Spectrum Sharing: Fundamental Limits, Scaling Laws, and Self-Enforcing Protocols

by

Raúl Hernán Etkin

Graduate (University of Buenos Aires, Argentina) 1999
M.S. (University of California, Berkeley) 2003

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering-Electrical Engineering and Computer Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Adjunct Professor Abhay K. Parekh, Co-chair
Professor David N. C. Tse, Co-chair
Professor Robert W. Brodersen
Professor Howard A. Shelanski

Fall 2006

The dissertation of Raúl Hernán Etkin is approved:

Co-chair Date

Co-chair Date

Date

Date

University of California, Berkeley

Fall 2006

Spectrum Sharing: Fundamental Limits, Scaling Laws, and Self-Enforcing Protocols

Abstract

Spectrum Sharing: Fundamental Limits, Scaling Laws, and Self-Enforcing Protocols

by

Raúl Hernán Etkin

Doctor of Philosophy in Engineering-Electrical Engineering and Computer Sciences

University of California, Berkeley

Adjunct Professor Abhay K. Parekh, Co-chair

Professor David N. C. Tse, Co-chair

Spectrum sharing arises whenever multiple wireless systems operate in the same frequency band. Due to mutual interference, the performance of the systems is coupled and tradeoff occurs between the rates that can be simultaneously achieved. The fundamental characterization of this tradeoff is given by the capacity region of the interference channel, whose determination is an open problem in information theory. For the case of the two-system Gaussian interference channel we provide a characterization of this capacity region to within a single bit/s/Hz. This characterization is obtained by deriving new outer bounds, and by using simple communication schemes that are special cases of the ones introduced by Han and Kobayashi in [29].

When many systems share the band, the interference cancellation techniques used in the two-system case may not be feasible, and a large fraction of the interference must be treated as noise. As the number of systems grows, interference aggregates and limits performance. We study the statistical properties of the interference aggregation phenomenon in a random network model and determine how to mitigate the strongest interference components.

Frequency selective fading can provide gains in a multi-user system due to multi-user diversity. We investigate whether similar gains can be achieved in multi-system spectrum sharing situations. For this we fix the rate of each system and study how the required bandwidth scales as the number of systems M grows large. While for Rayleigh fading the multi-user diversity gain provides bandwidth savings of the order of $\log(\log M)$, the multi-system diversity gain can provide larger bandwidth savings, of order $\log M$.

We lastly consider the problem of incentives in spectrum sharing. Systems are often independent and selfish. We investigate whether efficiency and fairness can be obtained with self-enforcing spectrum sharing rules that do not require cooperation among the systems. Any self-enforcing protocol must correspond to an equilibrium of a game. We first analyze the possible outcomes of a one shot game, and notice many inefficient solutions. However, since systems often coexist for long periods, a repeated game is more appropriate to model their interaction. In the repeated game, the possibility of building reputations and applying punishments enables a larger set of self-enforcing outcomes. When this set includes the optimal operating point, efficient, fair, and incentive compatible spectrum sharing becomes possible. We prove that our results are tight and quantify the best achievable performance in non-cooperative scenarios.

Adjunct Professor Abhay K. Parekh
Dissertation Committee Co-chair

Professor David N. C. Tse
Dissertation Committee Co-chair

This thesis is dedicated to the memory of my father Roberto, who explored new horizons as a young man, and dreamed of me having the same opportunity. His belief in higher education, and enthusiasm toward accepting new challenges, gave me the courage to leave my home country and pursue the doctorate degree.

This dedication is extended to my mother Claudia, who, besides having a busy and successful professional and artistic career, always made her family a priority. She has been present through every step of my life, providing support in difficult times, and enjoying every one of my accomplishments.

Finally, this dissertation is dedicated to my loving wife Vesna, my friend, my companion, and my source of inspiration for life. Her patience and understanding extend beyond the development and the writing of this thesis. I am thankful for all of her love and support.

Contents

List of Figures

Preface

Spectrum sharing occurs in any situation where multiple wireless devices coexist in the same area and use the same frequency band to communicate. The wireless spectrum is a broadcast medium: any signal transmitted by a device is received with some attenuation by all other devices operating in the same band. This property is useful when the same signal must be transmitted to multiple receivers. However, when many systems communicate independent information they interfere with each other and the broadcast nature of spectrum becomes a disadvantage.

Since radio signals are attenuated by several factors (distance, shadowing, multi-path fading) the same portion of spectrum can be used simultaneously by multiple systems operating independently. If the attenuation between systems is large, the performance of one system is not affected by the other systems. In this case, the performance is said to be *noise limited*: the interference is small compared to the noise and each system operates as if there were no other systems present. On the other hand, if the interference between systems is not negligible the performance of the different systems are coupled. In this case the performance is said to be *interference limited*.

In this thesis we study several aspects of the spectrum sharing problem. In the first part of this work we study how interference affects the performance of the wireless systems sharing the band. Interference is not always harmful: in some situations the effects of interference can be completely removed and the systems operate as if they were alone in the band. However, in other cases interference has a negative impact on the systems. In order to obtain a fundamental characterization of how interference limits communication we study the *Gaussian interference channel* in Part I. The information theoretic analysis that we do in this part allows us to tightly characterize (to within 1 bit/s/Hz) the achievable rates in a spectrum sharing situation between two systems. In addition, we present communication schemes that suggest how to design efficient protocols for spectrum sharing.

The second part of this work considers networks where the number of systems M is large. Unfortunately the M-user interference channel for $M > 2$ is not very well understood, and little work, if any, has been done in the area. In particular, if M is large, a tight and complete characterization of the fundamental communication limits in spectrum sharing may be beyond reach. However, one can gain insight into the spectrum sharing problem by looking at asymptotic regimes in which the number of systems M is allowed to grow. In Chapter 7 we study how interference aggregates when the systems are randomly distributed on the plane with uniform density. We do a probabilistic analysis to study the statistical properties of the received interference. This analysis shows that unless the frequency assignments are carefully coordinated between neighboring systems, the aggregate received interference may be very large.

The channel gains between transmitters and receivers may vary over frequency due to frequency selective fading. In some situations, these fluctuations can be *opportunistically* exploited to obtain a performance gains. In Chapter 8 we study the asymptotic performance improvement that can be obtained from frequency selective fading as the number of systems M grows large. This analysis complements the insights provided in the previous chapters that assume a frequency flat fading model.

The third part of the thesis deals with the problems that arise when the systems operate as independent, rational, and selfish entities, that try to maximize their own objectives. The models of the first and second parts of the thesis implicitly assume that the systems cooperate to achieve a common goal. Unfortunately, if the systems behave selfishly, greedy behavior may lead to inefficient and unfair situations. We formulate the spectrum sharing problem between selfish systems as a game, and analyse how to obtain good performance without assuming cooperation. The result of our study shows how to design spectrum sharing protocols that are *self-enforcing*, that is, that do not require the intervention of a central authority to verify compliance. These protocols try to achieve an efficient and fair allocation of resources between the systems.

This brief overview is complemented with more complete introductions in each part of the thesis. In addition, each part contains a section that summarizes the main results, points out the insights gained from the analysis, and mentions open problems for future research.

Acknowledgments

I would like to thank my mentor Professor David Tse for selflessly sharing his knowledge, intuition and teaching talent with me. I am grateful for Professor Tse's guidance and support that expand well beyond his responsibility as a research advisor. Not only did he teach me how to formulate and analyze research problems, but Professor Tse also made me aware of the benefits of interpreting results intuitively. His creative approach to teaching and giving engaging presentations was particularly enlightening. His positive thinking and encouraging words in difficult times gave me the confidence to master challenging research problems. Professor Tse made me the researcher that I am today, and for that I will always be grateful.

I would further like to thank Professor Abhay Parekh, my co-advisor for the past three years, for sharing his knowledge and experience with me. Professor Parekh's expertise in networking, distributed algorithms and economics were of great help during both formulation and solution of many of the research problems investigated in this thesis. In addition, his industrial experience was invaluable for keeping a practical perspective throughout the research process. Equally, I would like to thank Abhay Parekh for his optimism and friendliness.

The research that led to this thesis was funded by the National Science Foundation grant "The 3 R's of Spectrum Sharing: Reduce, Reuse, and Recycle". I would like to extend my gratitude to the members of the "3 R's Research Group": professors Michael Gastpar, Kannan Ramchandran, Anant Sahai, and Howard Shelanski; graduate students Krishnan Eswaran, Niels Hoven, Andrea Martens, Rahul Tandra, Danijela Cabric, Shridhar Mubaraq Mishra, and Ben Wild, for their comments and suggestions.

I would also like to thank Hua Wang, who collaborated with Professor Tse and me in the development of the Gaussian interference channel results during his visit to U.C. Berkeley in the summer of 2006.

Last but not least, I would like to thank Professors Robert Brodersen and Howard Shelanski for their considerable participation in my qualifying exam and dissertation committees.

Part I

Two-User Gaussian Interference Channel

Chapter 1

Preliminaries

1.1 Introduction

Interference is a central phenomenon in wireless communication when multiple uncoordinated links share a common communication medium. Most state-of-the-art wireless systems deal with interference in one of two ways:

- orthogonalize the communication links in time or frequency, so that they do not interfere with each other at all;
- allow the communication links to share the same degrees of freedom, but treat each other's interference as adding to the noise floor.

It is clear that both approaches can be sub-optimal. The first approach entails an *a priori* loss of degrees of freedom in both links, no matter how weak the potential interference is. The second approach treats interference as pure noise while it actually carries information and has structure that can potentially be exploited in mitigating its effect.

These considerations lead to the natural question of what is the best performance one can achieve without making any *a priori* assumptions on how the common resource is shared. A basic information theory model to study this question is the two-user Gaussian interference channel, where two point-to-point links with additive white Gaussian noise interfere with each other (Figure 1.1). The capacity region of this channel is the set of all simultaneously achievable rate pairs (R_1, R_2)

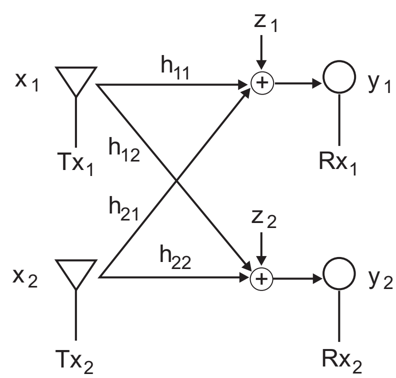


Figure 1.1: Two-user Gaussian interference channel.

in the two interfering links, and characterizes the fundamental tradeoff between the performance achievable in the two links in face of interference. Unfortunately, the problem of characterizing this region has been open for over thirty years. The only case in which the capacity is known is in the *strong* interference case, where each receiver has a better reception of the other user's signal than the intended receiver [7, 29]. The best known strategy for the other cases is due to Han and Kobayashi [29]. This strategy is a natural one and involves splitting the transmitted information of both users into two parts: private information to be decoded only at own receiver and common information that can be decoded at both receivers. By decoding the common information, part of the interference can be cancelled off, while the remaining private information from the other user is treated as noise. The Han-Kobayashi strategy allows arbitrary splits of each user's transmit power into the private and common information portions as well as time sharing between multiple such splits. Unfortunately, the optimization among such myriads of possibilities is not well-understood, so while it is clear that it will be no worse than the above-mentioned strategies as it includes them as special cases, it is not very clear how much improvement can be obtained and in which parameter regime would one get significant improvement. More importantly, it is also not clear how close to capacity can such a scheme get and whether there will be other strategies that can do significantly better.

In this part of the thesis, we make progress on this state of affairs by showing that a very simple Han-Kobayashi type scheme can in fact achieve rates within 1 bits/s/Hz of the capacity of the channel for *all* values of the channel parameters. That is, for all rate pairs (R_1, R_2) on the boundary of the achievable region, $(R_1 + 1, R_2 + 1)$ is not achievable. This result is particularly relevant in the high signal-to-noise-ratio (SNR) regime, where the achievable rates are high and grow unbounded as the noise level goes to zero. In fact, in some high SNR regimes, we can strengthen our results to show that our scheme is asymptotically optimal.

The high SNR regime is the interference-limited scenario: when the noise is small, interference from one link will have a significant impact on the performance of the other. The low SNR regime is less interesting since here the performance of each link is primarily noise-limited and interference is not having a significant effect.

The key feature of our proposed scheme is that the power of the private information of each user should be set such that it is received at the level of the Gaussian noise at the other receiver. In this way, the interference caused by the private information has a small effect on the other link beyond the impairment that is already caused by the noise. At the same time, quite a lot of private information can be conveyed to the intending receiver if the direct gain is appreciably larger than the cross gain.

To prove our result, we need good outer bounds on the capacity region of the interference channel. The best known outer bound is based on giving extra side information to one of the receivers so that it can decode all of the information from the other user (the one-sided channel and related bounds). It turns out that while this bound is sufficiently tight in some parameter regimes, it can get arbitrarily loose in others. We derive new outer bounds to cover for the other parameter ranges.

At high SNR, it is well known that the capacity of a point-to-point additive white Gaussian noise (AWGN) link, in bits/s/Hz, is approximately:

$$C_{\text{awgn}} \approx \log \mathsf{SNR} \tag{1.1}$$

Using our results, we can derive analogous approximations of the Gaussian interference channel capacity, accurate to within one bit/s/Hz. Just to give a flavor of the results, let us consider the symmetric case where the signal-to-noise ratios at the two receivers are the same (denoted by SNR)and the interference-to-noise ratios at the receivers are also the same (denoted by INR). The symmetric capacity, i.e. the best rate that both users can simultaneously achieve, is approximately:

$$C_{\text{sym}} \approx \begin{cases} \log(\frac{\mathsf{SNR}}{\mathsf{INR}}) & \log \mathsf{INR} < \frac{1}{2}\log \mathsf{SNR} & \text{(regime 1)} \\ \log \mathsf{INR} & \frac{1}{2}\log \mathsf{SNR} < \log \mathsf{INR} < \frac{2}{3}\log \mathsf{SNR} & \text{(regime 2)} \\ \log \frac{\mathsf{SNR}}{\sqrt{\mathsf{INR}}} & \frac{2}{3}\log \mathsf{SNR} < \log \mathsf{INR} < \log \mathsf{SNR} & \text{(regime 3)} \\ \log \sqrt{\mathsf{INR}} & \log \mathsf{SNR} < \log \mathsf{INR} < 2\log \mathsf{SNR} & \text{(regime 4)} \\ \log \mathsf{SNR} & \log \mathsf{INR} > 2\log \mathsf{SNR} & \text{(regime 5)} \end{cases} \tag{1.2}$$

Note that there are five regimes in which the qualitative behaviors of the capacity are different.

The fifth regime is the *very strong interference* regime [7]. Here the interference is so strong that each receiver can decode the other transmitter's information, treating its own signal as noise, before decoding its own information. Thus, interference has no impact on the performance of the other link. The fourth regime is the *strong interference* regime, where the optimal strategy is for both receivers to decode entirely each other's signal, i.e. all the transmitted information is common information. Here, the capacity increases monotonically with INR because increasing INR increases the common information rate.

The capacity in the fourth and fifth regimes follow from previous results. The first three regimes fall into the *weak interference* regime, and the capacity in these regimes is a consequence of the new results that we obtain. In these regimes, the interference is not strong enough to be decoded in its entirety. In fact, the capacity approximation (1.2) in regime 1 implies that if the interference is very weak, then treating interference as noise is optimal. The capacity expression for regimes 2 and 3 however imply that if the interference is not very weak, decoding it partially can significantly improve performance. Interestingly, the capacity is *not* monotonically decreasing with INR in the weak interference regime.

In point-to-point links, the notion of *degrees of freedom* is a fundamental measure of channel resources. It tells us how many signal dimensions are available for communication. In the (scalar) AWGN channel, there is one degree of freedom per second per Hz. When multiple links share the communication medium, one can think of the mutual interference as reducing the available degrees of freedom for useful communication. Our results quantify this reduction. Define

$$\alpha := \frac{\log \mathsf{INR}}{\log \mathsf{SNR}}$$

as the ratio of the interference-to-noise ratio and the signal-to-noise ratio in dB scale, and

$$d_{\text{sym}} := \frac{C_{\text{sym}}}{C_{\text{awgn}}}$$

as the *generalized degrees of freedom* per user, then (1.2) yields the following characterization:

$$d_{\text{sym}} = \begin{cases} 1-\alpha & 0 \le \alpha \le \frac{1}{2} \\ \alpha & \frac{1}{2} \le \alpha \le \frac{2}{3} \\ 1-\frac{\alpha}{2} & \frac{2}{3} \le \alpha \le 1 \\ \frac{\alpha}{2} & 1 \le \alpha \le 2 \\ 1 & \alpha \ge 2. \end{cases} \tag{1.3}$$

This is plotted in Figure 2.5 (shown later in Chapter 2), together with the performance of our baseline strategies of orthogonalizing and treating interference as noise. Note that orthogonalizing between the links, in which each link achieves half the degrees of freedom, is strictly sub-optimal except when $\alpha = \frac{1}{2}$ and $\alpha = 1$. Treating interference as noise, on the other hand, is strictly sub-optimal except for $\alpha \le \frac{1}{2}$. Note also the fundamental importance of comparing the signal-to-noise and the interference-to-noise ratios in dB scale.

The rest of Part I of the thesis is structured as follows. In Section 1.2, we describe the model. Chapter 2 focuses on the symmetric rate point in the symmetric interference channel, where the results can be described in the simplest form. Results on the entire capacity region for the general two-user channel in weak interference are derived in Chapter 3. Chapter 4 extends the results of Chapter 3 for channels where one of the links is in strong interference while the other is in weak interference. We conclude Part I with a summary and discussion in Chapter 5.

Regarding notation, we will use lowercase or uppercase letters for scalars, lowercase boldface letters for vectors, and calligraphic letters for sets. For example we write h or P for scalars, $\mathbf{x}$ for a vector, and $\mathcal{R}$ for a set. We use $H(\cdot)$ to denote binary entropy of a discrete random variable or vector, $h(\cdot)$ to denote binary differential entropy of a continuous random variable or vector, and $I(\cdot;\cdot)$ to denote mutual information. In addition, unless otherwise stated, all logarithms are to the base 2.

The results of Part I of the thesis are joint work with Hua Wang, University of Illinois at Urbana-Champaign. Preliminary versions of this research were presented in [24].

1.2 Model

In this section we describe the model to be used in the rest of this part. We consider a two-user Gaussian interference channel. In this model there are two transmitter-receiver pairs, where each transmitter wants to communicate with its corresponding receiver (cf. Figure 1.1).

This channel is represented by the equations:

$$\begin{aligned} y_1 &= h_{1,1}x_1 + h_{2,1}x_2 + z_1 \\ y_2 &= h_{1,2}x_1 + h_{2,2}x_2 + z_2 \end{aligned} \tag{1.4}$$

where for $i = 1, 2$, $x_i \in \mathbb{C}$ is subject to a power constraint P_i, i.e. $E[|x_i|^2] \leq P_i$, and the noise processes $z_i \sim \mathcal{CN}(0, N_0)$ are i.i.d. over time. For convenience we will denote the power gains of the channels by $g_{i,j} = |h_{i,j}|^2$, $i, j = 1, 2$.

It is easy to see that the capacity region of the interference channel depends only on four parameters: the signal to noise and interference to noise ratios. For $i = 1, 2$, let $\mathsf{SNR}_i = g_{i,i}P_i/N_0$ be the signal to noise ratio of user i, and $\mathsf{INR}_1 = g_{2,1}P_2/N_0$ ($\mathsf{INR}_2 = g_{1,2}P_1/N_0$) be the interference to noise ratio of user 1 (2). As will become apparent from our analysis, this parameterization in terms of SNR and INR is more natural for the interference channel, because it puts in evidence the main factors that determine the channel capacity.

For a given block length n, user i communicates a message $m_i \in \{1, \ldots, 2^{nR_i}\}$ by choosing a codeword from a codebook $\mathcal{C}_{i,n}$, with $|\mathcal{C}_{i,n}| = 2^{nR_i}$. The codewords $\{\mathbf{c}_i(m_i)\}$ of this codebook must satisfy the average power constraint:

$$\frac{1}{n}\sum_{t=1}^{n} |c_i(m_i)[t]|^2 \leq P_i$$

Receiver i observes the channel outputs $\{y_i[t] : t = 1, \ldots, n\}$ and uses a decoding function $f_{i,n} : \mathbb{C}^n \to \mathbb{N}$ to get the estimate $\hat{m}_i$ of the transmitted message m_i. The receiver is in error whenever $\hat{m}_i \neq m_i$. The average probability of error for user i is given by

$$\epsilon_{i,n} = E[P(\hat{m}_i \neq m_i)]$$

where the expectation is taken with respect to the random choice of the transmitted messages m_1 and m_2. Note that due to the interference among users, the probability of error of each user may depend on the codeword transmitted by the other user.

A rate pair (R_1, R_2) is achievable if there exists a family of codebook pairs $\{(\mathcal{C}_{1,n}, \mathcal{C}_{2,n})\}_n$ with codewords satisfying the power constraints P_1 and P_2 respectively, and decoding functions $\{(f_{1,n}(\cdot), f_{1,n}(\cdot)\}_n$, such that the average decoding error probabilities $\epsilon_{1,n}, \epsilon_{2,n}$ go to zero as the block length n goes to infinity.

The capacity region $\mathcal{R}$ of the interference channel is the closure of the set of achievable rate pairs.

Chapter 2

Symmetric Gaussian Interference Channel

2.1 Symmetric channel and symmetric rate point

In order to introduce the main ideas and results in the simplest possible setting, we start our analysis of the interference channel capacity region by considering a symmetric interference channel and the symmetric rate point[1].

In the symmetric interference channel we have $g_{1,1} = g_{2,2} = g_d$, $g_{1,2} = g_{2,1} = g_c$ and $P_1 = P_2 = P$, or equivalently, $\mathsf{SNR}_1 = \mathsf{SNR}_2$ and $\mathsf{INR}_1 = \mathsf{INR}_2$. In addition, the symmetric capacity is the solution to the following optimization problem:

$$C_{\text{sym}} = \begin{cases} \text{Maximize:} & \min\{R_1, R_2\} \\ \text{Subject to:} & (R_1, R_2) \in \mathcal{R} \end{cases}$$

where $\mathcal{R}$ is the capacity region of the interference channel.

Due to the convexity and symmetry of the capacity region of the symmetric channel, the

[1]This chapter is based on preliminary work that was presented in "Gaussian Interference Channel Capacity to Within One Bit: the Symmetric Case", by Raul Etkin, David Tse, and Hua Wang, IEEE Information Theory Workshop, October 22-26, 2006, Chengdu, China. ©2006 IEEE.

symmetric capacity is attained at the point of the capacity region that maximizes the sum rate $R_1 + R_2$. Since the capacity region is known in the strong interference case when $\mathsf{INR}/\mathsf{SNR} \geq 1$, we will focus on the case where $0 < \mathsf{INR}/\mathsf{SNR} < 1$ (i.e. $0 < g_c/g_d < 1$). In addition, we will concentrate on the situation where $g_c P \geq N_0$, i.e. the interfering signal power is at least as large as the noise power. The case $g_c P < N_0$ is not so interesting because the communication is essentially limited by noise.

2.2 A simple communication scheme

We will use a simple communication scheme that is a special case of the general type of schemes introduced by Han and Kobayashi in [29]. For a given block length n user i chooses a private message from codebook $\mathcal{C}^u_{i,n}$ and a common message from codebook $\mathcal{C}^w_{i,n}$. These codebooks satisfy the power constraints P_u and P_w with $P_u + P_w = P$. The sizes of these codebooks are such that $|\mathcal{C}^u_{i,n}| \cdot |\mathcal{C}^w_{i,n}| = 2^{nR_i}$. After selecting the corresponding codewords user i transmits the signal $\mathbf{x}_i = \mathbf{c}^u_i + \mathbf{c}^w_i$ by adding the private and common codewords. The private codewords are meant to be decoded by receiver i, while the common codewords must be decoded by both receivers.

The general Han and Kobayashi scheme allows to generate the codebooks using arbitrary input distributions, and allows to do time sharing between multiple strategies. We will consider a simple scheme where the codebooks are generated by using i.i.d. random samples of a Gaussian $\mathcal{CN}(0, \sigma^2)$ random variable with $\sigma^2 = P_u, P_w$. We choose P_u such that $g_c P_u = N_0$, i.e. the interference created by the private message has the same power as the Gaussian noise[2]. In addition, we use a fixed strategy, i.e. we don't do time sharing.

We will show that this simple scheme allows us to achieve a symmetric rate close to the symmetric rate capacity of the channel. In order to determine the symmetric rate that we can achieve with this scheme, it is useful to think of each user as being split into two virtual users: private user

[2]Note that this is possible with the available power under the assumption $g_c P \geq N_0$. If $g_c P < N_0$ one can choose $P_u = P$, but will not consider this case in this section.

U_i and common user W_i. Let MAC_1 be the multiple access channel formed by virtual users U_1, W_1, and W_2, and receiver 1, with the signal from virtual user U_2 being treated as noise. In a similar way, let MAC_2 be the multiple access channel formed by virtual users U_2, W_1, and W_2, and receiver 2, with the signal from virtual user U_1 being treated as noise. Since the common messages must be decoded by both receivers, while the private messages must be decoded only by the intending receiver, we see that the rates achievable by a the Han and Kobayashi scheme correspond to the intersection of the capacity regions of MAC_1 and MAC_2.

In general, it is possible to choose any operating point that lies in the intersection of MAC_1 and MAC_2. Different points may correspond to different decoding orders at the receivers, and also may result in different rates for the private and common messages of the users. In order to simplify the analysis we choose the private rates of both users, as well as the common rate of both users to be equal, i.e. $R_{u,1} = R_{u,2}$ and $R_{w,1} = R_{w,2}$. We also fix a decoding order at each receiver, so that the common messages are decoded first, while the private message is decoded last.

Since the private message is decoded last, while the private message of the other user is treated as noise, the private rate of each user is given by:

$$R_u = \log\left(1 + \frac{g_d P_u}{N_0 + g_c P_u}\right) = \log\left(1 + \frac{g_d P_u}{2N_0}\right) = \log\left(1 + \frac{g_d}{2g_c}\right).$$

Since each receiver decodes the common messages first, both private messages are treated as noise when decoding the common messages. With this decoding order, the sum rate of the common messages must satisfy two constraints:

$$\begin{aligned} R_{w,1} + R_{w,2} &\leq \log\left(1 + \frac{P_w(g_d + g_c)}{N_0 + P_u(g_d + g_c)}\right) \\ &= \log\left(1 + \frac{(P - N_0/g_c)(g_d + g_c)}{N_0 + N_0 g_d/g_c + N_0}\right) \\ &= \log\left(1 + \frac{(g_c P - N_0)(g_d + g_c)}{N_0(2g_c + g_d)}\right) \end{aligned} \tag{2.1}$$

and

$$R_{w,1} + R_{w,2} \leq 2\log\left(1 + \frac{g_c P_w}{N_0 + P_u(g_c + g_d)}\right)$$

$$= \quad 2\log\left(1+\frac{g_cP-N_0}{N_0+N_0/g_c(g_c+g_d)}\right)$$
$$= \quad 2\log\left(1+\frac{g_c(g_cP-N_0)}{N_0(2g_c+g_d)}\right) \tag{2.2}$$

where (2.1) arises from the sum rate constraint of the MAC formed by virtual users W_1 and W_2 at receiver 1 (or receiver 2) when the messages from virtual users U_1 and U_2 are treated as noise, and (2.2) arises from the individual rate constraint of decoding the message of virtual user W_1 at receiver 2 and virtual user W_2 at receiver 1, treating the messages from virtual users U_1 and U_2 as noise (see Figure 2.1).

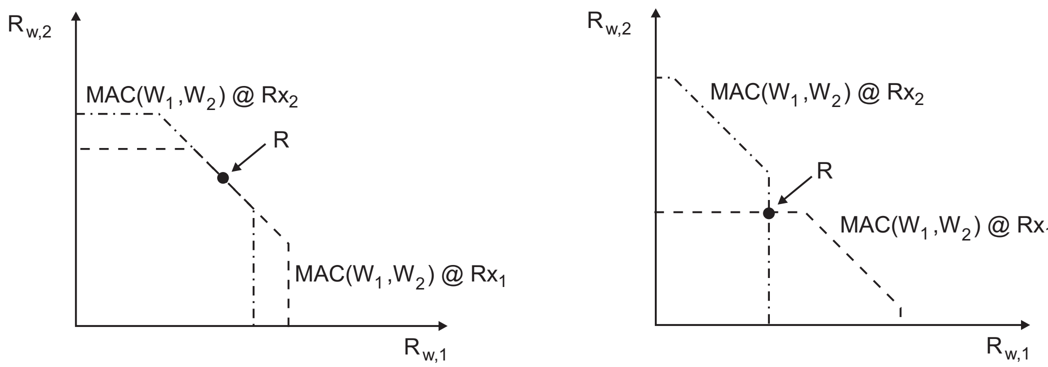


Figure 2.1: Intersection of the multiple access channel regions corresponding to virtual users W_1 and W_2 at receivers 1 and 2, when the signals from virtual users U_1 and U_2 are treated as noise. The left figure corresponds to the case in which the sum rate constraint (2.1) is active, while the right figure corresponds to the case in which the sum rate constraint(2.2) is active. In both cases, the symmetric rate point is indicated.

Therefore, with the simple Han and Kobayashi scheme we obtain a symmetric rate:

$$R_{HK} = \log\left(1+\frac{g_d}{2g_c}\right) + \min\left\{\frac{1}{2}\log\left[1+\frac{(g_cP-N_0)(g_d+g_c)}{N_0(2g_c+g_d)}\right], \log\left[1+\frac{g_c(g_cP-N_0)}{N_0(2g_c+g_d)}\right]\right\} \tag{2.3}$$

By comparing (2.1) and (2.2) we can determine the parameter ranges in which each of the terms of the $\min\{\cdot,\cdot\}$ in (2.3) is active. Define:

$$\mathcal{B}_1 \quad = \quad \left\{(g_c, g_d, P, N_0) : g_cP \geq N_0 \text{ and } g_d(g_c+g_d) < g_c^2(g_cP/N_0+1)\right\}$$
$$\mathcal{B}_2 \quad = \quad \left\{(g_c, g_d, P, N_0) : g_cP \geq N_0 \text{ and } g_d(g_c+g_d) \geq g_c^2(g_cP/N_0+1)\right\}. \tag{2.4}$$

Then, the first (second) term of the $\min\{\cdot,\cdot\}$ is active in $\mathcal{B}_1$ ($\mathcal{B}_2$). We denote by R_{HK_1} (R_{HK_2}) the symmetric rate expression that results in $\mathcal{B}_1$ ($\mathcal{B}_2$).

2.3 Known upper bounds

In order to evaluate the performance of our communication scheme, we can compare the symmetric rate achieved with an upper bound. We can obtain this upper bound by considering any outer bound to the interference channel capacity region evaluated at $R_1 = R_2$. The best known outer bound to the interference channel capacity region is that given in [35] Theorem 2. This result is derived for the normalized interference channel, i.e. $g_d = 1$, $N_0 = 1$, so in order to use the results in [35], we need to replace: $g_c \to g_c/g_d$, $P \to g_d P/N_0$. Specializing this result to the complex symmetric interference channel with $0 < g_c/g_d < 1$ one obtains the symmetric rate upper bound:

$$R_{UB_1} = \log\left[1 + \frac{-(g_c+g_d) + \sqrt{(g_c+g_d)^2 + 4g_c g_d (P/N_0)(g_c+g_d)}}{2g_c}\right] \tag{2.5}$$

Consider the case in which the first term of the $\min\{\cdot,\cdot\}$ of (2.3) is active. Then we can write:

$$\begin{aligned}
R_{UB_1} - R_{HK_1} &= \log\left[\frac{2g_c - (g_c+g_d) + \sqrt{(g_c+g_d)^2 + 4g_c g_d (P/N_0)(g_c+g_d)}}{2g_c}\right] \\
&\quad - \log\left(\frac{2g_c+g_d}{2g_c}\right) - \frac{1}{2}\log\left[\frac{2g_c + g_d + (g_c(P/N_0)-1)(g_d+g_c)}{2g_c+g_d}\right] \\
&= \frac{1}{2}\log\left\{\frac{\left[(g_c-g_d) + \sqrt{(g_c+g_d)^2 + 4g_c g_d (P/N_0)(g_c+g_d)}\right]^2}{(2g_c)^2}\right\} \\
&\quad -\frac{1}{2}\log\left\{\frac{(2g_c+g_d)[(2g_c+g_d) + (g_c(P/N_0)-1)(2g_c+g_d)]}{(2g_c)^2}\right\} \\
&\le \frac{1}{2}\log\left\{\frac{2(g_c^2+g_d^2) + 4g_c g_d P(g_c+g_d)/N_0 - 2(g_d-g_c)(g_c+g_d)}{(2g_c+g_d)g_c(1+(g_c+g+d)(P/N_0))}\right\} \\
&= \frac{1}{2}\log\left[\frac{4g_c + 4g_d(g_c+g_d)P/N_0}{(2g_c+g_d)(1+(g_c+g_d)P/N_0)}\right] \\
&= \frac{1}{2}\log\left[\left(\frac{4}{2+g_d/g_c}\right)\left(\frac{1+g_d(1+g_d/g_c)P/N_0}{1+(g_c+g_d)P/N_0}\right)\right] \\
&\le \frac{1}{2}\log\left[\left(\frac{4}{2+g_d/g_c}\right)\left(\frac{g_d}{g_c}\right)\right] \\
&= \frac{1}{2}\log\left[\frac{4}{2g_c/g_d+1}\right] \\
&< 1
\end{aligned} \tag{2.6}$$

where we used $\sqrt{1+x} \ge 1$ for $x \ge 0$ in the first inequality and the assumption $g_c \le g_d$ in the second and third inequalities. It is easy to check that for example, if $g_d = 1$, $g_c = 1/P^{1/4}$, and $N_0 = 1$, $R_{UB_1} - R_{HK_1} \to 1$ as $P \to \infty$. Therefore the worst case difference of 1 bit/s/Hz in (2.6) can actually occur.

We see that in the parameter range $\mathcal{B}_1$, the achievable strategy and the upper bound differ for at most 1 bit/s/Hz. Therefore, in this parameter range our simple scheme gives a bounded (and small) gap with respect to the upper bound.

We will now show that in the parameter range $\mathcal{B}_2$, the gap between our scheme and the upper bound can be arbitrarily large. In this parameter range we can write:

$$\begin{aligned} R_{UB_1} - R_{HK_2} &= \log\left[\frac{g_c - g_d + \sqrt{(g_c+g_d)^2 + 4g_c g_d (g_c+g_d)P/N_0}}{2g_c}\right] \\ &\quad - \log\left(\frac{2g_c+g_d}{2g_c}\right) - \log\left[\frac{2g_c+g_d+g_c(g_cP/N_0 - 1)}{2g_c+g_d}\right] \\ &\geq \log\left[\frac{g_c - g_d + 2\sqrt{g_c g_d (g_c+g_d)P/N_0}}{g_c+g_d+g_c^2 P/N_0}\right] \end{aligned} \tag{2.7}$$

where the inequality follows from discarding $(g_c+g_d)^2$ in the square root. To show that this difference can be unbounded, take $g_c = \sqrt{N_0/P}$ and $g_d = 1$. With this choice of parameters we get:

$$R_{UB_1} - R_{HK_2} \geq \log\left[\frac{\sqrt{N_0/P} - 1}{2(1+\sqrt{N_0/P})} + \frac{(P/N_0)^{1/4}\sqrt{1+\sqrt{N_0/P}}}{1+\sqrt{N_0/P}}\right] \tag{2.8}$$

where the right hand side goes to infinity for $P/N_0 \to \infty$.

We see that the gap between our scheme and the upper bound can be arbitrarily large in this case. This large gap could be due to a very suboptimal scheme, a loose upper bound, or both. It turns out that the large gap is due to the looseness of the upper bound. In order to prove this we need to derive a new upper bound for the sum rate of the interference channel.

As a first step toward deriving the new bound, in order to introduce some ideas that we will use later, and to better understand why (2.5) can be very loose, we will obtain a bound somewhat looser than (2.5) but that also gives a worst case 1-bit/s/Hz gap for the parameter range $\mathcal{B}_1$.

We will consider a general interference channel so that the upper bounds that we derive are not restricted to the symmetric interference channel. Consider a modified interference channel, where a genie provides the side information $v_2 = h_{2,1}x_2$ to receiver 1 (see Figure 2.2).

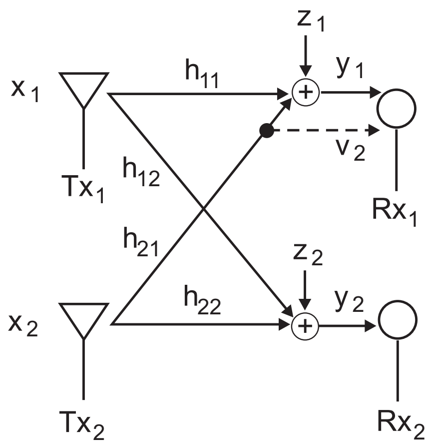


Figure 2.2: Genie-aided two-user interference channel. A genie provides signal $v_2 = h_{2,1}x_2$ to receiver 1.

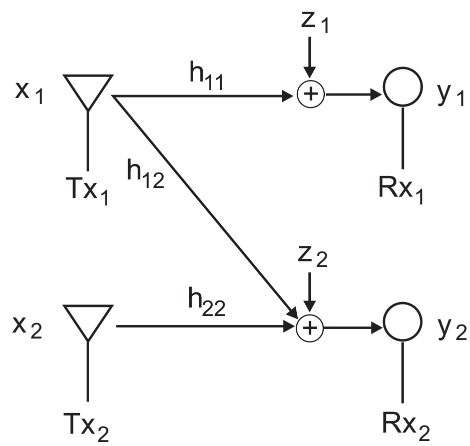


Figure 2.3: One-sided interference channel.

Since x_1 is independent of v_2 we can write for any block of length n:

$$I(\mathbf{x}_1^n; \mathbf{y}_1^n, \mathbf{v}_2^n) = I(\mathbf{x}_1^n; \mathbf{v}_2^n) + I(\mathbf{x}_1^n; \mathbf{y}_1^n | \mathbf{v}_2^n) = I(\mathbf{x}_1^n; h_{1,1}\mathbf{x}_1^n + \mathbf{z}_1^n)$$

and it follows that receiver 1 can get an interference-free signal by subtracting the interference v_2 provided by the genie. Therefore we obtain that the genie-aided channel is equivalent to the one-sided interference channel depicted in Figure 2.3.

The sum rate capacity of a one-sided interference channel for the case of $g_{1,2}/g_{1,1} < 1$ is given by [42][3]:

$$R_{sum}(\text{one-sided IC}) = \log\left(1 + \frac{g_{1,1}P_1}{N_0}\right) + \log\left(1 + \frac{g_{2,2}P_2}{N_0 + g_{1,2}P_1}\right) \tag{2.9}$$

and since the aid of the genie can only increase the capacity region of the interference channel, we obtain the upper bound for the symmetric rate:

$$R_{UB_2} = \frac{1}{2}\log\left(1 + \frac{g_d P}{N_0}\right) + \frac{1}{2}\log\left(1 + \frac{g_d P}{N_0 + g_c P}\right). \tag{2.10}$$

[3]We provide a derivation of the sum-rate capacity of the one sided interference channel in the proof of Theorem 3 in Section 3.1.

We will now check that this bound still achieves a 1-bit/s/Hz gap with respect to R_{HK_1}.

$$\begin{aligned}
R_{UB_2} - R_{HK_1} &= \frac{1}{2}\log\left(1+\frac{g_d P}{N_0}\right)+\frac{1}{2}\log\left(1+\frac{g_d P}{N_0+g_c P}\right)\\
&\quad -\log\left(\frac{2g_c+g_d}{2g_c}\right)-\frac{1}{2}\log\left[\frac{2g_c+g_d+(g_c(P/N_0)-1)(g_d+g_c)}{2g_c+g_d}\right]\\
&< \frac{1}{2}\log\left(1+\frac{g_d P}{N_0}\right)+\frac{1}{2}\log\left(1+\frac{g_d}{g_c}\right)\\
&\quad -\log\left(\frac{2g_c+g_d}{2g_c}\right)-\frac{1}{2}\log\left[\frac{2g_c+g_d+(g_c(P/N_0)-1)(g_d+g_c)}{2g_c+g_d}\right]\\
&= \frac{1}{2}\log\left(\frac{2g_c+2g_d}{2g_c+g_d}\right)+\frac{1}{2}\log\left[\frac{2(1+g_d P/N_0)}{1+(g_d+g_c)P/N_0}\right]\\
&\leq \frac{1}{2}\log\left(\frac{2g_c+2g_d}{2g_c+g_d}\right)+\frac{1}{2}\log\left[\frac{2(N_0/P+g_d)}{g_d+g_c}\right]\\
&\leq \frac{1}{2}\log\left(\frac{2g_c+2g_d}{2g_c+g_d}\right)+\frac{1}{2}\log\left[\frac{2(g_c+g_d)}{g_d+g_c}\right]\\
&\leq \frac{1}{2}\log(2)+\frac{1}{2}\log(2)\\
&= 1
\end{aligned} \tag{2.11}$$

where we used the assumptions $g_c P > N_0$ and $g_c/g_d \leq 1$.

Even though the bound (2.9) is not as good as the bounds presented in [35], all these bounds have the same worst case 1-bit/s/Hz gap with respect to our simple communication strategy in the parameter range $\mathcal{B}_1$. Also, in the parameter range $\mathcal{B}_2$ all these bounds are arbitrarily loose.

Why are all these bounds loose in $\mathcal{B}_2$? The problem is that they rely, in one way or another, in giving side information to receiver 1 so that he can eventually cancel the interfering signal from user 2. We can gain some intuition about why these bounds are loose in $\mathcal{B}_2$ by considering our simple communication scheme in the genie-aided channel of Figure 2.2. The side information provided by the genie allows receiver 1 to subtract the interference generated by transmitter 2. The rates of the virtual private users U_1 and U_2 are in this case:

$$\begin{aligned}
R_{u_1} &= \log\left(1+\frac{g_d P_u}{N_0}\right)=\log\left(1+\frac{g_d}{g_c}\right)\\
R_{u_2} &= \log\left(1+\frac{g_d P_u}{N_0+g_c P_u}\right)=\log\left(1+\frac{g_d}{2g_c}\right)
\end{aligned}$$

and we see that virtual user U_1 gains at most 1 bit/s/Hz due to the help of the genie. The sum rate of the MAC formed by virtual users W_1 and W_2 at receiver 2 does not change due to the aid of the genie. Therefore, the sum rate constraint (2.1) still holds. However, due to the aid of the genie receiver 1 can decode the message of virtual user W_2 and the sum rate constraint (2.2) does not appear in this case.

In $\mathcal{B}_1$ the sum rate constraint (2.1) is active in the original channel, and the aid of the genie does not allow to increase the sum rate by a large amount. In this regime, the bound obtained from the genie-aided channel is good. In contrast, in $\mathcal{B}_2$ the sum rate constraint (2.2) is active in the original channel, and the genie effectively releases this constraint by providing enough information to receiver 1 to decode the message of virtual user W_2. Since in $\mathcal{B}_2$ the constraint (2.1) is larger than (2.2) (and the gap between the two constraints can be made arbitrarily large), the bound obtained from the genie-aided channel is loose.

2.4 A new upper bound

In order to derive a tighter sum rate bound for the parameter range $\mathcal{B}_2$ we will make use of the help of genies, but will avoid giving too much information to either receiver. The information that we will provide will not allow either receiver to completely decode the message of the interfering transmitter. The new sum rate upper bound is given in the following theorem, which we state for a general (not necessarily symmetric) Gaussian interference channel.

Theorem 1 *For a Gaussian interference channel as defined in Section 1.2, equation (1.4), the sum rate is upper bounded by*

$$R_1 + R_2 \leq \log\left(1 + \frac{g_{2,1}P_2}{N_0} + \frac{g_{1,1}P_1}{N_0 + g_{1,2}P_1}\right) + \log\left(1 + \frac{g_{1,2}P_1}{N_0} + \frac{g_{2,2}P_2}{N_0 + g_{2,1}P_2}\right). \tag{2.12}$$

Proof: Define

$$\begin{aligned} s_1 &= h_{1,2}x_1 + z_2 \\ s_2 &= h_{2,1}x_2 + z_1 \end{aligned}$$

and consider the genie-aided channel where a genie provides s_1 to receiver 1 and s_2 to receiver 2 (see Figure 2.4). Clearly, the capacity region of this genie-aided channel is an outer bound to the

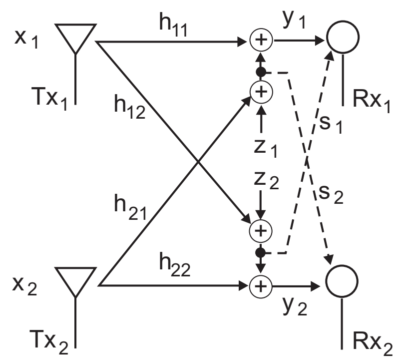


Figure 2.4: Genie-aided two-user Gaussian interference channel. A genie provides signals s_1 to receiver 1 and s_2 to receiver 2.

capacity region of the original interference channel. Therefore, we can obtain an upper bound for the sum rate of the original channel by computing an upper bound on the sum rate of the genie aided channel. For a block of length n we can bound the sum rate of the genie-aided channel in the following way:

$$\begin{aligned}
n(R_1+R_2) &\leq I(\mathbf{x}_1^n;\mathbf{y}_1^n,\mathbf{s}_1^n)+I(\mathbf{x}_2^n;\mathbf{y}_2^n,\mathbf{s}_2^n)+n\epsilon_n \\
&= I(\mathbf{x}_1^n;\mathbf{s}_1^n)+I(\mathbf{x}_1^n;\mathbf{y}_1^n|\mathbf{s}_1^n)+I(\mathbf{x}_2^n;\mathbf{s}_2^n)+I(\mathbf{x}_2^n;\mathbf{y}_2^n|\mathbf{s}_2^n)+n\epsilon_n \\
&= h(\mathbf{s}_1^n)-h(\mathbf{s}_1^n|\mathbf{x}_1^n)+h(\mathbf{y}_1^n|\mathbf{s}_1^n)-h(\mathbf{y}_1^n|\mathbf{x}_1^n,\mathbf{s}_1^n) \\
&\quad +h(\mathbf{s}_2^n)-h(\mathbf{s}_2^n|\mathbf{x}_2^n)+h(\mathbf{y}_2^n|\mathbf{s}_2^n)-h(\mathbf{y}_2^n|\mathbf{x}_2^n,\mathbf{s}_2^n)+n\epsilon_n \\
&= h(\mathbf{s}_1^n)-h(\mathbf{z}_2^n)+h(\mathbf{y}_1^n|\mathbf{s}_1^n)-h(\mathbf{s}_2^n) \\
&\quad +h(\mathbf{s}_2^n)-h(\mathbf{z}_1^n)+h(\mathbf{y}_2^n|\mathbf{s}_2^n)-h(\mathbf{s}_1^n)+n\epsilon_n \\
&= h(\mathbf{y}_1^n|\mathbf{s}_1^n)+h(\mathbf{y}_2^n|\mathbf{s}_2^n)-h(\mathbf{z}_1^n)-h(\mathbf{z}_2^n)+n\epsilon_n \\
&\leq \sum_{i=1}^{n}[h(y_{1i}|s_{1i})+h(y_{2i}|s_{2i})-h(z_{1i})-h(z_{2i})]+n\epsilon_n \qquad (2.13)
\end{aligned}$$

where the last inequality follows by the fact that removing conditioning cannot reduce differential entropy, and $\epsilon_n \to 0$ as $n \to \infty$.

