ *of the BS-MAC is given by*

$$\begin{aligned} \mathcal{C} \;=\; & \big\{(R_1, R_2) \in \mathbb{R}^2 \mid \{0 \le R_1 \le \log 2, 0 \le R_2 \le 0.5 \log 2\} & (5.54)\\ & \cup \{R_1 \le h(R_2/\log 2), 0.5 \log 2 < R_2 \le \log 2\}\big\}, & (5.55) \end{aligned}$$

where $h(x) \triangleq -x \log x - (1-x) \log(1-x)$ *is the binary entropy function. The point*

$$(R_1^\star(\theta), R_2^\star(\theta)) = \arg \max_{(R_1, R_2) \in \mathcal{C}} \theta R_1 + (1-\theta) R_2, \tag{5.56}$$

[10] Note that the inner and outer bounds are always tight for $\theta = 0$ and $\theta = 1$.

belonging to the boundary of the capacity region, is achieved by the input probability distributions

$$P^{\star}_{X_1}(1;\theta) = p(\theta) = 1 - P^{\star}_{X_1}(0;\theta), \quad P^{\star}_{X_2}(0;\theta) = P^{\star}_{X_2}(1;\theta) = 1/2, \tag{5.57}$$

where

$$p(\theta) \triangleq \left(1 + 2^{(1-1/\theta)}\right)^{-1} \tag{5.58}$$

and the weight $\theta \in [0,1]$.

Proposition 5.4 *Marginalization achieves the capacity region for the BS-MAC and* $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C} = \mathcal{R}^{\circ}$.

Proof. The proof follows from the application of Corollary 5.3 to this channel. See Appendix 5.C. □

Vinck showed in [Vin85b] that the capacity region of the BS-MAC with and without feedback were identical. Since this result is a necessary condition for allowing the inner and outer bounds to coincide, it could have been inferred a posteriori from Proposition 5.4.

Not for all the channels for which the capacity region can be efficiently computed using Algorithm 5.1 satisfy that both the inner and the outer bound are tight. In particular, for a wider class of channels the capacity region can be computed by showing only that the outer bound is tight through Corollary 5.1. This is the case of the non-deterministic noisy BS-MAC considered next.

Example 2: The noisy binary-switching MAC We denote by noisy binary-switching MAC (nBS-MAC) the binary-input ternary-output ($\mathcal{Y} = \{0, 1, \infty\}$) multiple-access channel characterized by the transition probability distribution

$$P^{\text{nBS}-\text{MAC}}_{Y|X_1X_2} \equiv \begin{bmatrix} \delta/2 & \delta/2 & 1-\delta \\ \delta/2 & \delta/2 & 1-\delta \\ 1-\epsilon & \epsilon & 0 \\ \epsilon & 1-\epsilon & 0 \end{bmatrix}, \tag{5.59}$$

where the columns represent the different elements of $\mathcal{Y}$ and the rows correspond to the natural ordering of the inputs (X_1, X_2). This channel is a non-deterministic extension of (5.53) that adds two random behaviors to the channel: i) a noisy switch, that with probability δ is not able maintain open the circuit and outputs equally likely bits, and ii) a binary symmetric channel with error probability ϵ when the switch is closed. Clearly, when $(\delta, \epsilon) = (0, 0)$ the nBS-MAC reduces to the BS-MAC addressed in Example 1.

Proposition 5.5 *The outer bound is tight for the nBS-MAC,* $\mathcal{R}^{\circ} = \mathcal{C}$, *and one capacity-achieving pair of distributions achieving the boundary point*

$$(R^{\star}_1(\theta), R^{\star}_2(\theta)) = \arg \max_{(R_1,R_2)\in\mathcal{C}} \theta R_1 + (1-\theta)R_2 \tag{5.60}$$

is

$$P^{\star}_{X_1}(1;\theta) = p(\delta,\epsilon;\theta) = 1 - P^{\star}_{X_1}(0;\theta), \quad P^{\star}_{X_2}(0;\theta) = P^{\star}_{X_2}(1;\theta) = 1/2, \tag{5.61}$$

where

$$p(\delta,\epsilon;\theta) = \frac{A(\delta,\epsilon;\theta) - \delta/(1-\delta)}{A(\delta,\epsilon;\theta)+1} \tag{5.62}$$

if $0 < \theta \leq 1/2$ *and*

$$p(\delta,\epsilon;\theta) = \{0 \leq p < 1 \; : \; B(p,\delta,\epsilon;\theta) = \theta(h(\delta) - h(\epsilon) + \log(2))\} \tag{5.63}$$

otherwise, where

$$A(\delta,\epsilon;\theta) = \left(\exp(h(\delta) + (1-1/\theta)h(\epsilon))2^{(1/\theta - 1)}\right)^{\frac{1}{1-\delta}} \tag{5.64}$$

$$B(p,\delta,\epsilon;\theta) = (1-\theta)(1-\delta)\log\frac{\delta + (1-\delta)p}{(1-\delta)(1-p)} \tag{5.65}$$

$$+ (2\theta - 1)\Big((1-\epsilon-\delta/2)\log\frac{\delta + (2-2\epsilon-\delta)p}{(1-\delta)(1-p)} + (\epsilon - \delta/2)\log\frac{\delta+(2\epsilon-\delta)p}{(1-\delta)(1-p)}\Big). \tag{5.66}$$

Proof. The proof follows from the application of Corollary 5.1 to this channel. See Appendix 5.D. □

In Figure 5.3 we show the capacity region of the nBS-MAC for several values of δ and ϵ. Typically, when δ is small, sender 1 has access to the channel in much better conditions than sender 2, which is reflected in the shape of the capacity regions. This fact can be qualitatively appreciated in Figure 5.4 also, where $p(\delta,\epsilon;\theta)$ is plotted versus θ, the weight of R_1. When θ is small, the rate of sender 2 is prioritized and hence sender 1 "opens the tap" for the transmission of information of X_2 by setting a large value for $p(\delta,\epsilon;\theta)$. When θ increases, the tap is progressively closed towards a value that maximizes R_1.

The derived analytical conditions for tightness of the marginalization bounds do not exhaust all the situations for which the marginalization approach is optimal (Lemma 5.3 is sufficient but not necessary for a capacity-achieving product distribution). Thus, it can happen for some channels that none of the previous results applies but marginalization is still optimal: it may occur that $\mathcal{C} \subset \mathcal{R}^{\circ}$ but after marginalization $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C}$ without satisfying the conditions in Corollary 5.2. A representative example of this subclass of channels is the Binary Adder MAC (BA-MAC) considered next.

Example 3: The binary adder MAC The binary adder MAC (BA-MAC) (or binary erasure MAC, as named in [Cov06, Example 14.3.3]), is the binary-input ternary-output ($\mathcal{Y} = \{0,1,2\}$) deterministic multiple-access channel whose input-output relationship is given by

$$Y = X_1 + X_2, \tag{5.67}$$

where the sum is taken over the natural numbers. In this case, none of the users has access to the channel in privileged conditions. There are two input combinations that can be correctly

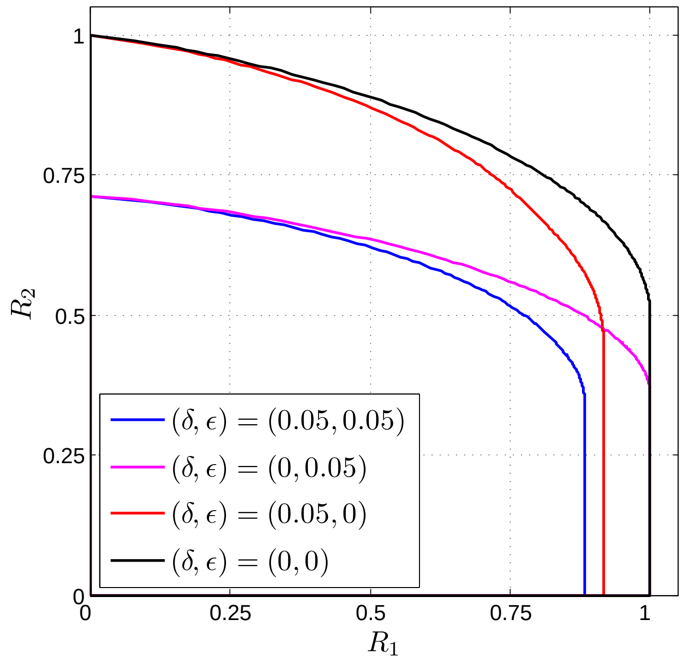

Figure 5.3: The capacity region $\mathcal{C}$ of the nBS-MAC, in [bit/ch. use], for different values of δ and ϵ. Note that $(\delta, \epsilon) = (0, 0)$ corresponds to the BS-MAC.

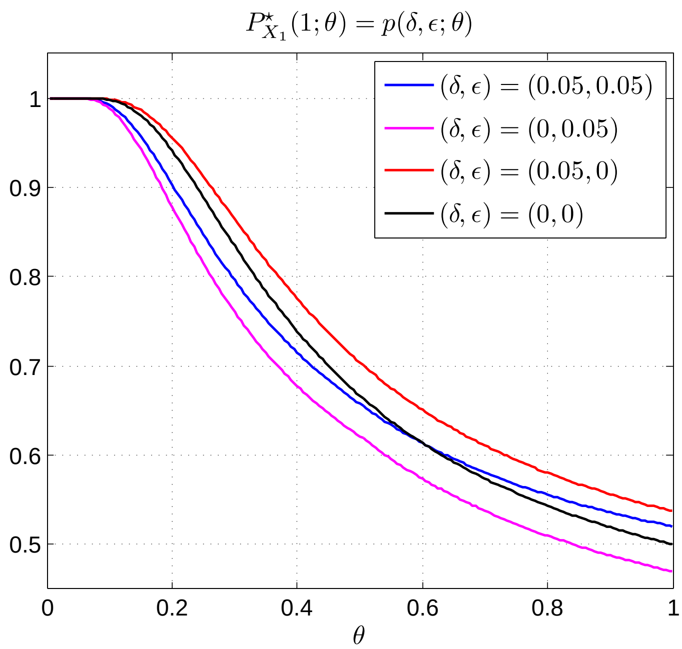

Figure 5.4: The probability $p(\delta, \epsilon; \theta)$ as a function of θ for different values of δ and ϵ. Note that $(\delta, \epsilon) = (0, 0)$ corresponds to $p(\theta)$ (5.58).

decoded without ambiguity ($(X_1, X_2) \in \{(0,0), (1,1)\}$), and two input combinations where the decoder has ambiguity about both users. Therefore, for this channel the capacity region is symmetric with respect to R_1 and R_2.

Proposition 5.6 (Liao [Lia72]) *The capacity region $\mathcal{C}$ of the BA-MAC is given by the set of rate pairs (R_1, R_2) satisfying*

$$0 \leq R_1 \leq \log 2 \tag{5.68}$$

$$0 \leq R_2 \leq \log 2 \tag{5.69}$$

$$R_1 + R_2 \leq \frac{3}{2}\log 2. \tag{5.70}$$

Furthermore, all the points of the boundary of the capacity region are achieved by a pair of uniform input probability distributions, i.e.,

$$P^{\star}_{X_i}(x_i;\theta) = 1/2, \quad x_i \in \{0,1\},\ i = 1,2,\ \forall\theta \in [0,1]. \tag{5.71}$$

The BA-MAC, as *noiseless multiple-access binary erasure channel*, was first studied by Liao [Lia72], who particularized his result of the capacity region of the general DMAC to this channel: a result that has been subsequently used in the literature [Kas76, Cha79, Cha81].

Proposition 5.7 *The achievable rate region of the proposed algorithm equals the capacity region for the BA-MAC. Furthermore, the inner and outer bounds on the capacity region provided by the proposed algorithm do not coincide, i.e., $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C} \subset \mathcal{R}^{\mathsf{o}}$.*

Proof. See Appendix 5.E. □

Interestingly, the BS-MAC and BA-MAC are the only two non-trivial channels that characterize the rest of binary-input ternary-output[11] deterministic DMACs, which can be obtained through isomorphisms of the input and output alphabets in one of these two canonical channels [Vin85a]. Hence, the proposed marginalization approach is tight for all of them. Finally, the following result extends the optimality of the relaxation to a wider class of channels.

Theorem 5.1 *Marginalization is tight in the sense that $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C}$, for all binary-input deterministic DMACs.*

Proof. See Appendix 5.F. □

5.2.4 Numerical results

In the following, we analyze the numerical performance of randomization and marginalization over three different binary-inputs ternary-output non-deterministic two-user DMACs, which we name DMAC_1, DMAC_2, and DMAC_3, characterized respectively by the transition probability

[11] When we refer to M-ary output DMACs we assume that $|\mathcal{Y}| = M$ and all the M values of the output alphabet can be exhausted by at least one input (there are not dummy output letters).

Algorithm 5.3 Random search

Set $(N, \sigma^2) = (500, 1/9)$.
for each value of $\theta \in [0, 1]$ **do**
  Set $(R_1(\theta), R_2(\theta), f^\star, \mathbf{p}_1(\theta), \mathbf{p}_2(\theta)) = (0, 0, 0, \mathbf{0}, \mathbf{0})$.
  Set $(\mathbf{p}_1^{(0)}, \mathbf{p}_2^{(0)}) = (\mathbf{1}/|\mathcal{X}_1|, \mathbf{1}/|\mathcal{X}_2|)$.
  for $j = 1 \dots N$ **do**
    Generate $(\mathbf{r}_1^{(j)}, \mathbf{r}_2^{(j)})$ i.i.d. $\sim \mathcal{N}(0, \sigma^2)$ of lengths $|\mathcal{X}_1|$ and $|\mathcal{X}_2|$, respectively.
    Update[12] $\mathbf{p}_k^{(j)} = [\mathbf{p}_k^{(j-1)} + \mathbf{r}_k^{(j)}]^+$ and normalize $\mathbf{p}_k^{(j)} := \mathbf{p}_k^{(j)}/(\mathbf{1}^T\mathbf{p}_k^{(j)})$, $k = 1, 2$.
    Evaluate (5.5)-(5.9) using $(\mathbf{p}_1^{(j)}, \mathbf{p}_2^{(j)})$: $(R_1^\star(\theta), R_2^\star(\theta))$.
    if $\theta R_1^\star + (1-\theta)R_2^\star > f^\star$ **then**
      $(R_1(\theta), R_2(\theta), f^\star, \mathbf{p}_1(\theta), \mathbf{p}_2(\theta)) = (R_1^\star, R_2^\star, \theta R_1^\star + (1-\theta)R_2^\star, \mathbf{p}_1^{(j)}, \mathbf{p}_2^{(j)})$.
    end if
  end for
end for
$\mathcal{R}^{\mathsf{i}}_{\mathsf{brute}} = \mathrm{Co}(\{(R_1(\theta), R_2(\theta)),\ \forall\theta\})$.

distributions

$$P^{(1)}_{Y|X_1X_2} \equiv \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.7 & 0.2 & 0.1 \\ 0.5 & 0.1 & 0.4 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}, \quad P^{(2)}_{Y|X_1X_2} \equiv \begin{bmatrix} 0.4 & 0.1 & 0.5 \\ 0.3 & 0.2 & 0.5 \\ 0.5 & 0.4 & 0.1 \\ 0.2 & 0.8 & 0 \end{bmatrix}, \quad P^{(3)}_{Y|X_1X_2} \equiv \begin{bmatrix} 0.1 & 0.2 & 0.7 \\ 0.3 & 0.5 & 0.2 \\ 0.3 & 0.4 & 0.3 \\ 0.8 & 0.1 & 0.1 \end{bmatrix}. \tag{5.72}$$

The columns represent the different elements of $\mathcal{Y} = \{0, 1, 2\}$, and the rows correspond to the natural ordering of the inputs. While DMAC_1 is the multiple-access channel example used in [Rez04] to illustrate the behavior of the algorithm for the computation of the sum-rate capacity, DMAC_2 and DMAC_3 have been chosen randomly. For each of the channels we compute the randomization and marginalization bounds described in Section 5.2.2, and use the algorithm in [Rez04] to compute $\mathcal{C}^{\mathsf{sum}}$. As for the randomization bound, $N = 500$ randomly generated product distributions have been tested for each θ using the approximation (5.40) with $n = 4$. Additionally, we consider the achievable region of a random search algorithm, denoted by $\mathcal{R}^{\mathsf{i}}_{\mathsf{rs}}$, as a benchmark (see Algorithm 5.3).

Figures 5.5, 5.6, and 5.7 show the bounds for DMAC_1, DMAC_2, and DMAC_3, respectively. DMAC_1 is another example of a non-deterministic channel for which the outer bound and the randomization and marginalization inner bounds coincide, and hence the capacity region can be effectively computed with the proposed methods. As for DMAC_2 and DMAC_3, which represent a more general situation, the bounds do not coincide and the capacity region cannot be directly evaluated.

[12] $[\mathbf{x}]^+$ denotes the component-wise application of the operator $[x]^+ \triangleq \max\{x, 0\}$.

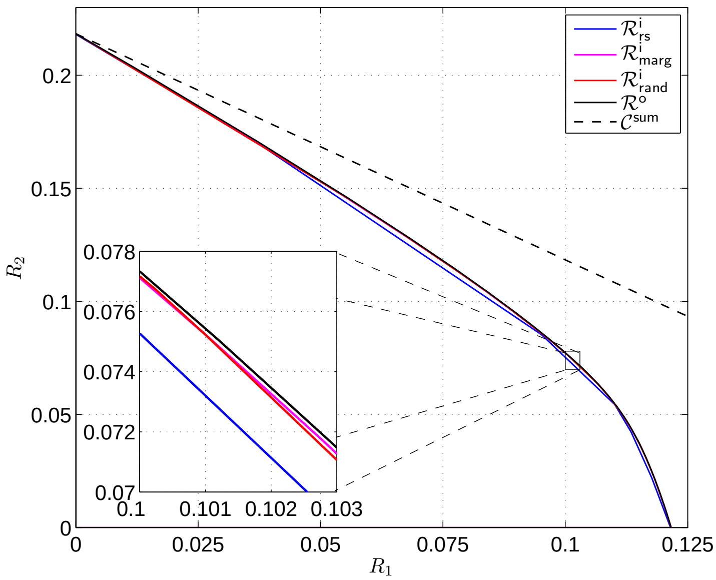

Figure 5.5: Bounds on the capacity region for DMAC$_1$. Units are [bit/ch. use].

However, if we consider $\mathcal{R}^{o} \cap \mathcal{C}^{sum}$ we are able to obtain a much tighter outer bound. Regardless of the tightness of $\mathcal{R}^{o}$, the accuracy of the sum-capacity offered by the proposed relaxation methods is remarkable for all the channels. The performance of marginalization and randomization is indistinguishable, as shown by appropriate zooms in Figures 5.5-5.7, while the behavior of the random search is irregular and, in general, much worse. To achieve similar performance, it requires a number of distributions under test orders of magnitude above that of randomization.

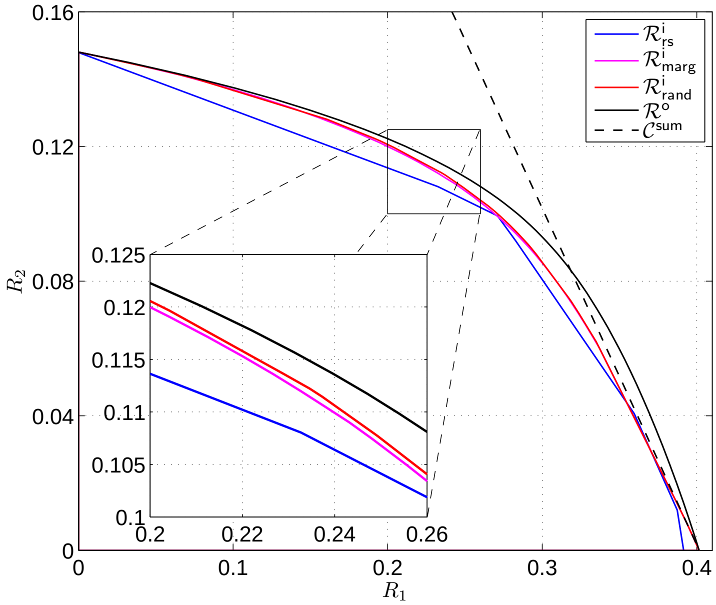

Figure 5.6: Bounds of the capacity region for DMAC$_2$. Units are [bit/ch. use].

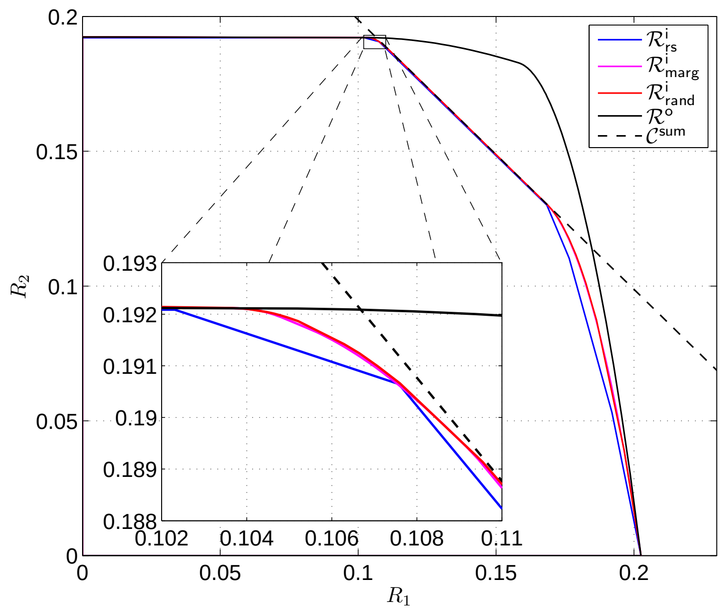

Figure 5.7: Bounds of the capacity region for DMAC_3. Units are [bit/ch. use].

5.3 The Degraded DMBC

5.3.1 The capacity region as a DC optimization problem

The capacity region $\mathcal{C}$ of the two-user dDMBC $X \to Y_1 \to Y_2$ is the convex hull of the set of rate pairs (R_1, R_2) satisfying

$$0 \leq R_1 \leq I(X; Y_1|U) \tag{5.73}$$

$$0 \leq R_2 \leq I(U; Y_2) \tag{5.74}$$

for some choice of the distribution $P_{UXY_1Y_2} = P_{UX}P_{Y_1|X}P_{Y_2|X}$ on $\mathcal{U} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2$, where P_{UX} is the joint probability distribution of the auxiliary random variable U and the transmitted codeword X, and $P_{Y_1|X}, P_{Y_2|X}$ are the given conditional distributions that depend on the nature of the channel. The region $\mathcal{C}$ is computable since it suffices to consider $|\mathcal{U}| = \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\}$ in the evaluation of (5.73)-(5.74). Relying on the fact that $\mathcal{C}$ is convex by time-sharing arguments, the supporting hyperplane theorem [Boy04, Sec. 2.5.2] allows us to parameterize its computation for some $\theta \in [0, 1]$ as

$$\underset{\{R_1, R_2, P_{UX}\}}{\text{maximize}} \quad \theta R_1 + (1-\theta) R_2 \tag{5.75}$$

$$\text{subject to} \quad 0 \leq R_1 \leq I(X; Y_1|U) \tag{5.76}$$

$$0 \leq R_2 \leq I(U; Y_2) \tag{5.77}$$

$$P_{UX}(u, x) \geq 0 \quad \forall (u, x) \in \mathcal{U} \times \mathcal{X} \tag{5.78}$$

$$\sum_{u,x} P_{UX}(u, x) = 1, \tag{5.79}$$

where the right hand sides of (5.76)-(5.77) amount to

$$I(X;Y_1|U) = \sum_{u,x,y_1} P_{UX}(u,x)P_{Y_1|X}(y_1|x)\log\frac{\left(\sum_{x'}P_{UX}(u,x')\right)P_{Y_1|X}(y_1|x)}{\sum_{x'}P_{UX}(u,x')P_{Y_1|X}(y_1|x')} \quad (5.80)$$

$$I(U;Y_2) = \sum_{u,y_2}\Big(\sum_{x}P_{UX}(u,x)P_{Y_2|X}(y_2|x)\Big)\times \quad (5.81)$$

$$\times\ \log\frac{\sum_{x'}P_{UX}(u,x')P_{Y_2|X}(y_2|x')}{\left(\sum_{x'}P_{UX}(u,x')\right)\left(\sum_{u',x'}P_{UX}(u',x')P_{Y_2|X}(y_2|x')\right)}. \quad (5.82)$$

Note that the solutions to (5.75)-(5.79), denoted by $(R_1^\star(\theta), R_2^\star(\theta), P_{UX}^\star(\theta))$, generally depend on θ.

For each θ, the optimal rates $(R_1^\star(\theta), R_2^\star(\theta))$ (which belong to the boundary of $\mathcal{C}$) satisfy the right hand side inequalities of (5.76)-(5.77) with equality. This allows us to rephrase (5.75)-(5.79) as

$$\underset{P_{UX}}{\text{maximize}} \quad \theta I(X;Y_1|U) + (1-\theta)I(U;Y_2) \quad (5.83)$$

$$\text{subject to} \quad P_{UX}(u,x)\geq 0 \quad \forall (u,x)\in\mathcal{U}\times\mathcal{X} \quad (5.84)$$

$$\sum_{u,x}P_{UX}(u,x) = 1, \quad (5.85)$$

where R_1, R_2 have been removed from the optimization since they can be computed evaluating (5.80)-(5.82) using $P_{UX}^\star(\theta)$, the solution to (5.83)-(5.85).

Lemma 5.4 *$I(X;Y_1|U)$ is concave in P_{UX}, whereas $I(U;Y_2)$ is a difference of concave functions of P_{UX}.*

Proof. See Appendix G. □

The computation of $\mathcal{C}$ (5.83)-(5.85) amounts hence to the maximization of the difference of two concave functions of P_{UX} (5.83) over the probability simplex. This falls within the class of DC problems[13], a wide class of non-convex optimization problems which can only be solved in cases where an underlying structure can be exploited. In general, their non-convexity makes them intractable and only brute-force or random search methods seem to be available. However, they provide no quantification on their incurred suboptimality.

5.3.2 Optimality conditions

Pursuing the computation of $\mathcal{C}$, we explore both local and global optimality conditions in order to either characterize completely the solutions to (5.83)-(5.85) or determine their structure in order to reduce the dimensionality of the search space. To that end, consider the following necessary optimality condition.

[13]It can equivalently be mapped to the minimization of the difference of two convex functions by minimizing the opposite of the objective in (5.83).

Lemma 5.5 *A necessary condition for global optimality (and sufficient condition for local optimality) of the joint probability distribution P_{UX} for any fixed $\theta \in [0,1]$ is*

$$\theta D(P_{Y_1|X=x}||P_{Y_1|U=u}) + (1-\theta)\mathbb{E}_{Y_2|X=x}\left\{\log \frac{P_{Y_2|U=u}(Y_2)}{P_{Y_2}(Y_2)}\right\} \begin{cases} = R(\theta) & \text{if } P_{UX}(u,x) > 0 \\ \leq R(\theta) & \text{if } P_{UX}(u,x) = 0 \end{cases} \tag{5.86}$$

for all $(u,x) \in \mathcal{U} \times \mathcal{X}$ and some non-negative real constant $R(\theta)$. Any distribution P_{UX} satisfying (5.86) yields an objective value in (5.83) of $\theta I(X;Y_1|U) + (1-\theta)I(U;Y_2) = R(\theta)$.

Proof. See Appendix H. □

Corollary 5.4 *A condition for global optimality of the joint probability distribution P_{UX} for any fixed $\theta \in [0,1]$ is to satisfy Lemma 5.5 with*

$$R(\theta) = C^\star(\theta) \triangleq \max_{(R_1,R_2)\in\mathcal{C}} \theta R_1 + (1-\theta)R_2, \tag{5.87}$$

where $C^\star(\theta)$ denotes the optimal value of (5.83)-(5.85) for the given θ.

Proof. The best among all the distributions satisfying a necessary condition for optimality must be optimal. □

5.3.3 The BEC-BSC degraded broadcast channel

We shall consider here the application of the necessary conditions of Lemma 5.5 to the BEC-BSC dDMBC, whose channel transition probabilities are shown in a diagram in Figure 5.8. This channel is such that receivers one and two see a Binary Erasure Channel (BEC) and a Binary Symmetric Channel (BSC) respectively. It models the situation in which the sender intends to convey independent information to one receiver equipped with erasure correction capabilities and another not equipped so. The second receiver copies the output of the first receiver except when there is an erasure (denoted by e), in which case the output is chosen uniformly. Hence, if ϵ denotes the erasure probability of the BEC seen by the first receiver, the error probability of the equivalent BSC seen by the second receiver is $\epsilon/2$.