Let $\mathbb{E}[x_{1i}^2] = P_{1i}$ and $\mathbb{E}[x_{2i}^2] = P_{2i}$, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} h(y_{1i}|s_{1i}) &\overset{(a)}{\leq} \frac{1}{n}\log\left[\pi e\left(N_0 + g_{2,1}P_{2i} + \frac{g_{1,1}P_{1i}N_0}{N_0 + g_{1,2}P_{1i}}\right)\right] \\
&\overset{(b)}{\leq} \log\left[\pi e\left(N_0 + g_{2,1}\left(\frac{1}{n}\sum_{i=1}^{n} P_{2i}\right) + \frac{g_{1,1}\left(\frac{1}{n}\sum_{i=1}^{n} P_{1i}\right)N_0}{N_0 + g_{1,2}\left(\frac{1}{n}\sum_{i=1}^{n} P_{1i}\right)}\right)\right] \\
&\overset{(c)}{\leq} \log\left[\pi e\left(N_0 + g_{2,1}P_2 + \frac{g_{1,1}P_1N_0}{N_0 + g_{1,2}P_1}\right)\right] \qquad (2.14)
\end{aligned}$$

where in step (a) we use the fact that the circularly symmetric complex Gaussian distribution maximizes conditional differential entropy for a given covariance constraint, in step (b) we use Jensen's inequality applied to a concave function, and in step (c) we used the fact that the function is increasing on P_1 and P_2. Similarly, we have

$$\frac{1}{n}\sum_{i=1}^{n} h(y_{2i}|s_{2i}) \leq \log\left[\pi e\left(N_0 + g_{1,2}P_1 + \frac{g_{2,2}P_2N_0}{N_0 + g_{2,1}P_2}\right)\right]. \qquad (2.15)$$

Thus we have

$$\begin{aligned}
R_1 + R_2 &\leq \frac{1}{n}\sum_{i=1}^{n}\left[h(y_{1i}|s_{1i}) + h(y_{2i}|s_{2i}) - h(z_{1i}) - h(z_{2i})\right] + \epsilon_n \\
&\leq \log\left[\pi e\left(N_0 + g_{2,1}P_2 + \frac{g_{1,1}P_1N_0}{N_0 + g_{1,2}P_1}\right)\right] - \log\left(\pi e N_0\right) \\
&\quad + \log\left[\pi e\left(N_0 + g_{1,2}P_1 + \frac{g_{2,2}P_2N_0}{N_0 + g_{2,1}P_2}\right)\right] - \log\left(\pi e N_0\right) + \epsilon_n \\
&= \log\left(1 + \frac{g_{2,1}P_2}{N_0} + \frac{g_{1,1}P_1}{N_0 + g_{1,2}P_1}\right) + \log\left(1 + \frac{g_{1,2}P_1}{N_0} + \frac{g_{2,2}P_2}{N_0 + g_{2,1}P_2}\right) + \epsilon_n. \qquad (2.16)
\end{aligned}$$

Letting $n \to \infty$, $\epsilon_n \to 0$ and we get the desired upper bound. ■

By analysing the proof of Theorem 1 one observes that the upper bound can be achieved by using codebooks generated from i.i.d. samples of circularly symmetric complex Gaussian random variables of variance P_1 and P_2, and by treating interference as noise at the decoders. Since treating interference as noise is optimal in the genie aided channel, the role of the side information provided by the genie is to compensate for the harm produced by interference. Making the signals more random results in added interference, but the side information compensates the negative effect of interference on the other link by enhancing the performance of the own link.

We specialize the bound of Theorem 1 to the symmetric interference channel to obtain the following upper bound on the symmetric rate:

$$R_{UB_{new}} = \log\left(1 + \frac{g_c P}{N_0} + \frac{g_d P}{N_0 + g_c P}\right). \tag{2.17}$$

This bound has a finite gap with respect to the achievable symmetric rate with our simple scheme in the parameter range $\mathcal{B}_2$. To verify this we compute:

$$\begin{aligned} R_{UB_{new}} - R_{HK_2} &= \log\left(1 + \frac{g_c P}{N_0} + \frac{g_d P}{N_0 + g_c P}\right) - \log\left(\frac{2g_c + g_d}{2g_c}\right) - \log\left[\frac{2g_c + g_d + g_c(g_c P/N_0 - 1)}{2g_c + g_d}\right] \\ &< \log\left(\frac{g_c + g_c^2 P/N_0 + g_d}{g_c}\right) - \log\left[\frac{g_c + g_d + g_c^2 P/N_0}{2g_c}\right] \\ &= \log(2) \\ &= 1 \end{aligned} \tag{2.18}$$

and we find that the gap in the symmetric rate with respect to the new upper bound is at most 1 bit/s/Hz in $\mathcal{B}_2$.

Using (2.11) and (2.18) we see that when $g_c P \geq N_0$ our simple scheme is at most 1 bit/s/Hz away from the symmetric rate channel capacity. Proving that the simple scheme is at most 1 bit/s/Hz away from capacity when $g_c P \leq N_0$ is straightforward. We can set $P_u = P$ and use as a symmetric rate upper bound the single user capacity. The difference between the achievable rate and the upper bound is:

$$\log\left(1 + \frac{g_d P}{N_0}\right) - \log\left(1 + \frac{g_d P}{N_0 + g_c P}\right) \leq \log\left(1 + \frac{g_d P}{N_0}\right) - \log\left(1 + \frac{g_d P}{2N_0}\right) \leq 1. \tag{2.19}$$

2.5 Generalized degrees of freedom

At high SNR, it is well known that the capacity of a point-to-point AWGN link, in bits/s/Hz, is approximately:

$$C_{\text{awgn}}(\mathsf{SNR}) \approx \log \mathsf{SNR} \tag{2.20}$$

and if n_t transmit and n_r receive antennas are available at the transmitter and receiver respectively, if the channel gain matrix is full rank, the capacity of the point-to-point multiple input-multiple output (MIMO) link, in bits/s/Hz, is approximately:

$$C_{\mathrm{MIMO}}(\mathsf{SNR}) \approx \min\{n_r, n_t\} \cdot \log \mathsf{SNR}. \tag{2.21}$$

In this equation, $\min\{n_r, n_t\}$ is the dimension of the signal space over which communication takes place, and is called the *degrees of freedom* of the channel. The degrees of freedom characterize to the first order the capacity of the channel, and therefore give a fundamental measure of the resources available for communication. Note that in the single antenna case, the channel has 1 degree of freedom.

When multiple links interfere with each other we can think of interference as effectively reducing the degrees of freedom of the channel. In order to make this notion precise we define the *generalized degrees of freedom* of the interference channel as:

$$d_{\mathrm{sym}}(\alpha) := \lim_{\mathsf{SNR},\mathsf{INR}\to\infty;\ \frac{\log \mathsf{INR}}{\log \mathsf{SNR}}=\alpha} \frac{C_{\mathrm{sym}}(\mathsf{INR},\mathsf{SNR})}{C_{\mathrm{awgn}}(\mathsf{SNR})}. \tag{2.22}$$

Since interference cannot help in communicating each user's message, it follows that $C_{\mathrm{sym}} \leq C_{\mathrm{awgn}}$ and therefore $d_{\mathrm{sym}} \leq 1$. Also note that the generalized degrees of freedom depend on the way in which INR and SNR go to infinity through the parameter α:

$$\alpha = \frac{\log \mathsf{INR}}{\log \mathsf{SNR}}.$$

Note that since $\mathsf{SNR}, \mathsf{INR} \to \infty$ in the definition of d_{sym}, we can assume without loss of generality that $\mathsf{SNR}, \mathsf{INR} > 1$, and hence $\alpha > 0$.

In the very strong interference case, each user can decode the interfering message before decoding his own message [7]. After decoding the interference and subtracting it from the received signal, the user effectively gets an AWGN channel for communicating his own message. It follows that the symmetric capacity in the very strong interference case is:

$$C_{\mathrm{sym}} = \log\left(1 + \frac{g_d P}{N_0}\right) = \log(1 + \mathsf{SNR}) \approx \log(\mathsf{SNR}). \tag{2.23}$$

The channel is in the very strong interference situation whenever:

$$\frac{g_c P}{N_0 + g_d P} \geq \frac{g_d P}{N_0} \Leftrightarrow \mathsf{INR} \geq \mathsf{SNR}^2 + \mathsf{SNR}. \tag{2.24}$$

Taking logs and assuming $\mathsf{SNR}, \mathsf{INR} \gg 1$ the very strong interference condition becomes $\log \mathsf{INR} \geq 2 \log \mathsf{SNR}$. In this regime we obtain $d_{\text{sym}} = 1$, and therefore, interference does not reduce the available degrees of freedom of the channel.

In the strong interference regime, each receiver is able to decode both messages. The capacity region of the interference channel is given by the intersection of the capacity regions of the two multiple access channels (MAC) formed by the two transmitters and each of the receivers. In the symmetric case, the sum capacity of both MACs is the same and the corresponding symmetric capacity is given by:

$$C_{\text{sym}} = \frac{1}{2} \log \left(1 + \frac{(g_d + g_c)P}{N_0} \right) = \frac{1}{2} \log \left(1 + \mathsf{SNR} + \mathsf{INR} \right). \tag{2.25}$$

The symmetric channel is in the strong interference situation whenever it is not in very strong interference and

$$\frac{g_c P}{N_0} \geq \frac{P}{N_0} \Leftrightarrow \mathsf{INR} \geq \mathsf{SNR} \tag{2.26}$$

which after taking logs becomes $\log \mathsf{SNR} \leq \log \mathsf{INR} < 2 \log \mathsf{SNR}$. This condition and (2.25) together with the assumption $\mathsf{SNR}, \mathsf{INR} \gg 1$ imply that $C_{\text{sym}} \approx \frac{1}{2} \log \mathsf{INR}$. It follows that under strong interference the generalized degrees of freedom are

$$d_{\text{sym}} = \frac{\log \sqrt{\mathsf{INR}}}{\log \mathsf{SNR}} = \frac{\alpha}{2}. \tag{2.27}$$

The previous results follow from the known capacity region of the interference channel in strong/very strong interference. If $\mathsf{INR} < \mathsf{SNR}$ the channel is in the weak interference regime. We can use the results of the previous subsection to fully characterize d_{sym} in this regime.

For $\mathsf{SNR}, \mathsf{INR} \gg 1$ the condition that defines $\mathcal{B}_1$ in (2.4) is $\mathsf{INR}^3 > \mathsf{SNR}^2$ which together with the weak interference condition implies $\frac{2}{3} \log \mathsf{SNR} < \log \mathsf{INR} < \log \mathsf{SNR}$. In this regime, which we call *moderate interference*, C_{sym} is characterized by the lower bound (2.3) (using the first term in the $\min\{\cdot, \cdot\}$) and the upper bound (2.10) which differ by at most by 1 bit/s/Hz. Using either of these bounds we obtain for $\mathsf{SNR}, \mathsf{INR} \gg 1$:

$$C_{\text{sym}} \approx \log \mathsf{SNR} - \frac{1}{2} \log \mathsf{INR} \tag{2.28}$$

and the generalized degrees of freedom under moderate interference are:

$$d_{\mathrm{sym}} = 1 - \frac{\log \sqrt{\mathsf{INR}}}{\log \mathsf{SNR}} = 1 - \frac{\alpha}{2}. \tag{2.29}$$

The parameter range $\mathcal{B}_2$ corresponds to $\log \mathsf{INR} \leq \frac{2}{3} \log \mathsf{SNR}$. In this regime the symmetric capacity is characterized by the lower bound (2.3) (using the second term in the $\min\{\cdot,\cdot\}$) and the upper bound (2.17) which also differ by at most by 1 bit/s/Hz. Using either of these bounds we obtain a symmetric capacity:

$$C_{\mathrm{sym}} \approx \log\left(\mathsf{INR} + \frac{\mathsf{SNR}}{\mathsf{INR}}\right). \tag{2.30}$$

Depending on how INR compares to $\frac{\mathsf{SNR}}{\mathsf{INR}}$ we can identify two subregimes. If $\mathsf{INR} \geq \frac{\mathsf{SNR}}{\mathsf{INR}}$, which implies $\frac{1}{2} \log \mathsf{SNR} \leq \log \mathsf{INR} \leq \frac{2}{3} \log \mathsf{SNR}$ we have what we call *weaker interference* and the symmetric capacity is given by:

$$C_{\mathrm{sym}} \approx \log \mathsf{INR} \tag{2.31}$$

and the generalized degrees of freedom are:

$$d_{\mathrm{sym}} = \frac{\log \mathsf{INR}}{\log \mathsf{SNR}} = \alpha. \tag{2.32}$$

Finally, if $\log \mathsf{INR} < \frac{1}{2} \log \mathsf{SNR}$ we have the *very weak interference* regime, where the symmetric capacity is:

$$C_{\mathrm{sym}} \approx \log \mathsf{SNR} - \log \mathsf{INR} \tag{2.33}$$

and the generalized degrees of freedom are:

$$d_{\mathrm{sym}} = 1 - \frac{\log \mathsf{INR}}{\log \mathsf{SNR}} = 1 - \alpha. \tag{2.34}$$

Note that in this regime we can also use the strategy that consists of both users transmitting their private messages with power P, treating the other user's message as noise. With this strategy we obtain a symmetric rate:

$$R = \log\left(1 + \frac{g_d P}{N_0 + g_c P}\right) \approx \log \mathsf{SNR} - \log \mathsf{INR} \tag{2.35}$$

which coincides with (2.33).

These results can be summarized as:

$$C_{\text{sym}} \approx \begin{cases} \log(\frac{\mathsf{SNR}}{\mathsf{INR}}) & \log \mathsf{INR} < \frac{1}{2}\log \mathsf{SNR} \\ \log \mathsf{INR} & \frac{1}{2}\log \mathsf{SNR} < \log \mathsf{INR} < \frac{2}{3}\log \mathsf{SNR} \\ \log \frac{\mathsf{SNR}}{\sqrt{\mathsf{INR}}} & \frac{2}{3}\log \mathsf{SNR} < \log \mathsf{INR} < \log \mathsf{SNR} \\ \log \sqrt{\mathsf{INR}} & \log \mathsf{SNR} < \log \mathsf{INR} < 2\log \mathsf{SNR} \\ \log \mathsf{SNR} & \log \mathsf{INR} > 2\log \mathsf{SNR} \end{cases} \tag{2.36}$$

and

$$d_{\text{sym}} = \begin{cases} 1-\alpha & 0 \le \alpha < \frac{1}{2} \\ \alpha & \frac{1}{2} \le \alpha < \frac{2}{3} \\ 1-\frac{\alpha}{2} & \frac{2}{3} < \alpha \le 1 \\ \frac{\alpha}{2} & 1 \le \alpha < 2 \\ 1 & \alpha \ge 2. \end{cases} \tag{2.37}$$

The generalized degrees of freedom are plotted in Figure 2.5, together with the performance of the baseline strategies of orthogonalizing the users (in frequency or time) and treating interference as noise. Note that orthogonalizing between the links, in which each link achieves half the degrees of freedom, is strictly sub-optimal except when $\alpha = \frac{1}{2}$ and $\alpha = 1$. Treating interference as noise, on the other hand, is strictly sub-optimal except for $\alpha \le \frac{1}{2}$. Note also the fundamental importance of comparing the signal-to-noise and the interference-to-noise ratios in dB scale.

Note that there are five regimes in which the qualitative behaviors of the capacity are different. The first three regimes fall into the weak interference regime, and the characterization of the symmetric capacity in these regimes is a consequence of the new results that we obtained. In these regimes, the interference is not strong enough to be decoded in its entirety. In fact, in regime 1 where the interference is very weak, treating interference as noise is optimal. In regimes 2 and 3 where the interference is not very weak, decoding it partially can significantly improve performance.

Interestingly, the capacity is *not* monotonically decreasing with INR in the weak interference regime. Increasing INR has two opposing effects: more common information can be decoded and

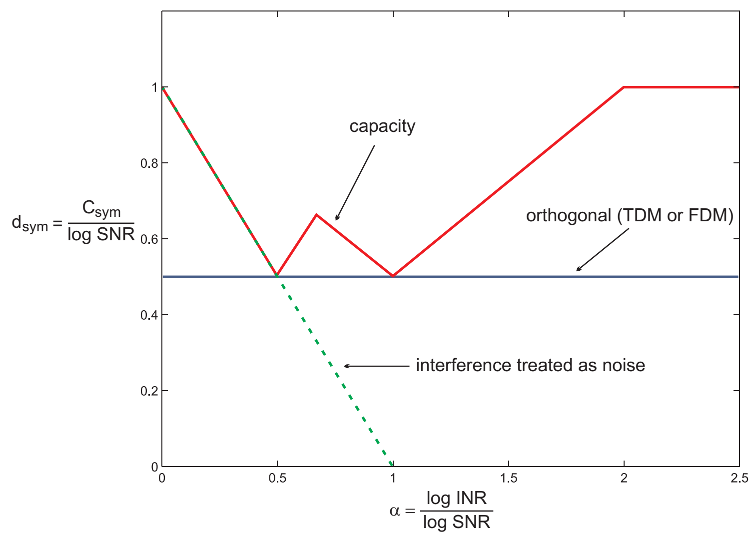


Figure 2.5: Generalized degrees of freedom for two suboptimal schemes vs. capacity. These suboptimal schemes are treating interference as noise and orthogonalizing the users over time or frequency.

cancelled but less private information can be sent under the constraint $\mathsf{INR}_p = 1$. Depending on which of these two effects dominates, the capacity increases or decreases with INR.

In the very weak interference regime where treating interference as noise is optimal, the common messages carry negligible information. In this regime, the loss in the private rate due to the increase in INR makes the capacity decrease with INR. However, once interference becomes strong enough to reach the weaker interference regime, the users can start using common information to partially cancel interference. As the interference level increases, more and more of this common information can be decoded and partially cancelled, and this effects dominates the behavior of capacity with INR. Therefore, capacity increases with INR in the weaker interference regime. However, as INR increases further to reach the moderate interference regime, the gains obtained by partially cancelling interference through the common messages are not enough to offset the loss of rate in the private information. In this regime capacity decreases with INR until the strong interference regime is reached. Since in the strong interference regime all the information is common information, increasing INR increases capacity. Finally, in the very strong interference regime all the interference can be cancelled before decoding the useful information, and interference does not have any effect on capacity.

2.6 One bit/s/Hz gap revisited

In Sections 2.3 and 2.4 we showed that the Han-Kobayashi scheme that sets the private message power so that the interference created is at noise level achieves a symmetric rate within one bit/s/Hz of the upper bounds. Therefore, we obtained a characterization of the symmetric capacity to within one bit/s/Hz. The finite and small gap between the lower and upper bounds on the symmetric capacity was obtained by direct calculation of the difference between the bounds. In this section we present an alternative analysis of the gap between the bounds that will provide a more intuitive explanation for the tightness of the bounds. We will assume $\mathsf{SNR}, \mathsf{INR} \gg 1$ to simplify some calculations.

We can split the gap analysis into two parts:

1. Fix Han-Kobayashi strategy (i.e. $Pu = N_0/g_c$, decode first the common messages (w_1, w_2) and then the private message u_1 or u_2[4]) and see how the symmetric rate changes when varying the channel from the given interference channel to the genie aided interference channel used in the bounds. Show that the rate does not change much by giving the side information for the fixed strategy.

2. Fix the channel to the genie aided interference channel, and change the Han-Kobayashi strategy by varying P_u from $P_u = N_0/g_c$ to $P_u = P$. Show that the rate does not change much by varying P_u.

Referring to Figure 2.6, the first part of the analysis corresponds to comparing the rates at points A and B, and showing that the gap Δ_1 is small. The second part corresponds to comparing the rates at points B and C, verifying that the gap Δ_2 is small.

[4]With some abuse of notation, we use u_i, w_i, $i = 1, 2$ to denote the private and common messages, and also to denote the symbols of the codewords actually sent over the channel.

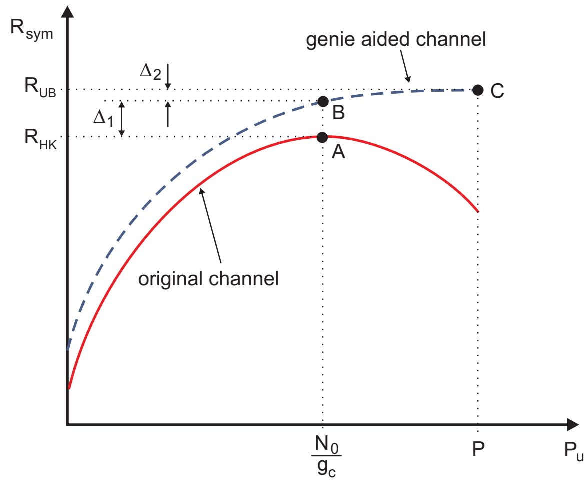


Figure 2.6: Gap between the achievable rate with the Han-Kobayashi scheme that sets $P_u = N_0/g_c$ and the symmetric capacity upper bound decomposed into two components: Δ_1 and Δ_2. Δ_1 results from fixing P_u and changing the channel; Δ_2 results from increasing P_u to P in the genie aided channel.

Since $P_u = P$ achieves the capacity of the genie aided channel (one can show that the sum rate upper bounds can be achieved by generating the codewords $\mathbf{x}_1^n$ and $\mathbf{x}_2^n$ with i.i.d. circularly symmetric complex Gaussian components of variance P, and treating interference as noise at the decoder) the above two-part analysis would show that by going from the initial Han-Kobayashi strategy in the original channel, to the capacity achieving strategy in the genie aided channel, the symmetric rate does not change much.

Since the argument depends on the regime of interest, we split the analysis into three regimes for the weak interference channel (see equation (2.37)):

(a) $0 < \alpha < 1/2$: very weak interference,

(b) $1/2 < \alpha < 2/3$: weaker interference,

(c) $2/3 < \alpha < 1$: moderate interference.

Step 1, regimes (a) and (b).

In these regimes, the outer bound is derived using the genie aided channel of Figure 2.4. In this genie aided channel receiver 1 has side information $s_1 = h_c x_1 + z_2$ and receiver 2 has side information $s_2 = g_c x_2 + z_1$. In these regimes the common message w_1 can be easily decoded at receiver 1 (and w_2 at receiver 2) because the binding constraint on the sum rate includes only the rates of the cross-common messages and the private messages. The fixed Han-Kobayashi strategy requires that w_2 be decoded at receiver 1 treating u_1 (and u_2) as noise.

The side information s_1 does not help much in decoding w_2 due to the following argument. After decoding w_1 and subtracting its effect from the received signals, the remaining side information is $\tilde{s}_1 = h_c u_1 + z_2$. The side information contains only information about u_1 which plays the role of interference at this stage. This side information would be useful in reducing the uncertainty about this interference, but since $g_c P_u = N_0$ this is a very noisy view of u_1. Therefore the rate achieved for w_2 in the genie aided channel is very similar to the rate achieved in the original channel.

With a similar argument we see that the rate at which w_1 can be decoded at receiver 2 in the genie aided channel is similar to the rate in the original channel.

After w_1 and w_2 have been decoded, the Han-Kobayashi strategy decodes u_1 at receiver 1 and u_2 at receiver 2. The side information provides a second view of u_1 at receiver 1 and of u_2 at receiver 2. Doing maximal ratio combining between the two received signals (after subtracting the common messages) we get that the total effective signal to noise ratio is the sum of the SNRs of the two receiver branches, which for $P_u = g_c/N_0$ is:

$$\mathsf{SNR}_{y_1,s_1} = \mathsf{SNR}_{y_1} + \mathsf{SNR}_{s_1} = \frac{g_d P_u}{N_0 + g_c P_u} + \frac{g_c P_u}{N_0} = \frac{g_d}{2g_c} + 1 = \frac{\mathsf{SNR}}{2\mathsf{INR}} + 1.$$

Therefore, the SNR of u_1 in the genie aided channel increases by at most 1 when compared to the SNR in the original channel, resulting in a maximum rate increase of 1 bit/s/Hz. However, when we take $\mathsf{SNR}, \mathsf{INR} \gg 1$, the rate increase is approximately zero.

The same argument can be used to show that the rate of u_2 can increase by at most 1 bit/s/Hz in the genie aided channel.

Step 1, regime (c)

In this regime, the symmetric rate upper bound is derived using the genie aided channel of Figure 2.2. In this genie aided channel receiver 1 has side information x_2 (x_2 can be obtained from

v_2 by dividing it by $h_{2,1}$), and therefore it can completely remove the interference from user 2 in his received signal y_1. The resulting channel is a one sided interference channel.

For the fixed Han-Kobayashi strategy, the rates of the messages u_2, w_1 and w_2 at receiver 2 cannot change, since receiver 2 does not receive any side information. To see this, note that in this regime the binding constraints on the rates of w_1 and w_2 are the MAC constraints at each receiver. Since MAC_2 did not change, the rates of w_1 and w_2 did not change either. In addition, after w_1 and w_2 are decoded at receiver 2, the maximum rate of u_2 does not change either since receiver 2 does not get any side information.

On the other hand the rate u_1 can increase, since receiver 1 can completely remove the interference $h_c u_2$. However this interference is at power level N_0, and removing it can only increase the SNR of u_1 by a factor of 2. Therefore u_1 can increase by at most 1 bit/s/Hz in the genie aided channel.

It follows that for the fixed Han-Kobayashi strategy, the symmetric rate increases by at most 1/2 bit/s/Hz due to the aid of the genie.

Step 2, regime (a)

In this regime the common messages have bounded (and small) rate. The actual rate is irrelevant since when we increase P_u from N_0/g_c to P, the common rate does not increase and, in fact, it goes to zero.

In addition, after maximal ratio combining, the SNR of the private message u_1 is given by:

$$\mathsf{SNR}_{u_1}(P_u) = \frac{g_d P_u}{N_0 + g_c P_u} + \frac{g_c P_u}{N_0}$$

and goes from $\mathsf{SNR}/(2\mathsf{INR}) + 1$ to:

$$\mathsf{SNR}_{u_1}(P) = \frac{g_d P}{N_0 + g_c P} + \frac{g_c P}{N_0} = \frac{\mathsf{SNR}}{\mathsf{INR}+1} + \mathsf{INR} < \frac{\mathsf{SNR}}{\mathsf{INR}} + \mathsf{INR} \tag{2.38}$$

when increasing P_u from N_0/g_c to P.

For $0 < \alpha < 1/2$, $\frac{\mathsf{SNR}}{\mathsf{INR}} > \mathsf{INR}$, the first term of (2.38) dominates, and the SNR of u_1 can increase by at most a factor of 2 when increasing P_u to full power, resulting in a rate gain of at most 1 bit/s/Hz.

The same argument can be used to show that the rate of u_2 cannot increase by more than 1 bit/s/Hz when varying P_u from N_0/g_c to P.

Therefore the symmetric rate can increase by at most 1 bit/s/Hz when varying P_u from our initial Han-Kobayashi strategy to the capacity achieving strategy in the genie aided channel.

Step 2, regime (b)

In this regime the common messages can provide non-negligible rate. When we increase P_u we affect the common message rates in two ways:

i) We reduce the common message power.

ii) We increase the interference generated by the private messages.

The effect of i) is negligible when $P_u \ll P$, since $P_w = P - P_u \approx P$. The effect of ii) is harder to analyze because increasing P_u increases the interference affecting the decoding of w_2, but at the same time increases the usefulness of the side information s_1 in reducing this interference.

We do a back of the envelope calculation to determine how P_u affects the rate of w_2. Receiver 1 can use the side information s_1 to get an estimate $\hat{u}_1$ of the private message u_1 which is treated as Gaussian noise at this stage. We can write $u_1 = \hat{u}_1 + \delta_{u_1}$, where for the minimum mean square error estimator $\hat{u}_1$, the estimation error δ_{u_1} is Gaussian, independent of $\hat{u}_1$, and has variance:

$$\sigma^2 = P_u \left(1 - \frac{1}{1 + N_0/(g_c P_u)}\right) \approx \frac{N_0}{g_c}.$$

In the above equation, the approximation is good for $N_0/(g_c P_u) \ll 1$ ($\mathsf{INR} \gg 1$). We see that the estimation error is approximately independent of the private message power P_u.

Receiver 1 can use $\hat{u}_1$ to reduce the interference plus noise in the received signal y_1, obtaining:

$$\tilde{y}_1 = y_1 - h_d \hat{u}_1 = h_c w_2 + h_c u_2 + z_1 - h_d \delta_{u_1}$$

The resulting SNR for decoding w_2 is given by

$$\mathsf{SNR}_{w_2} \approx \frac{g_c P/N_0}{g_c P_u/N_0 + g_d/g_c + 1} \approx \frac{g_c P/N_0}{g_c P_u/N_0 + g_d/g_c} \tag{2.39}$$

and the rate at which w_2 can be decoded is $R_{w_2} = \log(1 + \mathsf{SNR}_{w_2})$.

We see that increasing P_u results in a reduction in the rate of w_2, but surprisingly, this reduction is compensated by the increase in the rate of u_1. The SNR of u_1 after maximal ratio combining is:

$$\mathsf{SNR}_{u_1} = \frac{g_d P_u/N_0}{1 + g_c P_u/N_0} + \frac{g_c P_u}{N_0} \approx \frac{g_d}{g_c} + \frac{g_c P_u}{N_0} \tag{2.40}$$

where the last approximation holds for $g_c P_u/N_0 \gg 1$.

Comparing the denominator of (2.39) with (2.40) we see that the reduction in the rate of w_2 is approximately the same as the increase of the rate of u_1 as P_u is increased from N_0/g_c to P. Therefore, the sum of these rates (each of which depends logarithmically with SNR) remains approximately invariant when P_u increases.

Step 2, regime (c)

Increasing P_u increases the noise+interference power in MAC_2 from $2N_0$ to $N_0 + g_c P$, or approximately by a factor $g_c P/(2N_0) = \mathsf{INR}/2$. The sum rates of the messages u_2, w_1, and w_2 is reduced by a term $\log(g_c P/(2N_0)) = \log(\mathsf{INR}/2)$.

On the other hand increasing P_u produces an increase in the rate of the private message u_1 which does not see any interference after decoding w_1 due to the aid of the side information. The rate of u_1 is given by

$$R_{u_2}(P_u) = \log\left(1 + \frac{g_d P_u}{N_0}\right) \approx \log\left(\frac{g_d P_u}{N_0}\right).$$

When P_u is increased from N_0/g_c to P, this rate increases by a term $\log(g_c P/N_0) = \log(\mathsf{INR})$.

Comparing the reduction in the sum rate of the messages u_2 w_1 and w_2 to the increase in the rate of the message u_1 we see a net gain of 1 bit/s/Hz for $\mathsf{SNR}, \mathsf{INR} \gg 1$.

Therefore, the symmetric rate increases by at most 1/2 bit/s/Hz when we increase P_u from N_0/g_c to P in the genie aided channel.

Summary

The above analysis shows that the upper bounds are tight for two main reasons:

1. The side information essentially does not modify the binding constraints for the rates that can be achieved with the Han-Kobayashi strategy.

2. The side information provides *just enough help* to maintain the rates in the genie aided channel when varying the communication scheme from the Han-Kobayashi strategy to the capacity achieving scheme in the genie aided channel. Too little help would make $P_u = P$ suboptimal, and would complicate the computation of the capacity of the genie aided channel. Too much help would make the bounds loose.

Note that except in regime (a) where setting $P_u = P$ is optimal in the original channel, increasing P_u beyond N_0/g_c is strictly suboptimal. However, the side information in the genie aided channel compensates the negative effect of increasing P_u and essentially makes the rates invariant with P_u.

2.7 Tight characterization of symmetric capacity

Our simple Han and Kobayashi type scheme, together with the symmetric capacity upper bounds (2.10) and (2.17) allowed us to characterize the symmetric capacity to within one bit/s/Hz. We will now show that in some parameter ranges, the gap between the upper bound (2.17) and the rates achievable with some improved communication schemes vanishes for $\mathsf{SNR}, \mathsf{INR} \to \infty$.

The communication scheme that sets the private message power so that the interference generated onto the other receiver is at noise level (i.e. $g_c P_u = N_0$) is "universal" in the sense that the same scheme can be used to achieve a symmetric rate within one bit/s/Hz of capacity in the weak interference regime, regardless of the values of the parameters.

However, we can further improve the achievable symmetric rate by modifying the communication scheme for different parameter ranges. In the very weak interference regime, when $\log \mathsf{INR} < \frac{1}{2} \log \mathsf{SNR}$ we can simply assign $P_u = P$ and not use common messages at all. As stated in the previous subsection, this scheme achieves a symmetric rate:

$$R = \log\left(1 + \frac{g_d P}{N_0 + g_c P}\right) \tag{2.41}$$

and the gap between this rate and the upper bound (2.17) is:

$$\begin{aligned} R_{UB_{new}} - R &= \log\left(1 + \frac{g_c P}{N_0} + \frac{g_d P}{N_0 + g_c P}\right) - \log\left(1 + \frac{g_d P}{N_0 + g_c P}\right) \\ &= \log\left(1 + \frac{\mathsf{INR}(1+\mathsf{INR})}{1+\mathsf{INR}+\mathsf{SNR}}\right) \approx \log\left(1 + \frac{\mathsf{INR}^2}{\mathsf{SNR}}\right). \end{aligned} \tag{2.42}$$

Note that the same gap would be obtained with any scheme that uses private message power P_u such that $g_c P_u \to \infty$ as $\mathsf{SNR}, \mathsf{INR} \to \infty$.

Recall that $\alpha = \frac{\log \mathsf{INR}}{\log \mathsf{SNR}}$. The very weak interference regime corresponds to $0 < \alpha < \frac{1}{2}$. In this regime, (2.42) implies that for fixed α, $C_{\text{sym}} - R \to 0$ as $\mathsf{SNR}, \mathsf{INR} \to \infty$. Therefore, we have that for $0 < \alpha < 1/2$ the symmetric capacity is tightly characterized by:

$$C_{\text{sym}} = \log\left(\frac{\mathsf{SNR}}{\mathsf{INR}}\right). \tag{2.43}$$

For simplicity of presentation, from now on we assume without loss of generality that $g_d = 1$. In the weaker interference regime where $\frac{1}{2} < \alpha < \frac{2}{3}$, we can choose the private message power as $P_u = N_0(\mathsf{SNR}/\mathsf{INR})^{\gamma}$, where

$$0 < \frac{2\alpha - 1}{1 - \alpha} < \gamma < 1 \tag{2.44}$$

is fixed but arbitrary. This choice of P_u makes the received interference power corresponding to the private message $g_c P_u = g_d N_0 (\mathsf{INR}/\mathsf{SNR})^{(1-\gamma)} = N_0(\mathsf{INR}/\mathsf{SNR})^{(1-\gamma)}$ to go to zero as $\mathsf{SNR}, \mathsf{INR} \to \infty$. Note that in the weaker interference regime we have $\log \mathsf{SNR} < \log \mathsf{INR} < \frac{2}{3} \log \mathsf{SNR}$ and therefore $\mathsf{INR}/\mathsf{SNR} \to 0$ as $\mathsf{SNR}, \mathsf{INR} \to \infty$.

Fixing the decoding order so that the private messages are decoded last, this scheme achieves a symmetric rate:

$$\begin{aligned} R &= \log\left[1+\frac{g_d P_u}{N_0+g_c P_u}\right]+\min\left\{\frac{1}{2}\log\left[1+\frac{(g_d+g_c)(P-P_u)}{N_0+(g_d+g_c)P_u}\right],\log\left[1+\frac{g_c(P-P_u)}{N_0+(g_d+g_c)P_u}\right]\right\} \\ &\overset{(a)}{\simeq} \log\left[1+\left(\frac{\mathsf{SNR}}{\mathsf{INR}}\right)^{\gamma}\right]+\min\left\{\frac{1}{2}\log\left[1+\frac{\mathsf{SNR}}{(\mathsf{SNR}/\mathsf{INR})^{\gamma}}\right],\log\left[1+\frac{\mathsf{INR}}{(\mathsf{SNR}/\mathsf{INR})^{\gamma}}\right]\right\} \\ &\simeq \gamma\log\left(\frac{\mathsf{SNR}}{\mathsf{INR}}\right)+\min\left\{\frac{1}{2}\log\left[\mathsf{SNR}^{1-\gamma}\mathsf{INR}^{\gamma}\right],\log\left[\frac{\mathsf{INR}^{1+\gamma}}{\mathsf{SNR}^{\gamma}}\right]\right\} \\ &= \gamma(1-\alpha)\log(\mathsf{SNR})+\min\left\{\frac{1-\gamma+\alpha\gamma}{2},\alpha(1+\gamma)-\gamma\right\}\log(\mathsf{SNR}) \\ &\overset{(b)}{=} \alpha\log(\mathsf{SNR}) \end{aligned} \tag{2.45}$$

where $\simeq$ means that the the difference between the left and right hand sides goes to zero as $\mathsf{SNR},\mathsf{INR}\to\infty$, (a) follows from the fact that $g_c P_u \to 0$, $P_u/P \to 0$, and $g_c/g_d \to 0$ as $\mathsf{SNR},\mathsf{INR}\to\infty$, and (b) follows because the second term of the $\min\{\cdot,\cdot\}$ dominates due to (2.44).

From the upper bound (2.17) we obtain:

$$\begin{aligned} R_{UB_{new}} &= \log\left(1+\frac{g_c P}{N_0}+\frac{g_d P}{N_0+g_c P}\right) \\ &= \log\left(1+\mathsf{INR}+\frac{\mathsf{SNR}}{1+\mathsf{INR}}\right) \\ &\simeq \log\left(\mathsf{SNR}^{\alpha}+\mathsf{SNR}^{1-\alpha}\right) \\ &\simeq \max\{\alpha,1-\alpha\}\log(\mathsf{SNR}) \\ &= \alpha\log(\mathsf{SNR}). \end{aligned} \tag{2.46}$$

Comparing (2.45) with (2.46) we see that the difference $R_{UB_{new}} - R \to 0$ as $\mathsf{SNR},\mathsf{INR}\to\infty$ and therefore in the weaker interference regime the symmetric capacity is given by:

$$C_{\text{sym}} = \log(\mathsf{INR}). \tag{2.47}$$

We note that both in the very weak and weaker interference regimes we have some flexibility in setting the private message power to asymptotically achieve the symmetric capacity. In the very weak interference regime we can choose any private message power as long as $g_c P_u \to \infty$ when $\mathsf{SNR},\mathsf{INR}\to\infty$. In a similar way, in the weaker interference regime we can use any private message power that satisfies $g_c P_u \to 0$ as $\mathsf{SNR}, \mathsf{INR} \to \infty$. In both cases setting $g_c P_u = N_0$ does not asymptotically achieve the symmetric capacity, but results in a symmetric rate no smaller than 1 bit/s/Hz from it. Unfortunately, in the moderate interference regime the only choice of private message power that achieves a symmetric rate with bounded difference from the upper bound (2.10) is $g_c P_u =$ constant, and this choice of private message power does not result in a gap that vanishes as $\mathsf{SNR}, \mathsf{INR} \to \infty$.

In the strong interference regime the symmetric capacity is given by:

$$C_{\text{sym}} = \frac{1}{2} \log (1 + \mathsf{INR} + \mathsf{SNR}) \tag{2.48}$$

which asymptotically approaches $\log(\sqrt{\mathsf{INR}})$ for $1 < \alpha \leq 2$ as $\mathsf{SNR}, \mathsf{INR} \to \infty$.

Finally in the very strong interference regime the symmetric capacity is given by:

$$C_{\text{sym}} = \log (1 + \mathsf{SNR}) \tag{2.49}$$

which asymptotically approaches $\log(\mathsf{SNR})$ for $\mathsf{SNR} \to \infty$.

We summarize the results of this subsection in the following theorem.

Theorem 2 *Let* $\alpha = (\log \mathsf{INR} / \log \mathsf{SNR})$*. For* $0 < \alpha < 1/2$*,* $1/2 < \alpha < 2/3$ *and* $\alpha > 1$*, the approximation*

$$C_{\text{sym}} \approx \begin{cases} \log(\frac{\mathsf{SNR}}{\mathsf{INR}}) & \log \mathsf{INR} < \frac{1}{2} \log \mathsf{SNR} \\ \log \mathsf{INR} & \frac{1}{2} \log \mathsf{SNR} < \log \mathsf{INR} < \frac{2}{3} \log \mathsf{SNR} \\ \log \sqrt{\mathsf{INR}} & \log \mathsf{SNR} < \log \mathsf{INR} < 2 \log \mathsf{SNR} \\ \log \mathsf{SNR} & \log \mathsf{INR} \geq 2 \log \mathsf{SNR} \end{cases} \tag{2.50}$$

is asymptotically tight in the sense that the difference between C_{sym} *and the approximation goes to zero as* $\mathsf{SNR}, \mathsf{INR}$ *go to infinity with* α *fixed.*

Chapter 3

Gaussian Interference Channel in Weak Interference

In this chapter we will extend the results derived in the previous chapter for the symmetric Gaussian interference channel, where the parameters SNR_1, SNR_2, INR_1, and INR_2 are arbitrary. In this chapter we will relax the assumption of symmetry, but will still require that the channel be in weak interference. This means that the parameters satisfy $\mathsf{INR}_1 < \mathsf{SNR}_2$ and $\mathsf{INR}_2 < \mathsf{SNR}_1$, or equivalently, $g_{1,1} > g_{1,2}$ and $g_{2,2} > g_{2,1}$.

We start with a derivation of an outer bound to the capacity region. We then describe the Han-Kobayashi family of schemes and obtain a simplified expression for the rate region that can be achieved with this family. While computing the largest possible Han-Kobayashi rate region is in general hard due to the large number of combination of strategies that must be considered, we show that by time sharing between two or three simple Han-Kobayashi schemes one can achieve a rate region within one bit/s/Hz of the outer bound, and hence from capacity. Therefore, the inner bound resulting from these schemes and the outer bound together characterize within a single bit/s/Hz the capacity region. We conclude the chapter by extending the concept of generalized degrees of freedom (introduced in Section 2.5 for the symmetric rate of the symmetric channel) to the complete capacity region.

3.1 Outer bound to the capacity region

We present an outer bound to the capacity region of the Gaussian interference in weak interference in the following theorem.

Theorem 3 *For the Gaussian interference channel in weak interference, the capacity region is contained within the set of rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq \log\left(1 + \frac{g_{1,1}P_1}{N_0}\right) \tag{3.1}$$

$$R_2 \leq \log\left(1 + \frac{g_{2,2}P_2}{N_0}\right) \tag{3.2}$$

$$R_1 + R_2 \leq \log\left(1 + \frac{g_{2,2}P_2}{N_0}\right) + \log\left(1 + \frac{g_{1,1}P_1}{N_0 + g_{2,1}P_2}\right) \tag{3.3}$$

$$R_1 + R_2 \leq \log\left(1 + \frac{g_{1,1}P_1}{N_0}\right) + \log\left(1 + \frac{g_{2,2}P_2}{N_0 + g_{1,2}P_1}\right) \tag{3.4}$$

$$R_1 + R_2 \leq \log\left(1 + \frac{g_{2,1}P_2}{N_0} + \frac{g_{1,1}P_1}{N_0 + g_{1,2}P_1}\right) + \log\left(1 + \frac{g_{1,2}P_1}{N_0} + \frac{g_{2,2}P_2}{N_0 + g_{2,1}P_2}\right) \tag{3.5}$$

$$2R_1 + R_2 \leq \log\left(1 + \frac{g_{1,1}P_1}{N_0} + \frac{g_{2,1}P_2}{N_0}\right) + \log\left(1 + \frac{g_{1,2}P_1}{N_0} + \frac{g_{2,2}P_2}{N_0 + g_{2,1}P_2}\right) + \log\left(1 + \frac{g_{1,1}P_1}{N_0 + g_{1,2}P_1}\right) \tag{3.6}$$

$$R_1 + 2R_2 \leq \log\left(1 + \frac{g_{2,2}P_2}{N_0} + \frac{g_{1,2}P_1}{N_0}\right) + \log\left(1 + \frac{g_{2,1}P_2}{N_0} + \frac{g_{1,1}P_1}{N_0 + g_{1,2}P_1}\right) + \log\left(1 + \frac{g_{2,2}P_2}{N_0 + g_{2,1}P_2}\right) \tag{3.7}$$

which expressed in terms of SNR*s and* INR*s becomes:*

$$R_1 \leq \log\left(1 + \mathsf{SNR}_1\right) \tag{3.8}$$

$$R_2 \leq \log\left(1 + \mathsf{SNR}_2\right) \tag{3.9}$$

$$R_1 + R_2 \leq \log\left(1 + \mathsf{SNR}_2\right) + \log\left(1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_1}\right) \tag{3.10}$$

$$R_1 + R_2 \leq \log\left(1 + \mathsf{SNR}_1\right) + \log\left(1 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_2}\right) \tag{3.11}$$

$$R_1 + R_2 \leq \log\left(1 + \mathsf{INR}_1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_2}\right) + \log\left(1 + \mathsf{INR}_2 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_1}\right) \tag{3.12}$$

$$2R_1 + R_2 \leq \log\left(1 + \mathsf{SNR}_1 + \mathsf{INR}_1\right) + \log\left(1 + \mathsf{INR}_2 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_1}\right) + \log\left(1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_2}\right) \tag{3.13}$$

$$R_1 + 2R_2 \leq \log\left(1 + \mathsf{SNR}_2 + \mathsf{INR}_2\right) + \log\left(1 + \mathsf{INR}_1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_2}\right) + \log\left(1 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_1}\right). \tag{3.14}$$

Proof: We prove outer bounds (3.1)-(3.7). The equivalent outer bounds (3.8)-(3.14) can be easily derived by change of variables.

The bounds on R_1 (3.1) and R_2 (3.2) follow from the single user capacity of an AWGN channel.

The bounds (3.3) and (3.4) follow from the sum rate of the one-sided interference channel, which was given in (2.9) (a similar bound can be obtained by considering the one-sided channel that results from taking $g_{1,2} = 0$). Since the derivation of the sum-rate capacity of the one sided channel that follows from [15, 42, 45] is rather convoluted, we provide an alternative proof.

For a block of length n we can write:

$$\begin{aligned}
n(R_1 + R_2) &\leq I(\mathbf{x}_1^n; \mathbf{y}_1^n) + I(\mathbf{x}_2^n; \mathbf{y}_2^n) + n\epsilon_n \\
&\stackrel{(a)}{\leq} I(\mathbf{x}_1^n; \mathbf{y}_1^n) + I(\mathbf{x}_2^n; \mathbf{y}_2^n, \mathbf{x}_1^n) + n\epsilon_n \\
&\stackrel{(b)}{=} I(\mathbf{x}_1^n; \mathbf{y}_1^n) + I(\mathbf{x}_2^n; \mathbf{y}_2^n | \mathbf{x}_1^n) + n\epsilon_n \\
&= h(\mathbf{y}_1^n) - h(\mathbf{y}_1^n | \mathbf{x}_1^n) + h(\mathbf{y}_2^n | \mathbf{x}_1^n) - h(\mathbf{y}_2^n | \mathbf{x}_1^n \mathbf{x}_2^n) + n\epsilon_n \\
&= h(\mathbf{y}_1^n) - h(h_{2,1}\mathbf{x}_2^n + \mathbf{z}_1^n) + h(h_{2,2}\mathbf{x}_2^n + \mathbf{z}_2^n) - h(\mathbf{z}_2^n) + n\epsilon_n \\
&= h(h_{2,1}h_{2,2}\mathbf{x}_2^n + h_{2,1}\mathbf{z}_2^n) - h(h_{2,1}h_{2,2}\mathbf{x}_2^n + h_{2,2}\mathbf{z}_1^n) - n\log g_{2,1} + n\log g_{2,2} \\
&\quad + h(\mathbf{y}_1^n) - h(\mathbf{z}_2^n) + n\epsilon_n \\
&= h(\tilde{\mathbf{x}}_2^n + \tilde{\mathbf{z}}_2^n) - h(\tilde{\mathbf{x}}_2^n + \tilde{\mathbf{z}}_1^n) - n\log g_{2,1} + n\log g_{2,2} \\
&\quad + h(\mathbf{y}_1^n) - h(\mathbf{z}_2^n) + n\epsilon_n
\end{aligned} \tag{3.15}$$

where $\epsilon_n \to 0$ as $n \to \infty$, (a) follows from having a genie providing $\mathbf{x}_1^n$ to receiver 2, and (b) is due to the fact that $\mathbf{x}_1^n$ and $\mathbf{x}_2^n$ are independent. In (3.15) we defined

$$\begin{aligned}\tilde{\mathbf{x}}_2^n &= h_{2,1}h_{2,2}\mathbf{x}_2^n\\ \tilde{\mathbf{z}}_1^n &= h_{2,2}\mathbf{z}_1^n\\ \tilde{\mathbf{z}}_2^n &= h_{2,1}\mathbf{z}_2^n.\end{aligned}$$

In the weak interference regime we have $g_{2,1} < g_{2,2}$, which implies

$$g_{2,1}N_0 = \mathbb{E}[\tilde{z}_{2,i}^2] < \mathbb{E}[\tilde{z}_{1,i}^2] = g_{2,2}N_0.$$

Since the capacity region of the Gaussian interference channel only depends on the marginal distributions of z_1 and z_2, we can choose $\tilde{\mathbf{z}}_1^n$ as:

$$\tilde{\mathbf{z}}_1^n = \tilde{\mathbf{z}}_2^n + \mathbf{z}^n \tag{3.16}$$

where $\mathbf{z}^n \sim \mathcal{CN}(\mathbf{0}, (g_{2,2} - g_{2,1})N_0\mathbf{I}_2)$[1] is independent of $\tilde{\mathbf{z}}_2^n$.

Using the entropy power inequality [17] we can write:

$$\begin{aligned}h(\tilde{\mathbf{x}}_2^n + \tilde{\mathbf{z}}_1^n) - h(\tilde{\mathbf{x}}_2^n + \tilde{\mathbf{z}}_2^n) &= h(\tilde{\mathbf{x}}_2^n + \tilde{\mathbf{z}}_2^n + \mathbf{z}^n) - h(\tilde{\mathbf{x}}_2^n + \tilde{\mathbf{z}}_2^n)\\ &\geq n\log\left[1 + \frac{2^{(1/n)h(\mathbf{z}^n)}}{2^{(1/n)h(\tilde{\mathbf{x}}_2^n+\tilde{\mathbf{z}}_2^n)}}\right]\\ &\geq n\log\left[1 + \frac{(g_{2,2}-g_{2,1})N_0}{g_{2,1}(g_{2,2}P_2+N_0)}\right]\end{aligned} \tag{3.17}$$

where the last inequality is obtained by choosing $\tilde{\mathbf{x}}_2^n \sim \mathcal{CN}(\mathbf{0}, g_{2,2}g_{2,1}P_2\mathbf{I}_2)$ and using the entropy maximizing property of the circularly symmetric complex Gaussian distribution.

Furthermore, we can upper bound $h(\mathbf{y}_1^n)$ by choosing $\mathbf{x}_i^n \sim \mathcal{CN}(\mathbf{0}, P_i\mathbf{I}_2)$, $i = 1, 2$, and using again the entropy maximizing property of the circularly symmetric complex Gaussian distribution we can write:

$$h(\mathbf{y}_1^n) - h(\mathbf{z}_2^n) \leq n\log\left(1 + \frac{g_{1,1}P_1 + g_{2,1}P_2}{N_0}\right). \tag{3.18}$$

[1] $\mathbf{I}_2$ is the identity matrix in $\mathbb{R}^2$

Replacing (3.17) and (3.18) in (3.15) and after some simple calculations we obtain (3.3).

In a similar way we obtain (3.4).

Bound (3.5) was obtained in Theorem 1.

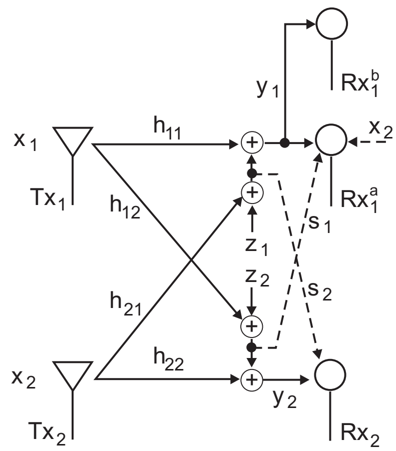


Figure 3.1: Genie-aided two-user Gaussian interference channel with virtual receivers. A genie provides signals s_1 and x_2 to virtual receiver 1a, and s_2 to receiver 2. Virtual receiver 1b only observes y_1.

Next we bound $2R_1 + R_2$. As we did in the proof of Theorem 1 we define

$$\begin{aligned} s_1 &= h_{1,2}x_1 + z_2 \\ s_2 &= h_{2,1}x_2 + z_1 \end{aligned}$$

and use a genie to provide additional information to the receivers. We can think of the term $2R_1$ as arising from the sum rate that transmitter 1 can achieve when communicating with two virtual receivers, 1a and 1b. Both virtual receivers observe y_1, but only receiver 1a gets help from the genie. In particular, the genie provides signals s_1 and x_2 to receiver 1a, and s_2 to receiver 2 (see Figure 3.1). Since the rates achievable with the help of the genie are no smaller than the achievable rates in the original channel, we can write:

$$\begin{aligned} n(2R_1 + R_2) &\leq I(\mathbf{x}_1^n;\mathbf{y}_1^n) + I(\mathbf{x}_1^n;\mathbf{y}_1^n) + I(\mathbf{x}_2^n;\mathbf{y}_2^n) + n\epsilon_n \\ &\leq I(\mathbf{x}_1^n;\mathbf{y}_1^n) + I(\mathbf{x}_1^n;\mathbf{y}_1^n,\mathbf{s}_1^n,\mathbf{x}_2^n) + I(\mathbf{x}_2^n;\mathbf{y}_2^n,\mathbf{s}_2^n) + n\epsilon_n \\ &\stackrel{(a)}{=} I(\mathbf{x}_1^n;\mathbf{y}_1^n) + I(\mathbf{x}_1^n;\mathbf{y}_1^n,\mathbf{s}_1^n|\mathbf{x}_2^n) + I(\mathbf{x}_2^n;\mathbf{y}_2^n,\mathbf{s}_2^n) + n\epsilon_n \end{aligned}$$

$$
\begin{aligned}
&= h(\mathbf{y}_1^n) - h(\mathbf{s}_2^n) + h(\mathbf{y}_1^n, \mathbf{s}_1^n|\mathbf{x}_2^n) - h(\mathbf{y}_1^n, \mathbf{s}_1^n|\mathbf{x}_1^n, \mathbf{x}_2^n) + I(\mathbf{x}_2^n; \mathbf{y}_2^n) + I(\mathbf{x}_2^n; \mathbf{s}_2^n|\mathbf{y}_2^n) + n\epsilon_n \\
&= h(\mathbf{y}_1^n) - h(\mathbf{s}_2^n) + h(\mathbf{s}_1^n|\mathbf{x}_2^n) + h(\mathbf{y}_1^n|\mathbf{s}_1^n, \mathbf{x}_2^n) - h(\mathbf{y}_1^n|\mathbf{x}_1^n, \mathbf{x}_2^n) - h(\mathbf{s}_1^n|\mathbf{x}_1^n, \mathbf{x}_2^n, \mathbf{y}_1^n) \\
&\quad + h(\mathbf{y}_2^n) - h(\mathbf{s}_1^n) + h(\mathbf{s}_2^n|\mathbf{y}_2^n) - h(\mathbf{s}_2^n|\mathbf{x}_2^n, \mathbf{y}_2^n) + n\epsilon_n \\
&= h(\mathbf{y}_1^n) - h(\mathbf{s}_2^n) + h(\mathbf{s}_1^n) + h(\mathbf{y}_1^n|\mathbf{s}_1^n, \mathbf{x}_2^n) - h(\mathbf{z}_1^n) - h(\mathbf{z}_2^n) \\
&\quad + h(\mathbf{y}_2^n) - h(\mathbf{s}_1^n) + h(\mathbf{s}_2^n|\mathbf{y}_2^n) - h(\mathbf{z}_1^n) + n\epsilon_n \\
&= h(\mathbf{y}_1^n) - I(\mathbf{s}_2^n, \mathbf{y}_2^n) + h(\mathbf{y}_1^n|\mathbf{s}_1^n, \mathbf{x}_2^n) + h(\mathbf{y}_2^n) - 2h(\mathbf{z}_1^n) - h(\mathbf{z}_2^n) + n\epsilon_n \\
&= h(\mathbf{y}_1^n) + h(\mathbf{y}_2^n|\mathbf{s}_2^n) + h(\mathbf{y}_1^n|\mathbf{s}_1^n, \mathbf{x}_2^n) - 2h(\mathbf{z}_1^n) - h(\mathbf{z}_2^n) + n\epsilon_n \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} [h(y_{1,i}) + h(y_{2,i}|s_{2,i}) + h(y_{1,i}|s_{1,i}, x_{2,i}) - 2h(z_{1,i}) - h(z_{2,i}) + \epsilon_n] \qquad (3.19)
\end{aligned}
$$

where (a) follows from the independence of $\mathbf{x}_1^n$ and $\mathbf{x}_2^n$, and (b) is obtained from the chain rule for differential entropies and removing conditioning, noting that the components of the noise vectors $\mathbf{z}_1^n$ and $\mathbf{z}_2^n$ are independent.

Let $\mathbb{E}[|x_{1,i}|^2] = p_{1,i}$, and $\mathbb{E}[|x_{2,i}|^2] = p_{2,i}$. For a given covariance matrix, the circularly symmetric complex Gaussian distribution not only maximizes the differential entropy, but also maximizes any conditional differential entropy. It follows that the individual terms of (3.19) are maximized by choosing $x_{1,i} \sim \mathcal{CN}(0, p_{1,i})$ and $x_{2,i} \sim \mathcal{CN}(0, p_{2,i})$, $i = 1, \ldots, n$. Using these distributions for $x_{1,i}$ and $x_{2,i}$, we can use Jensen's inequality to write:

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} h(y_{1,i}) &\leq \frac{1}{n}\sum_{i=1}^{n} \log\left[\pi e\left(g_{1,1}p_{1,i} + g_{2,1}p_{2,i} + N_0\right)\right] \\
&\leq \log\left[\pi e\left(g_{1,1}\frac{1}{n}\sum_{i=1}^{n} p_{1,i} + g_{2,1}\frac{1}{n}\sum_{i=1}^{n} p_{2,i} + N_0\right)\right] \\
&\leq \log\left[\pi e\left(g_{1,1}P_1 + g_{2,1}P_2 + N_0\right)\right] \qquad (3.20)
\end{aligned}
$$

together with

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} h(y_{2,i}|s_{2,i}) &= \frac{1}{n}\sum_{i=1}^{n} h(h_{1,2}x_{1,i} + h_{2,2}x_{2,i} + z_{2,i}|s_{2,i}) \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \log\left[\pi e\left(N_0 + g_{1,2}p_{1,i} + \frac{g_{2,2}N_0p_{2,i}}{N_0 + g_{2,1}p_{2,i}}\right)\right]
\end{aligned}
$$

$$\leq \quad \log\left[\pi e\left(N_0+g_{1,2}\left(\frac{1}{n}\sum_{i=1}^{n}p_{1,i}\right)+\frac{g_{2,2}N_0\left(\frac{1}{n}\sum_{i=1}^{n}p_{2,i}\right)}{N_0+g_{2,1}\left(\frac{1}{n}\sum_{i=1}^{n}p_{2,i}\right)}\right)\right]$$

$$\leq \quad \log\left[\pi e\left(N_0+g_{1,2}P_1+\frac{g_{2,2}N_0P_2}{N_0+g_{2,1}P_2}\right)\right] \tag{3.21}$$

and

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}h(y_{1,i}|s_{1,i}x_{2,i}) &= \frac{1}{n}\sum_{i=1}^{n}h(h_{1,1}x_{1,i}+z_{1i}|s_{1i})\\ &\leq \frac{1}{n}\sum_{i=1}^{n}\log\left[\pi e\left(N_0+\frac{g_{1,1}N_0p_{1,i}}{N_0+g_{1,2}p_{1,i}}\right)\right]\\ &\leq \log\left[\pi e\left(N_0+\frac{g_{1,1}N_0\left(\frac{1}{n}\sum_{i=1}^{n}p_{1,i}\right)}{N_0+g_{1,2}\left(\frac{1}{n}\sum_{i=1}^{n}p_{1,i}\right)}\right)\right]\\ &\leq \log\left[\pi e\left(N_0+\frac{g_{1,1}N_0P_1}{N_0+g_{1,2}P_1}\right)\right] \end{aligned}\tag{3.22}$$

where in all cases Jensen's inequality was applied to concave functions.

Thus we have

$$\begin{aligned}2R_1+R_2 &\leq \log\left[\pi e\left(g_{1,1}P_1+g_{2,1}P_2+N_0\right)\right]-\log\left[\pi eN_0\right]\\ &\quad +\log\left[\pi e\left(N_0+g_{1,2}P_1+\frac{g_{2,2}P_2N_0}{N_0+g_{2,1}P_2}\right)\right]-\log\left[\pi eN_0\right]\\ &\quad +\log\left[\pi e\left(N_0+\frac{g_{1,1}N_0P_1}{N_0+g_{1,2}P_1}\right)\right]-\log\left[\pi eN_0\right]+\epsilon_n\\ &= \log\left(1+\frac{g_{1,1}P_1}{N_0}+\frac{g_{2,1}P_2}{N_0}\right)+\log\left(1+\frac{g_{1,2}P_1}{N_0}+\frac{g_{2,2}P_2/N_0}{1+g_{2,1}P_2/N_0}\right)\\ &\quad +\log\left(1+\frac{g_{1,1}P_1/N_0}{1+g_{1,2}P_1/N_0}\right)+\epsilon_n. \end{aligned}\tag{3.23}$$

Finally, we can derive the bound (3.7) for R_1+2R_2 in a similar way. ■

3.2 Achievable rate region with Han-Kobayashi scheme

The best known achievable region for interference channel is the Han-Kobayashi region [29]. To define this region, Han and Kobayashi use auxiliary random variables U_1, U_2, W_1, W_2, and Q.

The U variables can be interpreted as representing private information, to be decoded only by the intended receiver, while the W variables represent common information to be decoded by both receivers. In addition, the auxiliary random variable Q plays the role of a time-sharing parameter.

For a fixed random vector $\mathbf{Z} = (Q, U_1, W_1, U_2, W_2, X_1, X_2, Y_1, Y_2)$ the Han-Kobayashi scheme allows to achieve a rate region $\mathcal{R}(\mathbf{Z})$, given in [29, Theorem 4.1]. Furthermore, [29, Theorem 3.2] shows that $\mathcal{R}^* = \text{Closure}\{\cup_{\mathbf{Z}\in\mathcal{P}^*}\mathcal{R}(\mathbf{Z})\}$ can be achieved, where $\mathcal{P}^*$ is the set of probability distributions for $\mathbf{Z}$ that satisfy:

- U_1, U_2, W_1, W_2 are conditionally independent given Q;
- $X_1 = f(U_1, W_1|Q)$ and $X_2 = f(U_2, W_2|Q)$;
- the conditional distribution of (Y_1, Y_2) given (X_1, X_2) matches the channel conditional distribution.

Unfortunately, the region $\mathcal{R}(\mathbf{Z})$ given in [29, Theorem 4.1] is not easy to evaluate and contains many redundant bounds. In order to simplify our future calculations we provide in this section a simplified expression for the region given in [29, Theorem 4.1]. [2]

Lemma 1 *The Han-Kobayashi region $\mathcal{R}(\mathbf{Z})$ is equivalent to the polyhedron consisting of all rate pairs (R_1, R_2) satisfying*

$$R_1 \leq I(Y_1; W_1|W_2Q) + I(Y_1; U_1|W_1W_2Q) \tag{3.24}$$

$$R_1 \leq I(Y_2; W_1|U_2W_2Q) + I(Y_1; U_1|W_1W_2Q) \tag{3.25}$$

$$R_2 \leq I(Y_2; W_2|W_1Q) + I(Y_2; U_2|W_1W_2Q) \tag{3.26}$$

$$R_2 \leq I(Y_1; W_2|U_1W_1Q) + I(Y_2; U_2|W_1W_2Q) \tag{3.27}$$

$$R_1 + R_2 \leq I(Y_1; W_1W_2|Q) + I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) \tag{3.28}$$

$$R_1 + R_2 \leq I(Y_2; W_1W_2|Q) + I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) \tag{3.29}$$

[2] [12] obtained independently a similar simplification of the Han-Kobayashi achievable region, derived using Fourier-Motzkin elimination to remove the redundant rate constraints.

$$R_1 + R_2 \leq I(Y_2; W_1|W_2Q) + I(Y_1; W_2|W_1Q) + I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) \quad (3.30)$$

$$2R_1 + R_2 \leq I(Y_1; W_1W_2|Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_2; W_1|W_2Q) \quad (3.31)$$

$$R_1 + 2R_2 \leq I(Y_2; W_1W_2|Q) + I(Y_1; U_1|W_1W_2Q) + 2I(Y_2; U_2|W_1W_2Q) + I(Y_1; W_2|W_1Q) \quad (3.32)$$

Proof: We use the definitions and notation of [29] in this proof.

The bounds on R_1 (3.24)(3.25) and R_2 (3.26)(3.27) are the same as those in [29, theorem 4.1].

Besides the three sum rate bounds (3.28)(3.29)(3.30), there is a fourth bound on $R_1 + R_2$ in [29, theorem 4.1]:

$$R_1 + R_2 \leq I(Y_1; W_1|W_2Q) + I(Y_2; W_2|W_1Q) + I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q).$$

However, this bound is equal to the sum of (3.24) and (3.26), and hence is redundant.

Next we evaluate the $2R_1 + R_2$ bounds in [29, theorem 4.1]. Let σ_1^* be as defined in [29]:

$$\sigma_1^* = \min \{I(Y_1; W_1|W_2Q), I(Y_2; W_1|U_2W_2Q)\}.$$

Depending on the value of $[\sigma_1^* - I(Y_2; W_1|W_2Q)]^+$, there are two cases.

<u>**C**ase I</u>: $\sigma_1^* - I(Y_2; W_1|W_2Q) \geq 0$

In this case, evaluating [29, equation (4.5)], we have

$$\begin{aligned} 2R_1 + R_2 &\leq 2\sigma_1^* + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) - \sigma_1^* + I(Y_2; W_1|W_2Q) \\ &\quad + \min\left\{I(Y_2; W_2|W_1Q), I(Y_2; W_2|Q), I(Y_1; W_2|W_1Q), I(Y_1; W_1W_2|Q) - \sigma_1^*\right\} \\ &= \sigma_1^* + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_2; W_1|W_2Q) \\ &\quad + \min\left\{I(Y_2; W_2|Q), I(Y_1; W_2|W_1Q), I(Y_1; W_1W_2|Q) - \sigma_1^*\right\} \end{aligned}$$

where the last equality follows from the fact that $I(Y_2; W_2|W_1Q) \geq I(Y_2; W_2|Q)$ due to the independence between W_1 and W_2.

If $\sigma_1^* = I(Y_1; W_1|W_2Q)$, we have

$$\begin{aligned} 2R_1 + R_2 &\leq I(Y_1; W_1|W_2Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_2; W_1|W_2Q) \\ &\quad + \min\left\{I(Y_2; W_2|Q), I(Y_1; W_2|W_1Q), I(Y_1; W_2|Q)\right\} \\ &= I(Y_1; W_1|W_2Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_2; W_1|W_2Q) \\ &\quad + \min\left\{I(Y_2; W_2|Q), I(Y_1; W_2|Q)\right\} \end{aligned}$$

where the last equality again follows from the independence between W_1 and W_2.

So we have two bounds:

$$\begin{aligned} 2R_1 + R_2 &\leq I(Y_2; W_1W_2|Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_1; W_1|W_2Q) \\ 2R_1 + R_2 &\leq I(Y_1; W_1W_2|Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_2; W_1|W_2Q) \end{aligned} \tag{3.33}$$

where the first bound is redundant since it is the summation of (3.24) and (3.29).

If $\sigma_1^* = I(Y_2; W_1|U_2W_2Q)$, we have

$$\begin{aligned} 2R_1 + R_2 &\leq I(Y_2; W_1|U_2W_2Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_2; W_1|W_2Q) \\ &\quad + \min\left\{I(Y_2; W_2|Q), I(Y_1; W_2|W_1Q), I(Y_1; W_1W_2|Q) - I(Y_2; W_1|U_2W_2Q)\right\} \end{aligned}$$

So we have three bounds:

$$\begin{aligned} 2R_1 + R_2 &\leq I(Y_2; W_1W_2|Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_2; W_1|U_2W_2Q) \\ 2R_1 + R_2 &\leq I(Y_2; W_1|W_2Q) + I(Y_1; W_2|W_1Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) \\ &\quad + I(Y_2; W_1|U_2W_2Q) \\ 2R_1 + R_2 &\leq I(Y_1; W_1W_2|Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_2; W_1|W_2Q) \end{aligned}$$

We can see that the first bound on $2R_1 + R_2$ is the summation of (3.25) and (3.29), and the second bound on $2R_1 + R_2$ is the summation of (3.25) and (3.30), so these two bounds are redundant. Also, the third bound is equal to (3.33).

It follows that in case I the only bound for $2R_1 + R_2$ is (3.33) which is (3.31).

Case II: $\sigma_1^* - I(Y_2; W_1|W_2Q) < 0$

Since $I(Y_2; W_1|W_2U_2Q) \geq I(Y_2; W_1|W_2Q)$ is always true, in this case, we have $\sigma_1^* = I(Y_1; W_1|W_2Q)$. Evaluating [29, equation (4.5)], we have

$$\begin{aligned} 2R_1 + R_2 \quad \leq \quad & 2I(Y_1; W_1|W_2Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) \\ & + \min\{I(Y_2; W_2|W_1Q), I(Y_2; W_2|Q) + I(Y_2; W_1|W_2Q) - I(Y_1; W_1|W_2Q), \\ & I(Y_1; W_2|W_1Q), I(Y_1; W_1W_2|Q) - I(Y_1; W_1|W_2Q)\} \end{aligned}$$

So we have the following four bounds:

$$\begin{aligned} 2R_1 + R_2 &\leq 2I(Y_1; W_1|W_2Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_2; W_2|W_1Q) \\ 2R_1 + R_2 &\leq I(Y_2; W_1W_2|Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_1; W_1|W_2Q) \\ 2R_1 + R_2 &\leq 2I(Y_1; W_1|W_2Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_1; W_2|W_1Q) \\ 2R_1 + R_2 &\leq I(Y_1; W_1W_2|Q) + 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_1; W_1|W_2Q). \end{aligned}$$

The first bound can be obtained by adding two times (3.24) and (3.26); the second bound can be obtained by adding (3.24) and (3.29); while the fourth bound can be obtained by adding (3.24) and (3.28). In addition, the third bound on $2R_1 + R_2$ is also redundant. To see this we rewrite the fourth bound as:

$$2R_1 + R_2 \leq 2I(Y_1; U_1|W_1W_2Q) + I(Y_2; U_2|W_1W_2Q) + I(Y_1; W_1|W_2Q) + I(Y_1; W_1|W_2Q) + I(Y_1; W_2|Q)$$

which is not larger than the third bound above, due to the fact that $I(Y_1; W_2|W_1Q) \geq Y(Y_1; W_2|Q)$. Therefore, in case II the four bounds on $2R_1 + R_2$ are redundant.

Summarizing cases I and II, we have only one non-redundant $2R_1 + R_2$ bound which is (3.31).

Finally, we can simplify the bounds on $R_1 + 2R_2$ in a similar way. ■

For Gaussian interference channel, we choose U_1, W_1, U_2, W_2 to be independent Gaussian random variables, and $X_1 = U_1 + W_1$, $X_2 = U_2 + W_2$. We define the parameters k_1 and k_2,

$0 \leq k_1 \leq \mathsf{INR}_2$ and $0 \leq k_2 \leq \mathsf{INR}_1$, such that the interference to noise ratio of user 1's (2's) private message at receiver 2 (1) is k_1 (k_2). Equivalently,

$$\begin{aligned} P_{u1} &= \frac{k_1 N_0}{g_{1,2}} \\ P_{u2} &= \frac{k_2 N_0}{g_{2,1}} \end{aligned}$$

where P_{ui} is the power of user i's private message, $i = 1, 2$. With this definition, the signal to noise ratio of user 1's private message at receiver 1 is $k_1 \frac{\mathsf{SNR}_1}{\mathsf{INR}_2} = k_1 \frac{g_{1,1}}{g_{1,2}}$ and the signal to noise ratio of user 2's private message at receiver 2 is $k_2 \frac{\mathsf{SNR}_2}{\mathsf{INR}_1} = k_2 \frac{g_{2,2}}{g_{2,1}}$. In addition, we assign the remaining power of each user $(P_i - P_{ui})$ to the common message.

This choice of random variables defines a Han-Kobayashi scheme parameterized in (k_1, k_2), which we denote $HK(k_1, k_2)$. Also, we can evaluate the rate region that this scheme achieves using Lemma 1, which we denote by $\mathcal{R}_{HK}(k_1, k_2)$.

We have the following expressions for relevant mutual information terms:

$$\begin{aligned}
I(Y_1; U_1|W_1W_2) &= \log\left(1 + \frac{g_{1,1}P_{u1}}{N_0 + g_{2,1}P_{u2}}\right) = \log\left[1 + \frac{g_{1,1}}{g_{1,2}}\left(\frac{k_1}{1+k_2}\right)\right] \\
I(Y_2; U_2|W_1W_2) &= \log\left(1 + \frac{g_{2,2}P_{u2}}{N_0 + g_{1,2}P_{u1}}\right) = \log\left[1 + \frac{g_{2,2}}{g_{2,1}}\left(\frac{k_2}{1+k_1}\right)\right] \\
I(Y_1; W_1|W_2) &= \log\left(1 + \frac{g_{1,1}(P_1 - P_{u1})}{N_0 + g_{2,1}P_{u2} + g_{1,1}P_{u1}}\right) = \log\left(1 + \frac{\frac{g_{1,1}P_1}{N_0} - k_1\frac{g_{1,1}}{g_{1,2}}}{1 + k_2 + k_1\frac{g_{1,1}}{g_{1,2}}}\right) \\
I(Y_1; W_2|W_1) &= \log\left(1 + \frac{g_{2,1}(P_2 - P_{u2})}{N_0 + g_{2,1}P_{u2} + g_{1,1}P_{u1}}\right) = \log\left(1 + \frac{\frac{g_{2,1}P_2}{N_0} - k_2}{1 + k_2 + k_1\frac{g_{1,1}}{g_{1,2}}}\right) \\
I(Y_1; W_1W_2) &= \log\left(1 + \frac{g_{1,1}(P_1 - P_{u1}) + g_{2,1}(P_2 - P_{u2})}{N_0 + g_{2,1}P_{u2} + g_{1,1}P_{u1}}\right) \\
&= \log\left(1 + \frac{\frac{g_{1,1}P_1}{N_0} - k_1\frac{g_{1,1}}{g_{1,2}} + \frac{g_{2,1}P_2}{N_0} - k_2}{1 + k_2 + k_1\frac{g_{1,1}}{g_{1,2}}}\right) \\
I(Y_2; W_2|W_1) &= \log\left(1 + \frac{g_{2,2}(P_2 - P_{u2})}{N_0 + g_{1,2}P_{u1} + g_{2,2}P_{u2}}\right) = \log\left(1 + \frac{\frac{g_{2,2}P_2}{N_0} - k_2\frac{g_{2,2}}{g_{2,1}}}{1 + k_1 + k_2\frac{g_{2,2}}{g_{2,1}}}\right) \\
I(Y_2; W_1|W_2) &= \log\left(1 + \frac{g_{1,2}(P_1 - P_{u1})}{N_0 + g_{1,2}P_{u1} + g_{2,2}P_{u2}}\right) = \log\left(1 + \frac{\frac{g_{1,2}P_1}{N_0} - k_1}{1 + k_1 + k_2\frac{g_{2,2}}{g_{2,1}}}\right) \\
I(Y_2; W_1W_2) &= \log\left(1 + \frac{g_{2,2}(P_2 - P_{u2}) + g_{1,2}(P_1 - P_{u1})}{N_0 + g_{1,2}P_{u1} + g_{2,2}P_{u2}}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \log\left(1+\frac{\frac{g_{2,2}P_2}{N_0}-k_2\frac{g_{2,2}}{g_{2,1}}+\frac{g_{1,2}P_1}{N_0}-k_1}{1+k_1+k_2\frac{g_{2,2}}{g_{2,1}}}\right)\\
I(Y_1;W_2|U_1W_1) &= \log\left(1+\frac{g_{2,1}(P_2-P_{u2})}{N_0+g_{2,1}P_{u2}}\right)=\log\left(1+\frac{\frac{g_{2,1}P_2}{N_0}-k_2}{1+k_2}\right)\\
I(Y_2;W_1|U_2W_2) &= \log\left(1+\frac{g_{1,2}(P_1-P_{u1})}{N_0+g_{1,2}P_{u1}}\right)=\log\left(1+\frac{\frac{g_{1,2}P_1}{N_0}-k_1}{1+k_1}\right)
\end{aligned}
$$

$$
\begin{aligned}
I(Y_1;U_1|W_1W_2) &= \log\left(1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\frac{k_1}{1+k_2}\right)\\
I(Y_2;U_2|W_1W_2) &= \log\left(1+\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}\frac{k_2}{1+k_1}\right)\\
I(Y_1;W_1|W_2) &= \log\left(1+\frac{\mathsf{SNR}_1-k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}{1+k_2+k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}\right)\\
I(Y_1;W_2|W_1) &= \log\left(1+\frac{\mathsf{INR}_1-k_2}{1+k_2+k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}\right)\\
I(Y_1;W_1W_2) &= \log\left(1+\frac{\mathsf{SNR}_1-k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}+\mathsf{INR}_1-k_2}{1+k_2+k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}\right)\\
I(Y_2;W_2|W_1) &= \log\left(1+\frac{\mathsf{SNR}_2-k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}{1+k_1+k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}\right)\\
I(Y_2;W_1|W_2) &= \log\left(1+\frac{\mathsf{INR}_2-k_1}{1+k_1+k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}\right)\\
I(Y_2;W_1W_2) &= \log\left(1+\frac{\mathsf{SNR}_2-k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}+\mathsf{INR}_2-k_1}{1+k_1+k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}\right)\\
I(Y_1;W_2|U_1W_1) &= \log\left(1+\frac{\mathsf{INR}_1-k_2}{1+k_2}\right)\\
I(Y_2;W_1|U_2W_2) &= \log\left(1+\frac{\mathsf{INR}_2-k_1}{1+k_1}\right)
\end{aligned}
$$

Replacing this mutual information expressions in the bounds of Lemma 1 we have that $\mathcal{R}_{HK}(k_1,k_2)$ is given by the rate pairs (R_1,R_2) that satisfy:

$$
\begin{aligned}
R_1 &\le \log\left(1+\frac{\mathsf{SNR}_1-k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}{1+k_2+k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}\right)+\log\left(1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\frac{k_1}{1+k_2}\right)\\
R_1 &\le \log\left(1+\frac{\mathsf{INR}_2-k_1}{1+k_1}\right)+\log\left(1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\frac{k_1}{1+k_2}\right)\\
R_2 &\le \log\left(1+\frac{\mathsf{SNR}_2-k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}{1+k_1+k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}\right)+\log\left(1+\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}\frac{k_2}{1+k_1}\right)
\end{aligned}
$$

$$
\begin{aligned}
R_2 &\leq \log\left(1+\frac{\mathsf{INR}_1-k_2}{1+k_2}\right)+\log\left(1+\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}\frac{k_2}{1+k_1}\right)\\
R_1+R_2 &\leq \log\left(1+\frac{\mathsf{SNR}_1-k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}+\mathsf{INR}_1-k_2}{1+k_2+k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}\right)\\
&\quad+\log\left(1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\frac{k_1}{1+k_2}\right)+\log\left(1+\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}\frac{k_2}{1+k_1}\right)\\
R_1+R_2 &\leq \log\left(1+\frac{\mathsf{SNR}_2-k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}+\mathsf{INR}_2-k_1}{1+k_1+k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}\right)+\log\left(1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\frac{k_1}{1+k_2}\right)\\
&\quad+\log\left(1+\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}\frac{k_2}{1+k_1}\right)\\
R_1+R_2 &\leq \log\left(1+\frac{\mathsf{INR}_2-k_1}{1+k_1+k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}\right)+\log\left(1+\frac{\mathsf{INR}_1-k_2}{1+k_2+k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}\right)\\
&\quad+\log\left(1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\frac{k_1}{1+k_2}\right)+\log\left(1+\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}\frac{k_2}{1+k_1}\right)\\
2R_1+R_2 &\leq \log\left(1+\frac{\mathsf{SNR}_1-k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}+\mathsf{INR}_1-k_2}{1+k_2+k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}\right)+2\log\left(1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\frac{k_1}{1+k_2}\right)\\
&\quad+\log\left(1+\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}\frac{k_2}{1+k_1}\right)+\log\left(1+\frac{\mathsf{INR}_2-k_1}{1+k_1+k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}\right)\\
R_1+2R_2 &\leq \log\left(1+\frac{\mathsf{SNR}_2-k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}+\mathsf{INR}_2-k_1}{1+k_1+k_2\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}}\right)+\log\left(1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\frac{k_1}{1+k_2}\right)\\
&\quad+2\log\left(1+\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}\frac{k_2}{1+k_1}\right)+\log\left(1+\frac{\mathsf{INR}_1-k_2}{1+k_2+k_1\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}}\right) \qquad (3.34)
\end{aligned}
$$

We will use this region and the outer bound derived in Section 3.1 to characterize the capacity region of the Gaussian interference channel in weak interference with an error no larger than one bit/s/Hz for each user.

3.3 Characterization of the capacity region to within one bit/s/Hz

In Chapter 2 we showed that a simple Han-Kobayashi scheme could achieve a symmetric rate close to the symmetric capacity of the symmetric Gaussian interference channel. More specifically, we showed that the scheme achieved a rate that was at most one bit/s/Hz below a capacity upper bound. In this section we prove an equivalent result for the whole capacity region of the

Gaussian interference channel in weak interference. In particular, we will show that by time-sharing between at most three simple Han-Kobayashi schemes we can achieve a rate region that is *within one bit/s/Hz* of the capacity region.

We make the "within one bit/s/Hz" notion precise with the following definition.

Definition 1 *An achievable region is* within one bit/s/Hz of the capacity region *if for any rate pair* (R_1, R_2) *on the Pareto boundary* [3] *of the achievable region, the rate pair* $(R_1 + 1, R_2 + 1)$ *is not achievable.*

Our goal is to find an achievable region such that for any rate point (R_1, R_2) of its boundary, the rate pair (R_1+1, R_2+1) exceeds some outer bound of the capacity region. By the above definition, such a region would be within one bit/s/Hz of capacity.

For a fixed distribution of the vector $\mathbf{Z} = (Q, U_1, W_1, U_2, W_2, X_1, X_2, Y_1, Y_2)$ the Han-Kobayashi scheme achieves a rate region delimited by straight lines of slope 0, $-1/2$, -1, -2, and ∞ (some of these lines may be missing in some degenerate cases). These lines are defined by the bounds on R_1, R_2, $R_1 + R_2$, $2R_1 + R_2$, and $2R_2 + R_1$ given in Lemma 1. Let these bounds be denoted by HK_{R_1}, HK_{R_2}, $HK_{R_1+R_2}$, $HK_{2R_1+R_2}$, and $HK_{R_1+2R_2}$ respectively.

On the other hand, the outer bound presented in Theorem 3 consists of a region delimited by straight lines of slope 0, $-1/2$, -1, -2, and ∞. In other words, the outer bound is defined in terms of upper bounds on R_1, R_2, $R_1 + R_2$, $2R_1 + R_2$, and $2R_2 + R_1$. Let these bounds be denoted by UB_{R_1}, UB_{R_2}, $UB_{R_1+R_2}$, $UB_{2R_1+R_2}$, and $UB_{R_1+2R_2}$ respectively.

Let $\Delta_{R_1} = UB_{R_1} - HK_{R_1}$, $\Delta_{R_2} = UB_{R_2} - HK_{R_2}$, $\Delta_{R_1+R_2} = UB_{R_1+R_2} - HK_{R_1+R_2}$, $\Delta_{2R_1+R_2} = UB_{2R_1+R_2} - HK_{2R_1+R_2}$, and $\Delta_{R_1+2R_2} = UB_{R_1+R_2} - HK_{R_1+2R_2}$ denote the differences between the boundaries of the Han-Kobayashi region and the outer bound. If $\Delta_{R_1} < 1$, increasing R_1 by one bit/s/Hz in any point on the vertical boundary of the Han-Kobayashi region results in a point that exceeds the R_1 upper bound. Similarly, $\Delta_{R_2} < 1$ implies that we cannot increase R_2 in any point on the horizontal boundary of the Han-Kobayashi region without exceeding the

[3] Refer to Chapter 11 in Part III for a definition of the Pareto boundary of a rate region.

corresponding upper bound. In addition, $\Delta_{R_1+R_2} < 2$, $\Delta_{2R_1+R_2} < 3$, and $\Delta_{R_1+2R_2} < 3$ imply that if we increased R_1 and R_2 by one bit/s/Hz in any point on the boundaries of slope $-1/2$, -1, or -2, the resulting rate pair would fall out of the achievable region. It follows that:

$$\begin{aligned}
\Delta_{R_1} &< 1 \\
\Delta_{R_2} &< 1 \\
\Delta_{R_1+R_2} &< 2 \\
\Delta_{2R_1+R_2} &< 3 \\
\Delta_{R_1+2R_2} &< 3 \qquad (3.35)
\end{aligned}$$

is a sufficient condition for the achievable region to be within one bit/s/Hz of the capacity region.

In Section 2.4 we showed that the Han-Kobayashi scheme that sets the private message power so that the interference created onto the other user's receiver is at the noise level (i.e. $g_c P_u = N_0$) achieved a symmetric rate that is at most one bit/s/Hz away from the symmetric capacity. We now explore how the same scheme performs over the whole achievable rate region. We first note that we can only set the private message power $P_{u1} = N_0/g_{1,2}$ and $P_{u2} = N_0/g_{2,1}$ when $\mathsf{INR}_1 \geq 1$ and $\mathsf{INR}_2 \geq 1$. Therefore in the following analysis, we will assume $\mathsf{INR}_1 \geq 1$ and $\mathsf{INR}_2 \geq 1$. Setting $k_1 = 1$ and $k_2 = 1$ in (3.34) we obtain that the achievable region $\mathcal{R}_{HK}(1,1)$ is given by the set of rate pairs (R_1, R_2) that satisfy:

$$\begin{aligned}
R_1 &\leq \log\left(1 + \frac{\mathsf{SNR}_1}{2}\right) \text{ if } 2\mathsf{INR}_2 \geq \mathsf{SNR}_1 && (3.36) \\
R_1 &\leq \log\left(\frac{(1+\mathsf{INR}_2)(2\mathsf{INR}_2+\mathsf{SNR}_1)}{4\mathsf{INR}_2}\right) \text{ if } 2\mathsf{INR}_2 < \mathsf{SNR}_1 && (3.37) \\
R_2 &\leq \log\left(1 + \frac{\mathsf{SNR}_2}{2}\right) \text{ if } 2\mathsf{INR}_1 \geq \mathsf{SNR}_2 && (3.38) \\
R_2 &\leq \log\left(\frac{(1+\mathsf{INR}_1)(2\mathsf{INR}_1+\mathsf{SNR}_2)}{4\mathsf{INR}_1}\right) \text{ if } 2\mathsf{INR}_1 < \mathsf{SNR}_2 && (3.39) \\
R_1 + R_2 &\leq \log\left(\frac{(1+\mathsf{INR}_1+\mathsf{SNR}_1)(2\mathsf{INR}_1+\mathsf{SNR}_2)}{4\mathsf{INR}_1}\right) && (3.40) \\
R_1 + R_2 &\leq \log\left(\frac{(1+\mathsf{INR}_2+\mathsf{SNR}_2)(2\mathsf{INR}_2+\mathsf{SNR}_1)}{4\mathsf{INR}_2}\right) && (3.41)
\end{aligned}$$

$$R_1 + R_2 \leq \log\left(\frac{(\mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 + \mathsf{SNR}_1)(\mathsf{INR}_1 + \mathsf{INR}_1\mathsf{INR}_2 + \mathsf{SNR}_2)}{4\mathsf{INR}_1\mathsf{INR}_2}\right) \quad (3.42)$$

$$2R_1 + R_2 \leq \log\left(\frac{(1 + \mathsf{INR}_1 + \mathsf{SNR}_1)(2\mathsf{INR}_2 + \mathsf{SNR}_1)(\mathsf{INR}_1 + \mathsf{INR}_1\mathsf{INR}_2 + \mathsf{SNR}_2)}{8\mathsf{INR}_1\mathsf{INR}_2}\right) \quad (3.43)$$

$$R_1 + 2R_2 \leq \log\left(\frac{(1 + \mathsf{INR}_2 + \mathsf{SNR}_2)(2\mathsf{INR}_1 + \mathsf{SNR}_2)(\mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 + \mathsf{SNR}_1)}{8\mathsf{INR}_1\mathsf{INR}_2}.\right) \quad (3.44)$$

We will now compute the gaps between the boundary lines of the achievable region and the outer bound. If $2\mathsf{INR}_2 \geq \mathsf{SNR}_1$ we can write:

$$\begin{aligned}
\Delta_{R_1} &= \log(1 + \mathsf{SNR}_1) - \log\left(1 + \frac{\mathsf{SNR}_1}{2}\right) \\
&= \log\left(\frac{1 + \mathsf{SNR}_1}{2 + \mathsf{SNR}_1}\right) + \log 2 \\
&< 1
\end{aligned}$$

while for $2\mathsf{INR}_2 < \mathsf{SNR}_1$ we have:

$$\begin{aligned}
\Delta_{R_1} &= \log(1 + \mathsf{SNR}_1) - \log\left(\frac{(1 + \mathsf{INR}_2)(2\mathsf{INR}_2 + \mathsf{SNR}_1)}{4\mathsf{INR}_2}\right) \\
&\leq \log\left(\frac{1 + \mathsf{SNR}_1}{2 + \mathsf{SNR}_1}\right) + \log\left(\frac{4\mathsf{INR}_2}{1 + \mathsf{INR}_2}\right) \\
&< 2.
\end{aligned}$$

In a similar way we obtain $\Delta_{R_2} < 1$ for $2\mathsf{INR}_1 \geq \mathsf{SNR}_2$, and $\Delta_{R_2} < 2$ for $2\mathsf{INR}_1 < \mathsf{SNR}_2$.

To compute $\Delta_{R_1+R_2}$ we need to consider the three sum rate bounds (3.40),(3.41), and (3.42) separately, using an appropriate upper bound from Theorem 3. For (3.40) we use the upper bound (3.10) to obtain a gap:

$$\begin{aligned}
\Delta^{(A)}_{R_1+R_2} &= \log(1 + \mathsf{SNR}_2) + \log\left(1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_1}\right) - \log\left(\frac{(1 + \mathsf{INR}_1 + \mathsf{SNR}_1)(2\mathsf{INR}_1 + \mathsf{SNR}_2)}{4\mathsf{INR}_1}\right) \\
&\leq \log\left(\frac{1 + \mathsf{SNR}_2}{2 + \mathsf{SNR}_2}\right) + \log\left(\frac{1 + \mathsf{INR}_1 + \mathsf{SNR}_1}{1 + \mathsf{INR}_1}\right) + \log\left(\frac{4\mathsf{INR}_1}{1 + \mathsf{INR}_1 + \mathsf{SNR}_1}\right) \\
&< \log\left(\frac{4\mathsf{INR}_1}{1 + \mathsf{INR}_1}\right) \\
&= 2
\end{aligned}$$

while for (3.41) we use the upper bound (3.11) to obtain in a similar way:

$$\Delta^{(B)}_{R_1+R_2} = \log(1 + \mathsf{SNR}_1) + \log\left(1 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_2}\right) - \log\left(\frac{(1 + \mathsf{INR}_2 + \mathsf{SNR}_2)(2\mathsf{INR}_2 + \mathsf{SNR}_1)}{4\mathsf{INR}_2}\right)$$

$$< \quad 2. \tag{3.45}$$

Finally, for (3.42) we use the upper bound (3.12) to get:

$$\begin{aligned}
\Delta^{(C)}_{R_1+R_2} &= \log\left(1+\mathsf{INR}_1+\frac{\mathsf{SNR}_1}{1+\mathsf{INR}_2}\right)+\log\left(1+\mathsf{INR}_2+\frac{\mathsf{SNR}_2}{1+\mathsf{INR}_1}\right)\\
&\quad -\log\left(\frac{(\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1)(\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2)}{4\mathsf{INR}_1\mathsf{INR}_2}\right)\\
&< \log\left(1+\mathsf{INR}_1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\right)+\log\left(1+\mathsf{INR}_2+\frac{\mathsf{SNR}_2}{\mathsf{INR}_1}\right)\\
&\quad -\log\left(\frac{(\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1)(\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2)}{4\mathsf{INR}_1\mathsf{INR}_2}\right)\\
&= \log\left(\frac{\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1}{\mathsf{INR}_2}\right)+\log\left(\frac{\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2}{\mathsf{INR}_1}\right)\\
&\quad -\log\left(\frac{(\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1)(\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2)}{4\mathsf{INR}_1\mathsf{INR}_2}\right)\\
&= 2.
\end{aligned} \tag{3.46}$$

If follows that regardless of whether (3.40), (3.41), or (3.42) is active, there is a matching sum rate upper bound with a gap $\Delta_{R_1+R_2} < 2$.