Proposition 5.8 *The capacity region of the BEC-BSC degraded broadcast channel, $\mathcal{C}^{\text{BEC}-\text{BSC}}$, is given by the convex hull of the points $\big((1-\epsilon)\log 2, 0\big)$, $(0,(1-h(0.5\epsilon))\log 2)$, and the set of rate pairs $(R_1(\theta),R_2(\theta))$, $0 < \theta < \frac{1}{2}\frac{1-\epsilon}{(1-0.5\epsilon)^2}$, achieved by the distribution*

$$P_{UX}(0,0;\theta) = \frac{\gamma(\theta)-\beta(\theta)}{(1-\alpha(\theta)\beta(\theta))(1+\gamma(\theta))} \tag{5.88}$$

$$P_{UX}(1,1;\theta) = \frac{1-\alpha(\theta)\gamma(\theta)}{(1-\alpha(\theta)\beta(\theta))(1+\gamma(\theta))} \tag{5.89}$$

$$P_{UX}(0,1) = \alpha(\theta)P_{UX}(0,0;\theta) \tag{5.90}$$

$$P_{UX}(1,0) = \beta(\theta)P_{UX}(1,1;\theta), \tag{5.91}$$

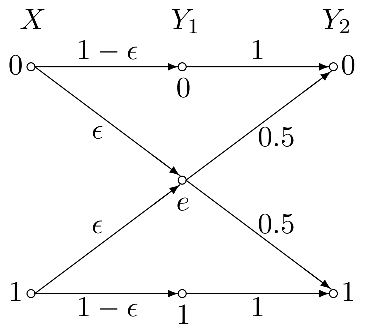

Figure 5.8: The BEC-BSC degraded broadcast channel

where $h(x) \triangleq -x \log x - (1-x)\log(1-x)$ *is the binary entropy function.*

The parameters $\alpha(\theta), \beta(\theta)$, *and* $\gamma(\theta)$ *are obtained as described next. Given* $\alpha(\theta) > 0$,

$$\gamma(\theta) = \frac{0.5\epsilon \exp\big(g(\alpha(\theta);\theta,\epsilon)/(1-\theta)\big) - (1-0.5\epsilon)}{0.5\epsilon - (1-0.5\epsilon)\exp\big(g(\alpha(\theta);\theta,\epsilon)/(1-\theta)\big)}, \tag{5.92}$$

where

$$g(\alpha;\theta,\epsilon) = (1-\theta)\log\frac{(1-0.5\epsilon)\alpha + 0.5\epsilon}{1-0.5\epsilon+0.5\epsilon\alpha} - \theta\log\alpha, \tag{5.93}$$

and $\beta(\theta) > 0$ *is such that* $g(\beta(\theta);\theta,\epsilon) = -g(\alpha(\theta);\theta,\epsilon)$ *(there are at most three such possible values of* $\beta(\theta)$ *for each given* $\alpha(\theta)$*). Finally,* $\alpha(\theta)$ *is determined from the following unidimensional maximization*

$$\alpha(\theta) = \arg \max_{\substack{P_{UX}:\alpha>0,\\ |g(\alpha;\theta,\epsilon)|<(1-\theta)\log(2/\epsilon-1)}} \theta I(X;Y_1|U) + (1-\theta)I(U;Y_2), \tag{5.94}$$

where it has been explicitly denoted that $P_{UX}(\theta)$ *in* (5.88)-(5.91) *and hence* $I(X;Y_1|U), I(U;Y_2)$ *exclusively depend on* α.

Proof. See Appendix I. □

The capacity region of the BEC-BSC degraded broadcast is shown in Figure 5.9. Note that the application of Lemma 5.5 and Corollary 5.4 allows us to compute $\mathcal{C}^{\text{BEC}-\text{BSC}}$ performing the maximization of a one-dimensional function instead of maximizing the achievable rates over the probability simplex containing the joint distributions on $\mathcal{U} \times \mathcal{X}$, for which three degrees of freedom would need to be explored for each θ.

5.4 Conclusions

The computation of the channel capacity of discrete memoryless channels is a convex problem that can be efficiently solved. The evaluation of capacity regions of multiterminal networks is not such an straightforward problem since, when capacity results are known, it gives rise to unavoidable non-convexities. In this context, we have addressed the evaluation of the capacity regions of the DMAC and the dDMBC.

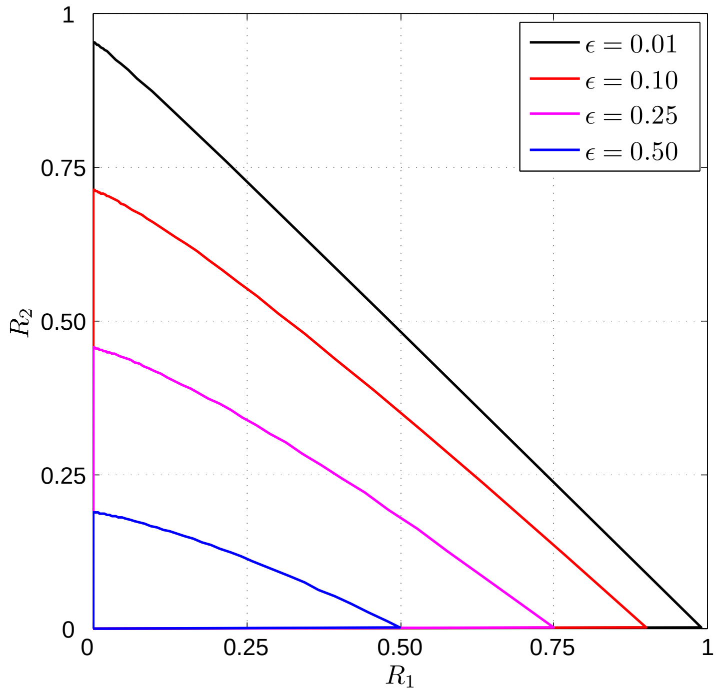

Figure 5.9: The capacity region $\mathcal{C}^{\text{BEC}-\text{BSC}}$ of the BEC-BSC degraded broadcast channel, in [bit/ch. use], for different values of ϵ.

An alternative reformulation of the capacity region of the DMAC condensed all the non-convexities of the problem into a single rank-one constraint. Since problems with this type of constraint cannot be solved optimally with the current state of the art, we proposed efficient methods to compute outer and inner bounds on the capacity region by solving a relaxed version of the problem and projecting its solution onto the original feasible set. Targeting numerical results, we adopted a randomization approach which generates random solutions to the original problem with distribution prescribed by the solution to the relaxed problem. Focusing on analytical results, we studied projection via minimum divergence, which amounts to the marginalization of the relaxed solution. In this latter case we derived sufficient conditions and necessary and sufficient conditions for the bounds to be tight. Furthermore, we were able to show that the class of channels for which the marginalization bounds matched exactly the capacity region included all the two-user binary-input deterministic DMACs as well as other non-deterministic channels (e.g., the nBS-MAC). In general, however, both methods are able to compute very tight bounds as shown for various examples.

On the other hand, the computation of the capacity region of the two-user dDMBC was characterized as a DC problem. Since this class of problems cannot be optimally solved in general, we focused on obtaining local and global optimality conditions. The KKT conditions of the problem were found to be necessary and sufficient for an input distribution to achieve a local maximum. An immediate consequence is that, among all the distributions satisfying the KKT, those attaining the largest objective value are capacity-achieving (i.e., globally optimal). These results enabled the maximization of the achievable rates over a candidate set of potentially much lower dimensionality than the original feasible set, where the BEC-BSC dDMBC served as an example.

5.A Appendix: Proof of Proposition 5.1

First, note that the functions f_1 (5.20), f_2 (5.21), and f_{12} (5.22) simplify to the right hand sides of (5.6), (5.7), and (5.8), respectively, when the vector-matrix formulation is used for some $\mathbf{P} \in \mathcal{P}_{\mathsf{prod}}$ with marginal distributions $\mathbf{p}_1$ and $\mathbf{p}_2$. Equivalence between (5.5)-(5.9) and (5.13)-(5.19) is hence proved thanks to constraint (5.17) and the equivalence of (5.18)-(5.19) and $\mathbf{P} \in \mathcal{P}_{\mathsf{prod}}$.

Regarding the concavity of the function $f_1(\mathbf{P}, \mathbf{p}_2)$ we can rewrite (5.20) as

$$
\begin{aligned}
f_1(\mathbf{P}, \mathbf{p}_2) &= \sum_{i,j} [\mathbf{P}]_{i,j} \underbrace{\left(\sum_y P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \log P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \right)}_{-H(Y|X_1=x_1^{(i)},X_2=x_2^{(j)})} \\
&+ \sum_{j,y} \underbrace{\left(\sum_i [\mathbf{P}]_{i,j} P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)}) \right)}_{\equiv P_{X_2Y}(x_2^{(j)},y)} \log \frac{[\mathbf{p}_2]_j}{\sum_{i'} [\mathbf{P}]_{i',j} P_{Y|X_1X_2}(y|x_1^{(i')}x_2^{(j)})} \\
&= -H(Y|X_1X_2) - \sum_{j,y} P_{X_2Y}(x_2^{(j)}, y) \log \frac{P_{X_2Y}(x_2^{(j)}, y)}{[\mathbf{p}_2]_j} \qquad (5.95)
\end{aligned}
$$

The term $H(Y|X_1, X_2)$ is linear in $\mathbf{P}$ and thus concave. The second term of (5.95) is jointly concave in $(P_{X_2Y}, \mathbf{p}_2)$ (by the same arguments that ensure the convexity of the divergence [Cov06, Thm. 2.7.2]), but since P_{X_2Y} is linear in $\mathbf{P}$, it is also concave in $(\mathbf{P}, \mathbf{p}_2)$, so is the function $f_1(\mathbf{P}, \mathbf{p}_2)$.

It follows by symmetry of (5.20) and (5.21) that the function $f_2(\mathbf{P}, \mathbf{p}_1)$ is jointly concave in $(\mathbf{P}, \mathbf{p}_1)$. Finally, the function f_{12} (5.22) can be shown to be concave resorting to the concavity of the mutual information of a DMAC with respect to the input distribution, in this case represented by $\mathbf{P}$. □

5.B Appendix: Proof of Lemma 5.2

Consider the relaxed problem (5.13)-(5.18). Since it is convex and satisfies Slater's conditions, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for optimality of any $(\mathbf{P}, \mathbf{p}_1, \mathbf{p}_2)$ [Boy04]. Taking its partial Lagrangian without relaxing the constraints (5.17)-(5.18),

$$
\begin{aligned}
\widetilde{\mathcal{L}}(R_1, R_2, \mathbf{p}_1, \mathbf{p}_2, \mathbf{P}; \boldsymbol{\lambda}) &= \theta R_1 + (1-\theta) R_2 + \lambda_1 (f_1(\mathbf{P}, \mathbf{p}_2) - R_1) \qquad (5.96) \\
&+ \lambda_2 (f_2(\mathbf{P}, \mathbf{p}_1) - R_2) + \lambda (f_{12}(\mathbf{P}) - R_1 - R_2), \qquad (5.97)
\end{aligned}
$$

and setting its derivatives with respect to R_1 and R_2 equal to zero we obtain

$$
\lambda_1^\star = \theta - \lambda \;,\; \lambda_2^\star = (1-\theta) - \lambda. \qquad (5.98)
$$

Using (5.98), (5.96)-(5.97) admits a simplified form that does not show dependency on (R_1, R_2),

$$\widetilde{\mathcal{L}}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{P}; \lambda) \equiv \widetilde{\mathcal{L}}(R_1, R_2, \mathbf{p}_1, \mathbf{p}_2, \mathbf{P}; [\lambda_1^\star \ \lambda_2^\star \ \lambda]) \tag{5.99}$$

$$= \lambda(f_{12}(\mathbf{P}) - f_1(\mathbf{P}, \mathbf{p}_2) - f_2(\mathbf{P}, \mathbf{p}_1)) + \theta f_1(\mathbf{P}, \mathbf{p}_2) + (1-\theta) f_2(\mathbf{P}, \mathbf{p}_1) \tag{5.100}$$

By grouping the primal optimization variables in $\mathbf{y} \triangleq (\mathbf{p}_1, \mathbf{p}_2, \mathbf{P})$, an explicit definition of the feasibility domain $\mathcal{D} \triangleq \{\mathbf{y} \mid (5.17)\text{-}(5.18) \text{ are satisfied}\}$ simplifies the application of the saddle-point property (strong duality holds) as

$$\min_{0 \le \lambda \le \min\{\theta, 1-\theta\}} \max_{\mathbf{y} \in \mathcal{D}} \widetilde{\mathcal{L}}(\mathbf{y}; \lambda) = \max_{\mathbf{y} \in \mathcal{D}} \min_{0 \le \lambda \le \min\{\theta, 1-\theta\}} \widetilde{\mathcal{L}}(\mathbf{y}; \lambda), \tag{5.101}$$

where we have used $0 \le \lambda \le \min\{\theta, 1-\theta\}$ according to dual feasibility ($\boldsymbol{\lambda} \succeq \mathbf{0}$) and (5.98). The dependence of $\widetilde{\mathcal{L}}(\mathbf{y}; \lambda)$ on λ in the inner minimization of the right hand side of (5.101) is linear (5.100) and hence its optimal value satisfies

$$\lambda^\star(\mathbf{y}) \equiv \lambda^\star(\varphi(\mathbf{y})) = \begin{cases} 0 & \text{if } \varphi(\mathbf{y}) > 0 \\ \min\{\theta, 1-\theta\} & \text{if } \varphi(\mathbf{y}) < 0 \end{cases}, \tag{5.102}$$

where $\varphi(\mathbf{y}) \triangleq f_{12}(\mathbf{P}) - f_1(\mathbf{P}, \mathbf{p}_2) - f_2(\mathbf{P}, \mathbf{p}_1)$[14]. Thus, the optimal value of the problem is

$$\max_{\mathbf{y} \in \mathcal{D}} \widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star(\mathbf{y})) = \widetilde{\mathcal{L}}(\mathbf{y}^\star; \lambda^\star(\mathbf{y}^\star)). \tag{5.103}$$

Let us now restrict the proof to the $0 \le \theta < 1/2$ case for the sake of simplicity (similar results are obtained for the case $1/2 \le \theta < 1$, and are included in the conditions of Lemma 5.2). To compute the optimal value (5.103) we need to know the sign of $\varphi(\mathbf{y}^\star)$ and adjust $\lambda^\star$ accordingly. However, $\mathbf{y}^\star$ depends in turn on $\lambda^\star$ through the maximization of $\widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star(\mathbf{y}))$. Therefore, to obtain the optimal $\mathbf{y}^\star$ we should first maximize $\widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star = \min\{\theta, 1-\theta\} = \theta)$ subject to $\varphi(\mathbf{y}) \le 0$, then maximize $\widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star = 0)$ subject to $\varphi(\mathbf{y}) > 0$, and select the $\mathbf{y}$ yielding the maximum objective value among both hypothesis.