It remains to compute the gaps for the lines of slope $-1/2$ and -2. We have:

$$\begin{aligned}
\Delta_{2R_1+R_2} &= \log(1+\mathsf{SNR}_1+\mathsf{INR}_1)+\log\left(1+\mathsf{INR}_2+\frac{\mathsf{SNR}_2}{1+\mathsf{INR}_1}\right)+\log\left(1+\frac{\mathsf{SNR}_1}{1+\mathsf{INR}_2}\right)\\
&\quad -\log\left(\frac{(1+\mathsf{INR}_1+\mathsf{SNR}_1)(2\mathsf{INR}_2+\mathsf{SNR}_1)(\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2)}{8\mathsf{INR}_1\mathsf{INR}_2}\right)\\
&< \log\left(\frac{\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2}{\mathsf{INR}_1}\right)+\log\left(\frac{1+\mathsf{INR}_2+\mathsf{SNR}_1}{1+\mathsf{INR}_2}\right)\\
&\quad +\log\left(\frac{2\mathsf{INR}_2}{2\mathsf{INR}_2+\mathsf{SNR}_1}\right)+\log\left(\frac{4\mathsf{INR}_1}{\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2}\right)\\
&= \log(4)+\log\left(\frac{1+\mathsf{INR}_2+\mathsf{SNR}_1}{2\mathsf{INR}_2+\mathsf{SNR1}}\right)+\log\left(\frac{2\mathsf{INR}_2}{1+\mathsf{INR}_2}\right)\\
&\le 3.
\end{aligned}$$

In a similar way, we obtain $\Delta_{R_1+2R_2} < 3$.

In summary, the region $\mathcal{R}_{HK}(1,1)$ satisfies:

$$\Delta_{R_1} < \begin{cases} 1 \text{ if } 2\mathsf{INR}_2 \ge \mathsf{SNR}_1 \\ 2 \text{ if } 2\mathsf{INR}_2 < \mathsf{SNR}_1 \end{cases}$$

$$\Delta_{R_2} < \begin{cases} 1 \text{ if } 2\mathsf{INR}_1 \geq \mathsf{SNR}_2 \\ 2 \text{ if } 2\mathsf{INR}_1 < \mathsf{SNR}_2 \end{cases}$$

$$\Delta_{R_1+R_2} < 2$$

$$\Delta_{2R_1+R_2} < 3$$

$$\Delta_{R_1+2R_2} < 3.$$

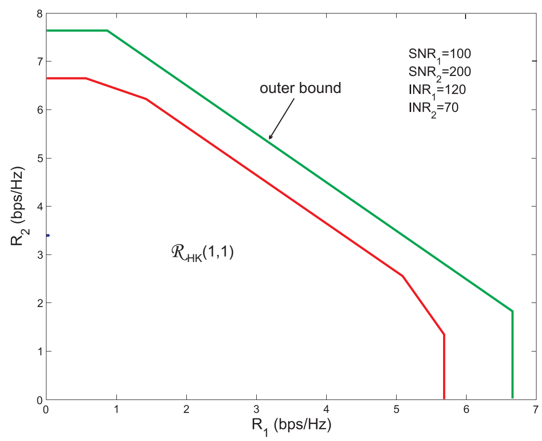


Figure 3.2: Comparison of Han-Kobayashi achievable region $\mathcal{R}_{HK}(1,1)$ and outer bound of Theorem 3. The achievable region is within one bit/s/Hz of the outer bound.

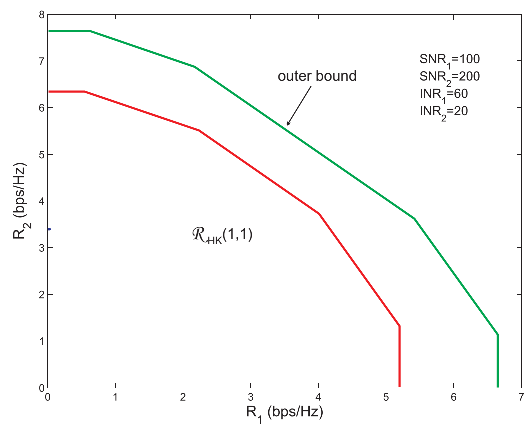


Figure 3.3: Comparison of Han-Kobayashi achievable region $\mathcal{R}_{HK}(1,1)$ and outer bound of Theorem 3. The achievable region fails to satisfy the one bit/s/Hz condition on the vertical and horizontal boundaries.

We can see that when $2\mathsf{INR}_2 \geq \mathsf{SNR}_1$ and $2\mathsf{INR}_1 \geq \mathsf{SNR}_2$, $\mathcal{R}_{HK}(1,1)$ satisfies the sufficient conditions (3.35) for being within one bit/s/Hz of capacity (see Figure 3.2).

However, when $2\mathsf{INR}_2 < \mathsf{SNR}_1$ or $2\mathsf{INR}_1 < \mathsf{SNR}_2$, the gap between the achievable region $\mathcal{R}_{HK}(1,1)$ and the outer bound may be larger than one bit/s/Hz on the vertical and horizontal boundaries, but in any case, the gap is smaller than two bits/s/Hz (see Figure 3.3).

From the above discussion, it follows that the $HK(1,1)$ scheme does not achieve large enough R_1 when $2\mathsf{INR}_2 < \mathsf{SNR}_1$, or large enough R_2 when $2\mathsf{INR}_1 < \mathsf{SNR}_2$ to be within one bit/s/Hz of the capacity region. However, we can increase the maximum achievable R_1 (R_2) rate by considering schemes that assign more power to the private message of user 1 (2), i.e. $k_1 > 1$ ($k_2 > 1$). In particular, we can assign all the available power P_1 (P_2) to P_{u1} (P_{u2}) by choosing $k_1 = \mathsf{INR}_2$

($k_2 = \mathsf{INR}_1$). The following lemma shows that by time sharing between the schemes $HK(1,1)$, $HK(\mathsf{INR}_2, 1)$, and $HK(1, \mathsf{INR}_1)$ we can obtain an achievable rate region that is within one bit/s/Hz of the capacity region.

Lemma 2 *When* $\mathsf{INR}_1 \geq 1$ *and* $\mathsf{INR}_2 \geq 1$*, the following achievable region*

$$\textit{Convex hull} \ \ \{\mathcal{R}_{HK}(1,1) \cup \mathcal{R}_{HK}(\mathsf{INR}_2, 1) \cup \mathcal{R}_{HK}(1, \mathsf{INR}_1)\}$$

is within one bit/s/Hz of the capacity region of the Gaussian interference channel with weak interference.

Proof: We have shown above that when $2\mathsf{INR}_2 \geq \mathsf{SNR}_1$ and $2\mathsf{INR}_1 \geq \mathsf{SNR}_2$, the achievable region $\mathcal{R}_{HK}(1,1)$ satisfies the sufficient conditions (3.35), and as a result, is within one bit/s/Hz of capacity. Also, when $2\mathsf{INR}_2 < \mathsf{SNR}_1$ and/or $2\mathsf{INR}_1 < \mathsf{SNR}_2$ the gaps $\Delta_{R_1+R_2}$, $\Delta_{2R_1+R_2}$, and $\Delta_{R_1+2R_2}$ still satisfy (3.35), so we only need to reduce the gaps Δ_{R_1} and/or Δ_{R_2}.

When $2\mathsf{INR}_2 < \mathsf{SNR}_1$, we will show that we can shift the vertical boundary of $\mathcal{R}_{HK}(1,1)$ to the line $R_1 = \log(1 + \mathsf{SNR}_1/2)$ by time sharing between the scheme $HK(\mathsf{INR}_2, 1)$ and $HK(1,1)$. This will reduce the gap Δ_{R_1} to a number strictly smaller than 1.

By taking $k_1 = \mathsf{INR}_2$ and $k_2 = 1$ in (3.34) and removing redundant bounds we obtain the region $\mathcal{R}_{HK}(\mathsf{INR}_2, 1)$ as the set of rate pairs (R_1, R_2) that satisfy:

$$\begin{aligned}
R_1 &\leq \log\left(1 + \frac{\mathsf{SNR}_1}{2}\right) \\
R_2 &\leq \log\left(1 + \frac{\mathsf{SNR}_2}{1+\mathsf{INR}_2}\right) \ \text{ if } \ \mathsf{INR}_1 + \mathsf{INR}_1\mathsf{INR}_2 \geq \mathsf{SNR}_2 \\
R_2 &\leq \log\left(\frac{(1+\mathsf{INR}_1)(\mathsf{INR}_1 + \mathsf{INR}_1\mathsf{INR}_2 + \mathsf{SNR}_2)}{2(1+\mathsf{INR}_2)\mathsf{INR}_1}\right) \ \text{ if } \ \mathsf{INR}_1 + \mathsf{INR}_1\mathsf{INR}_2 < \mathsf{SNR}_2 \\
R_1 + R_2 &\leq \log\left(\frac{(\mathsf{INR}_1 + \mathsf{INR}_1\mathsf{INR}_2 + \mathsf{SNR}_2)(1 + \mathsf{INR}_1 + \mathsf{SNR}_1)}{2(1+\mathsf{INR}_2)\mathsf{INR}_1}\right)
\end{aligned} \tag{3.47}$$

Note that there are no boundary lines with slope $-1/2$ or -2 in $\mathcal{R}(\mathsf{INR}_2, 1)$.

Consider the point B with coordinates (R_1^B, R_2^B) which maximizes R_2 within $\mathcal{R}_{HK}(\mathsf{INR}_2, 1)$ subject to $R_1 = \log(1+\mathsf{SNR}_1/2)$ (see Figure 3.4). By setting $R_1 = \log(1+\mathsf{SNR}_1/2)$, and finding the smallest upper bound on R_2 from the three resulting bounds in (3.47) we obtain:

$$R_2^B = \log\left(\frac{(1+\mathsf{INR}_1+\mathsf{SNR}_1)(\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2)}{\mathsf{INR}_1(1+\mathsf{INR}_2)(2+\mathsf{SNR}_1)}\right).$$

Note that the point B is on the intersection of the vertical bound and the bound with slope -1 of $\mathcal{R}_{HK}(\mathsf{INR}_2, 1)$, and therefore, the sum rate bound is always active.

We will show that the region obtained by taking the convex hull of $\mathcal{R}_{HK}(1,1) \cup \{(R_1^B, R_2^B)\}$ is larger than the region obtained by shifting the vertical bound of $\mathcal{R}_{HK}(1,1)$ to the line $R_1 = \log(1+\mathsf{SNR}_1/2)$.

In $\mathcal{R}_{HK}(1,1)$ consider the point A where the vertical bound intersects the bound with slope -2 (see Figure 3.4). This point A has coordinates:

$$\begin{aligned} R_1^A &= \log\left(\frac{(1+\mathsf{INR}_2)(2\mathsf{INR}_2+\mathsf{SNR}_1)}{4\mathsf{INR}_2}\right) \\ R_2^A &= \log\left(\frac{2\mathsf{INR}_2(1+\mathsf{INR}_1+\mathsf{SNR}_1)(\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2)}{\mathsf{INR}_1(1+\mathsf{INR}_2)^2(2\mathsf{INR}_2+\mathsf{SNR}_1)}\right) \end{aligned}$$

and achieves a sum rate:

$$R_1^A + R_2^A = \log\left(\frac{(1+\mathsf{INR}_1+\mathsf{SNR}_1)(\mathsf{INR}_1+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_2)}{2\mathsf{INR}_1(1+\mathsf{INR}_2)}\right)$$

which satisfies with equality the sum rate bound of (3.47).

If point A is part of $\mathcal{R}_{HK}(1,1)$ (which requires the $2R_1+R_2$ bound of $\mathcal{R}_{HK}(1,1)$ to be active) the segment $A-B$ obtained by time sharing between points A and B is strictly above the segment $A-C$ obtained when the region $\mathcal{R}_{HK}(1,1)$ is extended by shifting the vertical bound to the line $R_1 = \log(1+\mathsf{SNR}_1/2)$ (see Figure 3.4).

If on the other hand point A is outside $\mathcal{R}_{HK}(1,1)$ (which requires the $2R_1+R_2$ bound of $\mathcal{R}_{HK}(1,1)$ to be inactive), the maximum sum rate achievable in $\mathcal{R}_{HK}(1,1)$ is smaller than $R_1^A+R_2^A = R_1^B+R_2^B$. In addition, in this case the sum rate bound of $\mathcal{R}_{HK}(1,1)$ is active[4]. Therefore, point B is above the corner point C obtained by intersecting the sum rate bound of $\mathcal{R}_{HK}(1,1)$ with the vertical line

[4]This can be checked by comparing the sum of the individual R_1, R_2 bounds (i.e. (3.37) and (3.38) or (3.39)), and either (3.41) or (3.42).

$R_1 = \log(1 + \mathsf{SNR}_1/2)$ (see Figure 3.5). It follows that the region obtained by taking the convex hull of $\mathcal{R}_{HK}(1,1) \cup \{(R_1^B, R_2^B)\}$ is larger than the region obtained by shifting the vertical bound of $\mathcal{R}_{HK}(1,1)$ to the line $R_1 = \log(1 + \mathsf{SNR}_1/2)$.

A similar analysis can be done for the case when $2\mathsf{INR}_1 < \mathsf{SNR}_2$. We can time share between the strategies $HK(1,1)$ and $HK(1, \mathsf{INR}_1)$ to obtain a region that is the convex hull of $\mathcal{R}_{HK}(1,1) \cup \mathcal{R}_{HK}(1, \mathsf{INR}_1)$, which is larger than the region obtained from $\mathcal{R}_{HK}(1,1)$ by shifting the horizontal bound on R_2 to the line $R_2 = \log(1 + \mathsf{SNR}_2/2)$.

If follows that the region defined in the statement of the lemma is within one bit/s/Hz of the capacity region (see Figure 3.6).

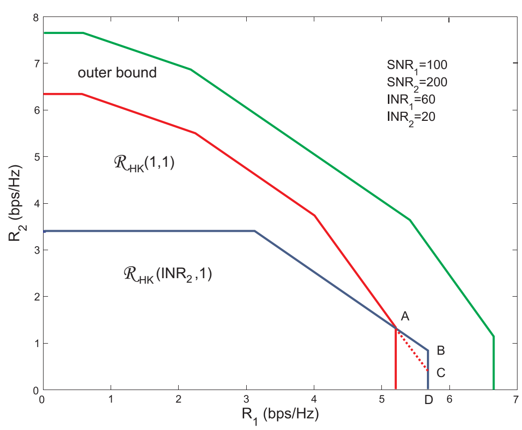


Figure 3.4: Achievable regions $\mathcal{R}_{HK}(1,1)$ and $\mathcal{R}_{HK}(\mathsf{INR}_2, 1)$, and outer bound for a case where $2\mathsf{INR}_2 < \mathsf{SNR}_1$ and the $2R_1 + R_2$ bound of $\mathcal{R}_{HK}(1,1)$ is active.

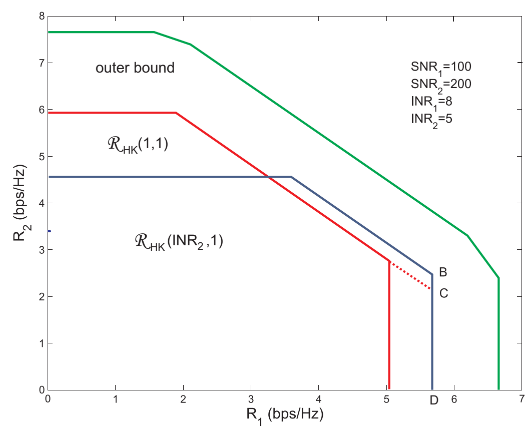


Figure 3.5: Achievable regions $\mathcal{R}_{HK}(1,1)$ and $\mathcal{R}_{HK}(\mathsf{INR}_2, 1)$, and outer bound for a case where $2\mathsf{INR}_2 < \mathsf{SNR}_1$ and the $2R_1 + R_2$ bound of $\mathcal{R}_{HK}(1,1)$ is not active.

■

Lemma 2 applies only when $\mathsf{INR}_1 \geq 1$ and $\mathsf{INR}_2 \geq 1$. If either $\mathsf{INR}_1 < 1$ or $\mathsf{INR}_2 < 1$ the interference is so weak that we cannot choose the private message power so that the interference created by the private message is at noise level, i.e. we cannot implement the scheme $HK(1,1)$. The closest we can get to the $HK(1,1)$ scheme is to set $P_{u1} = P_1$ (i.e. $k_1 = \mathsf{INR}_2$) when $\mathsf{INR}_2 < 1$ or $P_{u2} = P_2$ (i.e. $k_2 = \mathsf{INR}_1$) when $\mathsf{INR}_1 < 1$. The following lemma shows that this modification of

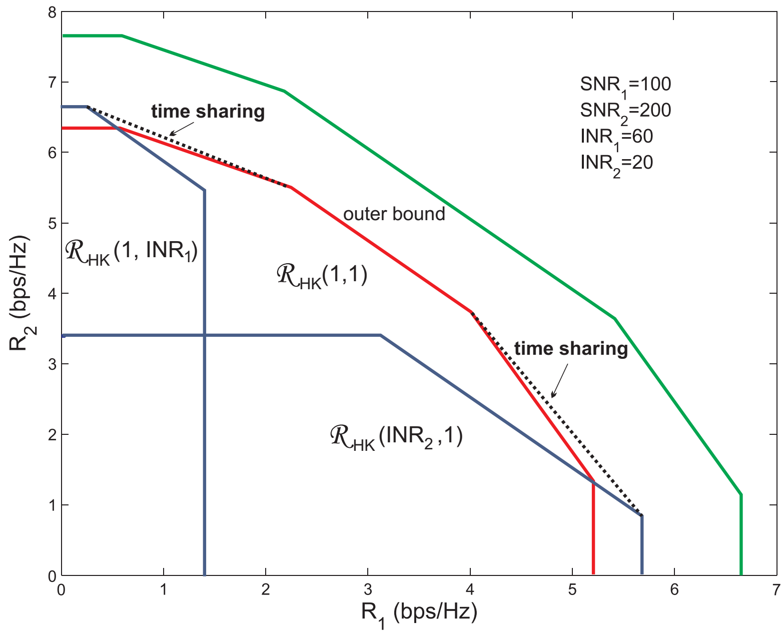


Figure 3.6: Region convex hull $\{\mathcal{R}_{HK}(1,1) \cup \mathcal{R}_{HK}(\mathsf{INR}_2, 1) \cup \mathcal{R}_{HK}(1, \mathsf{INR}_1)\}$ and outer bound when the $2R_1 + R_2$ and $R_1 + 2R_2$ bounds in $\mathcal{R}_{HK}(1,1)$ are active.

our simple scheme is enough to achieve a rate region within one bit/s/Hz of capacity.

Lemma 3 *When* $\mathsf{INR}_1 < 1$ *and* $\mathsf{INR}_2 \geq 1$*, the following achievable region*

$$\textit{Convex hull} \;\; \{\mathcal{R}_{HK}(1, \mathsf{INR}_1) \cup \mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)\}$$

is within one bit/s/Hz of the capacity region of the Gaussian interference channel with weak interference.

Similarly, when $\mathsf{INR}_1 \geq 1$ *and* $\mathsf{INR}_2 < 1$*, the following achievable region*

$$\textit{Convex hull} \;\; \{\mathcal{R}_{HK}(\mathsf{INR}_2, 1) \cup \mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)\}$$

is within one bit/s/Hz of the capacity region of the Gaussian interference channel with weak interference.

Proof: We will prove only the first part of the lemma. The same argument can be used to prove the second part.

Setting $k_1 = 1$ and $k_2 = \mathsf{INR}_1$ in (3.34) and removing redundant bounds we obtain that $\mathcal{R}_{HK}(1, \mathsf{INR}_1)$ is the set of rate pairs (R_1, R_2) that satisfies:

$$R_1 \leq \log\left(1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_1}\right) \text{ if } \mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 \geq \mathsf{SNR}_1 \tag{3.48}$$

$$R_1 \leq \log\left(\frac{(1 + \mathsf{INR}_2)(\mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 + \mathsf{SNR}_1)}{2(1 + \mathsf{INR}_1)\mathsf{INR}_2)}\right) \text{ if } \mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 < \mathsf{SNR}_1 \tag{3.49}$$

$$R_2 \leq \log\left(1 + \frac{\mathsf{SNR}_2}{2}\right) \tag{3.50}$$

$$R_1 + R_2 \leq \log\left(\frac{(\mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 + \mathsf{SNR}_1)(1 + \mathsf{INR}_2 + \mathsf{SNR}_2)}{2(1 + \mathsf{INR}_1)\mathsf{INR}_2}\right). \tag{3.51}$$

Subtracting (3.48) from the upper bound (3.8), subtracting (3.50) from the upper bound (3.9), and subtracting (3.51) from the upper bound (3.11), and noting that $\mathsf{INR}_1 < 1$ we obtain that:

$$\Delta_{R_1} < 1 \text{ if } \mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 \geq \mathsf{SNR}_1$$

$$\Delta_{R_2} < 1$$

$$\Delta_{R_1+R_2} < 2.$$

It follows that when $\mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 \geq \mathsf{SNR}_1$, the region $\mathcal{R}_{HK}(1, \mathsf{INR}_1)$ is within one bit/s/Hz of the capacity region (see Figure 3.7).

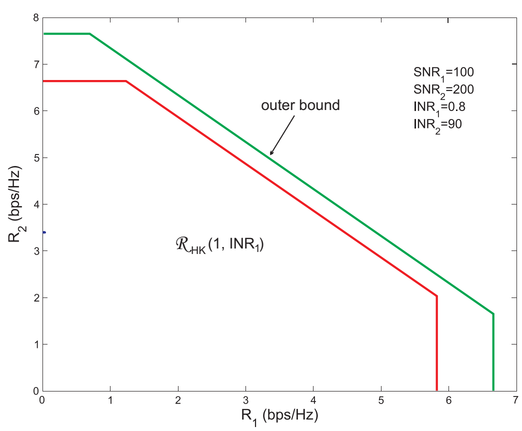


Figure 3.7: Han-Kobayashi achievable region $\mathcal{R}_{HK}(1, \mathsf{INR}_1)$ and capacity outer bound for a case where $\mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 \geq \mathsf{SNR}_1$.

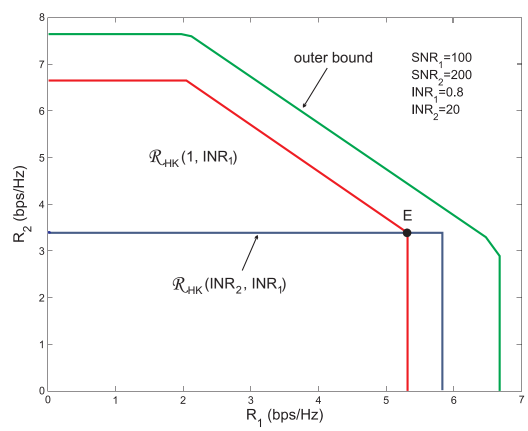


Figure 3.8: Han-Kobayashi achievable region $\mathcal{R}_{HK}(1, \mathsf{INR}_1) \cup \mathcal{R}(\mathsf{INR}_2, \mathsf{INR}_1)$ and capacity outer bound for a case where $\mathsf{INR}_2 + \mathsf{INR}_1\mathsf{INR}_2 < \mathsf{SNR}_1$.

If $\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2 < \mathsf{SNR}_1$ the gap Δ_{R_1} can exceed one, so in this case we extend the achievable rate region by considering the strategy $HK(\mathsf{INR}_2, \mathsf{INR}_1)$. Noting that this strategy consists of sending private messages at full power (and no common messages) we immediately get that the achievable region $\mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)$ consists of the set of rate pairs (R_1, R_2) that satisfy:

$$\begin{aligned} R_1 &\leq \log\left(1+\frac{\mathsf{SNR}_1}{1+\mathsf{INR}_1}\right) \\ R_2 &\leq \log\left(1+\frac{\mathsf{SNR}_2}{1+\mathsf{INR}_2}\right). \end{aligned}$$

Noting that $\mathsf{INR}_1 < 1$ it follows that $\Delta_{R_1} < 1$ for $\mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)$.

We will now show that when $\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2 < \mathsf{SNR}_1$, the corner point E of region $\mathcal{R}_{HK}(1, \mathsf{INR}_1)$ where the boundary of slope (-1) intersects the vertical boundary falls on the horizontal boundary of $\mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)$ (see Figure 3.8). This fact implies that every point on the boundary of the region $\mathcal{R}_{HK}(1, \mathsf{INR}_1) \cup \mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)$ is within one bit/s/Hz of outer bound. Toward this end we subtract (3.49) from (3.51) to obtain:

$$R_2^E = \log\left(\frac{1+\mathsf{INR}_2+\mathsf{SNR}_2}{1+\mathsf{INR}_2}\right) = \log\left(1+\frac{\mathsf{SNR}_2}{1+\mathsf{INR}_2}\right)$$

where we used the fact that $\mathsf{INR}_2 \geq 1$ to get that R_2^E is not determined by (3.50). This shows that point E falls in the horizontal boundary of $\mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)$, concluding the proof. ■

Finally, for the case when $\mathsf{INR}_1 < 1$ and $\mathsf{INR}_2 < 1$ we have the following result.

Lemma 4 *When* $\mathsf{INR}_1 < 1$ *and* $\mathsf{INR}_2 < 1$*, the following achievable region*

$$\mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)$$

is within one bit/s/Hz of the capacity region of the Gaussian interference channel with weak interference.

Proof: As in the proof of Lemma 3 have that the region $\mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)$ is given by the set of rate pairs (R_1, R_2) that satisfy:

$$R_1 \leq \log\left(1+\frac{\mathsf{SNR}_1}{1+\mathsf{INR}_1}\right) \tag{3.52}$$

$$R_2 \quad \leq \quad \log\left(1 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_2}\right). \tag{3.53}$$

Subtracting (3.52) from (3.8) and subtracting (3.53) from (3.9) we obtain:

$$\Delta_{R_1} \quad < \quad 1$$

$$\Delta_{R_2} \quad < \quad 1.$$

Therefore, the region $\mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)$ is within one bit/s/Hz of the capacity region (see figure 3.9).

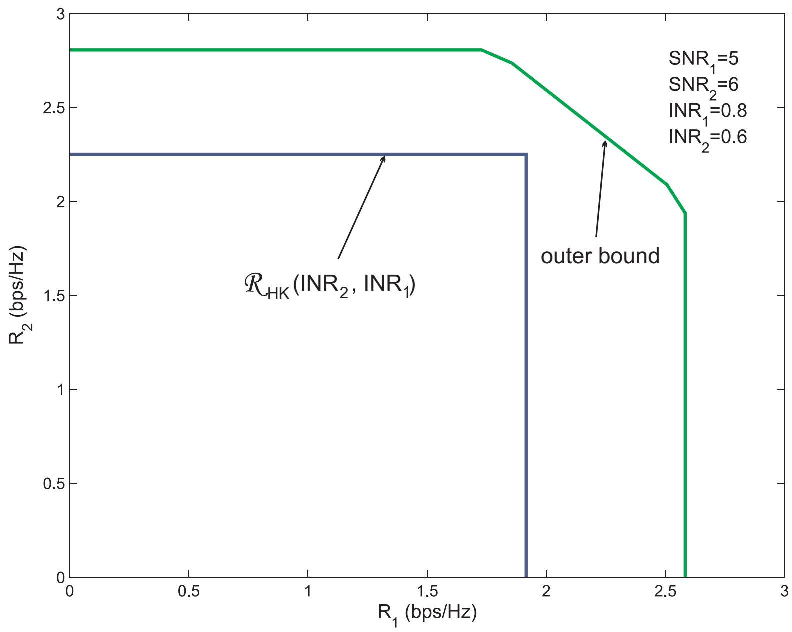


Figure 3.9: Achievable region $\mathcal{R}_{HK}(\mathsf{INR}_2, \mathsf{INR}_1)$ and capacity outer bound.

■

Combining the three previous lemmas we reach the main result of this section.

Theorem 4 *The achievable region*

Convex hull $\left\{\mathcal{R}_{HK}\left(\min(1, \mathsf{INR}_2), \min(1, \mathsf{INR}_1)\right) \cup \mathcal{R}_{HK}\left(\mathsf{INR}_2, \min(1, \mathsf{INR}_1)\right) \cup \mathcal{R}_{HK}\left(\min(1, \mathsf{INR}_2), \mathsf{INR}_1\right)\right\}$

is within one bit/s/Hz of the capacity region of the Gaussian interference channel with weak interference.

This theorem has two important implications for the Gaussian interference channel with weak interference:

- The outer bound of Theorem 3 is tight to within one bit/s/Hz.
- Fairly simple Han-Kobayashi schemes provide performance within one bit/s/Hz of capacity.

These remarks imply that from the practical point of view, especially when operating at moderate or large rates, there is not much to be gained by using more complicated communication schemes than the simple ones proposed here. However, in some cases, it is possible to improve upon our simple schemes with a performance gain of a fraction of a bit/s/Hz. For example, it is possible to approach the single user bound (i.e. point to point AWGN rate) of one user by increasing the private message power of that user, while decreasing the private message power of the other user.

3.4 Generalized degrees of freedom

In Section 2.5 we obtained the generalized degrees of freedom of the symmetric Gaussian interference channel for the symmetric rate point, d_{sym}. As we argued, the generalized degrees of freedom characterize to first order the capacity of the channel, and provide a fundamental measure of the resources available for communication. In this section we use the results of Sections 3.1 and 3.3 to characterize the generalized degrees of freedom for the complete capacity region of the Gaussian interference channel in weak interference.

Since the degrees of freedom depend on the way in which $\mathsf{SNR}_1, \mathsf{SNR}_2, \mathsf{INR}_1$, and INR_2 scale, we define[5]:

$$\begin{aligned} \alpha_1 &:= \frac{\log(\mathsf{INR}_2)}{\log(\mathsf{SNR}_1)} \\ \alpha_2 &:= \frac{\log(\mathsf{INR}_1)}{\log(\mathsf{SNR}_2)} \\ \beta &:= \frac{\log(\mathsf{SNR}_2)}{\log(\mathsf{SNR}_1)} \end{aligned} \tag{3.54}$$

[5]We assume that $\mathsf{SNR}_1, \mathsf{SNR}_2, \mathsf{INR}_1, \mathsf{INR}_2 > 1$ since we will let all these parameters to go to infinity. With this assumption $\alpha_1, \alpha_2, \beta > 0$.

which we can use to parameterize the capacity region $C(\alpha_1, \alpha_2, \beta, \mathsf{SNR}_1)$. Note that the weak interference conditions $\mathsf{INR}_1 < \mathsf{SNR}_2$ and $\mathsf{INR}_2 < \mathsf{SNR}_1$ imply that $\alpha_1 < 1$ and $\alpha_2 < 1$.

We define the degrees of freedom capacity region of the interference channel as:

$$D(\alpha_1, \alpha_2, \beta) := \lim_{\mathsf{SNR}_1 \to \infty} \frac{C(\alpha_1, \alpha_2, \beta, \mathsf{SNR}_1)}{\log \mathsf{SNR}_1}. \tag{3.55}$$

The region D is composed of degrees of freedom pairs (d_1, d_2) that characterize how the achievable rate pairs (R_1, R_2) scale with $\log \mathsf{SNR}_1$ for fixed α_1, α_2, and β.

Since D is defined asymptotically dividing the rates in the capacity region by $\log \mathsf{SNR}_1$, we can use either the inner or outer bounds to the capacity region, since the difference between them is bounded and goes to zero when dividing it by $\log \mathsf{SNR}$ for $\mathsf{SNR} \to \infty$.

We can approximate the bounds of Theorem 3 for large $\mathsf{SNR}_1, \mathsf{SNR}_2, \mathsf{INR}_1$, and INR_2 in the following way:

$$\begin{aligned}
R_1 &\lesssim \log(\mathsf{SNR}_1) \\
R_2 &\lesssim \log(\mathsf{SNR}_2) \\
R_1 + R_2 &\lesssim \min\Big\{ \log(\mathsf{SNR}_2) + [\log(\mathsf{SNR}_1) - \log(\mathsf{INR}_1)]^+, \log(\mathsf{SNR}_1) + [\log(\mathsf{SNR}_2) - \log(\mathsf{INR}_2)]^+, \\
&\quad \max\{\log(\mathsf{INR}_1), \log(\mathsf{SNR}_1) - \log(\mathsf{INR}_2)\} + \max\{\log(\mathsf{INR}_2), \log(\mathsf{SNR}_2) - \log(\mathsf{INR}_1)\}\Big\} \\
2R_1 + R_2 &\lesssim \max\{\log(\mathsf{SNR}_1), \log(\mathsf{INR}_1)\} + \max\{\log(\mathsf{INR}_2), \log(\mathsf{SNR}_2) - \log(\mathsf{INR}_1)\} + \log(\mathsf{SNR}_1) \\
&\quad - \log(\mathsf{INR}_2) \\
R_1 + 2R_2 &\lesssim \max\{\log(\mathsf{SNR}_2), \log(\mathsf{INR}_2)\} + \max\{\log(\mathsf{INR}_1), \log(\mathsf{SNR}_1) - \log(\mathsf{INR}_2)\} + \log(\mathsf{SNR}_2) \\
&\quad - \log(\mathsf{INR}_1)
\end{aligned}$$

where $(x)^+ = \max\{x, 0\}$. Dividing these bounds by $\log \mathsf{SNR}_1$ and using (3.54) we obtain that the region $D(\alpha_1, \alpha_2, \beta)$ is given by the set of generalized degrees of freedom pairs (d_1, d_2) that satisfy:

$$\begin{aligned}
d_1 &\leq 1 \\
d_2 &\leq \beta
\end{aligned}$$

$$\begin{aligned} d_1 + d_2 &\leq \min\left\{\beta + (1-\alpha_2\beta)^+, 1 + (\beta-\alpha_1)^+,\right. \\ &\qquad \left.\max\{\alpha_2\beta, 1-\alpha_1\} + \max\{\alpha_1, \beta - \alpha_2\beta\}\right\} \\ 2d_1 + d_2 &\leq \max\{1, \alpha_2\beta\} + \max\{\alpha_1, \beta-\alpha_2\beta\} + 1 - \alpha_1 \\ d_1 + 2d_2 &\leq \max\{\beta, \alpha_1\} + \max\{\alpha_2\beta, 1-\alpha_1\} + \beta - \alpha_2\beta. \end{aligned} \tag{3.56}$$

Having derived the generalized d.o.f. region in (3.56), we can specialize it for the symmetric interference channel by setting $\alpha_1 = \alpha_2 = \alpha$, and $\beta = 1$. We obtain that $D(\alpha, \alpha, 1)$ is given by the set of d.o.f. pairs (d_1, d_2) that satisfy:

$$\begin{aligned} d_1 &\leq 1 \\ d_2 &\leq 1 \\ d_1 + d_2 &\leq \min\left\{1 + (1-\alpha)^+, 2\max\{\alpha, 1-\alpha\}\right\} \\ 2d_1 + d_2 &\leq \max\{1, \alpha\} + \max\{\alpha, 1-\alpha\} + 1 - \alpha \\ d_1 + 2d_2 &\leq \max\{1, \alpha\} + \max\{\alpha, 1-\alpha\} + 1 - \alpha. \end{aligned}$$

We can now evaluate the generalized d.o.f region for the three weak interference regimes that we studied in Chapter 2. For $0 < \alpha < 1/2$ we have the very weak interference regime, which has d.o.f region given by:

$$D(\alpha, \alpha, 1) = \{(d_1, d_2) : 0 \leq d_1 \leq 1 \ , \ 0 \leq d_2 \leq 1 \ , \ d_1 + d_2 \leq 2 - 2\alpha\}.$$

Figure 3.10 shows how this region varies for different values of $\alpha \in (0, 1/2)$. Note that in this regime, the bounds for $2d_1 + d_2$, and $d_1 + 2d_2$ are inactive. Similarly for $1/2 < \alpha < 2/3$ we have the weaker interference regime, which has d.o.f region:

$$D(\alpha, \alpha, 1) = \{(d_1, d_2) : 0 \leq d_1 \ , \ 0 \leq d_2 \ , \ d_1 + d_2 \leq 2\alpha \ , \ 2d_1 + d_2 \leq 2 \ , \ d_1 + 2d_2 \leq 2\}$$

which is shown in Figure 3.11 for some values of $\alpha \in (1/2, 2/3)$. Finally, for $2/3 < \alpha < 1$ we have the moderate interference regime, with d.o.f. region:

$$D(\alpha, \alpha, 1) = \{(d_1, d_2) : 0 \leq d_1 \ , \ 0 \leq d_2 \ , \ d_1 + d_2 \leq 2 - \alpha \ , \ 2d_1 + d_2 \leq 2 \ , \ d_1 + 2d_2 \leq 2\}$$

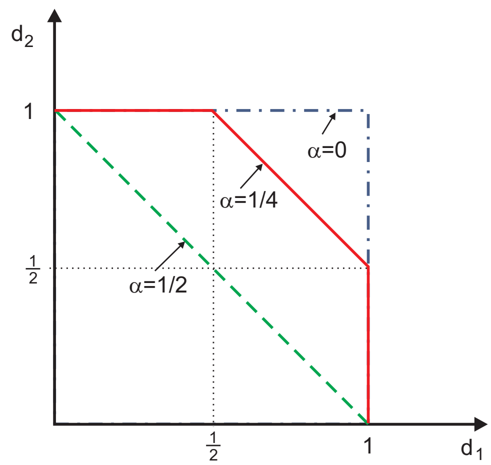


Figure 3.10: Generalized degrees of freedom region $D(\alpha, \alpha, 1)$ (symmetric channel) for the very weak interference regime ($0 < \alpha < 1/2$).

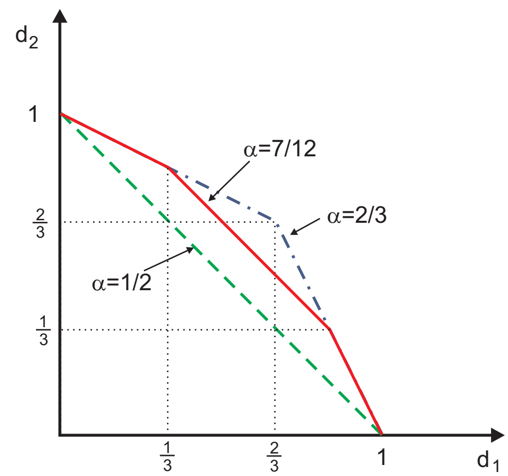


Figure 3.11: Generalized degrees of freedom region $D(\alpha, \alpha, 1)$ (symmetric channel) for the weaker interference regime ($1/2 < \alpha < 2/3$).

which is shown in Figure 3.12 for some values of $\alpha \in (2/3, 1)$. For completeness, we also present

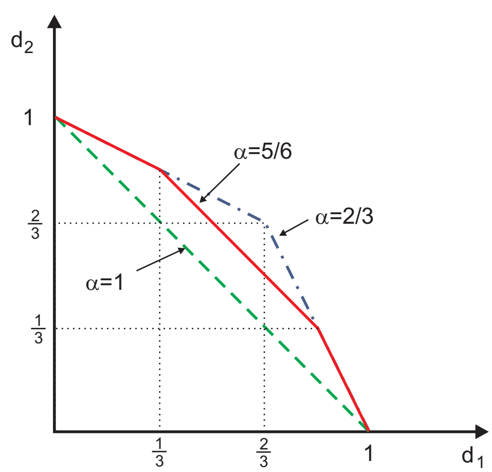


Figure 3.12: Generalized degrees of freedom region $D(\alpha, \alpha, 1)$ (symmetric channel) for the moderate interference regime ($2/3 < \alpha < 1$).

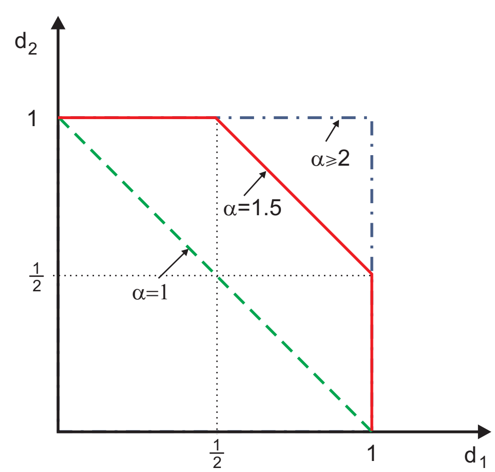


Figure 3.13: Generalized degrees of freedom region $D(\alpha, \alpha, 1)$ (symmetric channel) for the strong interference regime ($\alpha \geq 1$).

the generalized degrees of freedom region for the symmetric channel with strong interference. The characterization of this region follows from the known capacity region of the interference channel with strong interference. In this case we have:

$$D(\alpha, \alpha, 1) = \{(d_1, d_2) : 0 \leq d_1 \leq 1 \ , \ 0 \leq d_2 \leq 1 \ , \ d_1 + d_2 \leq \min\{\alpha, 2\}\}$$

which is shown in Figure 3.13.

From these figures we see that in the weaker interference regime the generalized d.o.f. region shrinks for increasing α (increasing interference), it grows for increasing α in the weaker interference regime, it shrinks again with increasing interference in the moderate interference regime, it grows again for increasing α in the strong interference regime, and finally remains constant when $\alpha \geq 2$ in the very strong interference regime. A similar effect was observed in Figure 2.5 for the symmetric capacity generalized degrees of freedom d_{sym}, which is given by the intersection of the boundary of the generalized d.o.f. region with the line $d_1 = d_2$.

Chapter 4

Gaussian Interference Channel with One-sided Strong Interference

4.1 Characterization of the capacity region

In the last chapter we analyzed the Gaussian interference channel in weak interference. The weak interference condition can be expressed as $\mathsf{INR}_1 < \mathsf{SNR}_2$ and $\mathsf{INR}_2 < \mathsf{SNR}_1$. In the weak interference case, we were able to characterize the capacity region to within one bit/s/Hz, and showed that simple Han-Kobayashi schemes could achieve any rates within one bit/s/Hz of capacity. On the other hand, the capacity of the interference channel under strong interference, i.e. $\mathsf{INR}_1 \geq \mathsf{SNR}_2$ and $\mathsf{INR}_2 \geq \mathsf{SNR}_1$, is well known [7, 29]. Therefore, it remains to consider the case where one of the users receives strong interference while the other receives weak interference. In this chapter we analyze such channel, which we call "with one-sided strong interference".

Most of ideas and techniques used in the previous chapter apply directly to the current setting. In particular, Lemma 1 was derived for a general interference channel (not necessarily Gaussian), and the achievable rate region for the Gaussian channel $\mathcal{R}_{HK}(k_1, k_2)$ given by the set of rate pairs (R_1, R_2) that satisfy (3.34) is applicable in the channel with one-sided strong interference.

However, the outer bound given in Theorem 3 needs a small revision. This is because the bound that is derived using the one-sided channel sum rate capacity needs to be changed when the corresponding one-sided channel is in strong interference. In addition, some of the bounds in Theorem 3 become redundant. We provide an outer bound for the Gaussian interference channel with one-sided strong interference in the following theorem.

Theorem 5 *For the Gaussian interference channel with one-sided strong interference, such that* $\mathsf{INR}_1 \geq \mathsf{SNR}_2$ *and* $\mathsf{INR}_2 < \mathsf{SNR}_1$*, the capacity region is contained within the set of rate pairs* (R_1, R_2) *satisfying:*

$$R_1 \leq \log(1 + \mathsf{SNR}_1) \tag{4.1}$$

$$R_2 \leq \log(1 + \mathsf{SNR}_2) \tag{4.2}$$

$$R_1 + R_2 \leq \log(1 + \mathsf{SNR}_1 + \mathsf{INR}_1) \tag{4.3}$$

$$R_1 + R_2 \leq \log(1 + \mathsf{SNR}_1) + \log\left(1 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_2}\right) \tag{4.4}$$

$$R_1 + R_2 \leq \log\left(1 + \mathsf{INR}_1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_2}\right) + \log\left(1 + \mathsf{INR}_2 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_1}\right) \tag{4.5}$$

$$R_1 + 2R_2 \leq \log(1 + \mathsf{SNR}_2 + \mathsf{INR}_2) + \log\left(1 + \mathsf{INR}_1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_2}\right) + \log\left(1 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_1}\right). \tag{4.6}$$

Similarly, when $\mathsf{INR}_2 \geq \mathsf{SNR}_1$ *and* $\mathsf{INR}_1 < \mathsf{SNR}_2$*, the capacity region is contained within the set of rate pairs* (R_1, R_2) *satisfying:*

$$R_1 \leq \log(1 + \mathsf{SNR}_1)$$

$$R_2 \leq \log(1 + \mathsf{SNR}_2)$$

$$R_1 + R_2 \leq \log(1 + \mathsf{SNR}_2 + \mathsf{INR}_2)$$

$$R_1 + R_2 \leq \log(1 + \mathsf{SNR}_2) + \log\left(1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_1}\right)$$

$$R_1 + R_2 \leq \log\left(1 + \mathsf{INR}_2 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_1}\right) + \log\left(1 + \mathsf{INR}_1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_2}\right)$$

$$2R_1 + R_2 \leq \log\left(1 + \mathsf{SNR}_1 + \mathsf{INR}_1\right) + \log\left(1 + \mathsf{INR}_2 + \frac{\mathsf{SNR}_2}{1 + \mathsf{INR}_1}\right) + \log\left(1 + \frac{\mathsf{SNR}_1}{1 + \mathsf{INR}_2}\right).$$

Proof: We base the proof of this theorem on the proof of Theorem 3. We will only prove the first part, since the proof of the second part follows along the same lines.

The bounds (4.1) and (4.1) are the single user bounds.

To derive (4.3) we convert the interference channel into a one-sided channel by giving side information x_1 to receiver 2, which he can use to completely remove the interference from transmitter 1. The resulting one-sided interference channel is in strong interference, which means that receiver 1 can decode the messages sent by both transmitters. In this case, the sum rate of the channel is upper bounded by the sum rate of the multiple access channel between transmitters 1 and 2, and receiver 1. This sum rate bound is (4.1).

The bounds (4.4),(4.5), and (4.6) are exactly (3.11),(3.12), and (3.14) of Theorem 3, which apply to the present setting with no change. ■

In order to obtain an inner bound to the capacity region we will consider Han-Kobayashi schemes as we did in Chapter 3. We will then compare the achievable performance to the outer bound of Theorem 5 to show that these schemes achieve a rate region that is within one bit/s/Hz of capacity.

In the Gaussian interference channel with one sided strong interference, the receiver that has strong interference can decode the messages sent by both transmitters. Without loss of generality consider the case where receiver 1 has strong interference (i.e. $\mathsf{INR}_1 \geq \mathsf{SNR}_2$). In this case the message sent by transmitter 2 is common information decodable by both receivers. It follows that in the $HK(k_1, k_2)$ scheme we can choose $k_2 = 0$ to set the power of the private message of user 2 to zero. In addition, we will see that setting the private message power of user 1 so that the interference created at receiver 2 is as close to the background noise level as possible achieves a rate region within one bit/s/Hz of capacity. We prove this result in the following two lemmas.

Lemma 5 *For the Gaussian interference channel with one-sided strong interference such that* $\mathsf{INR}_1 \geq$

SNR_2, $\mathsf{INR}_2 < \mathsf{SNR}_1$, *and* $\mathsf{INR}_2 > 1$, *the Han-Kobayashi rate region* $\mathcal{R}_{HK}(1,0)$ *is given by the set of rate pairs* (R_1, R_2) *that satisfy:*

$$\begin{aligned} R_1 &\leq \log\left(\frac{(1+\mathsf{INR}_2)(\mathsf{INR}_2+\mathsf{SNR}_1)}{2\mathsf{INR}_2}\right) && (4.7)\\ R_2 &\leq \log\left(1+\frac{\mathsf{SNR}_2}{2}\right) && (4.8)\\ R_1+R_2 &\leq \log\left(1+\mathsf{INR}_1+\mathsf{SNR}_1\right) && (4.9)\\ R_1+R_2 &\leq \log\left(\frac{(\mathsf{INR}_2+\mathsf{SNR}_1)(1+\mathsf{INR}_2+\mathsf{SNR}_2)}{2\mathsf{INR}_2}\right) && (4.10)\\ R_1+R_2 &\leq \log\left(\frac{(1+\mathsf{INR}_2)(\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1)}{2\mathsf{INR}_2}\right) && (4.11)\\ R_1+2R_2 &\leq \log\left(\frac{(\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1)(1+\mathsf{INR}_2+\mathsf{SNR}_2)}{2\mathsf{INR}_2}\right) && (4.12) \end{aligned}$$

and is within one bit/s/Hz of capacity.

Similarly, if $\mathsf{INR}_2 \geq \mathsf{SNR}_1$, $\mathsf{INR}_1 < \mathsf{SNR}_2$, *and* $\mathsf{INR}_1 > 1$ *the Han-Kobayashi rate region* $\mathcal{R}_{HK}(0,1)$ *is given by the set of rate pairs* (R_1, R_2) *that satisfy:*

$$\begin{aligned} R_1 &\leq \log\left(1+\frac{\mathsf{SNR}_1}{2}\right)\\ R_2 &\leq \log\left(\frac{(1+\mathsf{INR}_1)(\mathsf{INR}_1+\mathsf{SNR}_2)}{2\mathsf{INR}_1}\right)\\ R_1+R_2 &\leq \log\left(1+\mathsf{INR}_2+\mathsf{SNR}_2\right)\\ R_1+R_2 &\leq \log\left(\frac{(\mathsf{INR}_1+\mathsf{SNR}_2)(1+\mathsf{INR}_1+\mathsf{SNR}_1)}{2\mathsf{INR}_1}\right)\\ R_1+R_2 &\leq \log\left(\frac{(1+\mathsf{INR}_1)(\mathsf{INR}_1+\mathsf{INR}_2\mathsf{INR}_1+\mathsf{SNR}_2)}{2\mathsf{INR}_1}\right)\\ 2R_1+R_2 &\leq \log\left(\frac{(\mathsf{INR}_1+\mathsf{INR}_2\mathsf{INR}_1+\mathsf{SNR}_2)(1+\mathsf{INR}_1+\mathsf{SNR}_1)}{2\mathsf{INR}_1}\right) \end{aligned}$$

and is within one bit/s/Hz of capacity.

Proof: We will prove the first part of the lemma. The second part can be proved in a similar way.

The rate constraints of the region $\mathcal{R}_{HK}(1,0)$ can be obtained by setting $k_1 = 1$ and $k_2 = 0$ in (3.34). After removing redundant constraints we get the constraints given in the statement of the lemma.

In order to prove that the region $\mathcal{R}_{HK}(1,0)$ is within one bit/s/Hz of the capacity region we follow the same approach of chapter 3 and compute the gaps between the constraints (4.7)-(4.12) and the outer bound given in Theorem 5.

For (4.7) we use the upper bound (4.1) to get:

$$\begin{aligned}\Delta_{R_1} &= \log\left(1+\mathsf{SNR}_1\right) - \log\left(\frac{(1+\mathsf{INR}_2)(\mathsf{INR}_2+\mathsf{SNR}_1)}{2\mathsf{INR}_2}\right)\\ &\overset{(a)}{<} \log\left(\frac{1+\mathsf{SNR}_1}{1+\mathsf{SNR}_1}\cdot\frac{2\mathsf{INR}_2}{1+\mathsf{INR}_2}\right)\\ &\leq 1\end{aligned}$$

where (a) follows from the assumption $\mathsf{INR}_2 > 1$.

For (4.8) we use the upper bound (4.2) to get:

$$\begin{aligned}\Delta_{R_2} &= \log\left(1+\mathsf{SNR}_2\right) - \log\left(1+\frac{\mathsf{SNR}_2}{2}\right)\\ &= \log\left(2\cdot\frac{1+\mathsf{SNR}_2}{2+\mathsf{SNR}_2}\right)\\ &< 1.\end{aligned}$$

Noting that (4.9) equals (4.3) we get that when (4.9) is active, $\Delta_{R_1+R_2} = 0$.

When (4.10) is active, we use (4.4) to obtain:

$$\begin{aligned}\Delta_{R_1+R_2} &= \log\left(1+\mathsf{SNR}_1\right) + \log\left(1+\frac{\mathsf{SNR}_2}{1+\mathsf{INR}_2}\right) - \log\left(\frac{(\mathsf{INR}_2+\mathsf{SNR}_1)(1+\mathsf{INR}_2+\mathsf{SNR}_2)}{2\mathsf{INR}_2}\right)\\ &< \log\left(\frac{1+\mathsf{SNR}_1}{1+\mathsf{SNR}_1}\cdot\frac{1+\mathsf{INR}_2+\mathsf{SNR}_2}{1+\mathsf{INR}_2+\mathsf{SNR}_2}\cdot\frac{2\mathsf{INR}_2}{1+\mathsf{INR}_2}\right)\\ &\leq 1.\end{aligned}$$

If on the other hand (4.11) is active, we use (4.5) to compute:

$$\begin{aligned}\Delta_{R_1+R_2} &= \log\left(1+\mathsf{INR}_1+\frac{\mathsf{SNR}_1}{1+\mathsf{INR}_2}\right) + \log\left(1+\mathsf{INR}_2+\frac{\mathsf{SNR}_2}{1+\mathsf{INR}_1}\right)\\ &\quad - \log\left(\frac{(1+\mathsf{INR}_2)(\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1)}{2\mathsf{INR}_2}\right)\\ &\overset{(a)}{<} \log\left(1+\mathsf{INR}_1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\right) + \log\left(\frac{1+\mathsf{INR}_2}{1+\mathsf{INR}_2}+\frac{\mathsf{INR}_1}{(1+\mathsf{INR}_1)(1+\mathsf{INR}_2)}\right)\\ &\quad - \log\left(\frac{\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1}{2\mathsf{INR}_2}\right)\\ &\leq 2\end{aligned}$$

where we used the assumption $\mathsf{INR}_1 \geq \mathsf{SNR}_2$ in (a).

Therefore, regardless of which of the sum rate constraints is active in the achievable region, we obtain that $\Delta_{R_1+R_2} < 2$.

Finally, for (4.12) we use (4.6) to get:

$$\begin{aligned}
\Delta_{R_1+2R_2} &= \log\left(1+\mathsf{SNR}_2+\mathsf{INR}_2\right) + \log\left(1+\mathsf{INR}_1+\frac{\mathsf{SNR}_1}{1+\mathsf{INR}_2}\right) + \log\left(1+\frac{\mathsf{SNR}_2}{1+\mathsf{INR}_1}\right) \\
&\quad - \log\left(\frac{(\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1)(1+\mathsf{INR}_2+\mathsf{SNR}_2)}{2\mathsf{INR}_2}\right) \\
&< \log\left(1+\mathsf{INR}_1+\frac{\mathsf{SNR}_1}{\mathsf{INR}_2}\right) + \log\left(1+\frac{\mathsf{INR}_1}{1+\mathsf{INR}_1}\right) - \log\left(\frac{\mathsf{INR}_2+\mathsf{INR}_1\mathsf{INR}_2+\mathsf{SNR}_1}{2\mathsf{INR}_2}\right) \\
&\leq 2.
\end{aligned}$$

From Chapter 3 we know that if the gaps satisfy $\Delta_{R_1} < 1$, $\Delta_{R_2} < 1$, $\Delta_{R_1+R_2} < 2$ and $\Delta_{R_1+2R_2} < 3$, the corresponding Han-Kobayashi region is within one bit/s/Hz of capacity[1]. The region $\mathcal{R}_{HK}(1,0)$ satisfies these conditions, and therefore is within one bit/s/Hz of capacity. ■

Lemma 6 *For the Gaussian interference channel with one-sided strong interference such that* $\mathsf{INR}_1 \geq \mathsf{SNR}_2$, $\mathsf{INR}_2 < \mathsf{SNR}_1$, *and* $\mathsf{INR}_2 \leq 1$, *the Han-Kobayashi rate region* $\mathcal{R}_{HK}(\mathsf{INR}_2, 0)$ *is given by the set of rate pairs* (R_1, R_2) *that satisfy:*

$$R_1 \leq \log\left(1+\mathsf{SNR}_1\right) \tag{4.13}$$

$$R_2 \leq \log\left(1+\frac{\mathsf{SNR}_2}{1+\mathsf{INR}_2}\right) \tag{4.14}$$

$$R_1+R_2 \leq \log\left(1+\mathsf{INR}_1+\mathsf{SNR}_1)\right) \tag{4.15}$$

$$R_1+R_2 \leq \log\left(\frac{(1+\mathsf{SNR}_1)(1+\mathsf{INR}_2+\mathsf{SNR}_2)}{1+\mathsf{INR}_2}\right) \tag{4.16}$$

and is within one bit/s/Hz of capacity.

Similarly, if $\mathsf{INR}_2 \geq \mathsf{SNR}_1$, $\mathsf{INR}_1 < \mathsf{SNR}_2$, *and* $\mathsf{INR}_1 \leq 1$ *the Han-Kobayashi rate region* $\mathcal{R}_{HK}(0, \mathsf{INR}_1)$ *is given by the set of rate pairs* (R_1, R_2) *that satisfy:*

$$R_1 \leq \log\left(1+\frac{\mathsf{SNR}_1}{1+\mathsf{INR}_1}\right)$$

[1]The gap $\Delta_{2R_1+R_2}$ does not need to be computed when the Han-Kobayashi region does not have an explicit constraint on $2R_1+R_2$ as is the case here.

$$R_2 \leq \log(1+\mathsf{SNR}_2)$$

$$R_1+R_2 \leq \log(1+\mathsf{INR}_2+\mathsf{SNR}_2))$$

$$R_1+R_2 \leq \log\left(\frac{(1+\mathsf{SNR}_2)(1+\mathsf{INR}_1+\mathsf{SNR}_1)}{1+\mathsf{INR}_1}\right)$$

and is within one bit/s/Hz of capacity.

Proof: We will prove the first part of the lemma. The second part can be proved in a similar way.

The rate constraints of the region $\mathcal{R}_{HK}(\mathsf{INR}_2, 0)$ can be obtained by setting $k_1 = \mathsf{INR}_2$ and $k_2 = 0$ in (3.34). After removing redundant constraints we get the constraints given in the statement of the lemma.

We then compute the gaps between the constraints (4.13)-(4.16) and the outer bound given in Theorem 5.

Since (4.13) equals (4.1) we have $\Delta_{R_1} = 0$.

For (4.14) we use the bound (4.2) to get:

$$\begin{aligned}\Delta_{R_2} &= \log(1+\mathsf{SNR}_2) - \log\left(1+\frac{\mathsf{SNR}_2}{1+\mathsf{INR}_2}\right)\\ &\leq \log(1+\mathsf{SNR}_2) - \log\left(1+\frac{\mathsf{SNR}_2}{2}\right)\\ &= \log\left(\frac{1+\mathsf{SNR}_2}{2+\mathsf{SNR}_2}\right)+\log(2)\\ &< 1.\end{aligned}$$

Since the sum rate constraint (4.15) equals the upper bound (4.3), and also the sum rate constraint (4.16) equals the upper bound (4.4) we have that $\Delta_{R_1+R_2} = 0$.

If follows that $\mathcal{R}_{HK}(\mathsf{INR}_2, 0)$ is within one bit/s/Hz of capacity. ■

Figure 4.1 shows the achievable region $\mathcal{R}_{HK}(1, 0)$ and the outer bound of Theorem 3 for two examples where $\mathsf{INR}_1 \geq \mathsf{SNR}_2$, $\mathsf{INR}_2 < \mathsf{SNR}_1$, and $\mathsf{INR}_2 > 1$. Figure 4.2 shows the achievable region $\mathcal{R}_{HK}(\mathsf{INR}_2, 0)$ and the outer bound of Theorem 3 for an example where $\mathsf{INR}_1 \geq \mathsf{SNR}_2$, $\mathsf{INR}_2 \leq \mathsf{SNR}_1$, and $\mathsf{INR}_2 \leq 1$. In all cases we see that the Han-Kobayashi region is within one bit/s/Hz of the outer bound.

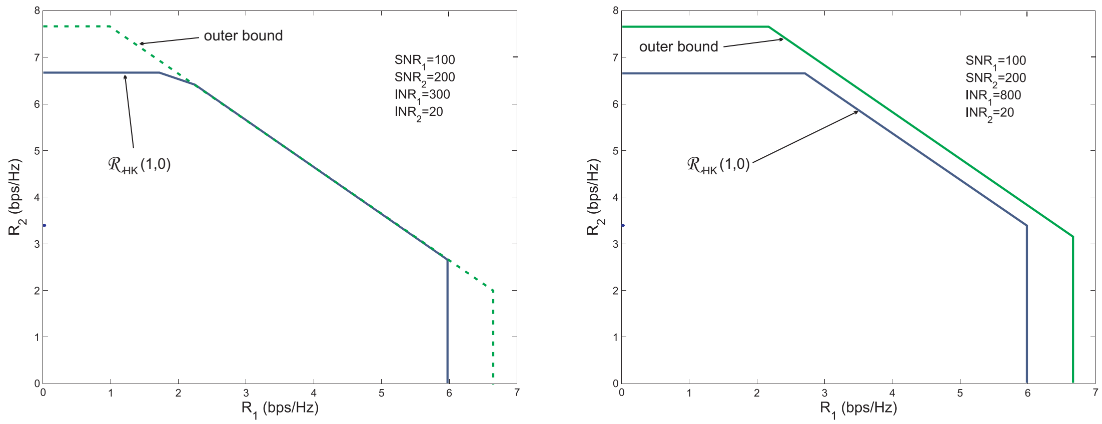


Figure 4.1: Achievable region $\mathcal{R}_{HK}(1,0)$ and outer bound of Theorem (5) for two examples where $\mathsf{INR}_1 \geq \mathsf{SNR}_2$, $\mathsf{INR}_2 < \mathsf{SNR}_1$, and $\mathsf{INR}_2 > 1$. In the left figure, the $R_1 + 2R_2$ inner bound is active. In the right figure, the $R_1 + 2R_2$ inner bound is not active.

In summary, Lemmas 5 and 6 imply together the following theorem.

Theorem 6 *The achievable region* $\mathcal{R}_{HK}\big(\min(1, \mathsf{INR}_2), 0\big)$ *is within one bit/s/Hz of the capacity region of the Gaussian interference channel with one-sided strong interference when* $\mathsf{INR}_1 \geq \mathsf{SNR}_2$, $\mathsf{INR}_2 < \mathsf{SNR}_1$.

In addition, the achievable region $\mathcal{R}_{HK}\big(0, \min(1, \mathsf{INR}_1)\big)$ *is within one bit/s/Hz of the capacity region of the Gaussian interference channel with one sided strong interference when* $\mathsf{INR}_2 \geq \mathsf{SNR}_1$, $\mathsf{INR}_1 < \mathsf{SNR}_2$.

4.2 Generalized degrees of freedom

In this section we extend the results of Section 3.4 to the Gaussian interference channel where one of the links is in strong interference while the other is in weak interference. We use the definitions of Section 3.4 for $\alpha_1, \alpha_2, \beta$, and $D(\alpha_1, \alpha_2, \beta)$.

We first consider the case when $\mathsf{INR}_1 \geq \mathsf{SNR}_2$ and $\mathsf{INR}_2 < \mathsf{SNR}_1$, which corresponds to $\alpha_2 \geq 1$ and $\alpha_1 < 1$. We use Theorems 5 and 6 to characterize the capacity region of the Gaussian

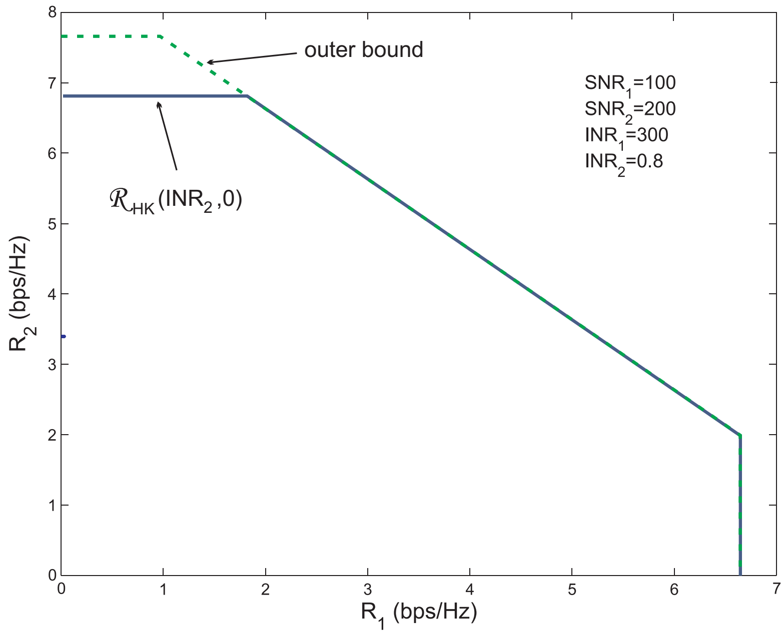


Figure 4.2: Achievable region $\mathcal{R}_{HK}(\mathsf{INR}_2, 0)$ and outer bound of Theorem (5) for an example where $\mathsf{INR}_1 \geq \mathsf{SNR}_2$, $\mathsf{INR}_2 < \mathsf{SNR}_1$, and $\mathsf{INR}_2 \leq 1$.

interference channel with one sided strong interference for large $\mathsf{SNR}_1, \mathsf{SNR}_2, \mathsf{INR}_1$, and INR_2. The capacity region is given by the set of rate pairs (R_1, R_2) that satisfy:

$$\begin{aligned} R_1 &\lesssim \log(\mathsf{SNR}_1) \\ R_2 &\lesssim \log(\mathsf{SNR}_2) \\ R_1 + R_2 &\lesssim \min\Big\{ \max\{\log(\mathsf{SNR}_1), \log(\mathsf{INR}_1)\}, \log(\mathsf{SNR}_1) + [\log(\mathsf{SNR}_2) - \log(\mathsf{INR}_2)]^+, \\ &\quad \max\{\log(\mathsf{INR}_1), \log(\mathsf{SNR}_1) - \log(\mathsf{INR}_2)\} + \max\{\log(\mathsf{INR}_2), \log(\mathsf{SNR}_2) - \log(\mathsf{INR}_1)\}\Big\} \\ R_1 + 2R_2 &\lesssim \max\{\log(\mathsf{SNR}_2), \log(\mathsf{INR}_2)\} + \max\{\log(\mathsf{INR}_1), \log(\mathsf{SNR}_1) - \log(\mathsf{INR}_2)\}. \end{aligned}$$

Dividing these bounds by $\log \mathsf{SNR}_1$ and using (3.54) we obtain that the region $D(\alpha_1, \alpha_2, \beta)$ is given by the set of generalized degrees of freedom pairs (d_1, d_2) that satisfy:

$$\begin{aligned} d_1 &\leq 1 \\ d_2 &\leq \beta \end{aligned}$$

$$\begin{aligned} d_1 + d_2 &\leq \min\Big\{\max\{1, \alpha_2\beta\}, 1 + (\beta - \alpha_1)^+, \\ &\quad \max\{\alpha_2\beta, 1 - \alpha_1\} + \max\{\alpha_1, \beta - \alpha_2\beta\}\Big\} \\ d_1 + 2d_2 &\leq \max\{\beta, \alpha_1\} + \max\{\alpha_2\beta, 1 - \alpha_1\}. \end{aligned}$$

In a similar way, for the Gaussian interference channel with one sided strong interference, such that $\mathsf{INR}_2 \geq \mathsf{SNR}_1$ and $\mathsf{INR}_1 < \mathsf{SNR}_2$, which corresponds to $\alpha_2 < 1$ and $\alpha_1 \geq 1$, the generalized degrees of freedom region $D(\alpha_1, \alpha_2, \beta)$ is given by the set of pairs (d_1, d_2) that satisfy:

$$\begin{aligned} d_1 &\leq 1 \\ d_2 &\leq \beta \\ d_1 + d_2 &\leq \min\Big\{\beta + (1 - \alpha_2\beta)^+, \max\{\beta, \alpha_1\}, \\ &\quad \max\{\alpha_2\beta, 1 - \alpha_1\} + \max\{\alpha_1, \beta - \alpha_2\beta\}\Big\} \\ 2d_1 + d_2 &\leq \max\{1, \alpha_2\beta\} + \max\{\alpha_1, \beta - \alpha_2\beta\}. \end{aligned}$$

Chapter 5

Discussion

In this part of the thesis we studied how interference affects communication and what is the best performance that can be achieved when two systems share the same portion of spectrum. We used the information theoretic model of the Gaussian interference channel and characterized to within one bit/s/Hz its capacity region.

Surprisingly, simple Han-Kobayashi schemes are enough to achieve any rate points that are within one bit/s/Hz of capacity. These schemes consist of splitting the message to be sent into two component messages: a private message to be decoded only by the intending receiver and a common message to be decoded by both receivers. These messages are sent using random Gaussian codebooks with powers appropriately chosen to satisfy the power constraint of each user. In most cases, the average power of the private codebook should be chosen such that the interference created onto the other user's receiver is as close as possible to the background noise variance. In this way, the most harmful part of the received interference (common message) can be decoded and cancelled from the received signal. The remaining interference cannot be cancelled, but being at noise level, it does not affect the communication rate substantially.

While the above simple scheme can achieve the capacity to within one bit/s/Hz, simple variations of the scheme can asymptotically achieve the symmetric capacity in some parameter regimes when the signal to noise, and interference to noise ratios grow large. More specifically, in the very weak interference regime sending the private messages with full power asymptotically achieves the symmetric capacity of the symmetric channel. In addition, in order to asymptotically achieve the symmetric capacity in the weaker interference regime, the private message power should be chosen such that the interference created at the other receiver goes to zero with $\mathsf{SNR}, \mathsf{INR} \to \infty$. The scheme that sets the private message power so that the interference created is at noise level has the nice property of performing well in a wide range of parameter values.

For the high SNR, INR regime we introduced the concept of generalized degrees of freedom to quantify to first order how the capacity region grows for increasing SNR and INR. We also identified the ratio $\alpha = \log \mathsf{INR}/\log \mathsf{SNR}$ as a measure of the interference level. The degrees of freedom view of capacity allows to clearly identify five different regimes of operation. These regimes can be defined based on the effects that interference has on capacity. Focusing on the symmetric degrees of freedom d_{sym} of the symmetric channel, we identified five regimes. The first three regimes correspond to the weak interference situation. The most surprising aspect of the behavior of capacity in these three regimes is the fact that d_{sym} is not monotonic on the interference level α. In the very weak interference regime where the interference is too weak to be partially cancelled, d_{sym} decreases with increasing α. On the other hand, in the weaker interference regime the increased interference helps decoding and subtracting part of the interfering signal, resulting in an increase of d_{sym}. However, as α continues growing we reach the moderate interference regime where d_{sym} decreases again. The capacity in the last two regimes was known from previous results. d_{sym} increases with increasing interference in the strong interference regime, and becomes constant and equal to 1 (capacity equal to the AWGN channel capacity) in the very strong interference regime. These results were extended to the complete capacity region of the interference channel with asymmetric parameters.

Even though our results are derived for the Gaussian interference channel we believe that they should easily extend to other continuous alphabet memoryless interference channels with input power constraints where the noise processes are i.i.d. with finite entropy. In essence, the capacity of the interference channel is governed by interference. The background noise has the only role of keeping the rates bounded. Therefore, while the exact capacity expressions may change with the noise distribution, we conjecture that the generalized degrees of freedom of the channel do not change. To support this conjecture, note that the capacity inner bounds that we derived for the Gaussian channel are also inner bounds for an interference channel with non-Gaussian i.i.d. noise of variance N_0 and finite entropy. This is because the inner bounds are determined by the intersection of the capacity regions of two multiple access channels, for which given a fixed noise variance, Gaussian noise is the worst type of noise. In addition, we believe that the same bounding techniques used to derive the outer bounds of the Gaussian interference channel capacity region can be used for interference channels with other finite entropy noise distributions. We finally conjecture that the gap between the inner and outer bounds would be bounded.

Part II

Spectrum Sharing in Large Networks

Chapter 6

Introduction

In this part of the thesis we study spectrum sharing situations between many systems. The model and analysis done in Part I were based on two basic assumptions:

- there are only two systems sharing the band;
- the channel gains are constant over frequency.

These assumptions allowed us to give a fundamental characterization of the achievable performance in the face of interference. Unfortunately, little is known about the capacity region of interference channels with more than two users, or when the fading gains are frequency selective. We will try to gain better understanding of the spectrum sharing problem between many systems by analyzing a limiting situation where the number of systems M grows to infinity. In addition to letting M grow large, we will make other simplifying assumptions that will lead to a more tractable analysis, at the expense of losing the information-theoretic character of the results.

When there are only two systems sharing the band, we have shown in Part I that most of the interference (essentially all the interference above the noise level) can be decoded and subtracted from the received signal. That is, splitting the transmitted message into a private and a common part, and choosing the private message power such that the interference created is at noise level is essentially optimal (to achieve capacity to within one bit/s/Hz). Consider now a situation where three systems share the band. One could try to mimic the analysis of Part I, assuming that the signal of the third system is treated as noise. Since this interfering signal may be received at a power level well above the noise power level, it may be required to increase the power of the private messages to adjust them to the new effective noise floor. But this increase in the power of the private message comes together with a decrease in the power of the common message, which is the part of the interfering signal that can be decoded and subtracted. We see that due to the presence of the third system, interference cancellation becomes more difficult. As the number of systems grows, we expect that interference cancellation may be feasible for only a couple of the strongest interferers. Therefore, since the interference received from the remaining systems cannot be cancelled, it must be treated as noise. In Chapter 7 we make the assumption that interference is treated as noise, and study the statistical properties of the aggregate received interference for a model where an infinite number of systems is distributed uniformly on the plane.

Using a standard path loss model and assuming finite and constant system density, with systems transmitting at finite power levels, we find that the probability density function of the aggregate interference is heavy tailed. This implies that the average received interference is infinite. Our study of the interference aggregation phenomenon shows that the large interference is not due to the infinite number of systems sharing the band, but results from the presence of a few interferers that can be arbitrarily close to the given system. Therefore, in order to mitigate interference and make it bounded, spectrum sharing protocols should coordinate the assignment of frequencies so that the closest neighbors use different frequency channels. We show that by coordinating the frequency allocation with the systems located within a distance r_0 of a given system one can achieve an aggregate interference with finite mean and variance. Moreover, as r_0 increases not only does the interference become smaller, but it also becomes more concentrated around its mean, which makes it more predictable. However, increasing r_0 comes at the expense of dividing the total bandwidth into more channels, resulting in a smaller available bandwidth for each system.

The model of Chapter 7 assumes a finite density of systems over the plane, and frequency flat channel gains. In practice, wireless channels often present gains that vary over frequency due to the presence of scatterers that reflect the signals and contribute to multipath fading [49]. The effect of frequency selective fading is important when the communication bandwidth is much larger than the coherence bandwidth of the channel. In a spectrum sharing situation between many systems the total bandwidth available for communication may be much larger than the bandwidth used by each individual system. This requires us to consider the influence of frequency selective fading over the spectrum sharing performance. In the broadcast channel, where a single transmitter communicates with many receivers, fading is beneficial[1]. This is because channel variations allow to exploit a phenomenon called multi-user diversity. In essence, users are scheduled in such a way that they receive information only when their channels are in a good fading state. Therefore, fading allows to obtain effective channel gains that are above their average value. In Chapter 8 we analyze whether fading can be exploited in a spectrum sharing situation between many systems and explore whether there is a *multi-system diversity* effect.

To isolate the effect of frequency selective fading from large scale path loss (whose effect is considered in Chapter 7) and to make the analysis tractable, we assume a mean path loss between systems of value 1. We further assume that the available bandwidth is divided into channels of fixed bandwidth equal to the coherence bandwidth of the channel, and that the fading gains are independent and identically distributed (i.i.d.) in all channels between all transmitters, and receivers. We finally assume that each system uses only one channel to communicate, and that it treats interference as noise. Under these assumptions we explore how the network bandwidth W scales with the number of systems M to achieve a minimum communication rate R in each system. Without frequency selective fading, the bandwidth should scale linearly with the number of systems, so linear bandwidth growth can be taken as a baseline to compare the multi-system diversity gains. By deriving lower and upper bounds on the required bandwidth that hold asymptotically with

[1] We implicitly assume that the channel gains can be fed back to the transmitter with little delay as compared to the coherence time of the channel.

large probability, we find that $W(M) \approx M/(2 \log M)$. Therefore, the required bandwidth can be reduced by a factor $\log(M)$ which can be interpreted as the multi-system diversity gain due to fading. We finally extend the result to spectrum sharing situations where each system is formed by multiple transmitters and a single receiver (multiple access channel), or multiple receivers and a single transmitter (broadcast channel).

Chapter 7

Interference Aggregation

7.1 Random network model

The observed interference at a given receiver depends on the signals transmitted at the different transmitters and on the path losses experienced by these signals. In order to have a statistical description of the interference aggregation phenomenon, we use a random network model where the different transmitters and receivers are placed randomly on the plane. To simplify the analysis we assume that each system is formed by a single transmitter-receiver pair.

In this model we assume that the receivers are distributed on the plane as a two-dimensional Poisson process with density λ, and that the transmitter of each system is uniformly distributed on a circle of radius r_{Mi} centered at the receiver i. It follows that the process formed by the location of the transmitters is also a two-dimensional Poisson process with density λ, but the transmitter and receiver processes are correlated. We assume that the channel power gain $c_{i,j}$ between the transmitter of system i and the receiver of system j is given by:

$$c_{i,j} = G_t G_r \left(\frac{c}{4\pi f} \right)^2 d_{i,j}^{-\alpha} \tag{7.1}$$

where f is the band's center frequency, c is the speed of light, and G_t and G_r are the transmit and receive antenna gains, $d_{i,j}$ is the distance between transmitter i and receiver j, and $\alpha > 2$ is the path loss exponent. Note that in our model we do not allow α to take value 2 which would correspond to communication in free space. This path loss formula allows us to adjust the path loss exponent α to different scenarios, depending on the antenna heights, range of communication, etc. We note that while (7.1) takes into account the attenuation due to large scale effects (distance, reflections on the ground, etc.) it does not model other effects such as shadowing or multipath fading (small scale effects). In cases where shadowing or multipath fading are present (7.1) can be interpreted as representing the average path loss over a period of time that is much larger than the time scale in which the small scale variations occur.

The interference generated by each transmitter is a function of the power spectral density (PSD) of its signal, so in order to characterize the interference distribution we need to make some assumption about the distribution of the transmission PSD. In this subsection we will assume that each system i uses a total power P_i drawn i.i.d. from a distribution with probability density function (pdf) $f_P(p)$ with finite mean, finite variance, and characteristic function $\phi_P(\omega)$. Also, we assume that the power spectral density of the transmitted signal is P_i/W over the bandwidth W.

The interference power received by a receiver placed at the origin is given by:

$$I = G_t G_r \left(\frac{c}{4\pi f}\right)^2 \sum_{k \in \mathcal{P}} P_k d_k^{-\alpha} \tag{7.2}$$

where $\mathcal{P}$ is the set of transmitters of the interfering systems distributed according to the Poisson process of density λ, c is the speed of light, d_k is the distance to the origin from transmitter k, and f is the center frequency of the given band. To simplify the analysis, we will normalize the interference by dividing it by the constant in front of the sum.

7.1.1 Stable laws

The aggregate interference (7.2) is a random variable that depends on the random location of the interferers, and their random transmission powers. In this section we will characterize the probability distribution of this random variable.

A useful result for evaluating the distributions of many random variables associated with Poisson processes is Campbell's Theorem, which can be stated as [34]:

Theorem 7 *(Campbell's Theorem) Let $\mathcal{P}$ be a Poisson process on $\mathcal{S}$ with mean measure μ, and let $f : S \to \mathbb{R}$ be measurable. Then the sum*

$$T = \sum_{X \in \mathcal{P}} f(X)$$

is absolutely convergent with probability if and only if

$$\int_{\mathcal{S}} \min\{|f(x)|, 1\}\mu(dx) < \infty.$$

If this condition holds, then

$$E\left(e^{j\omega T}\right) = \exp\left[\int_{\mathcal{S}} \left(e^{j\omega f(x)} - 1\right)\mu(dx)\right].$$

In order to use Theorem 7, let $\mathcal{S} = \mathbb{R}^2 \times [0, \infty)$, and define the Poisson process $\mathcal{P}$ on $\mathcal{S}$ as the Poisson process with density

$$\lambda^*(x, y, p) = \lambda \cdot f_P(p).$$

In this process, a point (x, y, p) represents a transmitter located in point (x, y) on the plane that transmits with power p.

In addition, define the function

$$f(x, y, p) = p(x^2 + y^2)^{-\alpha/2}$$

which can be interpreted as the (normalized) interference created at the origin by a transmitter located at point (x, y) that transmits with power p. With this definition we can write the normalized interference I as:

$$I = \sum_{(x,y,p) \in \mathcal{P}} f(x, y, p) = \sum_{(x,y,p) \in \mathcal{P}} p(x^2 + y^2)^{-\alpha/2}$$

To check the necessary condition of Theorem 7, we compute:

$$\begin{aligned}
\int_{\mathcal{S}} \min\{|f(x)|,1\}\mu(dx) &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \min\{|p\cdot(x^2+y^2)^{-\alpha/2}|,1\}\lambda\cdot f_P(p)dy\cdot dx\cdot dp \\
&= \int_0^\infty \int_0^\infty \min\{|p\cdot r^{-\alpha}|,1\}2\pi r\lambda f_P(p)\cdot dr\cdot dp \\
&= \int_0^\infty \int_0^{p^{1/\alpha}} 2\pi r\lambda f_P(p)\cdot dr\cdot dp + \int_0^\infty \int_{p^{1/\alpha}}^\infty p\cdot r^{-\alpha}2\pi r\lambda f_P(p)\cdot dr\cdot dp \\
&= \int_0^\infty p^{2/\alpha}\pi\lambda f_P(p)\cdot dp + \int_0^\infty \left(\frac{2}{\alpha-2}\right) p^{2/\alpha}\pi\lambda f_P(p)\cdot dp \\
&= \left(\frac{\alpha\pi\lambda}{\alpha-2}\right)\int_0^\infty p^{2/\alpha} f_P(p)dp \\
&\leq \left(\frac{\alpha\pi\lambda}{\alpha-2}\right)\left\{\int_0^1 f_P(p)dp + \int_1^\infty p\cdot f_P(p)dp\right\} \\
&\leq \left(\frac{\alpha\pi\lambda}{\alpha-2}\right)[1+E(P)] \\
&< \infty \qquad (7.3)
\end{aligned}$$

where we used the assumptions $\alpha > 2$ and $E(P) < \infty$. Therefore, we can use Theorem 7 to compute the log-characteristic function of the interference distribution:

$$\begin{aligned}
\Psi_I(\omega) &= \log E\left[\exp(j\omega I)\right] \\
&= \int_0^\infty \int_0^\infty \left[\exp\left(j\omega p r^{-\alpha}\right)-1\right] 2\pi\lambda r f_P(p)\cdot dp\cdot dr \\
&= \int_0^\infty \left[\phi_P(\omega r^{-\alpha})-1\right] 2\pi\lambda r dr \qquad (7.4)
\end{aligned}$$

After some calculations, under the assumption that $f_P(p)$ has support in $\mathbb{R}^+$, it is possible to rewrite (7.4) as:

$$\Psi_I(\omega) = \left[\lambda\pi\cos\left(\frac{\pi}{\alpha}\right)\Gamma\left(1-\frac{2}{\alpha}\right)\int_0^\infty f_P(x)x^{\frac{2}{\alpha}}dx\right]|\omega|^{\frac{2}{\alpha}}\left[-1+j\tan\left(\frac{\pi}{\alpha}\right)\operatorname{sgn}(\omega)\right] \qquad (7.5)$$

valid for $\alpha > 2$. This log-characteristic function corresponds to that of a stable distribution. The stable distributions have been extensively studied and can be parameterized by four parameters: an index of stability $\tilde{\alpha} \in (0,2]$, a skewness parameter $\beta \in [-1,1]$, a scale parameter $\gamma > 0$, and a location parameter $\delta \in \mathbb{R}$. With this parameterization, for $\tilde{\alpha} \neq 1$, the log-characteristic function of the stable distribution $S(\tilde{\alpha},\beta,\gamma,\delta;1)$ is given by [38]:

$$\Psi(\omega) = -\gamma^{\tilde{\alpha}}|\omega|^{\tilde{\alpha}}\left[1-j\beta\tan\left(\frac{\pi\tilde{\alpha}}{2}\right)\operatorname{sgn}(\omega)\right]+j\delta\omega \qquad (7.6)$$

We see that in our case we have $\tilde{\alpha} = 2/\alpha$, $\beta = 1$, $\delta = 0$, and $\gamma = \left[\lambda\pi \cos\left(\frac{\pi}{\alpha}\right) \Gamma\left(1 - \frac{2}{\alpha}\right) \cdot \int_0^\infty f_p(x) x^{\frac{2}{\alpha}} dx\right]^{\alpha/2}$.

Unfortunately there are only three cases in which the density of a stable distribution is known. These are for $\tilde{\alpha} = 2$ (Gaussian), $\tilde{\alpha} = 1$ (Cauchy), and $\tilde{\alpha} = 1/2$ (Levy). In our case, since $\alpha > 2$ we can obtain a closed expression for the pdf of the interference only when $\alpha = 4$, in which case it is given by the Levy distribution:

$$f(x) = \sqrt{\frac{\gamma}{2\pi}} x^{-3/2} \exp\left(-\frac{\gamma}{2x}\right) \ , x \geq 0 \tag{7.7}$$

For other values of α we can only evaluate the pdf of the interference numerically. However it is possible to obtain a tail approximation for any stable law [38] :

$$f(x|\tilde{\alpha}, \beta, \gamma, \delta) \sim \tilde{\alpha}\gamma^{\tilde{\alpha}} \sin\left(\frac{\pi\tilde{\alpha}}{2}\right) \frac{\Gamma[\tilde{\alpha}]}{\pi} (1+\beta) x^{-(\tilde{\alpha}+1)} = K x^{-\left(\frac{2}{\alpha}+1\right)} \tag{7.8}$$

where $h(x) \sim g(x)$ means $\lim_{x\to\infty} g(x)/h(x) = 1$. We see that in all cases the tail of the interference power distribution follows a power law. In particular, this distribution is always heavy tailed and the mean is infinite.

Note that the particular choice of the power pdf $f_P(p)$ only affects the distribution of the interference through the scale parameter γ, and any change of power pdf can be absorbed in the Poisson density constant λ.

The fact that the interference power pdf is heavy tailed means that, in our model, the expected interference power received in any point of the plane is infinite. Therefore, even though each system transmits a signal of finite power, the density of systems is finite, and the power decays with distance as a power law, the resulting interference in any point in the plane is very large.

7.2 Interference reduction

We would like to understand what effect makes the interference so large. One possible explanation is that since the total number of systems is infinite, the power law attenuation with distance is not enough to keep the aggregate interference power bounded. An alternative explanation is that few interferers, located in the proximity of the receiver create a large amount of interference. To see what is the main effect governing interference, we modify the model. We draw a circle of radius r_0 around the receiver of interest, and do not allow any other transmitter within this circle to transmit.

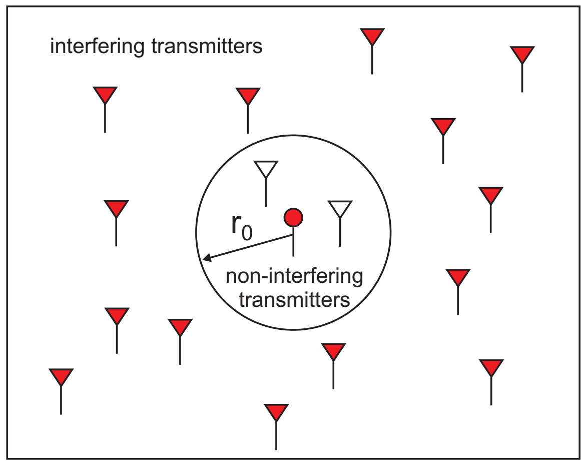


Figure 7.1: Aggregation of interference generated by transmitters located farther than r_0 from given receiver. The transmitters located in the circle of radius r_0 do not interfere with the given receiver.

We now characterize the distribution of the interference generated by transmitters farther away than r_0 from the given receiver (see Figure 7.1). It can be shown using techniques similar to those in [47] that the log-characteristic function of the (normalized) interference generated by transmitters out of a circle of radius r_0 centered at the receiver is given by:

$$\Psi_{r_0}(\omega) = \lim_{b\to\infty} \lambda\pi b^2 \left\{ \int_{r_0}^{b} \frac{2r}{b^2 - r_0^2} \left[\Phi_P\left(\frac{\omega}{r^\alpha}\right) - 1 \right] dr \right\} = \int_{r_0}^{\infty} 2\lambda\pi r \left[\Phi_P\left(\frac{\omega}{r^\alpha}\right) - 1 \right] dr \tag{7.9}$$

Exploiting the fact that the pdf of the power has support in $[0, \infty)$, writing the integral defining $\Phi_P(\omega)$, and making changes of variables, we can rewrite (7.9) as:

$$\Psi_{r_0}(\omega) = \lambda\pi|\omega|^{2/\alpha} \int_0^\infty f_P(x) x^{2/\alpha} \int_{r_0(x|\omega|)^{-1/\alpha}}^{\infty} 2y \left[-2\sin^2\left(\frac{1}{2y^\alpha}\right) + j \cdot \mathrm{sgn}(\omega) \sin\left(\frac{1}{y^\alpha}\right) \right] dy.dx \tag{7.10}$$

In the case when the transmission power is constant, with value P, (7.10) reduces to:

$$\Psi_{r_0}(\omega) = \lambda\pi|\omega|^{2/\alpha} P^{2/\alpha} \int_{r_0(P|\omega|)^{-1/\alpha}}^{\infty} 2y \left[-2\sin^2\left(\frac{1}{2y^\alpha}\right) + j \cdot \mathrm{sgn}(\omega) \sin\left(\frac{1}{y^\alpha}\right) \right] dy. \tag{7.11}$$

Furthermore, if we normalize the transmission power so that $P = 1$ we can obtain an expression for the log-characteristic function of the interference distribution that is defined in terms of well known functions:

$$\Psi_{r_0}(\omega) = \lambda\pi|\omega|^{2/\alpha}\left\{r_0^2|\omega|^{-2/\alpha} + \frac{2}{\alpha}\exp\left[-\frac{j\cdot\mathrm{sgn}(\omega)\pi}{a}\right]\left[\Gamma\left(-\frac{2}{\alpha}\right) - \Gamma\left(-\frac{2}{\alpha}, -jr_0^{-\alpha}\omega\right)\right]\right\} \quad (7.12)$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ is the Gamma function and $\Gamma(x,y) = \int_y^\infty t^{x-1}e^{-t}dt$ is the incomplete Gamma function.

In order to obtain the probability density function of the aggregate interference we take the Fourier transform of its characteristic function:

$$f_{I|r_0}(x|r_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega x}e^{\Psi_{r_0}(\omega)}d\omega \quad (7.13)$$

in which we can replace $\Psi_{r_0}(\omega)$ with (7.10), (7.11), or (7.12), depending on whether the transmission power is non-constant, constant, or constant with value $P = 1$ respectively.

When $r_0 > 0$ the interference distribution is not heavy tailed and we can compute the mean and variance of the the received interference power:

$$\begin{aligned}
E[I] &= \int_{r_0}^{\infty} 2\pi\lambda E[P]r^{1-\alpha}dr \\
&= \frac{2\pi\lambda}{\alpha-2}E[P]r_0^{2-\alpha} && (7.14) \\
Var[I] &= \int_{r_0}^{\infty} d\mathrm{Var}(I(r)) = \int_{r_0}^{\infty}\left\{E\left[\mathrm{Var}(I|dn(r))\right] + \mathrm{Var}\left[E(I|dn(r))\right]\right\} \\
&= \int_{r_0}^{\infty}\left\{E\left[\frac{\mathrm{Var}(P)}{r^{2\alpha}}dn(r)\right] + \mathrm{Var}\left[\frac{E(P)}{r^{\alpha}}dn(r)\right]\right\} \\
&= \int_{r_0}^{\infty} 2\pi\lambda\left[\mathrm{Var}(P) + E(P)^2\right]r^{1-2\alpha}dr \\
&= \frac{\pi\lambda}{\alpha-1}\left[\mathrm{Var}(P) + E(P)^2\right]r_0^{2-2\alpha} && (7.15)
\end{aligned}$$

where $I(r)$ is the interference power generated by interferers located at distance r, $n(r)$ is the number of interferers located at distance r, and $E(P)$ and $\mathrm{Var}(P)$ are the mean and variance of the transmitted power distribution. We can further specialize these expressions for the mean an variance to the case of constant transmission power ($\mathrm{Var}(P) = 0$), and constant transmission power of value 1 ($\mathrm{Var}(P) = 0$, $E(P) = 1$).

We see that the reason for the very large interference in the first model is the presence of few interferers located very close to the given receiver. By limiting the minimum transmitter-receiver distance to r_0 we made the expected interference finite and eliminated the most harmful interference. In our second model r_0 can be interpreted as a parameter that defines how close two radio devices can be. In addition, r_0 could be a protocol design parameter, where by appropriate spectrum management we don't allow any system within distance r_0 from the given receiver to transmit in the same band. Using (7.14) and (7.15) we can immediately see the benefit of increasing r_0 in reducing the moments of the received interference. Furthermore, we see that $\sqrt{\mathrm{Var}[I]}/E[I]$ decays as $1/r_0$ and therefore, the interference distribution becomes more concentrated around its mean when r_0 increases. This has the effect of making the interference more deterministic and easier to predict.

Chapter 8

Bandwidth Scaling for Large and Dense Networks [1]

The model considered in the previous chapter assumed that the path loss between transmitters and receivers is governed by the large scale path loss, ignoring small scale effects. We showed that in order to keep interference under control the spectrum sharing protocol should allocate different frequencies to systems that are close to each other. This orthogonal allocation of frequencies corresponds to having a frequency reuse factor larger than one. For example, if three systems are close to each other, they should use 1/3 of the band each. Comparing this to a situation where the frequency reuse factor is one, i.e. all systems use the same channel, we observe that the price paid for the reduction of the interference is a reduction of the available bandwidth for each system.

If we take into account small scale effects, and opportunistically assign the same channel to systems that interfere little with each other due to a fade in the channel gain, we may obtain some gain. The goal of this chapter is to quantify the performance improvement that can be attained by taking into account small scale effects. We do this by obtaining lower and upper bounds on the achievable performance, as the number of systems grows large.

[1] Revised on November 28, 2007

Let M be the total number of users, each formed by a transmitter receiver pair. Each user needs to communicate data at a rate R. We assume that each user can only use a fixed bandwidth B for communication and that interference is treated as noise. We investigate how the total required bandwidth scales with the number of users in the following situations:

1. Many SISO systems: transmitters and receivers are distributed arbitrarily (cf. Figure 8.1).

2. Many Multiple Access Systems: users are grouped into systems, and the receivers of the users belonging to the same system are collocated (cf. Figure 8.2).

3. Many Broadcast Systems: users are grouped into systems, and the transmitters of the users belonging to the same system are collocated (cf. Figure 8.3).