Let us start hypothesizing $\varphi(\mathbf{y}^\star) \le 0$, which implies $\lambda^\star = \theta$ (5.102) and simplifies (5.100) to

$$\widetilde{\mathcal{L}}(\mathbf{y}; \lambda^\star = \theta) = \theta f_{12}(\mathbf{P}) + (1 - 2\theta) f_2(\mathbf{P}, \mathbf{p}_1), \tag{5.104}$$

which should be maximized under the constraint $\varphi(\mathbf{y}) \le 0$. Instead, we shall perform an unconstrained maximization of (5.104) and later impose that the solution satisfies non-positivity of φ. Thus, we aim at finding a solution to the problem

$$\underset{\mathbf{p}_1, \mathbf{p}_2, \mathbf{P}}{\text{maximize}} \quad \theta f_{12}(\mathbf{P}) + (1-2\theta) f_2(\mathbf{P}, \mathbf{p}_1) \tag{5.105}$$

$$\text{subject to} \quad \mathbf{P}\mathbf{1} = \mathbf{p}_1 \ , \ \mathbf{P}^T\mathbf{1} = \mathbf{p}_2 \tag{5.106}$$

$$\mathbf{P} \succeq \mathbf{0} \ , \ \mathbf{1}^T\mathbf{P}\mathbf{1} = 1, \tag{5.107}$$

[14] If $\varphi(\mathbf{y}) = 0$, $\lambda^\star(\mathbf{y})$ can take any value, but it is irrelevant since it does not affect the final result.

whose Lagrangian is

$$\begin{aligned}\mathcal{L}(\mathbf{y};\boldsymbol{\Phi},\boldsymbol{\nu}_1,\boldsymbol{\nu}_2,\eta) &= \theta f_{12}(\mathbf{P}) + (1-2\theta) f_2(\mathbf{P},\mathbf{p}_1) + \mathbf{1}^T(\boldsymbol{\Phi}\odot\mathbf{P})\mathbf{1} \quad (5.108)\\ &+ \boldsymbol{\nu}_1^T(\mathbf{P}\mathbf{1}-\mathbf{p}_1) + \boldsymbol{\nu}_2^T(\mathbf{P}^T\mathbf{1}-\mathbf{p}_2) + \eta(\mathbf{1}^T\mathbf{P}\mathbf{1}-1), \quad (5.109)\end{aligned}$$

where $\boldsymbol{\Phi} \in \mathbb{R}_+^{|\mathcal{X}_1|\times|\mathcal{X}_2|}$, $\boldsymbol{\nu}_k \in \mathbb{R}^{|\mathcal{X}_k|}$ for $k=1,2$, $\eta \in \mathbb{R}$, and $\odot$ denotes Hadamard (element-wise) product. Setting the derivatives of (5.108)-(5.109) with respect to $\mathbf{p}_1$ and $\mathbf{p}_2$ equal to zero we obtain the optimal values of $\{\boldsymbol{\nu}_k\}$:

$$\boldsymbol{\nu}_1^\star = (1-2\theta)(\log\mathbf{p}_1 + \log(e)\mathbf{1}),\ \boldsymbol{\nu}_2^\star = \mathbf{0}. \quad (5.110)$$

Finally, plugging (5.110) into (5.108)-(5.109) and setting its derivative with respect to $[\mathbf{P}]_{i,j}$ equal to zero we arrive at

$$\begin{aligned}\sum_y P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})\Bigg(\theta\log\frac{P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})}{\sum_{i',j'}[\mathbf{P}]_{i',j'}P_{Y|X_1X_2}(y|x_1^{(i')}x_2^{(j')})}\\ + (1-2\theta)\log\frac{[\mathbf{p}_1]_i P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j)})}{\sum_{j'}[\mathbf{P}]_{i,j'}P_{Y|X_1X_2}(y|x_1^{(i)}x_2^{(j')})}\Bigg)\\ = \theta\log(e) - \eta - [\boldsymbol{\Phi}]_{i,j}. \quad (5.111)\end{aligned}$$

Now, by complementary slackness, we know that $[\boldsymbol{\Phi}]_{i,j}[\mathbf{P}]_{i,j} = 0$ which forces

$$[\boldsymbol{\Phi}]_{i,j}\begin{cases} = 0 & \text{if } [\mathbf{P}]_{i,j} > 0\\ \geq 0 & \text{if } [\mathbf{P}]_{i,j} = 0\end{cases}, \quad (5.112)$$

and allows us to rewrite (5.111) as

$$\begin{aligned}\theta D(P_{Y|X_1=x_1,X_2=x_2}||P_Y) + (1-2\theta)D(P_{Y|X_1=x_1,X_2=x_2}||P_{Y|X_1=x_1})\\ \begin{cases} = \theta\log(e) - \eta \text{ if } P_{X_1X_2}(x_1,x_2) > 0\\ \leq \theta\log(e) - \eta \text{ if } P_{X_1X_2}(x_1,x_2) = 0\end{cases}. \quad (5.113)\end{aligned}$$

It can be shown that any $\mathbf{P}$ with marginals $\mathbf{p}_1,\mathbf{p}_2$ satisfying (5.113) for some $\eta \in \mathbb{R}$ has an associated objective value $\theta\log(e) - \eta$. Let us impose that such a solution to the unconstrained maximization of (5.104) satisfies also $\varphi(\mathbf{y}) \leq 0$ (the constraint of the current hypothesis under test). We now show by contradiction that the objective value of any distribution with $\varphi(\mathbf{y}) > 0$ is strictly lower. Suppose that the optimal distribution, with optimal objective value $R_{\mathsf{o}}^\star$, is such that $\varphi(\mathbf{y}^\star) > 0$. In this case, $\lambda^\star = 0$ from (5.102), which results in the objective function

$$\widetilde{\mathcal{L}}(\mathbf{y};\lambda^\star=0) = \theta f_1(\mathbf{P},\mathbf{p}_2) + (1-\theta)f_2(\mathbf{P},\mathbf{p}_1), \quad (5.114)$$

which is maximized by $\mathbf{y}^\star$ under the constraint $\varphi > 0$. Then, since

$$f_1(\mathbf{P},\mathbf{p}_2) = f_{12}(\mathbf{P}) - f_2(\mathbf{P},\mathbf{p}_1) - \varphi(\mathbf{y}^\star) < f_{12}(\mathbf{P}) - f_2(\mathbf{P},\mathbf{p}_1) \quad (5.115)$$

it follows by assumption that

$$R_{\mathsf{o}}^{\star} = \widetilde{\mathcal{L}}(\mathbf{y}^{\star}; \lambda^{\star} = 0) = \max_{\mathbf{y} \in \mathcal{D}: \varphi(\mathbf{y}) > 0} \widetilde{\mathcal{L}}(\mathbf{y}; \lambda^{\star} = 0) < \max_{\mathbf{y} \in \mathcal{D}} (\theta f_{12}(\mathbf{P}) + (1-2\theta) f_2(\mathbf{P}, \mathbf{p}_1)) = \theta \log(e) - \eta, \tag{5.116}$$

which contradicts optimality. Therefore, a probability distribution $\mathbf{y}$ with $\varphi(\mathbf{y}) \leq 0$ satisfying (5.113) is optimal, its objective value is $\theta \log(e) - \eta = R_{\mathsf{o}}^{\star}$, and invalidates the existence of other optimal solutions with $\varphi(\mathbf{y}) > 0$. Thus, provided such distribution exists (5.113) and $\varphi(\mathbf{y}) \leq 0$ becomes necessary for optimality. As a final remark, note that (5.49) and $\varphi(\mathbf{y}) \leq 0$ are equivalent statements thanks to the equivalence of the functions f_1, f_2, and f_{12} and mutual information. □

5.C Appendix: Proof of Proposition 5.4

It is sufficient to show that the conditions of Corollary 5.3 apply to the BS-MAC. For the sake of brevity we will subsequently restrict to the $0 \leq \theta \leq 1/2$ case (the same results are obtained for $1/2 < \theta < 1$) and simplify the notation by using the convention $P_{X_1X_2}(x_1, x_2) = P_{x_1x_2}$, which yields

$$P_Y = \{P_{10}, P_{11}, P_{00} + P_{01}\} \tag{5.117}$$

$$P_{Y|X_1=0} = \{0, 0, 1\} \tag{5.118}$$

$$P_{Y|X_1=1} = \{P_{10}/(P_{10} + P_{11}), P_{11}/(P_{10} + P_{11}), 0\}. \tag{5.119}$$

The probabilities P_{10} and P_{11} cannot be zero simultaneously since that would imply $H(Y) = 0$ and hence $R_1 = R_2 = 0$, which is clearly suboptimal. We will hypothesize $P_{10} = 0$ and $P_{11} > 0$ (similar results are obtained under the hypothesis $P_{10} > 0$ and $P_{11} = 0$).

With respect to P_{00}, P_{01} we cannot argue to infer beforehand wether these variables are positive or not. However, when (5.48) is particularized for both $(x_1, x_2) = (0, 0)$ and $(x_1, x_2) = (0, 1)$ the same expression is obtained. This implies that either P_{00}, P_{01} are both positive or both zero. By hypothesizing $P_{00} = P_{01} = 0$, the conditions (5.48) of Lemma 5.2 particularize to

$$\theta \log \frac{1}{P_{00} + P_{01}} \leq R_{\mathsf{o}}^{\star} \tag{5.120}$$

$$\theta \log \frac{1}{P_{10}} + (1 - 2\theta) \log \frac{P_{10} + P_{11}}{P_{10}} \leq R_{\mathsf{o}}^{\star} \tag{5.121}$$

$$\theta \log \frac{1}{P_{11}} + (1 - 2\theta) \log \frac{P_{10} + P_{11}}{P_{11}} = R_{\mathsf{o}}^{\star}. \tag{5.122}$$

By rewriting (5.120) and (5.121) as

$$\theta \log \frac{1}{P_{00} + P_{01}} = R_{\mathsf{o}}^{\star} + (1 - 2\theta) \log(\ell) \tag{5.123}$$

$$\theta \log \frac{1}{P_{10}} + (1 - 2\theta) \log \frac{P_{10} + P_{11}}{P_{10}} = R_{\mathsf{o}}^{\star} + (1 - 2\theta) \log(\ell') \tag{5.124}$$

for some $0 < \ell, \ell' \leq 1$. It follows from (5.122) and (5.124) that $P_{11} = P_{10}\ell'^{1/\theta-2}$ regardless of ℓ', and hence $P_{10} = 0$ forces $P_{11} = 0$. Similarly, $P_{11} = 0$ forces $P_{10} = 0$. Since this is a suboptimal choice, the optimal distribution satisfies $P_{10}, P_{11} > 0$. This fact transforms (5.121) into an equality (or, equivalently, forces $\ell' = 1$ in (5.124)), and implies $P_{11} = P_{10}$. Combining (5.123) with (5.121) (as an equality), we obtain

$$P_{10} = P_{11} = (P_{00} + P_{01})(2\ell)^{1/\theta-2} \stackrel{(a)}{=} (2 + (2\ell)^{2-1/\theta})^{-1} \tag{5.125}$$

$$P_{00} + P_{01} = \frac{(2\ell)^{2-1/\theta}}{2 + (2\ell)^{2-1/\theta}}, \tag{5.126}$$

where (a) follows from $\sum_{i,j} P_{ij} = 1$. Expression (5.125) shows that $P_{00} + P_{01} > 0 \ \forall \ell \in (0, 1]$, which is inconsistent with the hypothesis $P_{00} = P_{01} = 0$. Since the optimal values of P_{00} and P_{01} must hence be positive, the family of potential optimal distributions can be obtained by setting $\ell = 1$ in (5.125)-(5.126), which results in

$$P^\star_{00} = \alpha \quad , \quad P^\star_{01} = 1 - p(\theta) - \alpha \tag{5.127}$$

$$P^\star_{10} = p(\theta)/2 \quad , \quad P^\star_{11} = p(\theta)/2, \tag{5.128}$$

where $\alpha \in (0, 1 - p(\theta))$, and $p(\theta)$ was defined in (5.58). Among all the distributions satisfying (5.127)-(5.128), some of them may also satisfy (5.49) and be optimal. In particular, $\alpha = (1 - p(\theta))/2$ results in the product distribution (5.57), which satisfies Lemma 5.3 (this implies $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C}$). Since that optimal product distribution is also inferred by the marginals of (5.127)-(5.128) regardless of the choice of α, it follows that $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C} = \mathcal{R}^{\mathsf{o}}$. □

5.D Appendix: Proof of Proposition 5.5

It is sufficient to show that there exists a product distribution satisfying Lemma 5.3 for any $0 < \theta < 1$. Let us start with the $0 < \theta \leq 1/2$ case first and simplify the notation by using the notation $P_{X_k}(1) = p_k = 1 - \overline{p}_k$, $k = 1, 2$, which yields

$$P_Y = \left\{\frac{\delta}{2}\overline{p}_1 + p_1((1-\epsilon)\overline{p}_2 + \epsilon p_2), \frac{\delta}{2}\overline{p}_1 + p_1(\epsilon\overline{p}_2 + (1-\epsilon)p_2), (1-\delta)\overline{p}_1\right\} \tag{5.129}$$

$$P_{Y|X_1=0} = \{\delta/2, \delta/2, 1-\delta\} \tag{5.130}$$

$$P_{Y|X_1=1} = \{(1-\epsilon)\overline{p}_2 + \epsilon p_2, \epsilon\overline{p}_2 + (1-\epsilon)p_2, 0\}. \tag{5.131}$$

We arbitrarily assume that $p_2 = \overline{p}_2 = 1/2$ and $p_1, \overline{p}_1 > 0$ such that the conditions in Lemma 5.3, after some algebraic manipulation, reduce to

$$\theta\big(\delta \log \frac{\delta}{\delta\overline{p}_1 + p_1} + (1-\delta)\log\frac{1-\delta}{(1-\delta)\overline{p}_1}\big) = C^\star(\theta) \tag{5.132}$$

for $(x_1, x_2) \in \{(0, 0), (0, 1)\}$ and

$$\theta\big(\epsilon \log \frac{2\epsilon}{\delta\overline{p}_1 + p_1} + (1-\epsilon)\log\frac{2(1-\epsilon)}{(1-\delta)\overline{p}_1}\big) + (1-2\theta)(\log 2 - h(\epsilon)) = C^\star(\theta) \tag{5.133}$$

for $(x_1, x_2) \in \{(1,0),(1,1)\}$. By setting the left hand sides of (5.132)-(5.133) equal to each other, we arrive at

$$\theta(1-\delta)\log\frac{\delta\overline{p}_1 + p_1}{(1-\delta)p_1} = \theta h(\delta) + (1-\theta)(\log 2 - h(\epsilon)), \tag{5.134}$$

which is satisfied by $p(\delta,\epsilon;\theta)$ as defined in (5.62).

The case $1/2 < \theta < 1$ is more cumbersome since we cannot find the analytical expression of the optimal p_1. The conditional distributions

$$P_{Y|X_2=0} = \left\{\frac{\delta}{2}\overline{p}_1 + (1-\epsilon)p_1, \frac{\delta}{2}\overline{p}_1 + \epsilon p_1, (1-\delta)\overline{p}_1\right\} \tag{5.135}$$

$$P_{Y|X_2=1} = \left\{\frac{\delta}{2}\overline{p}_1 + \epsilon p_1, \frac{\delta}{2}\overline{p}_1 + (1-\epsilon)p_1, (1-\delta)\overline{p}_1\right\} \tag{5.136}$$

and the arbitrary setting $p_2 = \overline{p}_2 = 1/2$ allow us to obtain, after some manipulations, the conditions in Lemma 5.3 as

$$(1-\theta)\big(\delta\log\frac{\delta}{\delta\overline{p}_1 + p_1} + (1-\delta)\log\frac{1-\delta}{(1-\delta)\overline{p}_1}\big) + (2\theta-1)\big(\frac{\delta}{2}\log\frac{\delta}{\delta\overline{p}_1 + 2(1-\epsilon)p_1} + \frac{\delta}{2}\log\frac{\delta}{\delta\overline{p}_1 + 2\epsilon p_1} + (1-\delta)\log\frac{1-\delta}{(1-\delta)p_1}\big) = C^\star(\theta) \tag{5.137}$$

for $(x_1, x_2) \in \{(0,0),(0,1)\}$ and

$$(1-\theta)\big((1-\epsilon)\log\frac{2(1-\epsilon)}{\delta\overline{p}_1 + p_1} + \epsilon\log\frac{2\epsilon}{(1-\delta)\overline{p}_1}\big) + (2\theta-1)\big((1-\epsilon)\log\frac{2(1-\epsilon)}{\delta\overline{p}_1 + 2(1-\epsilon)p_1} + \epsilon\log\frac{2\epsilon}{\delta\overline{p}_1 + 2\epsilon p_1}\big) = C^\star(\theta) \tag{5.138}$$

for $(x_1, x_2) \in \{(1,0),(1,1)\}$. By setting the left hand sides of (5.137)-(5.138) equal to each other and using $\overline{p}_1 = 1 - p_1$ we obtain

$$B(p_1,\delta,\epsilon;\theta) = \theta(h(\delta) - h(\epsilon) + \log 2), \tag{5.139}$$

where $B(p_1,\delta,\epsilon;\theta)$ is defined in (5.66). Since

$$B(0,\delta,\epsilon;\theta) = \theta(1-\delta)\log\frac{\delta}{1-\delta} \le \theta h(\delta) \overset{(a)}{\le} \theta(h(\delta) - h(\epsilon) + \log 2), \tag{5.140}$$

where (a) follows from Gibb's lemma, and

$$\lim_{p_1\to 1} B(p_1,\delta,\epsilon;\theta) = +\infty > \theta(h(\delta) - h(\epsilon) + \log 2), \tag{5.141}$$

it follows by continuity that an optimum $p_1 \in [0,1)$ satisfying (5.139) exists for any $1/2 \le \theta < 1$.