In particular for the latter two situations we will analyze what is the minimum number of systems required to support the M users while minimizing the total required bandwidth.

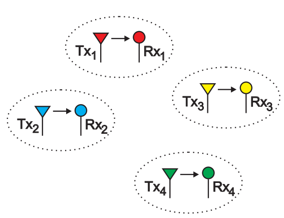


Figure 8.1: Many SISO systems.

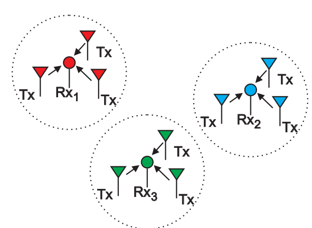


Figure 8.2: Many multiple access systems.

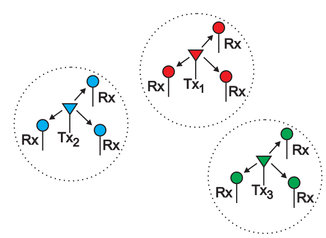


Figure 8.3: Many broadcast systems.

8.1 Model

We will use a model in which we discretize the spectrum into channels of bandwidth B. In this model B represents the coherence bandwidth of the channels. The power gain from the transmitter of user i to the receiver of user j in channel number k is denoted by $c_{i,j}(k)$. Except for the cases noted in the next paragraph, we assume that $\{c_{i,j}(k)\}_{i,j,k}$ are i.i.d. Exp(1) random variables. This assumption corresponds to a propagation environment where the large-scale path loss is of value 1. In other words, the path loss between transmitters and receivers is independent of the distance between them. The only effect that the model takes into account is the small-scale path loss due to multipath fading. Therefore, the model that we consider is appropriate for dense networks in a rich scattering environment, where the distance between transmitters and receivers has only a marginal effect on attenuation.

In situation 2-Many multiple access systems, we assume $c_{i,j}(\cdot) = c_{i,k}(\cdot)$ when users j and k belong to the same system. Similarly, in situation 3-Many broadcast systems, we assume that $c_{j,i}(\cdot) = c_{k,i}(\cdot)$ whenever users j and k belong to the same system.

We assume that each user can transmit in only one channel with power P, and that he treats the received interference as noise. We also assume that the users communicate using random Gaussian codebooks. Letting $f(i)$ denote the channel assigned to user i, the rate of this user is given by

$$R_i = B \log \left[1 + \frac{c_{i,i}(f(i))}{\frac{N_0 B}{P} + \sum_{\substack{j \neq i \\ f(j)=f(i)}} c_{j,i}(f(i))} \right]$$

where N_0 is the additive white Gaussian background noise PSD.

Let $W = \max_{1 \leq i \leq M} f(i)$ be the total number of channels used by the allocation $f(\cdot)$. We would like to minimize W subject to a minimum rate requirement for each user, i.e. $R_i \geq R$, where the optimization variables are $f(i) : i = 1, \ldots, M$. Let W^* be the optimal solution to this problem.

For each realization of the fading gains $\{c_{i,j}(k)\}_{i,j,k}$ the above optimization problem is an integer program, and seems very hard to solve. Also the solution W^* is a function of the channel gains which are random, and hence is a random variable too.

We will characterize W^* by deriving lower and upper bounds which will hold with probability approaching 1 as $M \to \infty$. When a given property or relation holds with probability approaching 1 for $M \to \infty$, we say that it holds a.a.s. (asymptotically almost sure).

Note that the rate constraint $R_i \geq R$ can be translated into a minimum SINR requirement

γ so that

$$R_i \geq R \Longleftrightarrow \frac{c_{i,i}(f(i))}{\frac{N_0 B}{P} + \sum_{\substack{j \neq i \\ f(j)=f(i)}} c_{j,i}(f(j))} \geq \gamma. \tag{8.1}$$

8.2 Many SISO systems

Consider first a channel assignment such that $f(i) \neq f(j), i \neq j$. This assignment allocates orthogonal channels to all the users, and requires a total bandwidth M. Since there is no interference, R_i only depends on $c_{i,i}(f(i))$, and it seems reasonable to believe that it is possible to find an assignment that achieves $R_i \geq R$ for all users. We just need to assign channel k to user i whenever $c_{i,i}(k)$ is large enough to achieve the required rate. This assignment does not allow for any sharing of channels among users, and can be taken as a baseline to compare the performance of other strategies.

Can we do better ? First we note that in order to reduce W we need to allocate multiple users per channel. If we assign a constant number of systems K to each channel we obtain $W(M) = M/K$, assuming that it is still possible to achieve the minimum rate requirements. In order to obtain a sublinear relation for $W(M)$, we need to let $K(M)$ grow with M. So a natural question to ask is how fast we can let $K(M)$ grow while achieving the required rates.

In order to keep interference under control, we need to assign users to the same channel only when they interfere little with each other. For example we may require that $f(i) = f(j) = k$ only when $c_{i,j}(k), c_{j,i}(k) \leq \alpha$. With this requirement, valid channel assignments can be seen as proper colorings of a random graph whose nodes represent the users and edges represent unacceptable interference among them. It is known that the chromatic number χ of an Erdös-Rényi random graph $\mathcal{G}(M,p)$ scales as $M \sim \frac{-\log(1-p)}{2} \frac{M}{\log(M)}$. Therefore for fixed α we get a scaling of $W(M) = \Theta(M/\log(M))$. However, this scaling requires that we allocate $\Theta(\log(M))$ users per channel, in which case interference may grow very large and we may not be able to achieve the required rates. To limit interference, as we increase M we need to reduce α, in which case $W(M) > \Theta(M/\log(M))$. It seems that our requirement of keeping the total interference bounded resulted in a worse scaling law than that of $\chi(G(M,p))$.

Should we conclude that $W(M) > \Theta(M/\log(M))$? The above reasoning has two flaws. First it does not take into account that $c_{i,i}(\cdot)$ varies across channels, and this variation may help in achieving a large SINR if we choose channels appropriately. Second, by reducing the channel assignment problem to a graph coloring problem we are losing the information about the specific values of the channel gains. When we do not draw an edge between users i and j we are implicitly replacing $c_{i,j}(k)$ and $c_{j,i}(k)$ by their upper bound α and therefore we obtain a worst case performance bound.

It turns out that $W^*(M) \sim \frac{\log(1+\gamma)\cdot M}{\log M}$. So even though our problem seems harder than the coloring problem in that we need to keep the total interference under control (instead of just satisfying pairwise constraints among nodes) we have just enough additional flexibility to keep $W^*(M)$ scaling as $\chi(G(M,p))$.

We next prove that $W^*(M) \sim \frac{\log(1+\gamma)\cdot M}{\log M}$ a.a.s. by deriving lower and upper bounds on $W^*(M)$, both in a.a.s. sense.

8.2.1 A lower bound on the required bandwidth

Theorem 8 *In the many SISO case with M users, under the assumptions described in Section 8.1, the optimal number of required channels to satisfy the rate constraints $R_i \geq R$ satisfies*

$$W^*(M) \geq \frac{\log(1+\gamma)\cdot M}{\log(M)}$$

asymptotically, with probability going to 1 as $M \to \infty$.

Proof: We prove the theorem by showing that the probability of finding a group of $\log(M)/\log(1+\gamma)$ users, each satisfying its rate constraint $R_i \geq R$, over the $\log(1+\gamma)\cdot M/(\log(M))$ available channels goes to zero as $M \to \infty$. This implies that any spectrum allocation $\{f(\cdot)\}$ that satisfies the rate constraints cannot assign more than $\log(M)/\log(1+\gamma)$ users per channel, and hence the total number of channels required cannot be smaller than $\log(1+\gamma)M/(\log(M))$.

We will say that the channel allocation $f(\cdot)$ is feasible if it satisfies (8.1) for $i = 1, \dots, M$. Let $\mathcal{A}_{f,k}$ be the set of users allocated to channel k under f:

$$\mathcal{A}_{f,k} = \{i : f(i) = k, i = 1, \dots, M\}$$

For a given realization of the fading gains we define $f^*(k)$ to be a feasible channel allocation rule that maximizes $|\mathcal{A}_{f,k}|$, i.e. it maximizes the number of users sharing channel k:

$$f^*(k) = \arg \max_{f(\cdot) \text{ feasible}} |\mathcal{A}_{f,k}|$$

Finally let $n^*(k) = |\mathcal{A}_{f^*(k),k}|$. Note that since $n^*(k)$ is a function of the fading gains $\{c_{i,j}(k)\}$ it is a random variable.

We will prove that $P(n^*(k) \geq \log(M)/\log(1+\gamma)) \leq \exp(-\log(M)\log\log(M)/\log(1+\gamma))$ for large M (to the first order in the exponent).

Fix the channel k, and let $X(n)$ be the number of groups of n users that can be assigned to channel k satisfying the rate requirements. That is, $X(n)$ is the number of feasible power allocations $f(\cdot)$ that result in $|\mathcal{A}_{f,k}| = n$. We can also write $X(n)$ as a sum of indicator functions:

$$X(n) = \sum_{\substack{\mathcal{S} \subseteq \{1,\dots,M\} \\ |\mathcal{S}|=n}} 1\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P} + \sum_{\substack{j \neq i \\ j \in \mathcal{S}}} c_{j,i}(k)} \geq \gamma \ , \forall i \in \mathcal{S} \right)$$

Clearly, we have

$$P(n^*(k) \geq n) = P(X(n) \geq 1)$$

We use the first moment method (Markov's inequality) to upper bound $P(X(n) \geq 1)$:

$$P(X(n) \geq 1) \leq E[X(n)] = \binom{M}{n} \left[P\left(\frac{c_{1,1}(k)}{\frac{N_0 B}{P} + \sum_{j=2}^{n} c_{j,1}(k)} \geq \gamma \right) \right]^n \tag{8.2}$$

where we exploited the fact that the $\{c_{i,j}(k)\}$ are i.i.d.

Since we assumed that $\{c_{i,j}(k)\}$ are i.i.d. Exp(1) random variables, we can compute the probability between brackets in closed form. This is easily done by noting that $\sum_{j=2}^{n} c_{j,1}(k) \sim \Gamma(n-1, 1)$, where the p.d.f. of $Z \sim \Gamma(k, \theta)$ is given by

$$f_Z(z) = \frac{z^{k-1}\theta^k}{\Gamma(k)} e^{-\theta z} \ , \ z, k, \theta > 0.$$

Then we have:

$$P\left(\frac{c_{1,1}(k)}{\frac{N_0B}{P}+\sum_{j=2}^{n}c_{j,1}(k)}\geq\gamma\right)=\int_0^\infty\int_{\gamma(N_0B/P+x)}^{\infty}e^{-c}\cdot\frac{x^{n-2}}{(n-2)!}e^{-x}dc\cdot dx=\frac{e^{-\gamma N_0B/P}}{(1+\gamma)^{n-1}}. \quad (8.3)$$

Replacing in (8.2) we get:

$$P(X(n)\geq 1)\leq\left(\frac{M\cdot e}{n}\right)^n\left[\frac{e^{-\gamma N_0B/P}}{(1+\gamma)^{n-1}}\right]^n=\left[M\exp\left(1-\frac{\gamma N_0B}{P}-\log(n)-(n-1)\log(1+\gamma)\right)\right]^n \quad (8.4)$$

If we set $n=\frac{\log(M)}{\log(1+\gamma)}$ we get

$$\begin{aligned}P\left[n^*(k)\geq\frac{\log(M)}{\log(1+\gamma)}\right] &= P\left[X\left(\frac{\log(M)}{\log(1+\gamma)}\right)\geq 1\right]\\ &\leq \exp\left\{\frac{\log(M)}{\log(1+\gamma)}\left[-\log\log(M)+C\right]\right\} \quad (8.5)\end{aligned}$$

where $C=1-\frac{\gamma N_0B}{P}+\log\log(1+\gamma)+\log(1+\gamma)$ is a constant independent of M.

We finally show that we cannot find a group of users of size $\log M/\log(1+\gamma)$ that satisfies the rate constraints in any of the $K=\log(1+\gamma)\cdot M/\log(M)$ channels using the union bound:

$$\begin{aligned}P\left(\max_{k\in\{1,\ldots,K\}}n^*(k)\geq\frac{\log(M)}{\log(1+\gamma)}\right) &\leq \frac{\log(1+\gamma)M}{\log(M)}P\left(n^*(k)\geq\frac{\log(M)}{\log(1+\gamma)}\right)\\ &\leq e^{[\log(M)+\log\log(1+\gamma)-\log(\log(M))]}e^{\frac{\log(M)}{\log(1+\gamma)}[-\log\log(M)+C]}\\ &\longrightarrow 0 \text{ as } M\to\infty \quad (8.6)\end{aligned}$$

■

8.2.2 An upper bound on the required bandwidth

Theorem 9 *In the many SISO case with M users, under the assumptions described in Section 8.1, for any $\epsilon>0$ the optimal number of required channels to satisfy the rate constraints $R_i\geq R$ satisfies*

$$W^*(M)\leq\frac{\log(1+\gamma)}{1-\epsilon}\cdot\frac{M}{\log(M)}\cdot(1+o(1))$$

asymptotically, with probability going to 1 as $M\to\infty$.

Proof: We will describe channel assignment algorithm that achieves the upper bound with probability going to 1 as M goes to infinity. The algorithm proceeds in stages, one for each channel k. Let $m(k)$

be the number of users that have not been allocated to a channel at stage k. Note that $m(1) = M$. Also, let $n = (1-\epsilon)\log(M)/\log(1+\gamma)$. The algorithm can be described as follows:

```
set k = 1;
while m(k) > M/log(M)^2
    {
    find a group of n users among the m(k) that satisfy R_i ≥ R in channel k;
    assign channel k to these n users;
    set k = k + 1;
    }
assign different M/log(M)^2 channels to the remaining users;
```

Assume first that the algorithm succeeds in each stage in finding n users satisfying the rate constraints. Then the algorithm finds a feasible channel allocation that uses no more than $M/n + M/\log(M)^2$ channels, and it follows that $W^*(M) \le M/n + M/\log(M)^2$.

We now need to prove that the algorithm can find groups of n users satisfying the rate constraints in all the stages where $m(k) \ge M/\log(M)^2$.

We start by computing a lower bound on the probability that the algorithm succeeds in finding a group of n users satisfying the rate constraints at stage k, when there are $m(k)$ users to choose from.

Without loss of generality, we label the $m(k)$ users that have not yet been assigned a channel at stage k from 1 to $m(k)$, and define $X(n)$ in a similar way as we did in the previous subsection:

$$X(n) = \sum_{\substack{\mathcal{S}\subseteq\{1,\ldots,m(k)\}\\|\mathcal{S}|=n}} 1\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P} + \sum_{\substack{j\neq i\\ j\in\mathcal{S}}} c_{j,i}(k)} \ge \gamma \ , \ \forall i \in \mathcal{S}\right)$$

Letting $\mathcal{G} = \{\mathcal{S} \subset \{1,\ldots,m(k)\} : |\mathcal{S}| = n\}$ and $I_{\mathcal{S}} = \prod_{i\in\mathcal{S}} 1\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P}+\sum_{\substack{j\neq i\\ j\in\mathcal{S}}} c_{j,i}(k)} \ge \gamma\right)$ we can express $X(n) = \sum_{\mathcal{S}\in\mathcal{G}} I_{\mathcal{S}}$.

We want to find a lower bound on $P(X(n) \geq 1)$ or equivalently, an upper bound on $P(X(n) = 0)$, for which we will use the following lemma.

Lemma 7 *Let* $\lambda = E[X(n)]$ *and* $\Delta = \sum_{\mathcal{S}\in\mathcal{G}} \sum_{\substack{\mathcal{T}\in\mathcal{G} \\ \mathcal{S}\cap\mathcal{T}\neq\emptyset}} E\left[I_{\mathcal{S}} I_{\mathcal{T}}\right]$. *Then,*

$$P(X(n) = 0) \leq \exp\left(-\frac{\lambda^2}{\Delta}\right)$$

Proof: We will follow the steps of the proof of Theorem 2.18 (ii) of [31] which states a similar result in a very different setting.

Let $\Psi(s) = E[e^{-sX}]$, $s > 0$. We will show first that

$$-(\log \Psi(s))' \geq \lambda e^{-s\Delta/\lambda} \ , \ s > 0 \tag{8.7}$$

which implies

$$-\log \Psi(s) \geq \int_0^\infty \lambda e^{-u\Delta/\lambda} du = \frac{\lambda^2}{\Delta}\left(1 - e^{-s\Delta/\lambda}\right) \tag{8.8}$$

By noting that $\lim_{s\to\infty} \Psi(s) = P(X = 0)$ we immediately obtain the result.

In order to obtain (8.7), we represent $-\Psi'(s)$ in the form

$$-\Psi'(s) = E\left(Xe^{-sX}\right) = \sum_{\mathcal{S}\in\mathcal{G}} E\left(I_{\mathcal{S}} e^{-sX}\right) \tag{8.9}$$

and for every $\mathcal{S} \in \mathcal{G}$ we split $X = Y_{\mathcal{S}} + Z_{\mathcal{S}}$, where $Y_{\mathcal{S}} = \sum_{\substack{\mathcal{T}\in\mathcal{G} \\ \mathcal{S}\cap\mathcal{T}\neq\emptyset}} I_{\mathcal{T}}$.

In order to proceed we need to prove the following lemma, which states a correlation inequality.

Lemma 8 *Let* $g, h : \mathbb{R} \to \mathbb{R}$ *be two increasing or decreasing functions. Then,*

$$E\left[g(Y_{\mathcal{S}})h(Z_{\mathcal{S}})|I_{\mathcal{S}} = 1\right] \geq E\left[g(Y_{\mathcal{S}})|I_{\mathcal{S}} = 1\right] E[h(Z_{\mathcal{S}})]$$

Proof: Focusing on a single channel k and dropping the corresponding indices, we use successive conditioning to write:

$$E\left[g(Y_{\mathcal{S}})h(Z_{\mathcal{S}})|I_{\mathcal{S}} = 1\right] = E\left\{ E\left[E\left(g(Y_{\mathcal{S}})h(Z_{\mathcal{S}}) \left| \begin{array}{c} \{c_{i,i} : i = 1,\ldots,m\} \\ \{c_{i,j} : i,j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right.\right) \left| \begin{array}{c} \{c_{i,j} : i,j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right.\right] \left| I_{\mathcal{S}} = 1 \right.\right\} \tag{8.10}$$

We will apply twice a continuous version of the FKG (Fortuin, Kasteleyn and Ginibre) [25] inequality for multivariate totally positive of order 2 (MTP_2) random vectors defined on $\mathbb{R}^n$ [32, 41], which we describe next.

Consider a pdf $f(\mathbf{x})$ defined on $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_n$, were each $\mathcal{X}_i$ is totally ordered, satisfying

$$f(\mathbf{x} \vee \mathbf{y}) f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x}) f(\mathbf{y})$$

where

$$\mathbf{x} \vee \mathbf{y} = (\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\})$$

and

$$\mathbf{x} \wedge \mathbf{y} = (\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\})$$

Any pdf $f(\cdot)$ satisfying this property is called *multivariate totally positive of order 2* (MTP_2). In particular, if $\mathcal{X} = \mathbb{R}$ and the random variables $\{X_i\}_{i=1}^n$ are independent, the corresponding joint distribution is MTP_2.

An important property satisfied by MTP_2 random vectors is the following correlation inequality.

Let $g(\cdot)$ and $h(\cdot)$ be simultaneously monotone increasing or decreasing, and let the random vector $\mathbf{X}$ be MTP_2. Then

$$E[g(\mathbf{X})h(\mathbf{X})] \geq E[g(\mathbf{X})]E[h(\mathbf{X})] \tag{8.11}$$

We now apply this inequality to the inner expectation of (8.10). When we condition on

$$\mathcal{C} = \left\{ \begin{array}{c} \{c_{i,i} : 1 = 1, \ldots, m\} \\ \{c_{i,j} : i, j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right\}$$

the only resulting randomness is due to $\{c_{i,j} : i, j \in \{1, \ldots, m\} \backslash \mathcal{S}, i \neq j\}$, which are still conditionally i.i.d., and hence MTP_2. Since larger cross-gains $c_{i,j}, i \neq j$, result in smaller SINR, both $Y_{\mathcal{S}}$ and $Z_{\mathcal{S}}$ are decreasing on $\{c_{i,j} : i, j \in \{1, \ldots, m\} \backslash \mathcal{S}, i \neq j\}$, and therefore $h(Y_{\mathcal{S}})$ and $g(Z_{\mathcal{S}})$ are both increasing or decreasing on $\{c_{i,j} : i, j \in \{1, \ldots, m\} \backslash \mathcal{S}, i \neq j\}$. Applying (8.11):

$$E\left(g(Y_{\mathcal{S}})h(Z_{\mathcal{S}})\,|\mathcal{C}\,\right) \geq E\left(g(Y_{\mathcal{S}})\,|\mathcal{C}\,\right) E\left(h(Z_{\mathcal{S}})\,|\mathcal{C}\,\right) \tag{8.12}$$

In the middle expectation of (8.10) we average over $\{c_{i,i} : i \in \{1, \ldots, m\} \backslash \mathcal{S}\}$ keeping $\{c_{i,j} : i, j \in \mathcal{S}\}$ fixed. Since conditioned on $\{c_{i,j} : i, j \in \mathcal{S}\}$ and $\{I_{\mathcal{S}} = 1\}$, $\{c_{i,i} : i \in \{1, \ldots, m\} \backslash \mathcal{S}\}$ are i.i.d., the corresponding random vector is MTP_2. Also, since larger direct gains $c_{i,i}$ result in larger SINR, the 2 expectations on the right side of (8.12) are both increasing or decreasing on $\{c_{i,i} : i \in \{1, \ldots, m\} \backslash \mathcal{S}\}$. Therefore, taking expectations on both sides of (8.12) and applying (8.11) we obtain:

$$\begin{aligned} E\left[E\left(g(Y_{\mathcal{S}})h(Z_{\mathcal{S}})\,|\mathcal{C}\,\right) \middle| \begin{array}{c} \{c_{i,j} : i, j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right] &\geq E\left[E\left(g(Y_{\mathcal{S}})\,|\mathcal{C}\,\right) E\left(h(Z_{\mathcal{S}})\,|\mathcal{C}\,\right) \middle| \begin{array}{c} \{c_{i,j} : i, j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right] \\ &\geq E\left[g(Y_{\mathcal{S}}) \middle| \begin{array}{c} \{c_{i,j} : i, j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right] E\left[h(Z_{\mathcal{S}}) \middle| \begin{array}{c} \{c_{i,j} : i, j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right] \end{aligned} \tag{8.13}$$

We finally note that $Z_{\mathcal{S}}$ is independent of $\{c_{i,j} : i, j \in \mathcal{S}\}$, and it follows that the second factor in the RHS of (8.13) is constant as a function of $\{c_{i,j} : i, j \in \mathcal{S}\}$ and can be pulled out from the outer expectation of (8.10). Therefore, by taking expectations on both sides of (8.13), averaging over $\{c_{i,j} : i, j \in \mathcal{S}\}$ conditioned on $\{I_{\mathcal{S}} = 1\}$, we obtain:

$$\begin{aligned} E\left[g(Y_{\mathcal{S}})h(Z_{\mathcal{S}})I_{\mathcal{S}} = 1\right] &\geq E\left\{ E\left[g(Y_{\mathcal{S}}) \middle| \begin{array}{c} \{c_{i,j} : i, j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right] E\left[h(Z_{\mathcal{S}}) \middle| \begin{array}{c} \{c_{i,j} : i, j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right] \middle| I_{\mathcal{S}} = 1 \right\} \\ &= E\left\{ E\left[g(Y_{\mathcal{S}}) \middle| \begin{array}{c} \{c_{i,j} : i, j \in \mathcal{S}\} \\ I_{\mathcal{S}} = 1 \end{array} \right] \middle| I_{\mathcal{S}} = 1 \right\} E\left[h(Z_{\mathcal{S}})\right] \\ &= E\left[g(Y_{\mathcal{S}})|I_{\mathcal{S}} = 1\right] E[h(Z_{\mathcal{S}})] \end{aligned} \tag{8.14}$$

■

Then, using Lemma 8 with $g(t) = h(t) = e^{-st}$ and letting $p_{\mathcal{S}} = E(I_{\mathcal{S}})$,

$$E\left(I_{\mathcal{S}}e^{-sX}\right) = p_{\mathcal{S}}E\left[e^{-sY_{\mathcal{S}}}e^{-sZ_{\mathcal{S}}}|I_{\mathcal{S}} = 1\right]$$

$$\geq \quad p_{\mathcal{S}} E\left[e^{-sY_{\mathcal{S}}} | I_{\mathcal{S}} = 1\right] E[e^{-sZ_{\mathcal{S}}}]$$

$$\geq \quad p_{\mathcal{S}} E\left[e^{-sY_{\mathcal{S}}} | I_{\mathcal{S}} = 1\right] \Psi(s) \tag{8.15}$$

Recall that $\lambda = E(X) = \sum_{\mathcal{S}\in\mathcal{G}} p_{\mathcal{S}}$. From (8.9) and (8.15), by applying Jensen's inequality twice, first to the conditional expectation and then to the sum, we obtain

$$\begin{aligned} -\left(\log\Psi(s)\right)' &= -\frac{\Psi'(s)}{\Psi(s)} \geq \sum_{\mathcal{S}\in\mathcal{G}} p_{\mathcal{S}} E\left(e^{-sY_{\mathcal{S}}} | I_{\mathcal{S}} = 1\right) \\ &\geq \lambda \sum_{\mathcal{S}\in\mathcal{G}} \frac{1}{\lambda} p_{\mathcal{S}} \exp\left[-E\left(sY_{\mathcal{S}} | I_{\mathcal{S}} = 1\right)\right] \\ &\geq \lambda \exp\left[-\sum_{\mathcal{S}\in\mathcal{G}} \frac{1}{\lambda} p_{\mathcal{S}} E\left(sY_{\mathcal{S}} | I_{\mathcal{S}} = 1\right)\right] \\ &= \lambda \exp\left[-\frac{s}{\lambda} \sum_{\mathcal{S}\in\mathcal{G}} E(Y_{\mathcal{S}} I_{\mathcal{S}})\right] = \lambda e^{-s\Delta/\lambda} \end{aligned} \tag{8.16}$$

■

In order to use the Lemma 7 we will get an upper bound on Δ/λ^2. We first compute an upper bound on $E(I_{\mathcal{S}} I_{\mathcal{T}})$ where $\mathcal{S}, \mathcal{T} \in \mathcal{G}$ and $|\mathcal{S} \cap \mathcal{T}| = i$.

$$\begin{aligned} E(I_{\mathcal{S}} I_{\mathcal{T}}) &= E\left[\prod_{i\in\mathcal{S}} 1\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P} + \sum_{\substack{t\neq i \\ t\in\mathcal{S}}} c_{t,i}(k)} \geq \gamma\right) \prod_{j\in\mathcal{T}} 1\left(\frac{c_{j,j}(k)}{\frac{N_0 B}{P} + \sum_{\substack{t\neq j \\ t\in\mathcal{T}}} c_{t,j}(k)} \geq \gamma\right)\right] \\ &\leq E\left[\prod_{i\in\mathcal{S}} 1\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P} + \sum_{\substack{t\neq i \\ t\in\mathcal{S}}} c_{t,i}(k)} \geq \gamma\right) \prod_{j\in\mathcal{T}\setminus\mathcal{S}} 1\left(\frac{c_{j,j}(k)}{\frac{N_0 B}{P} + \sum_{\substack{t\neq j \\ t\in\mathcal{T}}} c_{t,j}(k)} \geq \gamma\right)\right] \\ &= E\left[\prod_{i\in\mathcal{S}} 1\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P} + \sum_{\substack{t\neq i \\ t\in\mathcal{S}}} c_{t,i}(k)} \geq \gamma\right)\right] E\left[\prod_{j\in\mathcal{T}\setminus\mathcal{S}} 1\left(\frac{c_{j,j}(k)}{\frac{N_0 B}{P} + \sum_{\substack{t\neq j \\ t\in\mathcal{T}}} c_{t,j}(k)} \geq \gamma\right)\right] \\ &= \left[P\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P} + \sum_{\substack{t\neq i \\ t\in\mathcal{S}}} c_{t,i}(k)} \geq \gamma\right)\right]^{2n-i} \end{aligned} \tag{8.17}$$

And now we upper bound Δ/λ^2, letting $p = P\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P} + \sum_{\substack{t\neq i \\ t\in\mathcal{S}}} c_{t,i}(k)} \geq \gamma\right)$:

$$\begin{aligned} \frac{\Delta}{\lambda^2} &= \frac{\sum_{\mathcal{S}\in\mathcal{G}} \sum_{\substack{\mathcal{T}\in\mathcal{G} \\ \mathcal{S}\cap\mathcal{T}\neq\emptyset}} E\left[I_{\mathcal{S}} I_{\mathcal{T}}\right]}{E[X(n)]^2} \leq \frac{\binom{m}{n} \sum_{i=1}^{n} \binom{n}{i} \binom{m-n}{n-i} p^{2n-i}}{\binom{m}{n}^2 p^{2n}} \\ &= \sum_{i=1}^{n} \frac{\binom{n}{i}\binom{m-n}{n-i}}{\binom{m}{n}} p^{-i} = \sum_{i=1}^{n} a_i \end{aligned} \tag{8.18}$$

where $a_i = \frac{\binom{n}{i}\binom{m-n}{n-i}}{\binom{m}{n}} p^{-i}$. Now define $b_i = a_{i+1}/a_i$ and note that since $b_i = \frac{(n-i)^2}{(i+1)(m-2n+i+1)} p^{-1}$ decreases with i, $b_i \leq b_1$. Therefore $a_i \leq b_1^{i-1} a_1$.

Recalling the value of p computed in (8.3):

$$p = \frac{e^{-\gamma N_0 B/P}}{(1+\gamma)^{n-1}}, \tag{8.19}$$

and since $n = (1-\epsilon)\log(M)/\log(1+\gamma)$ and $m \geq M/[\log(M)]^2$ we have:

$$\begin{aligned} b_1 &= \frac{(n-1)^2}{2(m-2n+2)}p^{-1} \\ &\leq \frac{[\log(M)]^2}{2[\log(1+\gamma)]^2} \cdot \frac{1}{M/[\log(M)]^2 - 2\log(M)/\log(1+\gamma)+2} \cdot e^{\gamma N_0 B/P - \log(1+\gamma)} \cdot M^{1-\epsilon} \\ &\leq M^{-\epsilon'} \end{aligned} \tag{8.20}$$

for any $\epsilon' < \epsilon$, for M large enough. It follows that $b_1 \to 0$ as $M \to \infty$, and we have $b_1 < 1$ for M large enough.

Then we can write:

$$\begin{aligned} \frac{\Delta}{\lambda^2} &\leq \sum_{i=1}^{n} a_i \leq a_1 \sum_{i=0}^{n-1} b_1^i \leq \frac{a_1}{1-b_1} = \frac{\binom{n}{1}\binom{m-n}{n-1}}{\binom{m}{n}} \frac{p^{-1}}{1-b_1} = n^2 \frac{\prod_{i=0}^{n-2}(m-n-i)}{\prod_{i=0}^{n-1}(m-i)} \frac{p^{-1}}{1-b_1} \\ &\leq \frac{n^2}{m}\frac{p^{-1}}{1-b_1} \\ &\leq \frac{[\log(M)]^4}{M[\log(1+\gamma)]^2} \cdot M^{1-\epsilon} \cdot \frac{e^{\gamma N_0 B/P - \log(1+\gamma)}}{1-M^{-\epsilon'}} \\ &\leq M^{-\epsilon'} \end{aligned} \tag{8.21}$$

for any $\epsilon' < \epsilon$, for M large enough.

Using Lemma 7 we have that at any stage k such that $m(k) > \frac{M}{[\log(M)]^2}$

$$P(X(n)=0) \leq \exp\left(-M^{\epsilon'}\right)$$

Therefore, the probability of the algorithm succeeding in $\frac{\log(1+\gamma)}{1-\epsilon} \cdot \frac{M}{\log(M)}$ consecutive stages can be lower bounded by:

$$\begin{aligned} P(\text{algorithm succeeds}) &\geq \left[1-\exp\left(-M^{\epsilon'}\right)\right]^{\frac{\log(1+\gamma)}{1-\epsilon}\cdot\frac{M}{\log(M)}} \\ &\approx \exp\left[-\frac{\log(1+\gamma)}{1-\epsilon} \cdot \frac{M}{\log(M)} \exp\left(-M^{\epsilon'}\right)\right] \longrightarrow 1 \text{ as } M \to \infty \end{aligned} \tag{8.22}$$

Therefore, the algorithm succeeds in finding feasible n user groups while $m(k) > M/[\log(M)]^2$ and can find a feasible allocation with at most $\frac{\log(1+\gamma)}{1-\epsilon} \cdot \frac{M}{\log(M)}(1+o(1))$ channels with probability going to 1 as $M \to \infty$. ■

8.3 Many broadcast systems and multiple access systems

We now consider the cases where users are grouped into systems and either the transmitters, or the receivers of the users of the same system are collocated. The main difference between these cases and the many SISO case analyzed before is that the channels gains $\{c_{i,j}\}_{i,j}$ are no longer independent. To simplify the analysis, we will assume that γ, the minimum SINR required to achieve rate R, is larger than 1[2]. This assumption together with the dependence among the channel gains of the users belonging to the same system imply that users of the same system must be assigned to different channels. Since this last requirement is the same regardless of the kind of system (multiple access, or broadcast) the results of this subsection can be applied to any combination of multiple access and broadcast systems sharing the same spectrum.

Since we are adding further constraints on feasible channel assignments with respect to the many SISO case, it follows that the lower bound of Theorem 8 still holds for the multiple access and broadcast cases for any number of users per system.

Also, the requirement of using different channels for the users of the same system implies that the number of required channels is lower bounded by the size of the largest system. It follows that in order to have any hope of being able to obtain a bandwidth scaling $\Theta(M/\log(M))$ as in the many SISO case, the size of the largest system cannot exceed a scaling of $\Theta(M/\log(M))$.

We will show that if the systems have size $s = \frac{2\log(1+\gamma)}{1-\epsilon} \cdot \frac{M}{\log(M)}$ an approximately similar upper bound on $W^*(M)$ to Theorem 9 holds (the only difference in the bounds is a factor of 2). Note that smaller values of s put fewer restrictions on feasible channel assignments, and result in a similar upper bound on $W^*(M)$.

[2]We expect the bandwidth scaling law to remain $\Theta(M/\log(M))$ if $\gamma \leq 1$ since we conjecture that the maximum number of users of the same system that can be assigned to the same channel would remain bounded for any fixed $\gamma > 0$.

Theorem 10 *Let $\epsilon > 0$ be arbitrarily small, and γ be the minimum* SINR *required to achieve rate R as defined in (8.1). In the many MAC and many BC cases with M users and $s = \frac{2\log(1+\gamma)}{(1-\epsilon)}\frac{M}{\log(M)}$ users per system , under the assumptions described in Section 8.1, the optimal number of required channels to satisfy the rate constraints $R_i \geq R$ satisfies*

$$W^*(M) \leq \frac{2\log(1+\gamma)}{(1-\epsilon)}\frac{M}{\log(M)}(1+o(1))$$

asymptotically, with probability going to 1 as $M \to \infty$.

Proof: The proof of this theorem is very similar to the proof of Theorem 9. The main difference is that we must ensure that the channel assignment algorithm assigns different channels to the users of the same system.

We will describe a channel assignment algorithm that achieves the upper bound with probability going to 1 as M goes to infinity. The algorithm proceeds in stages, one for each channel k. Let $m(k)$ be the number of users of each system that have not been allocated to a channel at stage k. Note that $m(1) = s$, the number of users per system. Also, let $n = \frac{1-\epsilon}{2\log(1+\gamma)}\log(M)$ be the number of systems. The algorithm can be described as follows:

`set` $k = 1$;

`while` $m(k) > \frac{M}{\log(M)^3}$

{

`find a group of` n `users (with no channel assignment) that` \\

`satisfy` $R_i \geq R$ `in channel` k`;`

`assign channel` k `to these` n `users;`

`set` $k = k + 1$`;`

}

`assign different` $\frac{(1-\epsilon)}{2\log(1+\gamma)}\frac{M}{\log(M)^2}$ `channels to the remaining users;`

Assume first that the algorithm succeeds in each stage in finding n users, each belonging to a different system, satisfying the rate constraints. Then the algorithm finds a feasible channel allocation that uses no more than $M/n + (1-\epsilon)M/(2\log(1+\gamma)\log(M)^2)$ channels, and it follows that $W^*(M) \leq M/n + M/(2\log(1+\gamma)\log(M)^2)$.

We now need to prove that the algorithm can find groups of n users satisfying the rate constraints in all the stages where $m(k) \geq M/\log(M)^3$.

We start by computing a lower bound on the probability that the algorithm succeeds in finding a group of n users satisfying the rate constraints at stage k, when there are $m(k)$ users in each system to choose from.

Without loss of generality, we label the $n \cdot m(k)$ users that have not yet been assigned a channel at stage k from 1 to $n \cdot m(k)$. Let $t(i)$ denote the system user i belongs to. Define $X(n)$ as:

$$X(n) = \sum_{\substack{\mathcal{S} \in \{1,\ldots,nm(k)\} \\ |\mathcal{S}|=n}} 1\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P} + \sum_{\substack{j \neq i \\ j \in \mathcal{S}}} c_{j,i}(k)} \geq \gamma \ \forall i \in \mathcal{S} \right)$$

Letting $\mathcal{G} = \{\mathcal{S} \subset \{1,\ldots,nm(k)\} : |\mathcal{S}| = n, t(i) \neq t(j) \ \forall i,j \in \mathcal{S}\}$ and $I_{\mathcal{S}} = \prod_{i \in \mathcal{S}} 1\left(\frac{c_{i,i}(k)}{\frac{N_0 B}{P} + \sum_{\substack{j \neq i \\ j \in \mathcal{S}}} c_{j,i}(k)} \geq \gamma \right)$ we can express $X(n) = \sum_{\mathcal{S} \in \mathcal{G}} I_{\mathcal{S}}$.

We want to find a lower bound on $P(X(n) \geq 1)$ or equivalently, an upper bound on $P(X(n) = 0)$, for which we use Lemma 7 with the current definition of $\mathcal{G}$, where we require that the users in every element of $\mathcal{G}$ belong to different systems. Note that even though the channel gains are not independent across all the users, they are i.i.d. for all the users in $\mathcal{S}$ for any $\mathcal{S} \in \mathcal{G}$ with the new definition of $\mathcal{G}$. From this Lemma we have:

$$P(X(n) = 0) \leq \exp\left(-\frac{\lambda^2}{\Delta}\right)$$

where $\lambda = E[X(n)]$ and $\Delta = \sum_{\mathcal{S} \in \mathcal{G}} \sum_{\substack{\mathcal{T} \in \mathcal{G} \\ \mathcal{S} \cap \mathcal{T} \neq \emptyset}} E\left[I_{\mathcal{S}} I_{\mathcal{T}}\right]$.

In order to use the Lemma 7 we will get an upper bound on Δ/λ^2. We first compute an upper bound on $E(I_{\mathcal{S}} I_{\mathcal{T}})$ where $\mathcal{S}, \mathcal{T} \in \mathcal{G}$ and $|\mathcal{S} \cap \mathcal{T}| = i$:

$$E\left[I_{\mathcal{S}} I_{\mathcal{T}}\right] = E\left[\prod_{u \in \mathcal{S}} 1\left(\frac{c_{u,u}(k)}{\frac{N_0 B}{P} + \sum_{\substack{l \in \mathcal{S} \\ l \neq u}} c_{l,u}(k)} \geq \gamma \right) \prod_{j \in \mathcal{T}} 1\left(\frac{c_{j,j}(k)}{\frac{N_0 B}{P} + \sum_{\substack{l \in \mathcal{T} \\ l \neq j}} c_{l,j}(k)} \geq \gamma \right) \right]$$

$$
\begin{aligned}
&\leq E\left[\prod_{u\in\mathcal{S}} 1\left(\frac{c_{u,u}(k)}{\frac{N_0B}{P}+\sum_{\substack{l\in\mathcal{S}\\ l\neq u}} c_{l,u}(k)} \geq \gamma\right) \prod_{j\in\mathcal{T}\setminus\mathcal{S}} 1\left(\frac{c_{j,j}(k)}{\frac{N_0B}{P}+\sum_{\substack{l\in\mathcal{T}\setminus\mathcal{S}\\ l\neq j}} c_{l,j}(k)} \geq \gamma\right)\right] \\
&= E\left[\prod_{u\in\mathcal{S}} 1\left(\frac{c_{u,u}(k)}{\frac{N_0B}{P}+\sum_{\substack{l\in\mathcal{S}\\ l\neq u}} c_{l,u}(k)} \geq \gamma\right)\right] E\left[\prod_{j\in\mathcal{T}\setminus\mathcal{S}} 1\left(\frac{c_{j,j}(k)}{\frac{N_0B}{P}+\sum_{\substack{l\in\mathcal{T}\setminus\mathcal{S}\\ l\neq j}} c_{l,j}(k)} \geq \gamma\right)\right] \\
&= \left[P\left(\frac{c_{u,u}(k)}{\frac{N_0B}{P}+\sum_{\substack{l\in\mathcal{S}\\ l\neq u}} c_{l,u}(k)} \geq \gamma\right)\right]^n \left[P\left(\frac{c_{j,j}(k)}{\frac{N_0B}{P}+\sum_{\substack{l\in\mathcal{T}\setminus\mathcal{S}\\ l\neq j}} c_{l,j}(k)} \geq \gamma\right)\right]^{n-i} \\
&= p_n^n \cdot p_{n-i}^{n-i} \qquad (8.23)
\end{aligned}
$$

where we defined

$$
p_r \triangleq P\left(\frac{c_{u,u}(k)}{\frac{N_0B}{P}+\sum_{\substack{l\in\mathcal{R}\\ l\neq u}} c_{l,u}(k)} \geq \gamma\right)
$$

for $u \in \mathcal{R} \subseteq \mathcal{G}$, with $|\mathcal{R}| = r$. Using the fact that $\{c_{u,u}(k), c_{l,u}(k)\}_{\substack{l\in\mathcal{R}\\ l\neq u}}$ are i.i.d. Exp(1) random variables, we can compute p_r explicitly:

$$
p_r = \int_0^\infty \int_{\gamma(N_0B/P+x)}^\infty e^{-c} \cdot \frac{x^{r-2}}{(r-2)!} e^{-x} dc \cdot dx = \frac{e^{-\gamma N_0 B/P}}{(1+\gamma)^{r-1}}.
$$

Using (8.23) we upper bound Δ/λ^2:

$$
\begin{aligned}
\frac{\Delta}{\lambda^2} &= \frac{\sum_{\mathcal{S}\in\mathcal{G}} \sum_{\substack{\mathcal{T}\in\mathcal{G}\\ \mathcal{S}\cap\mathcal{T}\neq\emptyset}} E\left[I_{\mathcal{S}} I_{\mathcal{T}}\right]}{E[X(n)]^2} \leq \frac{m^n \sum_{i=1}^n \binom{n}{i}(m-1)^{n-i} p_n^n \cdot p_{n-i}^{n-i}}{m^{2n} p_n^{2n}} \\
&= \sum_{i=1}^n \frac{\binom{n}{i}(m-1)^{n-i}}{m^n} \frac{p_{n-i}^{n-i}}{p_n^n} = \sum_{i=1}^n a_i \qquad (8.24)
\end{aligned}
$$

where $a_i = \frac{\binom{n}{i}(m-1)^{n-i}}{m^n} \frac{p_{n-i}^{n-i}}{p_n^n}$. Now define $b_i = a_{i+1}/a_i$ and note that since

$$
b_i = \frac{(n-i)}{(i+1)(m-1)} e^{2(n-i-1)\log(1+\gamma)+\gamma N_0 B/P}
$$

decreases with i, $b_i \leq b_1$. Therefore $a_i \leq b_1^{i-1} a_1$.

Since $n = \frac{1-\epsilon}{2\log(1+\gamma)} \log(M)$, assuming that $m \geq M/[\log(M)]^3$ we have:

$$
\begin{aligned}
b_1 &= \frac{(n-1)}{2(m-1)} e^{2(n-2)\log(1+\gamma)+\gamma N_0 B/P} \\
&\leq \frac{\log(M)}{4\log(1+\gamma)\left(M/\log(M)^3 - 1\right)} e^{(1-\epsilon)\log(M)+\gamma N_0 B/P} \\
&\leq M^{-\epsilon'} \qquad (8.25)
\end{aligned}
$$

for any $\epsilon' < \epsilon$, for M large enough. It follows that $b_1 \to 0$ as $M \to \infty$, and we have $b_1 < 1$ for M large enough. Then we can write:

$$\begin{aligned}
\frac{\Delta}{\lambda^2} &\leq \sum_{i=1}^{n} a_i \leq a_1 \sum_{i=0}^{n-1} b_1^i \leq \frac{a_1}{1-b_1} = \frac{1}{1-b_1} \cdot \frac{\binom{n}{1}(m-1)^{n-1}}{m^n} \frac{p_{n-1}^{n-1}}{p_n^n} \\
&\leq \frac{1}{1-b_1} \left(\frac{n}{m}\right) e^{\gamma N_0 B/P} (1+\gamma)^{2(n-1)} \\
&\leq \frac{1}{1-M^{-\epsilon'}} \cdot \frac{\log(M)}{2\log(1+\gamma)} \cdot \frac{(\log(M))^3}{M} e^{\gamma N_0 B/P + (1-\epsilon)\log(M)} \\
&\leq M^{-\epsilon'} \qquad (8.26)
\end{aligned}$$

for any $\epsilon' < \epsilon$, for M large enough.

Using Lemma 7 we have that at any stage k such that $m(k) > \frac{M}{[\log(M)]^3}$

$$P(X(n) = 0) \leq \exp\left(-M^{\epsilon'}\right)$$

Therefore, the probability of the algorithm succeeding in $\frac{2\log(1+\gamma)}{(1-\epsilon)} \frac{M}{\log(M)}$ consecutive stages can be lower bounded by:

$$\begin{aligned}
P(\text{algorithm succeeds}) &\geq \left[1 - \exp\left(-M^{\epsilon'}\right)\right]^{\frac{2\log(1+\gamma)}{(1-\epsilon)} \frac{M}{\log(M)}} \\
&\approx \exp\left[-\frac{2\log(1+\gamma)}{(1-\epsilon)} \frac{M}{\log(M)} \exp\left(-M^{\epsilon'}\right)\right] \longrightarrow 1 \text{ as } M \to \infty \qquad (8.27)
\end{aligned}$$

Therefore, the algorithm succeeds in finding feasible n user groups while $m(k) > M/[\log(M)]^3$ and can find a feasible allocation with at most $\frac{2\log(1+\gamma)}{1-\epsilon} \cdot \frac{M}{\log(M)} \cdot (1 + o(1))$ channels with probability going to 1 as $M \to \infty$. ■

Chapter 9

Discussion

When interference cannot be cancelled the best that one can do to avoid it is to orthogonalize the systems in time or frequency. Orthogonalizing systems corresponds to allocating different frequency channels (or time-slots) to the interfering systems. As the number of systems M grows, for a fixed bandwidth requirement of each system, orthogonalizing all the systems in frequency requires a linear growth of the required bandwidth with M. In order to avoid such a large increase in the use of the spectrum resource, one can attempt to reuse the same frequency between two or more systems. The network designer then faces the problem of allocating channels to systems so that the aggregate received interference in each system is minimized. Our statistical analysis of interference aggregation shows that a reasonable spectrum allocation approach is to allocate different frequencies to systems close to each other. If the spectrum sharing protocol can guarantee that no other system within a distance r_0 from the given system is allocated the same channel, then the aggregate received interference has finite mean and variance. Moreover, as r_0 is chosen larger and larger, the aggregate interference becomes more concentrated around its mean. This makes the performance of the protocol more predictable and results in a better quality of service. The probability distributions for the aggregate interference derived in Chapter 7 can be used to analyze the performance of different spectrum sharing protocols that trade off the value of r_0 with the bandwidth assigned to each system for communication. In particular, the choice $r_0 = 0$ corresponds to a spectrum sharing protocol with full frequency reuse (i.e. code division multiple access (CDMA)). The results of Chapter 7 suggest that CDMA may have very poor performance in spectrum sharing applications where the location of transmitters and receivers cannot be controlled. This is due to the excessive interference created by interferers located in the proximity of a given receiver.

In multi-user systems, when the fading process can be accurately tracked at the transmitter, channel variations become beneficial. This is because channel fluctuations can be exploited by scheduling users when their fading states are good. In this way one can *ride the peaks* of the fading process obtaining what is called multi-user diversity gains [51]. For Rayleigh fading the average rate of the system grows like $\log\log M$ for an M user system due to multi-user diversity. Therefore, to keep the rate of each user constant the bandwidth of the system should scale as $W(M) = \Theta(M/\log\log M)$. We see that in the single system case multi-user diversity provides a gain of a factor $\log\log M$. In Chapter 8 we investigated if fading could provide similar gains in a multi-system scenario, where the multiple systems interfere with each other.

Under some assumptions, we derived asymptotic lower and upper bounds on the bandwidth required to guarantee a minimum rate R per system. These bounds state that the required bandwidth scales as $W(M) = \Theta(M/\log M)$. From this expression, we observe a *multi-system diversity* gain of the order of $\log M$. Therefore, under our assumptions, fading provides significantly larger gains in the multi-system scenario than in the single system case. This is because when multiple systems are present we can exploit not only the good states of the channel gains, but also the bad states for interference avoidance. In essence, we allocate the same channel to systems that do not interfere much with each other. As the number of systems grows, the chance of finding more systems that can share the same channel increases. However, as the number of systems sharing the same channel grows, the aggregate interference increases and cannot be neglected. Our analysis tightly quantifies how many systems can share the same channel while keeping the aggregate interference under control to satisfy the rate constraint R. We also showed that similar multi-system diversity gains can be achieved when each system is either a broadcast system (downlink) or a multiple access system (uplink) with at most $\Theta(M/\log M)$ users, where M is the total number of users in the network.

Part III

Spectrum Sharing Between Selfish Systems

Chapter 10

Preliminaries

10.1 Introduction

In this last part of the thesis we look at the spectrum sharing problem from a different angle than in the first two parts. So far we have been concerned with characterizing the performance that can be achieved in a spectrum sharing situation between systems that cooperate with each other. In addition, in Part I we characterized the whole capacity region of the interference channel without specifying how to choose an operating point in this region. In practice, systems can be selfish and may need incentives to comply with a given rule or protocol. In addition, the operating point should be chosen to satisfy some fairness objective. In the rest of this thesis we analyze how the lack of cooperation, and the need to achieve efficiency and fairness influences the spectrum sharing protocol design[1].

We study a scenario where multiple wireless systems share the same spectrum. For concreteness consider a typical urban area with 802.11 networks, bluetooth systems, walkie-talkies, etc. co-existing and operating in the same unlicensed band, e.g. ISM, UNII, etc. The systems do not have

[1]Based on preliminary work that was presented in "Spectrum Sharing for Unlicensed Bands", by Raul Etkin, Abhay Parekh, and David Tse, First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Baltimore, Maryland, November 8-11, 2005.

a common goal and do not cooperate with each other. We assume that spectrum is a scarce resource, so that efficiency is a concern. We are interested in designing *spectrum sharing rules* and protocols which allow the systems to share the bandwidth in a way that is *fair*, *efficient* and *compatible* with the *incentives* of the individual systems.

A resource allocation is *efficient* if it is not possible to improve the performance of a given system without degrading the performance of some other system. Usually, there are many efficient operating points, each representing a different performance trade-off among the systems. *Fairness* is related to the relative performance among the systems. It can be achieved by optimizing a global utility function over the possible resource allocations. Different utilities represent different fairness goals. Finally, an allocation is *incentive compatible* or *self-enforcing* if there is no incentive for an individual system to deviate from it.

Current regulations provide a spectral mask which limits the total power and power spectral density (PSD) that each transmitter can use. This is an attempt to limit the amount of interference generated by each transmitter at other systems' receivers. This approach, however, may severely constrain the data rate when there are no other systems around. In addition, it does not provide any performance guarantees nor does it avoid unfair situations. This is because the rate of a system is limited by the *aggregate* received interference. Limiting the transmission power of an individual system does not always result in acceptable received interference when the number of interfering systems grows or the distance between the interferer and the receiver is small.

In order to attain efficient resource allocations in diverse situations the spectrum sharing rules must be flexible. The rules should adapt to each specific scenario, taking into account the number of systems sharing the spectrum in a given time and location, and considering the particular interaction between the systems. However, adding flexibility to selfish systems may lead to inefficient and unfair situations too. The following examples illustrate this point:

1. Two systems of similar power levels, treating each other as noise, as a result of greedy behavior may spread their power over the available bandwidth. From the selfish perspective

of a greedy system, using more bandwidth results in better performance. Spreading may in fact be a globally optimal behavior if the background noise level for each system (without the interference from the other system) is high. However, if the background noise level *is small*, this could result in vastly inferior performance relative to the regime in which the two systems orthogonalize their signals. We analyze this situation in more detail in Chapter 11.

2. A bluetooth system might be forced to communicate at a negligible rate by the interference of a much higher powered 802.11 transmitter, who transmits at full power to maximize its rate. For example, if the bluetooth device is a headset attached to an 802.11 based VOIP (voice over IP) phone, this could result in unacceptable voice quality.

The first example shows that a spectrum allocation may be fair but inefficient. Both systems get similar rates, but these rates may be small compared to the best achievable performance. The second example shows how asymmetry between the systems (in this case due to different power capabilities) may lead to an unfair situation. We see that it is important for an allocation to *simultaneously* be fair and efficient. In both cases one could obtain an efficient and fair sharing by designing appropriate spectrum sharing rules (for example by assigning different frequency bands to the systems). However, as the above examples show, selfish systems may not have an incentive to follow these rules.

The issue of incentives will play a central role in our analysis. We model the utility of a system to be a function of rate, and incorporate power in the form of a maximum constraint. We also make the following assumptions:

A1 The allocation is determined as the outcome of a game.

A2 For the purposes of computing system rates, the interference from other systems is treated as noise.

A3 Each system calculates a set of operating points which are Nash Equilibria for the game.

A4 When there is more than one Nash Equilibrium point, the systems follow a convention (this could be an FCC mandate) to pick one of these points. Clearly, no system has any incentive to unilaterally deviate from this point.

The key problem is therefore to design spectrum sharing rules which lead to a Nash equilibrium that is fair and efficient. These rules are *self enforcing*, and as a result do not require the intervention of an external authority to verify compliance. This is of particular importance with the advent of new technologies like software-defined radios, which are inherently hard to certify and easy to alter.

We start by formulating a one shot game in which each system chooses its power allocation once and for all, and this yields the data rates at the operating point. We find, by extending a result of [13] that in low interference situations, the full-spread equilibrium is the only possible outcome of the game. And in many cases, the rates that result from the full-spread equilibrium are suboptimal (inefficient, unfair, or both). This is a negative result from the point of view of designing a standard, as it would be desirable to have multiple equilibrium points to choose from.

However, systems operate and have to co-exist over a long period of time (days, months, years), and in this context, it may be more reasonable to model the scenario as a repeated game where systems play multiple rounds, remembering the past experience in the choice of the power allocation in the next round. The situation brightens considerably once one considers repeated games. We show that in a repeated game, any vector of rates in the achievable region that is component-wise larger than the full-spread rates can be supported. Therefore, if the optimal rate vector (efficient and fair according to some global objective) is component-wise greater than the full-spread rate vector, there is no performance loss due to lack of cooperation.

In essence, the systems can enforce any such operating point by threatening to apply a punishment if any individual system deviates from the power allocation that achieves the desired operating point. The punishment is simply to apply the power allocation at the Nash equilibrium of the static game whose rates are component-wise smaller.

The computation of the fair and efficient operating point and the choice of strategies that achieve it require common knowledge of all the parameters (channel gains, power constraints, etc.).

These parameters have to be measured and communicated among the systems. A selfish system may have an incentive to falsify some of its parameters in order to obtain an advantage over the other systems. We show that in most cases, dishonest behaviors can be detected and punished. Therefore, by adding punishment to the parameter measurement process we can incentivize the systems to behave truthfully.

In our work we provide a unified framework to study the issues of efficiency, fairness and incentive compatibility in a non-cooperative spectrum sharing situation. Individual aspects of the problem have been considered separately in many related papers. For example, Cioffi et al. propose in [13] the iterative waterfilling algorithm (IWA) to obtain good power spectral allocations in digital subscriber lines (DSL)[2]. Game theory is used to find equilibrium points of the algorithm, but selfish behavior is never considered as part of the problem. Sufficient conditions for the uniqueness of the equilibrium of the IWA (which coincide with the Nash equilibria of the Gaussian Interference Game to be discussed in Chapter 12) have been presented in [36]. Popescu, et al. observed in [39] that iterative waterfilling may lead to inefficient solutions. Trying to overcome some of the difficulties of IWA, Huang, et al. propose in [30] to take into account the interference created onto other systems by exchanging "interference prices". Under some assumptions on the individual utilities of the systems and the global utility to be maximized, their algorithm converges to the optimal solution. Their analysis, however, assumes cooperation among the systems. Obtaining efficient and fair allocations requires solving optimization problems. The problem of optimizing resource allocations has been studied in [10, 11, 48, 53, 54]. Many of these works rely on relaxations (which may not be applicable to all scenarios) to solve the optimization with tractable complexity. Clemens and Rose study in [14] a repeated game between selfish players in a wireless model. They use a genetic algorithm to find *good* strategies in a limited strategy space. Their analysis, however, does not consider efficiency and fairness issues, and the strategies that they find may not be incentive compatible. [37] considers the issues of fairness and efficiency in non-cooperative wireless applications, and proposes the use of

[2] Since the channel models of the DSL and the wireless communication problems are essentially the same, the results for DSL are applicable to the wireless setting.

punishment to achieve a desired operating point. The games considered in [37], however, are games of complete information with small strategy spaces, and are very different from the games considered here.

The rest of Part III is organized as follows. In Section 10.2 we present the model to be used in the following Chapters. Chapter 11 focuses on the issues of fairness and efficiency in an ideal situation where systems cooperate with each other. Chapter 12 analyzes non-cooperative situations. Section 12.3 considers the problem of channel measurement and exchange between possibly dishonest systems. In Chapter 13 we introduce more general models than that of Section 10.2 and show that our main results hold in a broader setting where we allow the systems to use arbitrary codes (i.e. non Gaussian) and communication strategies (i.e. implement interference cancellation), and the systems are formed by more than one transmitter or receiver. Finally Chapter 14 presents some conclusions and open problems.

Preliminary versions of this research have been presented in [21] and [22].

10.2 Model

We model a situation in which M systems, each formed by a single transmitter-receiver pair, coexist in the same area. Consider an M user Gaussian interference channel in discrete time defined by:

$$y_i[n] = \sum_{j=1}^{M} h_{j,i} x_j[n] + z_i[n]; \text{ i=1,\ldots,M} \tag{10.1}$$

where $x_i, y_i, z_i \in \mathbb{C}$ and the noise processes are i.i.d. over time with $z_i \sim \mathcal{CN}(0, N_0)$. By assuming that the channel from each transmitter to each receiver has a single tap we are restricting attention to the case of flat fading. The input of user i has an average power constraint P_i.

We will assume that each system treats the received interference as noise. This leads to a tractable inner bound to the capacity region of the interference channel that we studied in Part I when the number of systems M is larger than two. In addition, practical limitations such as decoder

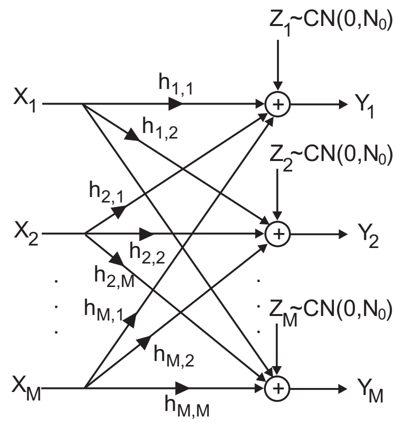


Figure 10.1: MxM Gaussian interference channel.

complexity, uncertainty in the estimation of $\{h_{j,i}\}$, delay constraints, etc., may preclude the use of interference cancellation techniques. Therefore the assumption of treating interference as noise may be realistic in many cases. Moreover, most of our game theoretic results of Chapter 12 extend to any achievable rate region of the interference channel. Therefore, the assumption of treating interference as noise is not essential for the validity of most of our game theoretic results.

Finally, we will assume that the systems use random Gaussian codebooks, which means that the transmitted signals look like white Gaussian processes. In Appendix A we consider a model where the codebooks of the systems are allowed to be non-Gaussian, and show that most of our results extend to this scenario.

Under these assumptions, using the capacity expression for the single user Gaussian channel, we can determine the maximum rate that system i can achieve for specific power allocations[3]:

$$R_i = \int_0^W \log\left(1 + \frac{c_{i,i}p_i(f)}{N_0 + \sum_{j\neq i} c_{j,i}p_j(f)}\right) df \tag{10.2}$$

where $p_i(f)$ is the power spectral density of the input signal of system i, and where for convenience we defined $c_{i,j} = |h_{i,j}|^2$. Note that due to the power constraints, $p_i(f)$ must satisfy:

$$\int_0^W p_i(f)df \leq P_i \tag{10.3}$$

[3]In all cases we use $\log(\cdot)$ for a base 2 logarithm.

The spectrum sharing problem that we consider is to determine a set of power allocations $\{p_i(f)\}$ for the M systems, that maximizes a given global utility function while satisfying the power constraints. This maximization results in allocations that are fair and efficient in a cooperative scenario, i.e. free from the problem of incentives. In the next section we study the structure of the optimal power allocations for any reasonable choice of global utility.

Chapter 11

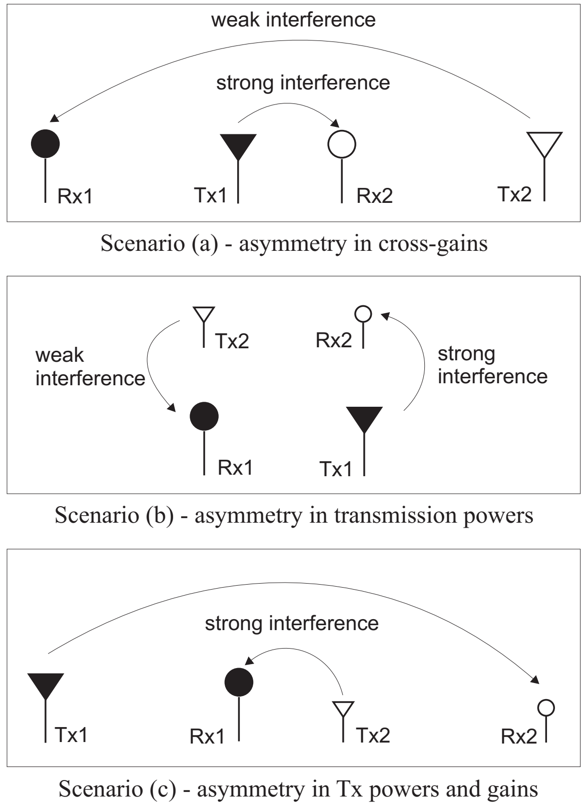
Cooperative Scenarios

11.1 Optimal Spectrum Allocations

Many of the fairness issues in spectrum sharing arise due to asymmetries between the systems. Figure 11.1 shows three different examples where two systems operate in asymmetric situations. In scenario (a) both systems have similar power capabilities (e.g. two 802.11 systems) but due to the locations of the transmitters and receivers, one system receives large interference while the other does not. Scenarios (b) and (c) describe situations where a high power system (e.g. 802.11 system) shares spectrum with a low power system (e.g. bluetooth system). In (b) all the gains are comparable, so intuitively the weak system is in disadvantage. In (c) due to asymmetry in the gains both systems can interfere with each other and one can imagine that a more fair situation may result.

For concreteness we assign specific parameter values to each scenario. Without loss of generality we can assume in all cases that $c_{1,1} = c_{2,2} = 1$, $N_0 = 1$ and $W = 1$. In scenario (a) we can assume $P_1 = P_2 = 10$, $c_{1,2} = 10$ and $c_{2,1} = 0.5$. For scenario (b) we set $P_1 = 10$, $P_2 = 1$, and $c_{1,2} = c_{2,1} = 1.1$. Finally in (c) we set $P_1 = 10$, $P_2 = 1$, $c_{1,2} = 0.5$ and $c_{2,1} = 10$.

Imagine that in these three scenarios we want to maximize some global utility function



Scenario (a) - asymmetry in cross-gains

Scenario (b) - asymmetry in transmission powers

Scenario (c) - asymmetry in Tx powers and gains

Figure 11.1: Three examples of asymmetric situations between two systems sharing the same band. The sizes of the antennas represent power capabilities, and smaller distances indicate higher gains.

$U(R_1, R_2)$, that represents some fairness objective. We are interested in determining the maximum value of U and the corresponding power spectral allocations that achieve it. In this section we will show how to solve this problem efficiently.

We assume that all the parameters are know to all the systems performing the optimization. In particular, we assume that the number of systems sharing the spectrum is common knowledge. The difficulty of detection has been pointed out in [40], however, we will assume in this paper that no such limitations exist. Practical algorithms for the estimation and exchange of parameters are presented in Section 12.3.

Let $\mathcal{R}$ be the achievable rate region:

$$\mathcal{R} = \left\{ \mathbf{R} : R_i = \int_0^W \log \left(1 + \frac{c_{i,i} p_i(f)}{N_0 + \sum_{j \neq i} c_{j,i} p_j(f)} \right) df \right.$$

$$\text{and} \int_0^W p_i(f)df \leq P_i \text{ with } p_i(f) \geq 0 \text{ for } i = 1, \ldots, M \Big\} \tag{11.1}$$

where $\mathbf{R} = (R_1, R_2, \ldots, R_M)$ and let $\mathcal{R}^*$ be the set of Pareto optimal points of $\mathcal{R}$:

$$\mathcal{R}^* = \Big\{(R_1, \ldots, R_M) \in \mathcal{R} : R_i \geq \tilde{R}_i \ \forall (R_1, \ldots, R_{i-1}, \tilde{R}_i, R_{i+1}, \ldots, R_M) \in \mathcal{R}, \text{ for } i = 1, \ldots, M\Big\} \tag{11.2}$$

In words, a rate allocation is Pareto optimal (or efficient) if it is not possible to increase the rate of any system without decreasing the rate of some other system.

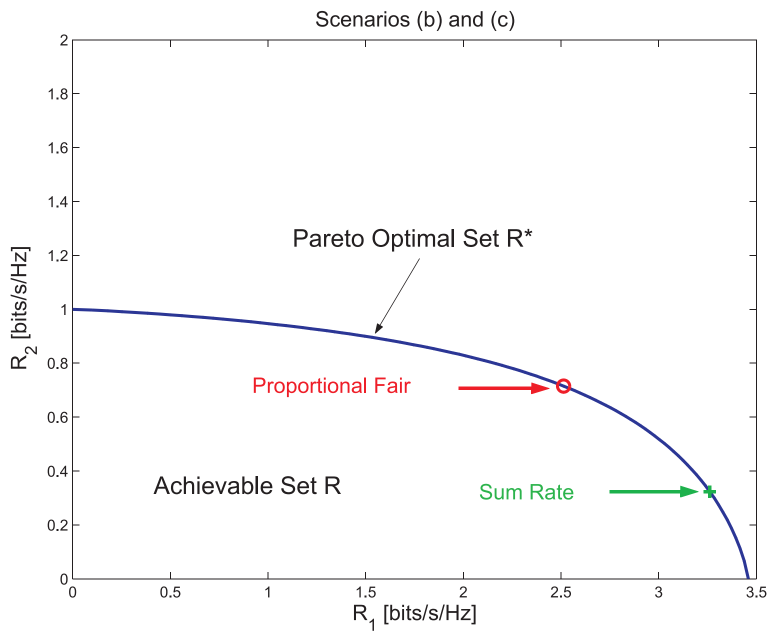


Figure 11.2: Achievable set $\mathcal{R}$ and Pareto efficient set $\mathcal{R}^*$ for scenarios (b) and (c). Indicated in the figure are the optimal sum rate, and proportional fair points.

Figure 11.2 shows the achievable set and Pareto optimal set for scenarios (b) and (c). The reason why for this specific choice of parameters in both scenarios we obtain the same sets will be explained later in this section.

The choice of the utility function will strongly influence the fairness in the resulting allocations. For example we may consider $U_{sum}(R_1, R_2) = R_1 + R_2$ if we are interested in maximizing the total sum rate. While in scenario (a) this choice of utility results in an optimal operating point where $R_1 = R_2$, in scenarios (b) and (c) the resulting optimal allocations are very unfair for system

2 ($R_2 \ll R_1$) (see Figure 11.2). A more fair allocation results from choosing the proportional fair metric $U_{PF}(R_1, R_2) = \log(R_1) + \log(R_2)$ proposed in [33]. By applying the $\log(\cdot)$ function to each rate, we give higher priority to the system in disadvantage. We can see in Figure 11.2 how in scenarios (b) and (c) the use of the proportional fair metric results in a more fair allocation. Note that in scenario (a) the use of U_{PF} results in the same rates as when U_{sum} is used.

For any utility function that is component-wise monotonically increasing in $(R_1, \ldots, R_M)$, the optimal rate allocation must occur in a point of the boundary $\mathcal{R}^*$. So it is of interest to obtain a simple characterization for $\mathcal{R}$ and $\mathcal{R}^*$. For example, if we knew $\mathcal{R}^*$ the problem of maximizing U_{sum} reduces to finding the point in $\mathcal{R}^*$ that is tangent to the line of slope (-1).

At first glance, computing $\mathcal{R}$ requires to search over all possible power allocations $p_i(f)$ that satisfy the power constraint. Since $p_i(f)$ are functions with arbitrarily many degrees of freedom, the computation of $\mathcal{R}$ seems to be an infinite dimensional problem. However the following theorem shows that we can restrict attention to piecewise constant power allocations, and as a result, the problem of computing $\mathcal{R}$ has finite dimension.

Theorem 11 *Any point in the achievable rate region $\mathcal{R}$ defined in (11.1) can be obtained with M power allocations that are piecewise constant in the intervals $[0, w_1), [w_1, w_2), \ldots, [w_{2M-1}, W]$, where $w_i \leq w_{i+1}, i = 1, \ldots, 2M-2$, for some choice of $\{w_i\}_{i=1}^{2M-1}$.*

Proof: We prove the theorem by defining a rate region $\tilde{\mathcal{R}}$ and showing that it can be achieved with piece-wise constant power allocations over at most $2M$ intervals. We then show that $\mathcal{R} \subseteq \tilde{\mathcal{R}}$.

Define the power vector

$$\mathbf{p} = (p(1), \ldots, p(M)) \in \mathbb{R}^M$$

and the rate vector

$$\mathbf{r} = (r(1), \ldots, r(M)) \in \mathbb{R}^M$$

where each rate is computed assuming frequency flat power allocations over the bandwidth W, i.e.

$$r(i) = W \log\left(1 + \frac{c_{i,i}p(i)}{N_0 W + \sum_{j\neq i} c_{j,i}p(j)}\right)$$

Also define $\mathcal{A} \subset \mathbb{R}^{2M}$ by

$$\mathcal{A} = \{(\mathbf{r}, \mathbf{p}) : \mathbf{p} \geq 0\}$$

We will first show that if $\mathcal{B} = \text{Convex Hull}\{\mathcal{A}\}$ any point in $\mathcal{B}$ can be achieved with a convex combination of at most $2M$ points of $\mathcal{A}$. This follows directly from Carathéodory's theorem and it's extension (see [18], Theorem 18) by noting that $\mathcal{A}$ lies in $\mathbb{R}^{2M}$ and it is connected.

We then define the rate region

$$\tilde{\mathcal{R}} = \{\mathbf{r} : (\mathbf{r}, \mathbf{p}) \in \mathcal{B} \text{ and } \mathbf{p} \leq (P_1, \ldots, P_M)\}$$

and show that $\mathcal{R} \subseteq \tilde{\mathcal{R}}$. Note that there is a one to one correspondence between points in $\tilde{\mathcal{R}}$ and piece-wise constant power allocations that satisfy the power constraints and are defined over at most $2M$ intervals.

We need to show that $\mathbf{r} \in \mathcal{R} \Rightarrow \mathbf{r} \in \tilde{\mathcal{R}}$. We will do this by first showing that if $\mathbf{r} \in \mathcal{R}$ then there is a sequence of rates $\{\mathbf{r}_n\}_n$ with $\mathbf{r}_n \in \tilde{\mathcal{R}}$ such that $\lim_{n\to\infty} \mathbf{r}_n = \mathbf{r}$. Finally, by showing that $\tilde{\mathcal{R}}$ is compact we will show that $\mathbf{r} \in \tilde{\mathcal{R}}$.

$\mathbf{r} \in \mathcal{R}$ implies that there exist power allocations $\{p_i(f)\}_{i=1}^M$ that satisfy the power constraints and that result in the rate vector $\mathbf{r} = (R_1, \ldots, R_M)$. Fix n, and partition the interval $[0, W)$ into n intervals $\{\mathcal{W}_k\}_{k=1}^n$ defined by $\mathcal{W}_k = [(k-1)W/n, k \cdot W/n)$. Let $\lambda_k = 1/n$ for $k = 1, \ldots, n$. Let $\tilde{p}_k(i) = W \cdot \inf_{f\in\mathcal{W}_k} p_i(f)$ and $\tilde{r}_k(i) = W \log\left(1 + \frac{c_{i,i}\tilde{p}_k(i)}{N_0 W + \sum_{j\neq i} c_{j,i}\tilde{p}_k(j)}\right)$ for $i = 1, \ldots, M$ and $k = 1, \ldots, n$. Finally let $\mathbf{p}_n = \sum_{k=1}^n \lambda_k \tilde{\mathbf{p}}_k$ and $\mathbf{r}_n = \sum_{k=1}^n \lambda_k \tilde{\mathbf{r}}_k$. Since $\int_0^W p_i(f)df = \sum_{k=1}^n \int_{\mathcal{W}_k} p_i(f)df \geq \sum_{k=1}^n \int_{\mathcal{W}_k} \tilde{p}_k(i)/W df = \sum_{k=1}^n W/n \cdot \tilde{p}_k(i)/W = \sum_{k=1}^n \lambda_k \tilde{p}_k(i) = p_n(i)$ for $i = 1, \ldots, M$, it follows that $\mathbf{p}_n \leq (P_1, \ldots, P_M)$ and hence $\mathbf{r}_n \in \tilde{\mathcal{R}}$. Also $\mathbf{r}_n$ is an approximation to the integral that defines $\mathbf{r}$ so it converges to it as the number of intervals n in which $[0, W]$ is partitioned goes to ∞.

To show that $\tilde{\mathcal{R}}$ is compact, we need to show that it is closed and bounded. We first show that $\tilde{\mathcal{R}}$ is bounded. Let $\mathbf{r}$ be any vector in $\tilde{\mathcal{R}}$. Then we have:

$$\begin{aligned} r(i) &= \sum_{k=1}^{2M} \lambda_k \tilde{r}_k(i) \\ &= \sum_{k=1}^{2M} \lambda_k W \log\left(1 + \frac{c_{i,i}\tilde{p}_k(i)}{N_0 W + \sum_{j\neq i} c_{j,i}\tilde{p}_k(j)}\right) \\ &\leq W \log\left(1 + \frac{c_{i,i}\sum_{k=1}^{2M}\lambda_k \tilde{p}_k(i)}{N_0 W}\right) \\ &\leq W \log\left(1 + \frac{c_{i,i}P_i}{N_0 W}\right) \end{aligned}$$

We finally show that $\tilde{\mathcal{R}}$ is closed by showing that it contains its boundary. Since $\tilde{\mathcal{R}}$ is convex, any point $\mathbf{r}^*$ in its boundary can be obtained as the solution to the optimization problem:

$$\begin{aligned} \text{Maximize} \quad & \sum_{i=1}^{M} u_i r(i) \\ \text{Subject to:} \quad & (\lambda_1, \ldots, \lambda_{2M}) \geq 0 \\ & \sum_{k=1}^{2M} \lambda_k = 1 \\ & r(i) = \sum_{k=1}^{2M} \lambda_k W \log\left(1 + \frac{c_{i,i}\tilde{p}_k(i)}{N_0 W + \sum_{j\neq i} c_{j,i}\tilde{p}_k(j)}\right) \\ & \tilde{\mathbf{p}}_k \geq 0,\ k = 1, \ldots, 2M \\ & \sum_{k=1}^{2M} \lambda_k \tilde{\mathbf{p}}_k \leq (P_1, \ldots, P_M) \end{aligned} \tag{11.3}$$

for some choice of $\mathbf{u} = (u_1, \ldots, u_M)$. We need to show that the optimal value $\sum_{i=1}^{M} u_i r^*(i)$ and the associated rate vector $\mathbf{r}^* = (r^*(1), \ldots, r^*(M))$ are achieved within the optimization region of (11.3).

We will show that we can restrict attention to bounded power vectors. If some vector $\tilde{\mathbf{p}}_k$ had an arbitrarily large component, $\tilde{p}_k(i)$ say, the corresponding weight λ_k would need to be arbitrarily small (at most $P_i/\tilde{p}_k(i)$) to satisfy the power constraint. Since $\lim_{x\to\infty}(1/x)\log(1+x) = 0$ this arbitrarily large power would have a negligible effect in the corresponding rate. It follows that the result of the optimization (11.3) is not altered by upper bounding the vectors $\tilde{\mathbf{p}}_k$. That is, for fixed $\mathbf{u}$ there exists a vector $\bar{\mathbf{p}}$ such that if we add the constraint $\tilde{\mathbf{p}}_k \leq \bar{\mathbf{p}}$ for $k = 1, \ldots, 2M$ in (11.3) the maximization problem remains the same. In this modified optimization problem the objective is a continuous function and the optimization region is a compact set, so the optimum is attained within the region. ■

Note that once we fix the choice of intervals to obtain a point in $\mathcal{R}$, the M power allocations are constant in the same intervals.

The result of Theorem 11 arises naturally from geometric considerations in our model where frequency is a continuous variable. In the digital subscriber line (DSL) and wireless communications literature the available bandwidth is often divided into N discrete channels, and frequency is treated as a discrete parameter [13, 30, 53, 54]. This approximation of continuous frequency into discrete channels makes the complexity of the resulting optimization scale with N. In the flat fading case this is an artifact of the discretization of frequency, since the true complexity of the problem is a function of the number of systems M and not the number of channels N.

For the special case of channels satisfying a pairwise high interference condition (which is satisfied with the choice of parameters in scenarios (a), (b) and (c)), it turns out that the optimal power allocations are orthogonal, and hence the characterization of $\mathcal{R}^*$ is further simplified.

Theorem 12 *Let* $(R_1, \ldots, R_M)$ *be a Pareto efficient rate vector achieved with power allocations* $\{p_i(f)\}_{i=1,\ldots,M}$. *If* $c_{i,j}c_{j,i} > c_{i,i}c_{j,j}$ *then the power allocations* $p_i(f)$ *and* $p_j(f)$ *are orthogonal, i.e.* $p_i(f)p_j(f) = 0$ *for* $f \in [0, W]$.

Proof: We use Theorem 11 to restrict attention to piece-wise constant power allocations.

We will prove the theorem by contradiction. Assume that $p_i(f)$ and $p_j(f)$ are not orthogonal in some interval, $\mathcal{W}_k$ say. Then, $p_i(f) = p_i > 0$ and $p_j(f) = p_j > 0$ for $f \in \mathcal{W}_k$. Also let $p_r(f) = p_r$ for $f \in \mathcal{W}_k$, $r \neq i, j$. Finally let $I_i = N_0 + \sum_{r \neq i,j} c_{r,i} p_r$ ($I_j = N_0 + \sum_{r \neq i,j} c_{r,j} p_r$) be the noise plus interference PSD received by system i (j) without taking into account the interference received from system j (i). We will show that the rates R_i and R_j can be increased simultaneously without decreasing the rates R_r, $r \neq i, j$, thus showing that $p_i(f)$ and $p_j(f)$ cannot result in Pareto efficient rates.

We first split $\mathcal{W}_k$ into two equal intervals $\mathcal{W}_{k,1}$ and $\mathcal{W}_{k,2}$. We then modify $p_i(f)$ in $\mathcal{W}_k$ as follows: $p_i(f) = p_i + \Delta_{i,1}/2$ for $f \in \mathcal{W}_{k,1}$ and $p_i(f) = p_i - \Delta_{i,1}/2$ for $f \in \mathcal{W}_{k,2}$. Note that the difference in power levels between the two intervals is $\Delta_{i,1}$, which we choose to be sufficiently small. We now run the iterative waterfilling algorithm[1] restricted to the interval $\mathcal{W}_k$ with power constraints $p_i|\mathcal{W}_k|$ and $p_j|\mathcal{W}_k|$, starting with the power allocation $p_j(f)$. When system j waterfills first, it observes a noise+interference with a difference in power levels between $\mathcal{W}_{k,1}$ and $\mathcal{W}_{k,2}$ of $c_{i,j}\Delta_{i,1}$. After waterfilling, its own power allocation $p_j(f)$ has a difference $\Delta_{j,1} = -\frac{c_{i,j}}{c_{j,j}}\Delta_{i,1}$ between the two sub-intervals. When system i waterfills, it observes a noise plus interference with difference $c_{j,i}\Delta_{j,1} = -\frac{c_{j,i}c_{i,j}}{c_{j,j}}\Delta_{i,1}$ and after waterfilling, its power allocation has a difference in power levels $\Delta_{i,2} = \frac{c_{i,j}c_{j,i}}{c_{i,i}c_{j,j}}\Delta_{i,1}$. Since by assumption $\frac{c_{i,j}c_{j,i}}{c_{i,i}c_{j,j}} > 1$, the power differences $\Delta_{i,n}$ and $\Delta_{j,n}$ increase in magnitude for increasing n, until the algorithm converges. Once the algorithm converges the resulting power allocations in the interval $\mathcal{W}_k$ satisfy the properties:

- The power differences have positive magnitude: $|\Delta_{i,\infty}| > 0$ and $|\Delta_{j,\infty}| > 0$.
- The power allocations are waterfilling solutions to each other.

We will now show that these allocations, restricted to the interval $\mathcal{W}_k$, result in higher rates for all systems than the initial flat allocations. Let $R_i(\Delta_j)$ be the rate associated to the interval $\mathcal{W}_k$ when system i uses a flat power allocation and system j uses a power allocation with a difference Δ_j between the intervals $\mathcal{W}_{k,1}$ and $\mathcal{W}_{k,2}$.

$$\begin{aligned} R_i(\Delta_j) &= \frac{|\mathcal{W}_k|}{2}\log\left[1 + \frac{c_{i,i}p_i}{I_i + c_{j,i}(p_j - \Delta_j/2)}\right] \\ &+ \frac{|\mathcal{W}_k|}{2}\log\left[1 + \frac{c_{i,i}p_i}{I_i + c_{j,i}(p_j + \Delta_j/2)}\right] \end{aligned}$$

Differentiating with respect to Δ_j:

$$R_i'(\Delta_j) = \frac{|\mathcal{W}_k|c_{i,i}c_{j,i}p_i}{4\log_e(2)} \cdot \left\{ \frac{1}{[I_i + c_{j,i}(p_j - \Delta_j/2) + c_{i,i}p_i]} \cdot \frac{1}{[I_i + c_{j,i}(p_j - \Delta_j/2)]} \right.$$

[1]The iterative waterfilling algorithm consists of letting each system distribute its available power over the noise+interference seen as if it was pouring liquid over a container. See [52] for a detailed explanation.

$$- \quad \frac{1}{[I_i + c_{j,i}(p_j + \Delta_j/2) + c_{i,i}p_i]} \cdot \frac{1}{[I_i + c_{j,i}(p_j + \Delta_j/2)]}\Bigg\}$$

so it follows that $R_i'(\Delta_j) > 0$ for $\Delta_j > 0$ and $R_i'(\Delta_j) < 0$ for $\Delta_j < 0$. Therefore $R_i(\Delta_j)$ is minimized for $\Delta_j = 0$. Equivalently, for fixed total noise plus interference power, assuming that system i uses a flat power allocation, its rate is minimized when the total noise plus interference PSD is white. In addition, if system i waterfills over the noise plus interference seen, its rate can only increase with respect to the one obtained with a flat power allocation. So we conclude that after the iterative waterfilling algorithm converges, the rate of system i is larger than the initial rate obtained with flat allocations for systems i and j. A similar conclusion can be made for the rate of system j after convergence of iterative waterfilling.

In addition, the rates of the other systems do not decrease with the modified power allocations of systems i and j. If for example system r had zero power in the band k (i.e. $p_k = 0$) then modifying the power allocations of systems i and j does not change its rate. If on the other hand system r had positive power in the band, the result of modifying the power allocations of systems i and j using iterative waterfilling is to change the total noise plus interference seen by system r from a white PSD to a colored PSD of equal power. By the analysis above, this change in the PSD of the noise plus interference can only increase R_r. ■

The condition $c_{i,j}c_{j,i} > c_{i,i}c_{j,j}$ means that for systems i and j, the product of the channel cross gains $c_{i,j}c_{j,i}$ is greater than the product of the channel direct gains $c_{i,i}c_{j,j}$. Note that the condition can be satisfied even if one of the cross gains is small, by having the other cross gain large enough. Also, note that the condition is independent of the power constraints $\{P_i, P_j\}$ and noise variance N_0. In our three examples, we chose $c_{1,1} = c_{2,2} = 1$, so to check whether the condition of Theorem 12 is satisfied, we only need to check that $c_{1,2}c_{2,1} > 1$. In scenario (a) we have $c_{1,2}c_{2,1} = 10 \cdot 0.5 = 5$, in scenario (b) we have $c_{1,2}c_{2,1} = 1.1 \cdot 1.1 = 1.21$, and in scenario (c) we have $c_{1,2}c_{2,1} = 0.5 \cdot 10 = 5$, so in the three cases the condition is met.

In particular, if $c_{i,j}c_{j,i} > c_{i,i}c_{j,j}$ for any $i \neq j$, $j = 1, \ldots, M$, we can achieve any Pareto efficient rate vector with frequency division multiplexing (FDM). In this case, the the maximization of any concave, non-decreasing function of $(R_1, \ldots, R_M)$ reduces to the optimization problem:

$$\begin{aligned} \text{Maximize} \quad & U\left[W_1 \log\left(1 + \frac{c_{1,1}P_1}{W_1 N_0}\right), \ldots, W_M \log\left(1 + \frac{c_{M,M}P_1}{W_M N_0}\right)\right] \\ \text{subject to:} \quad & \sum_{i=1}^{M} W_i = W \text{ and } W_i \geq 0, i = 1, \ldots, M \end{aligned} \tag{11.4}$$

which, as can be easily verified, is a concave optimization problem and can be efficiently solved. In particular, the weighted sum rate utility $U_{WS}(R_1, \ldots, R_M) = \sum_{i=1}^{M} u_i R_i$ for non-negative weights $\{u_i\}_i$, and the proportional fair utility $U_{PF} = \sum_{i=1}^{M} \log(R_i)$ result in concave problems. These results allow to easily compute the rates in Figure 11.2.

Note that since the Pareto efficient rates are obtained with orthogonal allocations when $c_{1,2}c_{2,1} > 1$ (for direct gains equal to 1), the actual values of the cross gains $c_{1,2}$ and $c_{2,1}$ have no influence on the achievable region. This explains why scenarios (b) and (c) result in the same achievable region and optimal rates.

When the conditions of Theorem 12 are not satisfied, we can use techniques such as Lagrangian methods to solve the problem of maximizing $U(R_1, \ldots, R_M)$ with tractable complexity.

Chapter 12

Non-cooperative Scenarios

Throughout the previous chapter we have implicitly assumed that the M systems cooperate to maximize a global utility function by choosing appropriate power allocations. This assumption may be realistic when the different systems are jointly designed with a common goal, are complying with some standard or regulation, or are in fact transmitter-receiver pairs of a single global system.

However, in a spectrum sharing scenario where regulations may be lax and systems may be competing with one another to gain access to the shared medium, assuming selfish behavior may be more realistic. In this chapter we analyze how the lack of cooperation among systems may affect the set of achievable rates.

We will consider the same model introduced in Section 10.2 under the assumption that the different systems behave selfishly and rationally. We associate to each system i a utility function $U_i(R_i)$, which we assume concave and increasing in R_i[1]. The systems are selfish in the sense that they only try to maximize their own utility. The rationality assumption means that each system will never choose a strictly dominated strategy[2]. We analyze the set of achievable rates in this

[1]Notice the difference between the utility $U_i(R_i)$ of each individual system vs. the global utility $U(R_1, \ldots, R_M)$ introduced in Chapter 11.

[2]A strategy s_i for player i is strictly dominated by strategy s_i' if $U_i(s_1, \ldots, s_{i-1}, s_i', s_{i+1}, \ldots, s_M) < U_i(s_1, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_M))$ for each $(s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_M)$ that can be constructed from the other players' strategy spaces.

non-cooperative scenario using non-cooperative game theory.

Once the set of non-cooperative achievable rates is determined, the operating point is chosen by the protocol to achieve efficiency and fairness. We can think of this protocol as a widely known standard that the systems can choose to follow or as a set of spectrum sharing rules imposed by the regulation authority. In either case, a system knows the protocol, which specifies how to act in every possible situation, but is free to comply with it or not. However, each system knows that all other systems comply with the protocol[3]. This last remark is key in our game theoretic analysis.

12.1 Short interaction between systems: one shot game

We first consider a static game of complete and perfect information, usually referred to in the literature as the Gaussian Interference Game [52]. The complete information assumption is justified in Section 12.3 where we show that we can incentivize the systems to measure and exchange their parameters truthfully.

The game has M players, the M systems. The strategy space $\mathcal{S}_i$ of system i is the set of power allocations $p_i(f),\ f \in [0, W]$ that satisfy the power constraint (10.3). A strategy s_i for user i is the choice of power allocation $p_i(f)$. For a given strategy profile $(s_1, \ldots, s_M)$ the rate of user i is given by (10.2). The players play simultaneously, and know the utility functions of all the other players ($N_0, \{c_{i,j}\}_{i,j}, \{P_i\}_{i=1}^M, W$ are common knowledge). A strategy profile $\{s_i^*\}_{i=1}^M$ is a Nash Equilibrium (N.E.) of the game if

$$R_i(s_1^*, \ldots, s_M^*) \geq R_i(s_1^*, \ldots, s_{i-1}^*, s_i, s_{i+1}^*, \ldots, s_M^*)$$
$$\text{for all } s_i \in \mathcal{S}_i,\ i = 1, \ldots, M \qquad (12.1)$$

A direct consequence of the flat-fading and white noise assumption is the following fact:

Fact 1 *The set of frequency-flat allocations $p_i(f) = P_i/W, f \in [0, W]$ for $i = 1, \ldots, M$ is a Nash Equilibrium of the Gaussian Interference Game.*

[3]This is a consequence of the rationality assumption and the choice of a protocol that operates in a Nash equilibrium of a game.

This means that the best possible strategy for a given system is to spread its available power over the total bandwidth whenever all the interfering systems are spreading their signals. Fact 1 can be understood by noting that the best response of a system to a strategy profile of the other systems is to waterfill the available power over the noise+interference seen. When all the other systems use flat allocation, the waterfilling power allocation is flat, and it follows that flat allocations are best responses to each other.

If the players randomize their actions, the (mixed) strategy of each player is the choice of probability distribution used for the randomization. The utility that each user gets is the expected utility, averaged over the random choices of actions of all the players. Taking into account these changes in the definition of the strategies and the utilities, the concept of a mixed strategy Nash equilibrium can be defined exactly as before.

When studying the set of N.E., one needs to consider both pure and mixed strategies. However, in the case of the Gaussian Interference Game it turns out that we need only consider pure strategies.

Theorem 13 *The Gaussian Interference Game can only have pure strategy Nash equilibria. That is, every mixed strategy N.E. of the game must consist of atomic distributions with a single atom, and therefore is a pure strategy N.E.*

Proof: In the Nash equilibrium, let F_P be the distribution used by system i to select the power allocation $p_i(f)$, and let F_I be the distribution function of the total noise plus interference seen by system i. The utility of system i at the N.E. is given by:

$$\begin{aligned} E[U_i] &= \int\int U_i \left\{ \int_0^W \log\left[1 + \frac{c_{i,i} p_i(f)}{I(f)}\right] df \right\} dF_P \; dF_I \\ &\overset{(a)}{\leq} \int U_i \left\{ \int\int_0^W \log\left[1 + \frac{c_{i,i} p_i(f)}{I(f)}\right] df \; dF_P \right\} dF_I \\ &\overset{(b)}{\leq} \int U_i \left\{ \int_0^W \log\left[1 + \frac{c_{i,i} \int p_i(f) dF_P}{I(f)}\right] df \right\} dF_I \end{aligned}$$

where (a) follows from the concavity of $U_i(\cdot)$, and (b) follows from the concavity of the $\log(\cdot)$ function, using Jensen's inequality in both cases. Since $\log(\cdot)$ is strictly concave, we can only have equality throughout when F_P corresponds to a constant random variable. Furthermore, $\int dF_P p_i(f)$ satisfies the power constraint P_i and hence system i can increase its utility by using the power allocation $\int dF_P p_i(f)$ whenever F_P is not atomic with a single atom. It follows that for F_P to be part of a N.E. it must correspond to a constant. ■

If the channel gains across systems are sufficiently small the full-spread N.E. is the only N.E. of the Gaussian game. The following theorem gives a sufficient condition for the uniqueness of the full-spread N.E.

Theorem 14 *If* $\sum_{\substack{j=1 \\ j\neq i}}^{M} \frac{c_{j,i}}{c_{i,i}} < 1$ *for* $i = 1, \ldots, M$ *then the full-spread N.E. is the only N.E. of the Gaussian Interference Game.*

Proof: Let $\mathcal{P}$ be the set of vectors of power allocations $(\{p_1(f)\}, \ldots, \{p_M(f)\})$ that satisfy the power constraints $P_1, \ldots, P_M$. The proof consists of showing that there exists a mapping $T : \mathcal{P} \to \mathcal{P}$ with two properties:

- The set of fixed points of T coincides with the set of Nash equilibria of the GIG.
- T is a pseudocontraction with respect to some norm.