Since we have been able to find product distributions satisfying Lemma 5.3 for all $0 \le \theta < 1$, Corollary 5.1 implies that $\mathcal{R}^\circ = \mathcal{C}$. □

5.E Appendix: Proof of Proposition 5.7

Thanks to the symmetry of the problem, we restrict without loss of generality to the $0 < \theta \leq 1/2$ case (similar results hold for $1/2 < \theta < 1$) and simplify the notation by using the convention $P_{X_1X_2}(x_1, x_2) = P_{x_1x_2}$, which yields

$$P_Y = \{P_{00}, P_{01} + P_{10}, P_{11}\} \tag{5.142}$$
$$P_{Y|X_1=0} = \{P_{00}/(P_{00} + P_{01}), P_{01}/(P_{00} + P_{01}), 0\} \tag{5.143}$$
$$P_{Y|X_1=1} = \{0, P_{10}/(P_{10} + P_{11}), P_{11}/(P_{10} + P_{11})\}. \tag{5.144}$$

Let us hypothesize $P_{01} = 0$. In this case, the rest of probabilities should satisfy $P_{00}, P_{10}, P_{11} > 0$ since both X_1 and X_2 can be perfectly decoded given Y. If we focus on the conditions (5.48) of Lemma 5.2 particularized to $(x_1, x_2) = (0, 0)$ and $(x_1, x_2) = (0, 1)$ we obtain

$$\theta \log \frac{1}{P_{00}} + (1 - 2\theta) \log \frac{P_{00} + P_{01}}{P_{00}} = R^\star_\circ \tag{5.145}$$
$$\theta \log \frac{1}{P_{01} + P_{10}} + (1 - 2\theta) \log \frac{P_{00} + P_{01}}{P_{01}} = R^\star_\circ + (1 - 2\theta) \log(\ell) \tag{5.146}$$

respectively. It follows from (5.145) and (5.146) that

$$P_{00} = (P_{01} + P_{10})^{\frac{\theta}{1-\theta}} (\ell P_{01})^{\frac{1-2\theta}{1-\theta}}, \tag{5.147}$$

which implies $P_{00} = 0$ regardless of the actual value of P_{10}. This is clearly suboptimal since $P_{00} > 0$ causes no penalty on R_2 (X_2 can still be uniquely decodable from Y) and allows user 1 to communicate at some positive rate. Similarly, if we hypothesize $P_{10} = 0$ we arrive at the suboptimal choice $P_{11} = 0$. Hence, the optimal distribution must satisfy $P_{01}, P_{10} > 0$ and, at the same time, $P_{00}, P_{11} > 0$ (since $(x_1, x_2) = (0, 0)$ and $(x_1, x_2) = (0, 0)$ are input values that allow for perfect decoding of both codewords). In this situations, the conditions (5.48) of Lemma 5.2 for all (x_1, x_2) particularize here to

$$\theta \log \frac{1}{P_{00}} + (1 - 2\theta) \log \frac{P_{00} + P_{01}}{P_{00}} = R^\star_\circ \tag{5.148}$$
$$\theta \log \frac{1}{P_{01} + P_{10}} + (1 - 2\theta) \log \frac{P_{00} + P_{01}}{P_{01}} = R^\star_\circ \tag{5.149}$$
$$\theta \log \frac{1}{P_{01} + P_{10}} + (1 - 2\theta) \log \frac{P_{10} + P_{11}}{P_{10}} = R^\star_\circ \tag{5.150}$$
$$\theta \log \frac{1}{P_{11}} + (1 - 2\theta) \log \frac{P_{10} + P_{11}}{P_{11}} = R^\star_\circ. \tag{5.151}$$

From (5.148) and (5.149) we obtain

$$P_{01} + P_{10} = P_{00}(P_{00}/P_{01})^{1/\theta - 2}, \tag{5.152}$$

an equivalence that can be plug in (5.150) and used with (5.148) to obtain

$$\frac{P_{00}}{P_{01}} = \frac{P_{11}}{P_{10}}. \tag{5.153}$$

Similarly, from (5.150)-(5.151)

$$P_{01} + P_{10} = P_{11}(P_{11}/P_{10})^{1/\theta-2}. \tag{5.154}$$

Expressions (5.152) and (5.154) imply $P_{00} = P_{11}$, which can be indistinctly used in (5.152) or (5.153) to arrive at the potential optimal solution

$$P^\star_{00} = P^\star_{11} = p(1-\theta)/2\ ,\ P^\star_{10} = P^\star_{01} = (1-p(1-\theta))/2, \tag{5.155}$$

which is not a product distribution and hence $\mathcal{C} \subset \mathcal{R}^\circ$ (recall that $p(\theta)$ was defined in (5.58)). To verify that (5.155) is indeed the unique solution to the relaxed problem (5.13)-(5.18) it remains to be checked that it satisfies (5.49). To that end, let us consider

$$I(X_2;Y|X_1) \;=\; H(Y|X_1) = h(p(1-\theta)) = I(X_1;Y|X_2) \tag{5.156}$$

$$I(X_1X_2;Y) \;=\; H(Y) = h(p(1-\theta)) + p(1-\theta)\log 2 \tag{5.157}$$

to show

$$I(X_1;Y|X_2) + I(X_2;Y|X_1) \;=\; 2h(p(1-\theta)) \overset{(a)}{\geq} h(p(1-\theta)) + 2(1-p(\theta))\log 2 \tag{5.158}$$

$$\overset{(b)}{\geq} h(p(1-\theta)) + 2(1-2/3)\log 2 = h(p(1-\theta)) + 2/3\log 2 \tag{5.159}$$

$$\overset{(c)}{\geq} h(p(1-\theta)) + p(1-\theta)\log 2 = I(X_1X_2;Y), \tag{5.160}$$

where inequality (a) follows from the linear bound $h(x) \geq 2(1-x)\log 2$ for $x \geq 1/2$ and the fact that $p(1-\theta) \geq 1/2$ (see Figure 5.4); inequalities (b) and (c) are a consequence of the fact that, for $0 < \theta \leq 1/2$, $p(1-\theta) \leq 2/3$.

Finally, the last step is to notice that the product distribution induced by the marginals of (5.155) are uniform and hence capacity achieving for any $0 < \theta \leq 1/2$ (recall Proposition 5.6)), implying $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C}$. □

5.F Appendix: Proof of Theorem 5.1

Consider each possible value of the size of the output alphabet $|\mathcal{Y}|$:

- *Binary output* - None of the binary-input binary-output deterministic DMACs has a capacity region dominating the timesharing line joining the points $(\log 2, 0)$ and $(0, \log 2)$[15]. Since marginalization is tight for $\theta \in \{0,1\}$ and the bounds are convexified using the convex hull operation it follows that $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{C}$.
- *Ternary output* - All the binary-input ternary-output DMACs can be obtained through isomorphisms of the input and output alphabets with respect to the transition matrices of the BS-MAC and the BA-MAC [Vin85a]. Hence, marginalization is tight for all of them as it is for the BS-MAC and the BA-MAC.

[15]The capacity region is upper-bounded by $R_k \leq H(Y|X_k) \leq H(Y) \leq \log 2$, $k = 1, 2$, and $R_1 + R_2 \leq H(Y) \leq \log 2$, which defines the triangle joining the points $(0,0), (\log 2, 0)$, and $(0, \log 2)$.

- *Quaternary output* - Quaternary output channels allow for perfect decoding of both codewords given the output. The unique arbitrary distribution maximizing simultaneously $I(X_1;Y|X_2), I(X_2;Y|X_1)$, and $I(X_1X_2;Y)$ is $P_{X_1X_2}(x_1,x_2) = 1/4\ \forall x_1,x_2 \in \{0,1\}$, which is a product distribution. Hence, $\mathcal{R}^{\mathsf{i}}_{\mathsf{marg}} = \mathcal{R}^{\mathsf{o}} = \mathcal{C}$.

- *Higher than four-dimensional output* - These channels cannot exist since there are only four possible input combinations. □

5.G Appendix: Proof of Lemma 5.4

The mutual information ruling the rate of the first link can be decomposed as[16]

$$I(X;Y_1|U) = \sum_{u,x,y_1} P_{UX}(u,x)P_{Y_1|X}(y_1|x)\Big(\log P_{Y_1|X}(y_1|x) + \log \frac{P_U(u)}{\sum_{x'} P_{UX}(u,x')P_{Y_1|X}(y_1|x')}\Big) \quad (5.161)$$

$$= \sum_x \Big(\sum_u P_{UX}(u,x)\Big) H(Y_1|X=x) - D(P_{UY_1}||Q_{UY_1}), \quad (5.162)$$

where

$$Q_{UY_1}(u,y_1) = P_U(u) \quad \forall (u,y_1) \in \mathcal{U}\times\mathcal{Y}_1 \quad (5.163)$$

is a dummy function which is not a probability distribution. While the first term is linear in P_{UX} and hence concave, the second is concave in (P_{UY_1}, Q_{UY_1}) thanks to the convexity of the divergence, which is based on the log-sum inequality, regardless of the fact that its two arguments may or may not be probability distributions[17]. Since (P_{UY_1}, Q_{UY_1}) are linear in P_{UX}, it follows that $I(X;Y_1|U)$ is concave in P_{UX}.

Regarding the other mutual information term,

$$\begin{aligned} I(U;Y_2) &= \sum_u \Big(\sum_{x,y_2} P_{UX}(u,x)P_{Y_2|X}(y_2|x)\Big) \log \frac{1}{P_U(u)} && (5.164) \\ &+ \sum_{u,y_2} \Big(\sum_x P_{UX}(u,x)P_{Y_2|X}(y_2|x)\Big) \log \frac{\sum_{x'} P_{UX}(u,x')P_{Y_2|X}(y_2|x')}{\sum_{x'} P_X(x')P_{Y_2|X}(y_2|x')} && (5.165) \\ &= H(U) - (-D(P_{UY_2}||Q_{UY_2})), && (5.166) \end{aligned}$$

where

$$Q_{UY_2}(u,y_2) = \sum_x P_X(x)P_{Y_2|X}(y_2|x)\ \forall (u,y_2) \in \mathcal{U}\times\mathcal{Y}_2 \quad (5.167)$$

is another dummy function not satisfying the properties of a probability distribution. By the concavity of the entropy function, the convexity of the divergence, and the fact that $(P_U, P_{UY_2}, Q_{UY_2})$ are linear in P_{UX} it follows that $I(U;Y_2)$ is a difference of concave functions of P_{UX}. □

[16] When required, we denote the marginals of P_{UX} by P_U and P_X to shorten notation.

[17] Actually, in this case $D(\cdot||\cdot)$ is not a proper divergence since its arguments are not probability distributions, but this is totally irrelevant for the convexity characterization of the problem.

5.H Appendix: Proof of Lemma 5.5

We will show that satisfaction of the KKT conditions, rephrased as in (5.86) is a necessary and sufficient condition for ensuring that P_{UX} achieves a local maximum of (5.83)-(5.85) and therefore (5.86) is a necessary optimality condition. From [Ber95, Sec. 3.4], we know that the KKT conditions of (5.83)-(5.85), are satisfied by any P_{UX} achieving a local maximum (hence proving the necessity part). The corresponding Lagrangian is

$$\begin{aligned}\mathcal{L}(P_{UX};\mathbf{\Phi}_{UX},\eta) &= \theta I(X;Y_1|U)+(1-\theta)I(U;Y_2) \qquad (5.168)\\ &+ \sum_{u,x}\Phi_{UX}(u,x)P_{UX}(u,x)+\eta\Big(\sum_{u,x}P_{UX}(u,x)-1\Big), \qquad (5.169)\end{aligned}$$

while the derivatives of mutual information are

$$\begin{aligned}\frac{\partial I(X;Y_1|U)}{\partial P_{UX}(u,x)} &= D(P_{Y_1|X=x}||P_{Y_1|U=u}) \qquad (5.170)\\ \frac{\partial I(U;Y_2)}{\partial P_{UX}(u,x)} &= \mathbb{E}_{Y_2|X=x}\left\{\log\frac{P_{Y_2|U=u}(Y_2)}{P_{Y_2}(Y_2)}\right\}-\log(e). \qquad (5.171)\end{aligned}$$

Setting the derivative of the Lagrangian with respect to $P_{UX}(u,x)$ equal to zero it follows that at any local maximum the following holds

$$\theta D(P_{Y_1|X=x}||P_{Y_1|U=u})+(1-\theta)\mathbb{E}_{Y_2|X=x}\left\{\log\frac{P_{Y_2|U=u}(Y_2)}{P_{Y_2}(Y_2)}\right\}=(1-\theta)\log(e)-\eta-\Phi_{UX}(u,x). \qquad (5.172)$$

Expression (5.172) can be rephrased as (5.86) using complementary slackness ($\Phi_{UX}(u,x)P_{UX}(u,x)=0$), dual feasibility ($\Phi_{UX}(u,x)\geq 0$), and noticing that an alternative formulation of the mutual informations involved is

$$\begin{aligned}I(X;Y_1|U) &= \sum_{u,x}P_{UX}(u,x)D(P_{Y_1|X=x}||P_{Y_1|U=u}) \qquad (5.173)\\ I(U;Y_2) &= \sum_{u,x}P_{UX}(u,x)\mathbb{E}_{Y_2|X=x}\left\{\log\frac{P_{Y_2|U=u}(Y_2)}{P_{Y_2}(Y_2)}\right\}. \qquad (5.174)\end{aligned}$$

To show that any P_{UX} satisfying the KKT conditions is indeed a local maximum of (5.83)-(5.85), let $\partial R_{P_{UX}}(u,x)\triangleq\theta\frac{\partial I(X;Y_1|U)}{\partial P_{UX}(u,x)}+(1-\theta)\frac{\partial I(U;Y_2)}{\partial P_{UX}(u,x)}$ denote a linear combination of (5.170)-(5.171) evaluated using P_{UX}. If

$$\sum_{u,x}\partial R_{P_{UX}}(u,x)(Q_{UX}(u,x)-P_{UX}(u,x))\leq 0 \qquad (5.175)$$

holds for any arbitrary distribution $Q_{UX}(u,x)$ it follows that P_{UX} is a local maximum of (5.83)-(5.85) [Ber95, Sec. 2.1]. Since any P_{UX} satisfying the KKT conditions has an associated $\partial R_{P_{UX}}(u,x)=-\eta-\Phi_{UX}(u,x)$ and satisfies complementary slackness ($\Phi_{UX}(u,x)P_{UX}(u,x)=0$), (5.175) becomes

$$-\sum_{u,x}\Phi_{UX}(u,x)Q_{UX}(u,x)\leq 0, \qquad (5.176)$$

which indeed holds for any Q_{UX} thanks to dual feasibility. □

5.I Appendix: Proof of Proposition 5.8

Let us simplify the notation by using the equivalence $P_{UX}(u,x) = P_{ux}$ and imposing $\mathcal{U} = \{0,1\}$ ($|\mathcal{U}| = 2$ suffices). In this case the distributions involved amount to

$$P_{Y_1|X} = \begin{bmatrix} 1-\epsilon & 0 \\ \epsilon & \epsilon \\ 0 & 1-\epsilon \end{bmatrix}, \; P_{Y_2|X} = \begin{bmatrix} 1-0.5\epsilon & 0.5\epsilon \\ 0.5\epsilon & 1-0.5\epsilon \end{bmatrix} \tag{5.177}$$

for the channel transition matrices,

$$P_{Y_1|U} = \begin{bmatrix} \frac{(1-\epsilon)P_{00}}{P_{00}+P_{01}} & \frac{(1-\epsilon)P_{10}}{P_{10}+P_{11}} \\ \epsilon & \epsilon \\ \frac{(1-\epsilon)P_{01}}{P_{00}+P_{01}} & \frac{(1-\epsilon)P_{11}}{P_{10}+P_{11}} \end{bmatrix}, \; P_{Y_2|U} = \begin{bmatrix} \frac{(1-0.5\epsilon)P_{00}+0.5\epsilon P_{01}}{P_{00}+P_{01}} & \frac{(1-0.5\epsilon)P_{10}+0.5\epsilon P_{11}}{P_{10}+P_{11}} \\ \frac{0.5\epsilon P_{00}+(1-0.5\epsilon)P_{01}}{P_{00}+P_{01}} & \frac{0.5\epsilon P_{10}+(1-0.5\epsilon)P_{11}}{P_{10}+P_{11}} \end{bmatrix} \tag{5.178}$$

for the output distributions conditioned on U, and

$$P_{Y_2} = \begin{bmatrix} (1-0.5\epsilon)(P_{00}+P_{10}) + 0.5\epsilon(P_{01}+P_{11}) \\ 0.5\epsilon(P_{00}+P_{10}) + (1-0.5\epsilon)(P_{01}+P_{11}) \end{bmatrix} \tag{5.179}$$

for the output distribution of the second receiver. In (5.177)-(5.179) we have used the convention that the columns of the matrices represent the elements of the output alphabets ($\mathcal{Y}_1$ or $\mathcal{Y}_2$) and the rows correspond to the natural ordering of the inputs (U or X). Expressions (5.177)-(5.179) can be used to formulate the conditions of Lemma 5.5 for the BEC-BSC dDMBC as follows

$$\begin{aligned} \theta(1-\epsilon)\log\frac{P_{00}+P_{01}}{P_{00}} + (1-\theta)\Big[(1-0.5\epsilon)\log\frac{(1-0.5\epsilon)P_{00}+0.5\epsilon P_{01}}{(P_{00}+P_{01})[(1-0.5\epsilon)(P_{00}+P_{10})+0.5\epsilon(P_{01}+P_{11})]} \\ +0.5\epsilon\log\frac{0.5\epsilon P_{00}+(1-0.5\epsilon)P_{01}}{(P_{00}+P_{01})[0.5\epsilon(P_{00}+P_{10})+(1-0.5\epsilon)(P_{01}+P_{11})]}\Big] = R(\theta) - \ell_{00} \end{aligned} \tag{5.180}$$

$$\begin{aligned} \theta(1-\epsilon)\log\frac{P_{10}+P_{11}}{P_{10}} + (1-\theta)\Big[(1-0.5\epsilon)\log\frac{(1-0.5\epsilon)P_{10}+0.5\epsilon P_{11}}{(P_{10}+P_{11})[(1-0.5\epsilon)(P_{00}+P_{10})+0.5\epsilon(P_{01}+P_{11})]} \\ +0.5\epsilon\log\frac{0.5\epsilon P_{10}+(1-0.5\epsilon)P_{11}}{(P_{10}+P_{11})[0.5\epsilon(P_{00}+P_{10})+(1-0.5\epsilon)(P_{01}+P_{11})]}\Big] = R(\theta) - \ell_{10} \end{aligned} \tag{5.181}$$

$$\begin{aligned} \theta(1-\epsilon)\log\frac{P_{00}+P_{01}}{P_{01}} + (1-\theta)\Big[0.5\epsilon\log\frac{(1-0.5\epsilon)P_{00}+0.5\epsilon P_{01}}{(P_{00}+P_{01})[(1-0.5\epsilon)(P_{00}+P_{10})+0.5\epsilon(P_{01}+P_{11})]} \\ +(1-0.5\epsilon)\log\frac{0.5\epsilon P_{00}+(1-0.5\epsilon)P_{01}}{(P_{00}+P_{01})[0.5\epsilon(P_{00}+P_{10})+(1-0.5\epsilon)(P_{01}+P_{11})]}\Big] = R(\theta) - \ell_{01} \end{aligned} \tag{5.182}$$

$$\begin{aligned} \theta(1-\epsilon)\log\frac{P_{10}+P_{11}}{P_{11}} + (1-\theta)\Big[0.5\epsilon\log\frac{(1-0.5\epsilon)P_{10}+0.5\epsilon P_{11}}{(P_{10}+P_{11})[(1-0.5\epsilon)(P_{00}+P_{10})+0.5\epsilon(P_{01}+P_{11})]} \\ +(1-0.5\epsilon)\log\frac{0.5\epsilon P_{10}+(1-0.5\epsilon)P_{11}}{(P_{10}+P_{11})[0.5\epsilon(P_{00}+P_{10})+(1-0.5\epsilon)(P_{01}+P_{11})]}\Big] = R(\theta) - \ell_{11}, \end{aligned} \tag{5.183}$$

where $\ell_{ij} \geq 0$ and $\ell_{ij} = 0$ if $P_{ij} > 0$. In order to find an optimal distribution satisfying (5.180)-(5.183) we need to hypothesize on the number of entries of P_{ij} that are equal to zero. To that end, consider the following situations:

- *Hypothesis 1 - P_{UX} has three zero entries.* This implies $H(U) = H(X) = 0$ and, consequently, $I(X;Y_1|U) = I(U;Y_2) = 0$, which is clearly suboptimal.

- *Hypothesis 2 - P_{UX} has two zero entries.* This class of distributions comprises the cases: i) $X = U$ and $X = \overline{U}$, where all the capacity-achieving distributions achieve $(0,(1-\theta)(1-h(0.5\epsilon))$, ii) $U = 0$ and $U = 1$, where all the capacity achieving distributions achieve $(\theta(1-\epsilon)\log 2, 0)$, and iii) $X = 0$ and $X = 1$, which imply $I(X;Y_1|U) = I(U;Y_2) = 0$.

- *Hypothesis 3 - P_{UX} has one zero entry.* Consider w.l.o.g $P_{00} = 0$ and $P_{ij} > 0\,\forall(i,j) \neq (0,0)$. This implies $\ell_{ij} = 0\ \forall(i,j) \neq (0,0)$ and $\ell_{00} \geq 0$. The conditions (5.180)-(5.183) cannot be satisfied simultaneously because the left hand side of (5.180) equals $+\infty$, while $R(\theta) \leq C^{\star}(\theta)$ is bounded and $\ell_{00} \geq 0$. Therefore, distributions with one zero entry are never optimal.

We subsequently focus on distributions P_{UX} with strictly positive entries, which imply $\ell_{ij} = 0$ $\forall i,j \in \{0,1\}$. Let us describe such distributions by

$$P_{UX} = \begin{bmatrix} p_\alpha & \alpha p_\alpha \\ \beta p_\beta & p_\beta \end{bmatrix}, \tag{5.184}$$

where $p_\alpha, p_\beta, \alpha, \beta > 0$ and $(1+\alpha)p_\alpha + (1+\beta)p_\beta = 1$. Setting the left hand sides of (5.180) and (5.182) equal to each other and using (5.184) it follows

$$(1-\theta)\log\frac{(1-0.5\epsilon)+0.5\epsilon\alpha}{0.5\epsilon+(1-0.5\epsilon)\alpha}\frac{0.5\epsilon(p_\alpha+\beta p_\beta)+(1-0.5\epsilon)(\alpha p_\alpha+p_\beta)}{(1-0.5\epsilon)(p_\alpha+\beta p_\beta)+0.5\epsilon(\alpha p_\alpha+p_\beta)} + \theta\log\alpha = 0. \tag{5.185}$$

Proceeding similarly with the left hand sides of (5.181) and (5.183) we arrive at

$$(1-\theta)\log\frac{(1-0.5\epsilon)\beta+0.5\epsilon}{0.5\epsilon\beta+(1-0.5\epsilon)}\frac{0.5\epsilon(p_\alpha+\beta p_\beta)+(1-0.5\epsilon)(\alpha p_\alpha+p_\beta)}{(1-0.5\epsilon)(p_\alpha+\beta p_\beta)+0.5\epsilon(\alpha p_\alpha+p_\beta)} - \theta\log\beta = 0. \tag{5.186}$$

Since both (5.185) and (5.186) must hold, we can equal their left hand sides, which imposes $g(\alpha;\theta,\epsilon) = -g(\beta;\theta,\epsilon)$, where g is defined in (5.93). On the other hand, considering that $g(1/\alpha;\theta,\epsilon) = -g(\alpha;\theta,\epsilon)$ it follows that $\beta = 1/\alpha'$ with α' such that $g(\alpha';\theta,\epsilon) = g(\alpha;\theta,\epsilon)$. Rewriting (5.185) as

$$(1-\theta)\log\frac{0.5\epsilon(p_\alpha+\beta p_\beta) + (1-0.5\epsilon)(\alpha p_\alpha+p_\beta)}{(1-0.5\epsilon)(p_\alpha+\beta p_\beta)+0.5\epsilon(\alpha p_\alpha+p_\beta)} = g(\alpha;\theta,\epsilon), \tag{5.187}$$

it follows that

$$\frac{p_\alpha+\beta p_\beta}{\alpha p_\alpha+\beta} = \frac{0.5\epsilon\exp\big(g(\alpha;\theta,\epsilon)/(1-\theta)\big) - (1-0.5\epsilon)}{0.5\epsilon - (1-0.5\epsilon)\exp\big(g(\alpha;\theta,\epsilon)/(1-\theta)\big)} \triangleq \gamma. \tag{5.188}$$

Since γ is the ratio of two strictly positive probabilities, we impose $\gamma > 0$ in (5.188) to obtain that

$$\gamma > 0 \Leftrightarrow |g(\alpha;\theta,\epsilon)| < (1-\theta)\log(2/\epsilon - 1). \tag{5.189}$$

An equivalent rephrasing of (5.188) is $p_\alpha/p_\beta = (\gamma - \beta)/(1 - \alpha\gamma)$, which induces a distribution of the form (5.88)-(5.91). Analyzing the monotony of $g(\alpha;\theta,\epsilon)$ over the interval of interest $\alpha \in (0, +\infty)$ it can be shown that

$$\operatorname{sign}\left\{\frac{\partial g(\alpha;\theta,\epsilon)}{\partial\alpha}\right\} = \operatorname{sign}\left\{-\alpha^2 + \frac{2}{\epsilon}\Big[\frac{1-\epsilon}{(1-0.5\epsilon)\theta} - 2(1-0.5\epsilon)\Big]\alpha - 1\right\}, \tag{5.190}$$

which shows that for $\theta \geq \frac{1-\epsilon}{2(1-0.5\epsilon)^2}$ the function $g(\alpha;\theta,\epsilon)$ is strictly decreasing and hence $\alpha' = \alpha$, $\beta = 1/\alpha$. It can be checked that (5.186)-(5.187) imply $\alpha = \beta = 1$ which causes $R_2 = 0$ and the optimal rate pair to be $((1-\epsilon)\log 2, 0)$.

When $0 < \theta < \frac{1-\epsilon}{2(1-0.5\epsilon)^2}$, the function $g(\alpha;\theta,\epsilon)$ has one local minimum and one local maximum, which bounds the number of different values of α' such that $g(\alpha';\theta,\epsilon) = g(\alpha;\theta,\epsilon)$ to a maximum of three. The maximization of the objective value $\theta I(X;Y_1|U) + (1-\theta)I(U;Y_2)$ over the distributions of the class (5.88)-(5.91) satisfying (5.189) yields the best distribution satisfying Lemma 2 with strictly positive probabilities, and its associated rate pair $(R_1(\theta), R_2(\theta))$. The convex hull of these rate pairs together with the extreme points $((1-\epsilon)\log 2, 0)$, $(0, (1-h(0.5\epsilon))\log 2)$ is hence $\mathcal{C}^{\mathrm{BEC-BSC}}$ since there is no other rate pair achieved by a distribution satisfying Lemma 5.5. □

Chapter 6

Conclusions

This dissertation has addressed the problem of multiuser interference in wireless networks under many different points of view. We started in Chapter 2 with the when and how of partial interference cancelation, a powerful technique to be used when the receivers of the network have full statistical knowledge of the interference. By giving special emphasis to the broadcast and interference channels, we addressed the network scenarios that suffer from multiuser interference in a more strict sense. The fact that a generalization of the best coding/decoding strategy for the broadcast channel allowing for simultaneous partial interference cancelation at the receivers did not achieve larger rates was rather surprising. This, in combination with the realization that a similar approach makes a big difference in the interference channel gave us the first conclusion of this thesis: coding and decoding complexity can be traded whenever the interference is under the control of the same source. Intuitively, appropriate coding techniques exploiting signal correlation can alleviate the complexity burden of the receivers.

While the study of the broadcast channel did not result in better achievability results, it was a key stepping stone that enabled the proposal of a novel transmission strategy for the interference channel: superposition coding and aided decoding. Compared to the best long-standing achievable region for the interference channel, superposition coding required less auxiliary random variables and aided decoding allowed to relax some rate inequalities. While these facts increase the chances of having a potentially larger achievable rate region, recent literature has shown that the proposed strategy and the best know result yield identical achievable rates. Thus, the only advantage of the proposed region is simplicity and potentially better performance in the finite blocklength regime via better error exponents.

In the context of Chapter 2, it is implicitly assumed that the receivers have full knowledge of the codebooks of the interfering users. However, that is not possible in applications backed by decentralized wireless networks with uncoordinated nodes. With receivers totally unaware of interference, partial cancelation becomes infeasible and leaves each sender-destination pair armed only with point-to-point (single user) strategies. This motivated the study of the totally asynchronous interference channel with single-user receivers in Chapter 3. Having a capacity region rather involved due to the need to resort to Information Spectrum formulation, the evaluation of achievable rates is tackled based on simpler single-letter inner bounds.

The study of the Gaussian case provides the second conclusion of the thesis: oppositely to what happens with frame synchronism, Gaussian distributed-codes fall short of maximizing the achievable rates for all channel instances. Despite their natural appeal, basically due to the fact that they always lead to closed-form characterizations of achievable regions, they are clearly suboptimal whenever the channel happens to be interference-limited. Indeed, this is the regime where appropriate statistical signal design can have a larger impact on the achievable rates. Analytical conditions determining the existence of non-Gaussian distributed codes yielding higher achievable rates are found to be in excellent agreement with the performance of explicit codes. Besides, the losses associated to the lack of transmission synchronism and the use of single-user decoders in the low- and high-power and low- and high-interference regimes have been quantified.

At this point, we adopted in Chapter 4 a rather different approach by studying how to deal with multiuser interference in a cellular network with an arbitrary number of users under practical constraints. By focusing on half-duplex MIMO terminals, OFDMA transmission, relaying infrastructure, and partial channel state information, most of the characteristics of upcoming wireless systems are taken into account for network utility maximization. With terminals unaware of interference, the rate degradation studied in Chapter 3 is mitigated thanks to the central coordinator role of the base station, which assigns transmission resources disjointly, in so facilitating link performance prediction for QoS provision.

Novel concave lower bounds on the ergodic capacities involved in the expression of the achievable rates in the cell enabled the proposal of two efficient algorithms for global (Pareto) optimal and sequential optimal resource allocation. These algorithms maximize network utility, a cell-wide aggregate indicator accounting for overall user satisfaction, and are able to deal with heterogeneous QoS requirements, diverse wireless equipment quality, and the use of non-ideal codes. The high-level conclusion of this chapter is that, by describing user satisfaction with a function depending only on the long-term throughput, the connection between physical layer operation and higher-layer needs becomes simpler and allows for feasible cross-layer optimization. The fewer the key performance indicators, the better.

Finally, Chapter 5 explored yet another different facet of multiuser interference: the complexity increase that it poses in the evaluation of fundamental performance limits. While the capacity evaluation of single-user channels is a problem that can efficiently solved, the evaluation of capacity regions of multiterminal networks is not such an amenable problem, as it often leads to unavoidable non-convexities. Conditioned by the lack of computable expressions, we focused on two specific network scenarios: the multiple access channel and the degraded broadcast channel.

For both channels, the computation of the capacity region implied solving a non-convex problem. Whereas in the multiple access channel the culprit was a rank-one constraint, in the degraded broadcast channel the difficulty arose from a difference of convex functions in the objective. Problems with these type of non-convexities cannot be solved optimally with the current state of the art. Our focus, hence, was to obtain efficient inner and outer bounds to their capacity region.

While for the multiple access channel we focused on relaxation methods that not only allowed us to obtain numerical results but also facilitated some analysis on their optimality, for the degraded broadcast channel the emphasis was on the characterization of the set of optimal input probability distributions. The remark of this chapter is that, while in the single-user case (no interference at all) a homogeneous approach can be taken in the evaluation of channel capacity, multiuser interference prevents us from doing that in a multiterminal network. By introducing non-convex constraints of different nature, the set of methods and strategies to be used in the evaluation of capacity regions must therefore be diverse.

Bibliography

[20005] IEEE Std. 802.16e 2005, “IEEE standard for local and metropolitan area networks, Part 16: Air interface for fixed and mobile broadband wireless access systems, Amendment 2: Physical and medium access control layers for combined fixed and mobile operation in licensed bands”, *Tech. Rep. IEEE Standards Dept.*, Dec. 2005.

[36.] 3GPP TS 36.201, *Evolved univeral terrestrial radio access (E-UTRA), Long term evolution (LTE) physical layer, General description*.

[Ahl68] R. Ahlswede, “The weak capacity of averaged channels”, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiet*, Vol. 11, pp. 61–73, 1968.

[Ahl71] R. Ahlswede, “Multi-way communication channels”, *Proc. IEEE Intl. Symp. Inform. Theory (ISIT)*, Tsakhadsor, Armenian SSR, 1971.

[Apk99] P. Apkarian, H. D. Tuan, “A sequential SDP/Gauss-Newton algorithm for rank-constrained LMI problems”, *Proc. IEEE Conf. on Decision and Control*, Phoenix, AZ, Dec. 1999.

[Ari72] S. Arimoto, “An algorithm for computing the capacity of arbitrary discrete memoryless channels”, *IEEE Trans. Inform. Theory*, Vol. IT-18, pp. 14–20, Jan. 1972.

[Bae06] C. Bae, D.-H. Cho, “Adaptive resource allocation based on channel information in multihop OFDM systems”, *Proc. IEEE VTC Fall*, Montreal, Canada, Sep. 2006.

[Ben79] R. Benzel, “The capacity region of a class of discrete additive degraded interference channels”, *IEEE Trans. Inform. Theory*, Vol. IT-25, pp. 228–231, March 1979.

[Ber73] P. Bergmans, “Random coding theorem for broadcast channels with degraded components”, *IEEE Trans. Inform. Theory*, Vol. IT-19, pp. 197–207, March 1973.

[Ber95] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, 1995.

[Bha06] S. R. Bhaskaran, E. Telatar, “Kurtosis constraints in communication over fading channels”, *Proc. IEEE Intl. Conf. on Communications (ICC)*, Istanbul, Turkey, June 2006.

[Bie79] M. Bierbaum, H.-M. Wallmeier, “A note on the capacity region of the multiple-access channel”, *IEEE Trans. Inform. Theory*, Vol. IT-25, pp. 484, July 1979.

[Big04] E. Biglieri, G. Taricco, *Transmission and reception with multiple antennas: theoretical foundations*, Now Publishers, 2004.

[Big07a] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, H. V. Poor, *MIMO Wireless Communications*, Cambridge University Press, 2007.

[Big07b] E. Biglieri, M. Lops, "Multiuser detection in a dynamic environment - Part I: User identification and data detection", *IEEE Trans. Inform. Theory.*, Vol. 53, pp. 3158–3170, Sep. 2007.

[Bla72] R. E. Blahut, "Computation of channel capacity and rate-distortion functions", *IEEE Trans. Inform. Theory*, Vol. IT-18, pp. 460–473, Jul. 1972.

[Boy04] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[Bro05] S. I. Bross, A. Lapidoth, "An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback", *IEEE Trans. Inform. Theory*, Vol. 51, pp. 811–833, Mar 2005.

[Cal05] E. Calvo, M. Stojanovic, "A coordinate descent algorithm for multichannel multiuser detection in underwater acoustic DS-CDMA systems", *Proc. IEEE OCEANS Europe Conf.*, Brest, France, June 2005.

[Cal07a] E. Calvo, J. R. Fonollosa, "Efficient resource allocation for orthogonal transmission in broadcast channels", *Proc. IEEE Workshop on Signal Process. Advances for Wireless Commun. (SPAWC)*, Helsinki, Finland, June 2007.

[Cal07b] E. Calvo, J. R. Fonollosa, J. Vidal, "Near-optimal joint power and rate allocation for OFDMA broadcast channels", *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Honolulu, HI, Apr. 2007.

[Cal07c] E. Calvo, D. P. Palomar, J. R. Fonollosa, J. Vidal, "The computation of the capacity region of the discrete MAC is a rank-one non-convex optimization problem", *Proc. IEEE Intl. Symp. on Inform. Theory (ISIT)*, Nice, France, June 2007.

[Cal07d] E. Calvo, D. P. Palomar, J. R. Fonollosa, J. Vidal, "On the computation of the capacity region of the discrete MAC", *submitted to IEEE Tran. Inform. Theory*, May 2007.

[Cal07e] E. Calvo, J. Vidal, J. R. Fonollosa, "Resource allocation in multihop OFDMA broadcast networks", *Proc. IEEE Workshop on Signal Process. Advances for Wireless Commun. (SPAWC)*, Helsinki, Finland, June 2007.

[Cal08a] E. Calvo, J. R. Fonollosa, J. Vidal, "The totally asynchronous interference channel with single-user receivers", *submitted to IEEE Trans. Inform. Theory*, Sep. 2008.

[Cal08b] E. Calvo, I. Kovács, L. García, J. R. Fonollosa, "A reconfigurable downlink air interface: design, simulation methodology, and performance evaluation", *Proc. ICT - Mobile Summit*, Stockholm, Sweden, June 2008.

[Cal08c] E. Calvo, D. P. Palomar, J. R. Fonollosa, J. Vidal, "The computation of the capacity region of the discrete degraded BC is a non-convex DC problem", *Proc. IEEE Intl. Symp. on Inform. Theory (ISIT)*, Toronto, Canada, July 2008.

[Cal08d] E. Calvo, M. Stojanovic, "Efficient channel estimation-based multi-user detection for underwater CDMA systems", *accepted for publication in IEEE Journal of Oceanic Engineering*, 2008.

[Cal08e] E. Calvo, J. Vidal, J. R. Fonollosa, "Optimal resource allocation in relay-assisted cellular networks with partial CSI", *submitted to IEEE Trans. Signal Process.*, May 2008.

[Car75] A. B. Carleial, "A case where interference does not reduce capacity", *IEEE Trans. Inform. Theory*, Vol. IT-21, pp. 569–570, Sep. 1975.

[Car78] A. B. Carleial, "Interference channels", *IEEE Trans. Inform. Theory*, Vol. IT-24, pp. 60–70, Jan 1978.

[Car83] A. B. Carleial, "Outer bounds on the capacity of interference channels", *IEEE Trans. Inform. Theory*, Vol. IT-29, pp. 602–606, July 1983.

[Cha79] S. Chang, E. J. Weldon Jr., "Coding for t-user multiple-access channels", *IEEE Trans. Inform. Theory*, Vol. IT-25, pp. 684–691, Nov. 1979.

[Cha81] S. Chang, J. K. Wolf, "On the t-user m-frequency noiseless multiple-access channel with and without intensity information", *IEEE Trans. Inform. Theory*, Vol. IT-27, pp. 41–48, Jan. 1981.

[Che93] R. S. Cheng, S. Verdú, "On limiting characterizations of memoryless multiuser capacity regions", *IEEE Trans. Inform. Theory.*, Vol. 39, pp. 609–612, Mar. 1993.

[Cho68] C. K. Chow, C. N. Liu, "Approximating discrete probability distributions with dependence trees", *IEEE Trans. Inform. Theory*, Vol. IT-14, pp. 462–467, May 1968.

[Cho06a] H. Chong, M. Motani, H. K. Garg, "A comparison of two achievable rate regions for the interference channel", *Proc. of the ITA Inaugural Workshop*, La Jolla, CA, Feb. 2006.

[Cho06b] H. F. Chong, M. Motani, H. K. Garg, H. El Gamal, "On the Han-Kobayashi region for the interference channel", *submitted to IEEE Trans. Inform. Theory*, Aug. 2006.

[Clo07] P. Closas, E. Calvo, J. Fernández, A. Pagès, "Coupling noise effect in self-synchronizing wireless sensor networks", *Proc. IEEE Workshop on Signal Process. Advances for Wireless Commun. (SPAWC)*, Helsinki, Finland, June 2007.

[Cos85] M. H. M. Costa, "On the Gaussian interference channel", *IEEE Trans. Inform. Theory*, Vol. IT-31, pp. 607–615, Sep. 1985.

[Cos87] M. H. M. Costa, A. El Gamal, "The capacity region of the discrete memoryless interference channel with strong interference", *IEEE Trans. Inform. Theory*, Vol. IT-33, pp. 710–711, Sep. 1987.

[Cov72] T. M. Cover, "Broadcast channels", *IEEE Trans. Inform. Theory*, Vol. IT-18, pp. 2–14, Jan. 1972.

[Cov75] T. M. Cover, "An achievable rate region for the broadcast channel", *IEEE Trans. Inform. Theory*, Vol. IT-21, pp. 399–404, July 1975.

[Cov79] T. M. Cover, A. A. El Gamal, "Capacity theorems for the relay channel", *IEEE Trans. Inform. Theory*, Vol. 25, pp. 572–584, Sep. 1979.

[Cov81a] T. M. Cover, C. S. K. Leung, "An achievable rate region for the multiple-access channel with feedback", *IEEE Trans. Inform. Theory*, Vol. IT-27, pp. 292–298, May 1981.

[Cov81b] T. M. Cover, R. J. McEliece, E. C. Posner, "Asynchronous multiple-access channel capacity", *IEEE Trans. Inform. Theory.*, Vol. IT-27, pp. 409–413, July 1981.

[Cov98] T. M. Cover, "Comments on broadcast channels", *IEEE Trans. Inform. Theory*, Vol. 44, pp. 2524–2530, Oct. 1998.

[Cov06] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2006.

[Dan97] G. B. Dantzig, M. N. Thapa, *Linear programming. 1: Introduction*, Springer, 1997.

[Dig01] S. N. Diggavi, T. M. Cover, "The worst additive noise under a covariance constraint", *IEEE Trans. Inform. Theory.*, Vol. 47, pp. 3072–3081, Nov. 2001.

[Doh04a] M. Dohler, A. Glekias, H. Aghvami, "Resource allocation for fdma-based regenerative multihop links", *IEEE Trans. Wireless Commun.*, Vol. 3, pp. 1989–1993, Nov. 2004.

[Doh04b] M. Dohler, A. Glekias, H. Aghvami, "A resource allocation strategy for distributed MIMO multi-hop communication systems", *IEEE Commun. Letters*, Vol. 8, pp. 99–101, Feb. 2004.

[Doh05] M. Dohler, H. Aghvami, "On the approximation of MIMO capacity", *IEEE Trans. Wireless Commun.*, Vol. 4, pp. 30–34, Jan. 2005.

[Dup04] F. Dupuis, W. Yu, F. M. J. Willems, "Blahut-Arimoto algorithms for computing channel capacity and rate-distortion with side information", *Proc. IEEE Intl. Symp. Inform. Theory (ISIT)*, Chicago, IL, June/July 2004.

[Eff08] M. Effros, A. Goldsmith, Y. Liang, "Capacity definitions for general channels with receiver side information", *submitted to IEEE Trans. Inform. Theory.*, Apr. 2008.

[Etk07] R. H. Etkin, D. N. C. Tse, H. Wang, "Gaussian interference channel capacity to within one bit", *submitted to IEEE Trans. Inform. Theory*, Feb. 2007.

[Fie71] M. Fiedler, "Bounds for the determinant of the sum of Hermitian matrices", *Proc. American Math. Society*, Vol. 30, pp. 27–31, Sep. 1971.

[Gal74] R. G. Gallager, "Capacity and coding for degraded broadcast channels", *Problemy Peredaci Informaccii*, Vol. 10, pp. 3–14, Oct. 1974.

[Gam81] A. El Gamal, E. C. van der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel", *IEEE Trans. Inform. Theory*, Vol. IT-27, pp. 120–122, Jan. 1981.

[Gam82] A. El Gamal, M. H. M. Costa, "The capacity region of a class of deterministic interference channels", *IEEE Trans. Inform. Theory*, Vol. IT-28, pp. 343–346, March 1982.

[Gra00] I. S. Gradshteyn, I. M. Ryzhik, *Table of integrals, series, and products*, Academic Press, 6th Ed., 2000.

[Gro06] IEEE 208.16 Relay Task Group, "Multi-hop relay system evaluation methodology (channel model and performance metric)", *IEEE 802.16j-06/013r2*, Nov. 2006.

[Gro07] IEEE 802.16 Broadband Wireless Access Group, "Ieee 802.16m system requirements", *IEEE 802.16m-07/002r4*, Oct. 2007.

[Gup00] P. Gupta, P. R. Kumar, "The capacity of wireless networks", *IEEE Trans. Inform. Theory.*, Vol. 46, pp. 388–404, March 2000.

[Haj79] B. E. Hajek, M. B. Pursley, "Evaluation of an achievable rate region for the broadcast channel", *IEEE Trans. Inform. Theory*, Vol. IT-25, pp. 36–46, Jan. 1979.

[Han81] T. S. Han, K. Kobayashi, “A new achievable rate region for the interference channel”, *IEEE Trans. Inform. Theory*, Vol. IT-27, pp. 49–60, Jan. 1981.

[Hel95] C. W. Helstrom, *Elements of signal detection and estimation*, Prentice Hall, 1995.

[Hen01] D. Henrion, G. Meinsma, “Rank-one LMIs and Lyapunov’s inequality”, *IEEE Trans. Autom. Control*, Vol. 46, pp. 1285–1288, Aug. 2001.

[HM05] A. Høst-Madsen, J. Zhang, “Capacity bounds and power allocation for wireless relay channels”, *IEEE Trans. Inform. Theory*, Vol. 51, pp. 2020–2040, June 2005.

[Hor99] R. Horst, N. V. Thoai, “DC programming: an overview”, *Jrnl. of Optim. Theory and Applications*, Vol. 103, pp. 1–43, Oct. 1999.

[Hui85] J. Y. N. Hui, P. A. Humblet, “The capacity region of the totally asynchronous multiple-access channel”, *IEEE Trans. Inform. Theory.*, Vol. IT-31, pp. 207–216, Mar. 1985.

[Iha78] S. Ihara, “On the capacity of channels with additive non-Gaussian noise”, *Info. Ctrl.*, Vol. 37, pp. 34–39, Apr. 1978.

[Jai84] R. Jain, D. Chiu, W. Hawe, “A quantitative measure of fairness and discrimination for resource allocation in shared systems”, *DEC Research Report TR-301*, 1984.

[Jal03] J. Jaldén, C. Martin, B. Ottersten, “Semidefinite programming for detection in linear systems - optimality conditions and space-time decoding”, *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong, Apr. 2003.

[Kas76] T. Kasami, S. Lin, “Coding for a multiple-access channel”, *IEEE Trans. Inform. Theory*, Vol. IT-22, pp. 129–136, March 1976.

[Kel98] F. P. Kelly, A. Maulloo, D. Tan, “Rate control for communication networks: shadow prices, proportional fairness, and stability”, *Jrnl. Operations Research Society*, Vol. 49, pp. 237–252, March 1998.

[Kie05] M. Kießling, “Unifying analysis of ergodic MIMO capacity in correlated Rayleigh fading environments”, *Europ. Trans. Telecomm.*, Vol. 16, pp. 17–35, Jan. 2005.

[Kra04] G. Kramer, “Outer bounds on the capacity of Gaussian interference channels”, *IEEE Trans. Inform. Theory*, Vol. 50, pp. 581–586, March 2004.

[Ku69] H. Ku, S. Kullback, “Approximating discrete probability distributions”, *IEEE Trans. Inform. Theory*, Vol. IT-15, pp. 444–447, July 1969.

[Lap03] A. Lapidoth, S. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels”, *IEEE Trans. Inform. Theory*, Vol. 49, pp. 2426–2467, Oct. 2003.

[Lee06] K.-D. Lee, V. C. M. Leung, “Fair allocation of subcarrier and power in an OFDMA wireless mesh network”, *IEEE Jrnl. Select. Areas Commun.*, Vol. 24, pp. 2051–2060, Nov. 2006.

[Les06] A. Leshem, E. Zehavi, “Bargaining over the interference channel”, *Proc. IEEE Intl. Symp. on Inform. Theory (ISIT)*, Seattle, WA, July 2006.

[Lia72] H. J. Liao, *Multiple access channels*, Ph.D. dissertation, Dep. Elec. Eng. Univ. of Hawaii, 1972.

[Lia07] Y. Liang, V. V. Veeravalli, H. V. Poor, "Resource allocation for wireless fading relay channels: max-min solution", *IEEE Trans. Inform. Theory*, Vol. 53, pp. 3432–3453, Oct. 2007.

[Lin06] X. Lin, N. B. Shroff, R. Srikant, "A tutorial on cross-layer optimization in wireless networks", *IEEE Jrnl. Selec. Areas Commun.*, Vol. 24, pp. 1452–1463, Aug. 2006.

[Liu07] T. Liu, P. Viswanath, "An extremal inequality motivated by multiterminal information theoretic problems", *IEEE Trans. Inform. Theory.*, Vol. 53, pp. 1839–1851, May 2007.

[Ma02] W. Ma, T. N. Davidson, K. M. Wong, Z. Luo, P. Ching, "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous cdma", *IEEE Trans. Signal Processing*, Vol. 50, pp. 912–922, Apr. 2002.

[Mar79] K. Marton, "A coding theorem for the discrete memoryless broadcast channel", *IEEE Trans. Inform. Theory*, Vol. IT-25, pp. 306–311, May 1979.

[Meu75] E. van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel", *IEEE Trans. Inform. Theory*, Vol. IT-21, pp. 180–190, March 1975.

[Meu77] E. C. van der Meulen, "A survey of multi-way channels in information theory: 1961-1976", *IEEE Trans. Inform. Theory*, Vol. IT-23, pp. 1–37, Jan. 1977.

[Mo00] J. Mo, J. Walrand, "Fair end-to-end window-based congestion control", *IEEE/ACM Trans. Networking*, Vol. 8, pp. 556–567, Oct. 2000.

[Muñ07] O. Muñoz, J. Vidal, A. Agustín, E. Calvo, A. Alcón, "Resource management for relaying-enhanced WiMAX: OFDM and OFDMA", *Workshop Trends in Radio Resource Manag.*, Barcelona, Spain, Nov. 2007.

[Nab04] R. Nabar, H. Bölcskei, F. Kneubühler, "Fading relay channels: performance limits and space-time signal design", *IEEE Jrnl. Select. Areas Commun.*, Vol. 22, pp. 1099–1109, Aug. 2004.

[Nai07] C. Nair, A. El Gamal, "An outer bound to the capacity region of the broadcast channel", *IEEE Trans. Inform. Theory*, Vol. 53, pp. 350–355, Jan. 2007.

[Ng07] T. C.-Y. Ng, W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks", *IEEE Jrnl. Select. Areas Commun.*, Vol. 25, pp. 328–339, Feb. 2007.

[Nik93] C. L. Nikias, P. A. Petropulu, *Higher order spectral analysis: A nonlinear signal processing framework*, Englewood Cliffs, New Jersey: Prentice Hall, 1993.

[Och06] H. Ochiani, P. Mitran, V. Tarokh, "Variable rate two phase collaborative communication protocols for wireless networks", *IEEE Trans. Inform. Theory*, Vol. 52, pp. 4299–4313, Sep. 2006.

[Ors04] R. Orsi, U. Helmke, J. B. Moore, "A Newton-like method for solving rank constrained linear matrix inequalities", *Proc. IEEE Conf. on Decision and Control*, Paradise Island, Bahamas, Dec. 2004.

[Oza84] L. H. Ozarow, "The capacity region of the white Gaussian multiple access channel with feedback", *IEEE Trans. Inform. Theory*, Vol. IT-30, pp. 623–629, July 1984.

[Pal05] D. P. Palomar, "Convex primal decomposition for multicarrier linear MIMO transceivers", *IEEE Trans. Signal Processing*, Vol. 53, pp. 4661–4674, Dec. 2005.

[Pal07] D. P. Palomar, M. Chiang, "Alternative distributed algorithms for network utility maximization: framework and applications", *IEEE Trans. Automatic Control*, Vol. 52, pp. 2254–2269, Dec. 2007.

[Pan08] J. Pang, G. Scutari, F. Facchinei, C. Wang, "Distributed power allocation with rate constraints in parallel interference channels", *IEEE Trans. Inform. Theory.*, Vol. 54, pp. 3471–3489, Aug. 2008.

[Pap84] A. Papoulis, *Probability, random variables, and stochastic processes*, McGraw.Hill, 2nd ed., 1984.

[Pea16] K. Pearson, "Mathematical contributions to the theory of evolution. XIX. Second suplement to a memoir on skew variation", *Philosophical Trans. of the Royal Society of London, Series A*, Vol. 216, pp. 429–457, 1916.

[Pin64] M. S. Pinsker, *Information and information stability of random variables and processes*, San Francisco: Holden-Day, 1964.

[Rez04] M. Rezaeian, A. Grant, "Computation of total capacity for discrete memoryless multiple-access channels", *IEEE Trans. Inform. Theory*, Vol. 50, pp. 2779–2784, Nov. 2004.

[RT05] J. R. Ruíz-Tolosa, E. Castillo, *From vectors to tensors*, Springer, 2005.

[Sas04] I. Sason, "On achievable rate regions for the Gaussian interference channel", *IEEE Trans. Inform. Theory*, Vol. 50, pp. 1345–1356, June 2004.

[Sat78] H. Sato, "On the capacity region of a discrete two-user channel for strong interference", *IEEE Trans. Inform. Theory*, Vol. IT-24, pp. 377–379, May 1978.

[Sat81] H. Sato, "The capacity of the Gaussian interference channel under strong interference", *IEEE Trans. Inform. Theory*, Vol. IT-27, pp. 786–788, Nov. 1981.

[SB06] A. Somekh-Barruch, S. Verdú, "General relayless networks: representation of the capacity region", *Proc. IEEE Intl. Symp. on Inform. Theory (ISIT)*, Seattle, WA, July 2006.

[Scu08a] G. Scutari, D. P. Palomar, S. Barbarossa, "Optimal linear precoding strategies for wideband non-cooperative systems based on game theory - Part I: Nash equilibria", *IEEE Trans. Signal Process.*, Vol. 56, pp. 1230–1249, March 2008.

[Scu08b] G. Scutari, D. P. Palomar, S. Barbarossa, "Optimal linear precoding strategies for wideband non-cooperative systems based on game theory - Part II: Algorithms", *IEEE Trans. Signal Process.*, Vol. 56, pp. 1250–1267, March 2008.

[Sha48] C. E. Shannon, "A mathematical theory of communication", *Bell Syst. Tech. J.*, Vol. 27, pp. 379–423, 1948.

[Sha07] X. Shang, G. Kramer, B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels", *submitted to IEEE Trans. Inform. Theory*, Dec 2007.

[Shi03] H. Shin, J. H. Lee, "Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole", *IEEE Trans. Inform. Theory*, Vol. 49, pp. 2636–2647, Oct. 2003.

[Sle73] D. Slepian, J. K. Wolf, "Noiseless coding of correlated information sources", *IEEE Trans. Inform. Theory*, Vol. IT-19, pp. 471–480, July 1973.

[Teo07] K. H. Teo, Z. Tao, J. Zhang, "The mobile broadband WiMAX standard", *IEEE Signal Process. Mag.*, Vol. 24, pp. 144–148, Sep. 2007.

[Tse05] D. Tse, P. Viswanath, *Fundamentals of wireless communications*, Cambridge University Press, 2005.

[Val03] M. C. Valenti, B. Zhao, "Distributed turbo codes: Towards the capacity of the relay channel", *Proc. IEEE VTC Fall*, Orlando, FL, Oct. 2003.

[Van86] P. Vanroose, E. C. van der Meulen, "Coding for the binary switching multiple access channel", *Proc. Symp. Inform. Theory in the Benelux*, Noordwijkerhout, 1986.

[Ver89] S. Verdú, "Multiple-access channels with memory with and without frame synchronism", *IEEE Trans. Inform. Theory.*, Vol. 35, pp. 605–619, May 1989.

[Ver94] S. Verdú, T. S. Han, "A general formula for channel capacity", *IEEE Trans. Inform. Theory.*, Vol. 40, pp. 1147–1157, July 1994.

[Vin85a] A. J. Vinck, "On the multiple access channel", *Proc. 2nd Joint Swedish-Soviet Intl. Workshop Inform. Theory*, Gränna, Sweden, 1985.

[Vin85b] A. J. Vinck, W. Hoeks, K. Post, "On the capacity of the 2-user m-ary multiple access channel with feedback", *IEEE Trans. Inform. Theory*, Vol. IT-31, pp. 540–543, July 1985.

[Wan05] B. Wang, J. Zhang, A. Høst-Madsen, "On the capacity of MIMO relay channels", *IEEE Trans. Inform. Theory*, Vol. 51, pp. 29–43, Jan. 2005.

[Wat96] Y. Watanabe, "The total capacity of two-user multiple-access channel with binary output", *IEEE Trans. Inform. Theory*, Vol. 42, pp. 1453–1465, Sep. 1996.

[Wat02] Y. Watanabe, "The total capacity of multiple-access channel", *Proc. IEEE Intl. Symp. Inform. Theory (ISIT)*, Lausanne, Switzerland, June/July 2002.

[Wei06] H. Weingarten, Y. Steinberg, S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel", *IEEE Trans. Inform. Theory*, Vol. 52, pp. 3936–3964, Sep. 2006.

[Wie05] A. Wiesel, Y. C. Eldar, S. Shamai, "Semidefinite relaxation for detection of 16-QAM signaling in MIMO channels", *IEEE Letters on Signal Processing*, Vol. 12, pp. 653–656, Sep. 2005.

[Wil82] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels", *IEEE Trans. Inform. Theory*, Vol. IT-28, pp. 93–95, Jan. 1982.

[Won99] C. Y. Wong, R. S. Cheng, K. B. Letaief, R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation", *IEEE Jrnl. Select. Areas Commun.*, Vol. 17, pp. 1747–1758, Oct. 1999.

[Wyn73] A. D. Wyner, "A theorem on the entropy of certain binary sequences and applications: Part II", *IEEE Trans. Inform. Theory*, Vol. IT-19, pp. 772–777, Nov. 1973.

[Xie04] L. Xie, P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation", *IEEE Trans. Inform. Theory.*, Vol. 50, pp. 748–767, May 2004.

[Xue06] Y. Xue, B. Li, K. Nahrstedt, "Optimal resource allocation in wireless ad hoc networks: a price-based approach", *IEEE Trans. Mobile Computing*, Vol. 5, pp. 347–364, Apr. 2006.

[YJC07] Y.-J-Chang, F.-T. Chien, C.-C. J. Kuo, "Cross-layer QoS analysis of opportunistic OFDM-TDMA and OFDMA networks", *IEEE Jrnl. Select. Areas Commun.*, Vol. 25, pp. 657–666, May 2007.

[Yu04] W. Yu, W. Rhee, S. Boyd, J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels", *IEEE Trans. Inform. Theory*, Vol. 50, pp. 145–152, Jan. 2004.

[Yu05] M. Yu, J. Li, "Is amplify-and-forward practically better than decode-and-forward or vice versa?", *Proc. IEEE Intl. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, March 2005.

[Yu06] W. Yu, R. Lui, "Dual methods for non-convex spectrum optimization of multicarrier systems", *IEEE Trans. Commun.*, Vol. 54, pp. 1310–1322, July 2006.