Let $\mathbf{p}^* \in \mathcal{P}$ be a fixed point of T. T is a pseudocontraction with respect to some norm $\|\cdot\|$ if there exists $\alpha \in [0, 1)$ such that:

$$\|T(\mathbf{p}) - \mathbf{p}^*\| \leq \alpha \|\mathbf{p} - \mathbf{p}^*\| \tag{12.2}$$

for all $\mathbf{p} \in \mathcal{P}$. If $\tilde{\mathbf{p}}^*$ is any fixed point of T, the condition (12.2) implies that $\tilde{\mathbf{p}}^* = \mathbf{p}^*$, and hence T has a unique fixed point.

Given a power allocation vector $\mathbf{p} = (\{p_1(f)\}, \ldots, \{p_M(f)\})$ we define $T(\mathbf{p})$ as the power allocation vector that results after each system waterfills its available power over the noise plus interference seen from the other systems, when they use the power allocations of $\mathbf{p}$. In other words, the ith component of $T(\mathbf{p})$ is the power allocation of system i when it waterfills its total power P_i

over the noise plus interference observed from the other systems, when they use power allocations $p_1(f), \ldots, p_{i-1}(f), p_{i+1}(f), p_M(f)$. Since waterfilling is the best response of a system to a set of power allocations of the other systems, $\mathbf{p}$ is a fixed point of T iff $\mathbf{p}$ is a N.E. of the GIG.

We define the norm $\|\cdot\|$ as the maximum component norm in the following way:

$$\|\mathbf{p}\| = \max_{i\in\{1,\ldots,M\}} \sup_{f\in[0,W]} |p_i(f)|$$

We make the technical assumption that the power allocations $\{p_i(f)\}$ are bounded, so that $\|\mathbf{p}\| < \infty$.

From Fact 1 we have that the set of flat allocations is a N.E. of the GIG, and hence is a fixed point $\mathbf{p}^*$ of T. We will now verify that T satisfies (12.2) with a modulus $\alpha = \max_i \sum_{j\neq i} \frac{c_{j,i}}{c_{i,i}}$ which, by assumption, is smaller than 1. Let $\mathbf{p} \in \mathcal{P}$ and $\Delta = \|\mathbf{p} - \mathbf{p}^*\|$. Then $|p_i(f) - p_i^*(f)| \leq \Delta$ for

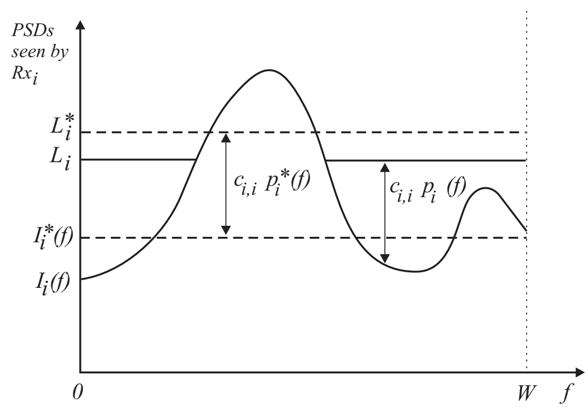


Figure 12.1: Power spectral densities seen at the receiver of system i when it waterfills over the noise+interference $I_i(f)$ and $I_i^*(f)$.

$f \in [0, W]$ and $i = 1, \ldots, M$. Let $I_i(f)$ ($I_i^*(f)$) be the noise+interference seen by system i when the other systems use power allocations from $\mathbf{p}$ ($\mathbf{p}^*$). Then it follows that $|I_i(f) - I_i^*(f)| \leq \Delta \sum_{j\neq i} c_{j,i}$ for $f \in [0, W]$. Let $\tilde{p}_i(f)$ be the power allocation of system i after waterfilling over $I_i(f)$. There are two cases to consider:

1. $\tilde{p}_i(f) > 0$ for all $f \in [0, W]$. In this case the "fluid" of system i covers all frequencies, and hence the total fluid level L_i is exactly the same as the one obtained when waterfilling over the noise plus interference profile $I_i^*(f)$. In this case we have $\tilde{p}_i(f) - p_i^*(f) = (I_i^*(f) - I_i(f))/c_{i,i}$

and it follows that $|\tilde{p}_i(f) - p_i^*(f)| \leq \Delta \sum_{j \neq i} c_{j,i}/c_{i,i}$ for all $f \in [0, W]$, $i = 1, \ldots, M$. It follows that $\|T(\mathbf{p}) - \mathbf{p}^*\| \leq \alpha \|\mathbf{p} - \mathbf{p}^*\|$.

2. $\tilde{p}_i(f) = 0$ in some frequencies. In this case, the "fluid" of system i does not cover all frequencies, and hence the total fluid level L_i is lower than L_i^*, the level obtained when waterfilling over the noise plus interference profile $I_i^*(f)$ (See Figure 12.1). In order to bound $|\tilde{p}_i(f) - p_i^*(f)|$ we look at the lowest valley and highest peak of $\tilde{p}_i(f)$. If the total fluid level L_i was equal to L_i^*, the analysis of 1) would imply that at the highest peak, $\tilde{p}_i(f) - p_i^*(f) < \Delta \sum_{j \neq i} c_{j,i}/c_{i,i}$. But since $L_i < L_i^*$, this difference is even smaller. On the other hand, $\tilde{p}_i(f_1) = 0$ at some f_1 implies that at the frequency f where $I_i(f)$ reaches its highest peak (i.e. $f \in \arg\max_{\tilde{f}} I_i(\tilde{f})$), $(I_i(f) - I_i^*(f))/c_{i,i} > p_i^*(f)$, and hence $\Delta \sum_{j \neq i} c_{j,i}/c_{i,i} > p_i^*(f)$. Therefore, since at the frequency f we have $\tilde{p}_i(f) = 0$, it follows that $p_i^*(f) - \tilde{p}_i(f) < \alpha\Delta$. The same condition holds for any frequency f where $\tilde{p}_i(f) = 0$ (lowest valleys of $\tilde{p}_i(f)$), since $p_i^*(f)$ is constant. Therefore, it follows that for all frequencies $|\tilde{p}_i(f) - p_i^*(f)| \leq \alpha\Delta$ for $i = 1, \ldots, M$ and we conclude that $\|T(\mathbf{p}) - \mathbf{p}^*\| \leq \alpha \|\mathbf{p} - \mathbf{p}^*\|$.

■

Theorem 14 does not give us any information about the uniqueness of the Nash equilibrium when the condition $\sum_{\substack{j=1 \\ j \neq i}}^{M} \frac{c_{j,i}}{c_{i,i}} < 1$ for all i is not met.

While studying the equilibria of the iterative waterfilling algorithm, Luo and Pang derived independently in [36] more general sufficient conditions for the uniqueness of the Nash equilibrium than the one given in Theorem 14. Our condition is derived using a different method and provides additional insight into the problem.

In many cases, the set of rates that results from the full-spread N.E. is not Pareto efficient (i.e. is not in $\mathcal{R}^*$) so there may be a significant performance loss if the M systems operate in this point due to lack of cooperation. And in many cases this inefficient outcome is the only possible outcome of the game. Consider for example a two system scenario (call it (d)) with $c_{1,1} = c_{2,2} = 1$,

$c_{2,1} = c_{1,2} = 1/4$, $W = 1$, $N_0 = 1$ and $P_1 = P_2 = P$. Note that in this case the condition of Theorem 14 is satisfied. If both users spread their signals, they obtain rates

$$R_1^{FS} = R_2^{FS} = \log\left(1 + \frac{P}{1 + P/4}\right) \text{[bits/s/Hz]} \tag{12.3}$$

which tends to $\log(5)$[bits/s/Hz] as $P \to \infty$. However, if the systems orthogonalize their power allocations using half of the bandwidth each, the resulting rates are:

$$R_1 = R_2 = \frac{1}{2}\log(1 + 2P) \text{[bits/s/Hz]} \tag{12.4}$$

which tends to ∞ as $P \to \infty$. The regime in which $P \gg N_0$ corresponds to the high SNR regime. In this regime, when the systems orthogonalize their power allocations they can communicate with an interference free channel, and achieve large data rates. If on the contrary both systems spread their signals, the signal to interference plus noise ratio becomes limited by interference, resulting in a reduced communication rate. This example shows that the inefficiency resulting from choosing the full-spread equilibrium can be arbitrarily large. The

If the Gaussian game has a Nash equilibrium in which the power allocations of the different systems are orthogonal the performance loss due to lack of cooperation is bounded. In the following theorem we present necessary and sufficient conditions for the existence of an orthogonal Nash equilibrium.

Theorem 15 *The Gaussian Interference game with M players has an orthogonal Nash equilibrium (i.e. $p_i(f) > 0 \Rightarrow p_j(f) = 0, \forall j \neq i$, for all $f \in [0, W]$) if for all sequences of indices $i_1, i_2, \ldots, i_k$, $2 \leq k \leq M$, $i_s \neq i_t$ for $s \neq t$ the channel gains $\{c_{i,j}\}$ satisfy:*

$$\frac{c_{i_1,i_2} c_{i_2,i_3} \cdots c_{i_k,i_1}}{c_{i_1,i_1} c_{i_2,i_2} \cdots c_{i_k,i_k}} > 1. \tag{12.5}$$

Furthermore, the existence of an orthogonal Nash equilibrium requires that for all sequences of indices $i_1, i_2, \ldots, i_k$, $2 \leq k \leq M$, $i_s \neq i_t$ for $s \neq t$ the channel gains $\{c_{i,j}\}$ satisfy:

$$\frac{c_{i_1,i_2} c_{i_2,i_3} \cdots c_{i_k,i_1}}{c_{i_1,i_1} c_{i_2,i_2} \cdots c_{i_k,i_k}} \geq 1. \tag{12.6}$$

Proof: See Appendix A. ■

The necessary and sufficient conditions of Theorem 15 are quite restrictive in practice. For example for three systems we need $c_{1,2}c_{2,1} > c_{1,1}c_{2,2}$, $c_{1,3}c_{3,1} > c_{1,1}c_{3,3}$, $c_{2,3}c_{3,2} > c_{2,2}c_{3,3}$, $c_{1,2}c_{2,3}c_{3,1} > c_{1,1}c_{2,2}c_{3,3}$, and $c_{1,3}c_{3,2}c_{2,1} > c_{1,1}c_{2,2}c_{3,3}$. For all these conditions to hold simultaneously, all system must interfere a lot with each other. We note that as the number of systems M grows, the likelihood of finding a real situation where an orthogonal Nash equilibrium exists becomes very small.

12.2 Long term interactions: a repeated game

Scenario (d) shows that there are situations in which the only possible outcome of the game is very inefficient, and as a result, there is a large performance degradation due to lack of cooperation. This negative result can be attributed to the static nature of the game that we defined.

Many wireless systems operate and co-exist with the same set of competing systems over a long period of time. In this context, it may be more reasonable to model the scenario as a repeated (or dynamic) game where systems play multiple rounds, remembering the past experience in the choice of the power allocation in the next round. We will consider an infinite horizon repeated game, where the Gaussian Interference Game is repeated forever. The utility of each player is defined by

$$U_i = (1-\delta)\sum_{t=0}^{\infty}\delta^t R_i(t) \tag{12.7}$$

where $R_i(t)$ is the utility of user i in the stage game at time t, and $\delta \in (0,1)$ is a discount factor that accounts for the delay sensitivity of the systems. At the end of each stage, all the players can observe the outcome of the stage-game and can use the complete history of play to decide on the future action. A strategy in the repeated game is a complete plan of action, that defines what the player will do in every possible contingency in which he may need to act.

One property of this repeated game is that sequences of strategy profiles that form a N.E.

in the stage game, form a N.E. in the dynamic game[4]. Furthermore, the dynamic game allows for a much richer set of N.E. This is an advantage from the point of view of policy making or standardization. The systems can agree through a standardization process to operate in any N.E. of the dynamic game. Having many equilibrium points to choose from gives more flexibility in obtaining a fair and efficient resource allocation. A natural question that arises is what set of rates can be supported as a N.E. of the repeated game. The following theorem, a general version of which is due to Friedman [26, 27], gives a sufficient condition for the rate vector $(R_1, \ldots, R_M)$ to be achievable as the resulting utilities in a N.E. of the repeated game.

Theorem 16 *Let R_i^{FS} be the rate of system i when all the systems spread their power over the bandwidth W, i.e. the rate obtained in the full-spread N.E. There exists a sub-game perfect N.E.[5] of the dynamic Gaussian Interference Game with utilities $(U_1, \ldots, U_M) = (R_1, \ldots, R_M)$ whenever $(R_1, \ldots, R_M) \in \mathcal{R}$ and $R_i > R_i^{FS}$ for $i = 1, \ldots, M$ for a discount factor δ sufficiently close to 1.*

Proof: Theorem C of [28] states that any utility vector that Pareto dominates the payoffs of a Nash equilibrium of the stage game can be supported by a sub-game perfect N.E. of the repeated game for a discount factor δ sufficiently close to 1. This Folk theorem is due to Friedman [26,27], although he considered only Nash equilibria instead of perfect equilibria in his work. In the Gaussian Interference Game the full-spread allocations form a N.E. (see Fact 1), and we can use $(R_1^{FS}, \ldots, R_M^{FS})$ as the payoff vector of the N.E. of the stage game in the Theorem above. ■

Let $\{p_i(f)\}_{i=1}^M$ be the power allocations that result in the rate vector $(R_1, \ldots, R_M)$ (which always exist since $(R_1, \ldots, R_M) \in \mathcal{R}$). The strategy that each system follows to obtain the rate vector $(R_1, \ldots, R_M)$ in Theorem 16 is the following trigger strategy:

- at $t = 1$: use power allocation $p_i(f)$.
- at $t = t_0$: if at time $t = t_0 - 1$ every user $j \in \{1, \ldots, M\}$ used the power allocation $p_j(f)$ then

[4]For the reader familiar with game theory, these equilibria are in fact sub-game perfect Nash equilibria.

[5]The sub-game perfect N.E. is a refined and stronger version of the N.E. concept defined before. It guarantees that the N.E. does not arise due to unbelievable threats.

use $p_i(f)$. Otherwise spread the power over the total band, i.e. use the power allocation P_i/W for $f \in [0, W]$

The idea behind this strategy, is to "cooperate" by using the required power allocation as long as all the other systems cooperated in the previous stages. As soon as at least one system deviates from the "good" behavior, a punishment is triggered where all the other systems spread their powers forever. Since the rates obtained by the systems once the punishment is triggered are lower than those obtained with cooperation, it is in the system's own interest to cooperate. Friedman's analysis shows that if δ is not too small, the above set of strategies forms a sub-game perfect N.E.. The sub-game perfection property of the N.E. guarantees that each system will indeed apply the punishment once the punishing situation arises. This property makes the threats believable.

Applying these ideas to scenario (d), we can define a trigger strategy where system 1 uses the first half of the bandwidth, and system 2 uses the second half, as long as in all the previous stages both systems complied with this frequency allocation. If at some stage any of the systems stops complying, a punishment is triggered where the systems spread their powers forever. For large enough P this pair of strategies forms a N.E. where each system obtains a utility $1/2 \log(1+2P)$. This shows how the punishment strategies within the dynamic game formulation allow us to overcome the inefficiency that we observed in the static game.

Theorem 16 gives us a sufficient condition for a rate vector $(R_1, \ldots, R_M)$ to be achievable through a N.E. But if the condition of the theorem is not met we may still have hope to find some other N.E. to support the desired set of rates. A natural question to ask is if there are other N.E. that result in utilities $(R_1, \ldots, R_M)$ with some $R_i < R_i^{FS}$. The following theorem answers this question negatively and provides a converse to Theorem 16.

Theorem 17 *The rate R_i^{FS} is the reservation utility of player i in the Gaussian Interference Game. That is, player i can obtain a utility at least as large as R_i^{FS} by using the power allocation $p_i(f) = P_i/W$, $f \in [0, W]$ regardless of the power allocations used by the other players. Therefore, the rate*

R_i obtained by user i in any N.E. of the Gaussian Interference Game must satisfy $R_i \geq R_i^{FS}$. The same statement holds for the repeated Gaussian Interference Game.

Proof: We will first prove that for a white Gaussian input, the worst possible Gaussian interference of given power is white. Since the power of each system is bounded to P_j, the total interference power seen by system i is bounded to $\sum_{j \neq i} c_{j,i} P_j$. We will prove that $I_i^*(f) = N_0 + \sum_{j \neq i} c_{j,i} P_j / W$ for $f \in [0, W]$ minimizes

$$R_i = \int_0^W \log \left[1 + \frac{c_{i,i} P_i}{W I_i(f)} \right] df$$

for noise+interference power bounded to $N_0 W + \sum_{j \neq i} c_{j,i} P_j$.

First, we need only consider $I_i(f)$ satisfying $\int_0^W I_i(f) df = N_0 W + \sum_{j \neq i} c_{j,i} P_j$, since increasing the interference power can only reduce R_i. We will consider only $I_i(f)$ that are continuous almost everywhere. Let f_1 and f_2 in $(0, W)$, be any continuity points of $I_i(f)$. Let $I = [I_i(f_1) + I_i(f_2)]/2$ and $\Delta = I_i(f_1) - I_i(f_2)$. Then for a small band δ around f_1 and f_2 the resulting rate of system i is:

$$\delta R_i = \delta \log \left[1 + \frac{c_{i,i} P_i}{W(I + \Delta/2)} \right] + \delta \log \left[1 + \frac{c_{i,i} P_i}{W(I - \Delta/2)} \right]$$

As was shown in the proof of Theorem 12 δR_i is minimized for $\Delta = 0$. Therefore, we conclude that for the minimizing $I_i(f)$, $I_i(f_1) = I_i(f_2)$. Since f_1 and f_2 are arbitrary continuity points of $I_i(f)$, we have that the optimal $I_i(f)$ must be constant almost everywhere. It follows that $I_i(f) = I_i^*(f)$ almost everywhere.

If system i uses a white input, the worst case interference is obtained when all the other systems spread their powers, and it follows that a rate at least as large as R_i^{FS} is always achieved. Therefore, there is no incentive for player i to play any strategy that results in a utility smaller than R_i^{FS}. ■

An immediate consequence of Theorems 16 and 17 is that if the desired operating point $(R_1, \ldots, R_M)$ (i.e. the maximizer of a desired global utility) is component-wise greater than the spreading rate vector $(R_1^{FS}, \ldots, R_M^{FS})$ there is no performance loss due to lack of cooperation. However, when this condition is not satisfied, the best that one can do is to find the point $(R_1, \ldots, R_M) \in \mathcal{R}^*$ that maximizes the global utility subject to $(R_1, \ldots, R_M) \geq (R_1^{FS}, \ldots, R_M^{FS})$.

Referring to Figure 12.2 we see that in scenario (b) the optimal sum rate point lies within the achievable region in the non-cooperative setting. However, the optimal proportional fair point lies outside of this set and cannot be supported without cooperation. The best that one can do in the non-cooperative setting is to operate in the point indicated in the figure. In scenario (c) both the optimal sum rate and optimal proportional fair rates are achievable in the non-cooperative setting. Note that while in the cooperative case the specific values of the cross gains had no influence on the achievable region (as long as the strong interference condition is satisfied) this is not true in the non-cooperative setting. This is because large cross gains enable the systems to apply punishments, and hence achieve a good N.E. through believable threats. In scenario (c) the large value of $c_{2,1}$ allows system 2 to punish system 1 whenever it departs from the proportional fair allocation.

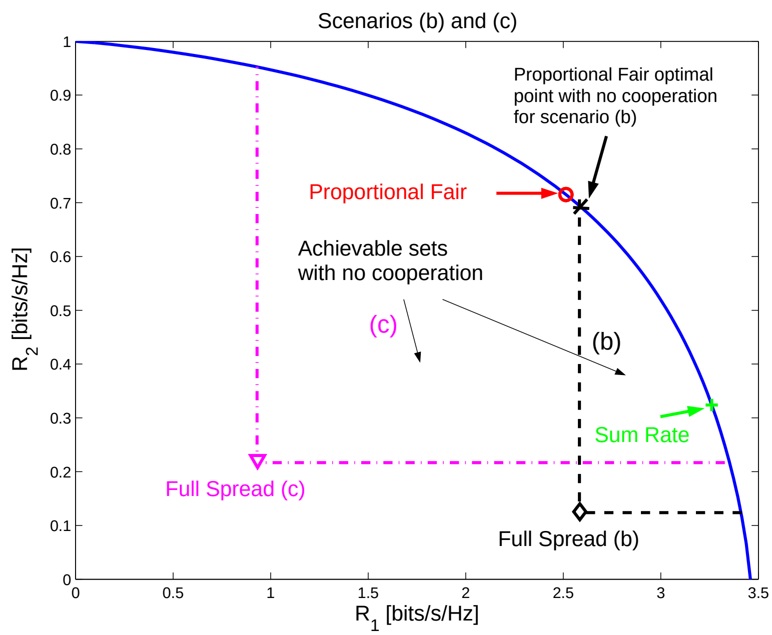


Figure 12.2: Achievable rates with no cooperation for scenarios (b) and (c).

To further illustrate the concepts introduced in this and the previous section, consider a two user scenario and assume that we use the proportional fair utility U_{PF} to measure the global performance. Without loss of generality we assume that $c_{1,1} = c_{2,2} = 1$, $W = 1$ and $N_0 = 1$. Also

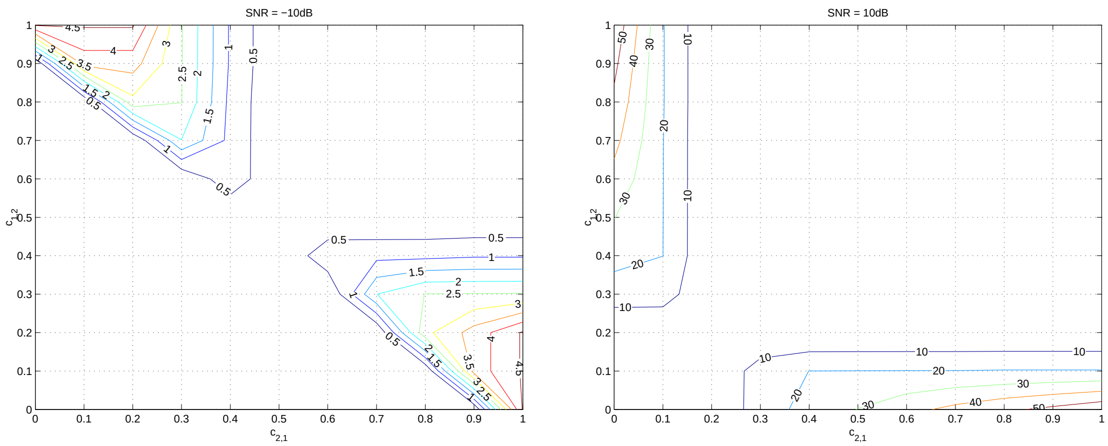


Figure 12.3: Contour plots of $(R_i - \tilde{R}_i)/R_i(\%)$ (percentage of rate loss with respect to the cooperative solution) as a function of the cross gains $c_{1,2}$ and $c_{2,1}$, for $\mathsf{SNR} = -10, 10$dB.

we take $P_1 = P_2 = P$ and analyze the results in terms of the $\mathsf{SNR} = P/N_0$. At a given SNR we can control the asymmetry between the two systems by varying the cross gains $c_{1,2}$ and $c_{2,1}$.

For a fixed set of parameters, using the results of Chapter 11 we optimize the power allocations to maximize the proportional fair metric, obtaining R_1^* and R_2^* as the resulting rates. In the non-cooperative scenario, R_1^* and R_2^* can only be supported by a N.E. if $R_1^* \geq R_1^{FS}$ and $R_2^* \geq R_2^{FS}$. If these inequalities are not satisfied, we obtain the best possible solution for the non-cooperative case by maximizing $\log(R_1) + \log(R_2)$ subject to the constraint $R_i \geq R_i^{FS}$, $i = 1, 2$, being $\tilde{R}_1$ and $\tilde{R}_2$ the corresponding optimal rates. If $R_i^* = \tilde{R}_i$ for $i = 1, 2$ we conclude that there is no loss due to lack of cooperation. If $R_i^* > \tilde{R}_i$ for $i = 1$ or $i = 2$ we measure the loss due to lack of cooperation using $\max_{i \in \{1,2\}} 100(R_i - \tilde{R}_i)/R_i$, i.e. the percentage loss in rate for one of the systems. Note that the other system will have a rate larger than the one obtained with cooperation. In Figure 12.3 we see that for low SNR, the region of rate pairs $(c_{1,2}, c_{2,1})$ for which there is a loss due to lack of cooperation is large, but this loss is quite small (about 5% in the worst case seen in the figure). As the SNR increases, this region becomes progressively smaller, but the corresponding performance loss is more significant. For $\mathsf{SNR} = 10$dB this performance loss can be as large as 50% of the proportional fair rates, but this only occurs in very asymmetric situations.

At low SNR performance is limited by noise not interference, so whether the systems cooperate or not does not have much influence in performance. At larger SNRs, interference becomes the dominant performance limiter. In very asymmetric situations the full-spread point (R_1^{FS}, R_2^{FS}) is such that R_1^{FS} or R_2^{FS} is large. In either of these cases, one system obtains a large enough rate with the full-spread allocations, and a threat of the other system to apply the spreading punishment is not effective to modify its behavior. Using Theorem 17 we see that if R_1^{FS} or R_2^{FS} is large, the set of rates achievable in the non-cooperative situation is quite limited, and often it does not include the optimal cooperative point. Fortunately, the level of asymmetry required to reach this unfair situation increases with SNR. The same qualitative behavior should be observed with other performance metrics and larger number of users.

The plots of Figure 12.3 illustrate the performance loss due to lack of cooperation when there is asymmetry in the cross gains between the systems. One can do a similar analysis for the case when the source of asymmetry is the transmission power instead of the cross gains.

12.3 Parameter measurement and exchange between selfish systems

The games considered in Section 12.2 were games of complete information. This means that all the parameters (power constraints, channel gains, etc.) are common knowledge to all the systems. In practice, some of these parameters have to be measured, and the corresponding measurements must be exchanged between the systems.

In a selfish environment it is possible that a system may try to tamper with the parameter measurement and exchange process in order to obtain an advantage. This situation can be modeled as a new game where we add in the action space of each system the different ways in which it can alter the parameter measurement and exchange process. In this section we show, under some assumptions, that this modified game has an efficient and fair Nash equilibrium for many reasonable fairness goals. We accomplish this by defining a parameter measurement and exchange protocol and showing that deviations from it can either be detected and punished or do not result in a better utility for the deviant system.

12.3.1 Channel measurement and exchange protocol

To measure the channel gains $\{c_{i,j}\}$ the transmitters of the different systems take turns in sending a pilot signal of normalized power. When transmitter i sends its pilot, all the receivers measure simultaneously the value of the channel gain $c_{i,j}$. After M time-slots all the channel gains are measured. We assume the existence of a low rate *control channel* that the systems can use to communicate with each other information about the parameters.

There are two ways in which a system may attempt to tamper with the measurement and exchange process: (a) a system may transmit the pilot with different power level than the nominal one; (b) a system may communicate fake channel measurement values. If these deviations pass undetected, the resource allocation algorithm may lead to solutions that are unfair and inefficient.

In order to detect such deviations we add some detection mechanisms to the protocol. These are:

- *Test messages*: System i transmits a test message using a predefined codebook of rate

  $$R = W \log\left(1 + c_{i,j} P_i / N_0 / W\right).$$

  If $c_{i,j}$ is the true value of the channel gain between transmitter i and receiver j, system j should be able to decode the message and feed it back to system i (using a low rate code) with negligible probability of error. If on the other hand system j has reported a value of $c_{i,j}$ larger than the true one, its error probability would be large, and the test would fail. Note that this test can only check whether the value of $c_{i,j}$ reported by system j is *larger* than the true value.

- *Multiple pilots*: All transmitters $j \neq i$ randomly select pilot powers P_j and transmit simultaneously white Gaussian signals. System j is required to broadcast the value of the total interference power received at its receiver. The other systems can then exchange the random power values and check whether the reported interference power matches $\sum_{j \neq i} c_{j,i} P_j$. Since system i does not know the values of P_j it can at most scale all values of $\{c_{j,i}\}_{j \neq i}$ by the same factor without being detected.

- *Triangulation*: This test uses a metric multi dimensional scaling technique [20] to determine the location of transmitters and receivers from the channel gain measurements. The values of the channel gains cannot be arbitrary; they must correspond to actual locations of transmitters and receivers in a two or three dimensional space. If the path loss model only takes into account the large-scale path loss (which is a function of the transmitter-receiver distance) we can translate each channel measurement $c_{i,j}$ into a transmitter-receiver distance $d_{i,j}$. Given these distances, the multi dimensional scaling technique finds locations for the transmitters and receivers that match these measurements. However, if for example system i scales all the channel gains $\{c_{i,j}\}$ (or $\{c_{j,i}\}$) by the same factor, the match between the locations of transmitters and receivers given by multi dimensional scaling and the distance measurements will be bad. This is because transmitter (receiver) i cannot be simultaneously closer or farther to all the other receivers (transmitters). In principle, this method works whenever the number of transmitters and receivers is larger than the embedding dimension (2 in the plane or 3 in the three-dimensional space).

  In practice, the path loss model may include other effects besides large-scale path loss, such as shadowing or multipath fading. In this case, the distance measurements obtained from the channel gain measurements may be subject to some "noise" that takes into account these path loss uncertainties. If we assume that the shadowing and multipath fading between different transmitters and receivers are independent random variables, the measurement errors in the distance measurements will be independent. If the systems report the true measured gains, due to statistical averaging the error between the measurements and the locations reconstructed by multi dimensional scaling will be moderate. In addition, the error will not change significantly if we remove sequentially the measurements of each system. If on the contrary system i scales all its measured gains simultaneously by the same factor (or equivalently, increases or decreases its nominal pilot power during the measurement phase) the error between the measurements and the locations reconstructed by multi dimensional scaling will be large and will decrease substantially once we remove the measurements of system i from the reconstruction. The accuracy of this detection technique will be larger when the number of systems is large, in which case the measurement errors due to fading are reduced by the law of large numbers.

Finally note that even if we could accurately detect the location of the transmitter and receiver of system i, we cannot make any statement about the direct channel gain $c_{i,i}$ in the case of shadowing and multipath fading. This is because there is no averaging that allows us to remove the effects of these attenuations in a single channel gain. Therefore, unless we use a simple path loss model that only takes into account transmitter and receiver distance, the triangulation technique can only detect falsely reported cross-gains scaled by a common factor.

- *Rate detection*: Some system j decodes some of system i's messages, determines its communication rate, and feeds back the rate measurement to the other systems. This allows to detect if a system is communicating at a rate smaller or larger than the one assigned by the resource allocation algorithm. This technique requires that the SINR between the transmitter of system i and the receiver of system j be larger than the SINR of system i. In addition, system j needs to know the codebook used by system i, and should have the computational power to decode the message. Note that the probability of finding a system satisfying these requirements increases with the number of systems and the network density. In addition, note that if system j fails in decoding system i's message, the protocol could require some other system k, say, to decode. It follows that the technique is robust against decoding failures of specific systems which may be due to fading or other effects.

12.3.2 A new repeated game

We modify the repeated game definition of Section 12.2 to add an initial stage of channel measurement, exchange, and verification. In this initial stage, each system can either comply with the protocol defined above or deviate from it. To deter selfish systems from deviating from the protocol we introduce punishment strategies as we did in Section 12.2. If in this initial stage any deviation is detected, a punishment is triggered in which all the systems spread their available power over the total bandwidth forever. In addition, if at any stage of the repeated game any system deviates from the power allocation or communication rate determined by the resource allocation algorithm, a similar punishment is triggered.

We want to show that the above punishment strategies form a Nash equilibrium in the modified repeated game whenever the corresponding game of Section 12.2 had such an efficient and fair Nash equilibrium. Therefore we assume that the spreading rates obtained with punishment are smaller than the rates allocated by the resource allocation algorithm under the true channel gains.

The test messages, multiple pilots, and triangulation techniques (under the assumption of having a large and/or dense network) can detect any misrepresentation of the channel cross gains by a single deviant system. In addition, if a system reports a direct channel gain larger than the true value, the rate detection technique can detect that its communication rate is lower than the one required by the resource allocation algorithm, which assumes a direct gain larger than the real one.

Since any detected deviation from the protocol results in punishment, there is no incentive for any individual system to misrepresent channel cross gains, or to announce a direct gain larger than the true one.

It remains to show that there is no incentive for a system to report that its direct gain is smaller than its true value. Since the rate detection technique can measure the communication rate, if a system underestimates its direct gain, it will have to communicate at a rate compatible with the reported gain to avoid being detected as a deviant. Therefore, we need to compare R_i^*, the rate assigned to system i by the resource allocation algorithm when system i reports its true direct gain $c_{i,i}$ to the rate $\tilde{R}_i^*$ assigned to system i when it reports a direct gain $\tilde{c}_{i,i}$ with $\tilde{c}_{i,i} < c_{i,i}$.

We state our results in the next two propositions.

Proposition 1 *Let the resource allocation algorithm maximize R_1 subject to $R_i/R_1 = \alpha_i > 0$, $i = 2, \ldots, M$. Let $\tilde{c}_{k,k} < c_{k,k}$ for some k. If $\mathbf{R}^*$ and $\tilde{\mathbf{R}}^*$ are the optimal rate vectors under $c_{k,k}$ and $\tilde{c}_{k,k}$ respectively, then $R_k^* \geq \tilde{R}_k^*$.*

Proof: We first note that $\tilde{\mathcal{R}}$, the achievable region when the direct gain of system k is $\tilde{c}_{k,k}$, is a strict subset of the achievable region $\mathcal{R}$, that results from the direct gain $c_{k,k}$ (all other parameters being equal in both cases). This is because for each feasible set of power allocations, the rate of system k is strictly smaller when its direct gain is $\tilde{c}_{k,k}$. The optimal rate vector $\mathbf{R}^*$ ($\tilde{\mathbf{R}}^*$) can be obtained

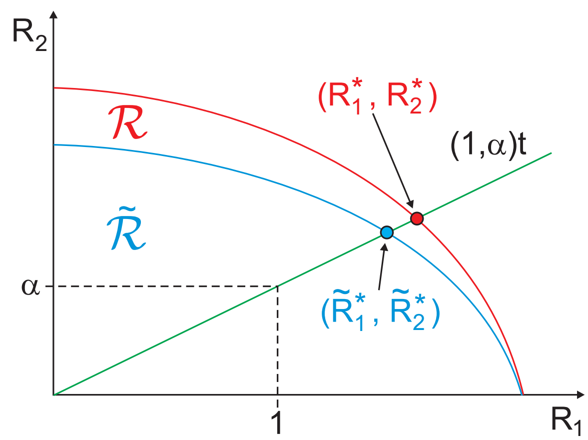


Figure 12.4: Graphical solution to the optimization problem that maximizes the rates of the systems, subject to $R_2/R_1 = \alpha$, for two different values of the direct gain of system 2 ($c_{2,2} > \tilde{c}_{2,2}$).

as the intersection of the line $(1, \alpha_2, \ldots, \alpha_M)t$, $t \in \mathbb{R}$, with the Pareto boundary of $\mathcal{R}$ ($\tilde{\mathcal{R}}$). Since $\tilde{\mathcal{R}} \subset \mathcal{R}$ the intersection of the line with the Pareto boundary occurs for smaller rates in $\tilde{\mathcal{R}}$ than in $\mathcal{R}$ (see Figure 12.4). ■

Proposition 1 shows that if the fairness goal is to maximize the rates of the systems for fixed ratios between the rates, there is no incentive for system k to announce a smaller direct channel gain. Note that if the Pareto boundaries of $\mathcal{R}$ and $\tilde{\mathcal{R}}$ are smooth (i.e. have continuous derivatives) the choice of $\alpha_i = 1$, $i = 2, \ldots, M$, corresponds to maximizing a Max-Min fair metric.

If there is high interference among the systems, we can state an equivalent result for a family of resource allocation algorithms that maximize a class of increasing and concave functions of the rates. Special cases of these functions are the weighted sum of rates $\sum_{i=1}^{M} \alpha_i R_i$ for non-negative weights $\{\alpha_i\}_i$ and the proportional fair metric $\sum_{i=1}^{M} \log(R_i)$.

Proposition 2 *Assume that the strong interference condition $c_{i,j}c_{j,i} > c_{i,i}c_{j,j}$ holds for all pair of systems $i \neq j$. Let $U(R_1, \ldots, R_M) = \sum_{i=1}^{M} U_i(R_i)$ where $U_i(R_i)$ is increasing and concave on R_i for $i = 1, \ldots, M$. Further assume that if $R_k = W_k \log(1 + c_{k,k}P_k/(W_kN_0))$ then $\log(1 + c_{k,k}P_k/(W_kN_0))\partial U_k(R_k)/\partial R_k$ decreases with increasing W_k and decreasing $c_{k,k}$. Let $\tilde{c}_{k,k} < c_{k,k}$ for some k. If $\mathbf{R}^*$ is the rate vector that maximizes $U(R_1, \ldots, R_M)$ under $c_{k,k}$ and $\tilde{\mathbf{R}}^*$ is the corresponding maximizer under $\tilde{c}_{k,k}$, then $R_k^* \geq \tilde{R}_k^*$.*

Proof: The strong interference assumption and Theorem 12 imply that the optimal power allocations are orthogonal, and the resulting optimization problem of maximizing $U(R_1, \ldots, R_M)$ is concave. Let $L(W_1, \ldots, W_M)$ be the Lagrangian of the maximization of $\sum_{i=1}^{M} U_i(R_i(W_i))$ subject to $\sum_{i=1}^{M} W_i = W$:

$$L(W_1, \ldots, W_M) = \sum_{i=1}^{M} U_i(R_i(W_i)) - \lambda \left(\sum_{i=1}^{M} W_M - W \right) \tag{12.8}$$

Taking derivatives with respect to W_i we obtain the optimality conditions:

$$\frac{\partial L(W_1, \ldots, W_M)}{\partial W_i} = \frac{\partial U_i(R_i(W_i))}{\partial W_i} - \lambda = 0 \tag{12.9}$$

for $i = 1, \ldots, M$. We first note that since $U_i(R_i(W_i)$ is concave on W_i its derivative $\partial U_i(R_i(W_i))/\partial W_i$ is decreasing on W_i. Therefore, an increase in λ results in a reduction of W_i. In addition, we can rewrite (12.9) as:

$$\begin{aligned}
\frac{\partial U_i(R_i(W_i))}{\partial W_i} &= \frac{\partial U_i(R_i)}{\partial R_i} \frac{\partial R_i(W_i)}{\partial W_i} = \frac{\partial U_i(R_i)}{\partial R_i} \left[\log\left(1 + \frac{c_{i,i}P_i}{W_iN_0}\right) - \frac{c_{i,i}P_i}{W_iN_0 + c_{i,i}P_i} \right] \\
&= \frac{\partial U_i(R_i)}{\partial R_i} \log\left(1 + \frac{c_{i,i}P_i}{W_iN_0}\right) \left\{ 1 - \left[\log\left(1 + \frac{c_{i,i}P_i}{W_iN_0}\right) \left(1 + \frac{W_iN_0}{c_{i,i}P_i}\right) \right]^{-1} \right\} \qquad (12.10) \\
&= \lambda \qquad (12.11)
\end{aligned}$$

Assume that we replace $c_{k,k}$ by a smaller value $\tilde{c}_{k,k}$ leaving all the other parameters unchanged. For the purpose of getting a contradiction, assume that this decrease in $c_{k,k}$ results in an increase of the optimal λ. Then, by the observation that we made above, the optimal W_i^* will decrease for all $i \neq k$, since the equations that define their values for given λ do not depend on $c_{k,k}$. In addition, since $W_k^* = W - \sum_{i \neq k} W_i^*$, the corresponding optimal W_k^* increases.

Let $f(x) = 1 - [(1+1/x)\log(1+x)]^{-1}$ for $x > 0$. Then, $f'(x) = [x - \log(1+x)]/[(1+x)^2 \log^2(1+x)] > 0$, and $f(x)$ increases with x. Letting $x = c_{k,k}P_k/(W_k N_0)$ we see that if $c_{k,k}$ decreases and W_k increases, x and $f(x)$ decrease. Since by hypothesis

$$\frac{\partial U_k(R_k)}{\partial R_k} \log\left(1 + \frac{c_{k,k}P_k}{W_k N_0}\right)$$

decreases with increasing W_k and decreasing $c_{k,k}$, it follows that (12.10) decreases for $i = k$, contradicting the assumption that we made that λ increases.

Therefore, for $\tilde{c}_{k,k} < c_{k,k}$, the optimal Lagrange multiplier λ decreases, W_i^* increase for $i \neq k$, and W_k^* decreases. Having a smaller direct gain and a smaller bandwidth, the kth system obtains a smaller rate $\tilde{R}_k^*$ for any $\tilde{c}_{k,k} < c_{k,k}$. ■

Note that in the case of the weighted sum metric with non-negative weights $\{\alpha_i\}_i$,$U_i(R_i) = \alpha_i R_i$, $\partial U_i/\partial R_i = \alpha_i$ and $\log(1 + c_{k,k}P_k/(W_k N_0))\partial U_k(R_k)/\partial R_k = \alpha_k \log(1 + c_{k,k}P_k/(W_k N_0))$ which decreases with increasing W_k and decreasing $c_{k,k}$. In the case of the proportional fair metric $U_i(R_i) = \log(R_i)$, $\partial U_i/\partial R_i = 1/R_i$ and $\log(1 + c_{k,k}P_k/(W_k N_0))\partial U_k(R_k)/\partial R_k = 1/W_i$ which decreases with increasing W_k. Therefore, if the resource allocation algorithm maximizes a weighted sum of rates or a proportional fair metric, and the strong interference condition holds, there is no incentive for any system to report a smaller value for its direct gain.

In summary, we have shown that in a dense network under some assumptions on the fairness metrics and/or the channel gains, there is no performance loss due to selfish behavior even when the channel gains are not known a priori to the systems.

Chapter 13

Extensions

The model that we considered in Section 10.2 includes a number of simplifying assumptions: systems have a single transmitter-receiver pair, the codebooks used are Gaussian, and interference is treated as noise. In this chapter we show that the main results of Chapter 12 hold in more general models that do not have some of these restrictions.

13.1 Non-Gaussian signals

The model introduced in Section 10.2 assumed that each system generated its codebook randomly using a Gaussian distribution. We now consider a more general model in which the different systems generate their codebooks using arbitrary distributions subject to a mean power constraint. We analyze whether Theorems 16 and 17 hold in this more general setting.

One can imagine that with more freedom in the choice of input distributions, it may be possible to exert a stronger punishment over a misbehaving system. In addition, a misbehaving system may want to find its most robust input signal to maximize its rate once the punishment is triggered. When system i misbehaves and the remaining systems apply a punishment, the interference observed at the receiver of system i has a maximum power given by $\sum_{j\neq i} c_{j,i}P_j$. It is possible to analyze the rates that system i gets for different input and punishment signals by considering a game between a sender x and an interferer z. Let $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, and assume that $\mathbf{X} \in \mathbb{C}^N$ and $\mathbf{Y} \in \mathbb{C}^N$ are independent, zero mean, circularly symmetric random vectors with covariance matrices $\mathbf{K_X}$ and $\mathbf{K_Z}$ subject to trace constraints $\mathrm{tr}(\mathbf{K_X}) \leq P_X$ and $\mathrm{tr}(\mathbf{K_Z}) \leq P_Z$. The strategy of each player is to choose a distribution subject to the power constraint. Define the payoff of player x to be $I(\mathbf{X};\mathbf{Y})$ and the payoff of player z to be $-I(\mathbf{X};\mathbf{Y})$. Since the payoffs sum to a constant (i.e. 0) the game is a zero-sum game. In our context $P_X = c_{i,i}P_i$ and $P_Z = \sum_{j\neq i} c_{j,i}P_j$.

The sender should choose its distribution $p_\mathbf{X}$ to solve $\sup_{p_\mathbf{X}} \inf_{p_\mathbf{Z}} I(\mathbf{X};\mathbf{Y})$ and the interferer should choose its distribution $p_\mathbf{Z}$ to solve $\inf_{p_\mathbf{Z}} \sup_{p_\mathbf{X}} I(\mathbf{X};\mathbf{Y})$. In both cases, the supremum and infimum are taken over all distributions satisfying the power constraints P_X and P_Z respectively. In this way the sender maximizes its payoff under the worst possible interference, and the interferer maximizes its payoff when the sender can adapt its signal to the observed interference. It turns out that the game has a saddle point and a saddle value, that is, $\sup_{p_\mathbf{X}} \inf_{p_\mathbf{Z}} I(\mathbf{X};\mathbf{Y}) = \inf_{p_\mathbf{Z}} \sup_{p_\mathbf{X}} I(\mathbf{X};\mathbf{Y})$ which is achieved when both the input and interfering signals are white and Gaussian [3].

It follows that the strongest punishment that can be applied over a misbehaving system is achieved when all the other systems use white Gaussian signals. Therefore, the punishments used in Theorem 16 should still be white and Gaussian even if we have the freedom to choose other distributions. Moreover, by using a white Gaussian signal any system i can obtain a rate at least as large as R_i^{FS} regardless of the distribution of the interfering signals. Therefore Theorem 17 still holds under the more general model that we consider here. In particular, it is not possible to achieve any rate vector with a component $R_i < R_i^{FS}$ even allowing for arbitrary input distributions.

13.2 General interference channel achievable regions

The model of Section 10.2 assumed that the systems treat interference as noise when decoding their messages. This assumption results in a tractable inner bound to the capacity region of the Gaussian interference channel. However, as we know from the analysis of Part I, treating interference as noise may be suboptimal in many cases. In particular, for the case of two systems we can obtain a larger achievable rate region by using simple Han-Kobayashi schemes[1]. These schemes require to decode part of the message transmitted by the interfering system, and once decoded, this part of the interference can be cancelled out. This scheme implicitly assumes that the codebooks used by the systems are common information. In this section we will extend Theorems 16 and 17 to allow for general communications strategies for the interference channel.

In order to extend our results, we need to be able to exert a strong punishment to a misbehaving system. As we discussed in Section 13.1, if the misbehaving system cannot cancel any part of the interference (i.e. treats interference as noise) the strongest possible punishment signal is white and Gaussian. Therefore, all that is required to extend our punishment strategies is to devise a mechanism that prevents a misbehaving system from implementing interference cancellation once the punishment is triggered. Toward this end, we design the spectrum sharing protocol so that each system has two codebooks for communication: a shared codebook, and a private codebook. The shared codebook is common knowledge to all the systems, and is used while all the systems comply with the spectrum sharing protocol. For example, this codebook may consist of two component codebooks to implement the Han-Kobayashi strategy. On the other hand the private codebook (which varies from system to system) is only known by each system, and is used during the punishment stage of the spectrum sharing protocol.

It is easy to see that Theorems 16 and 17 directly extend to this new setting. That is, for $\mathbf{R}$ in any achievable region of the Gaussian interference channel, $\mathbf{R}$ can be achieved by the modified spectrum sharing protocol without assuming cooperation among the systems if and only if $\mathbf{R} > \mathbf{R}^{FS}$, where $\mathbf{R}^{FS}$ is the rate vector in the Gaussian full-spread Nash equilibrium.

[1]Refer to Part I for a description of the Han-Kobayashi communication scheme and the corresponding inner bounds for the two-user Gaussian interference channel capacity region.

13.3 Multi-user systems

We now consider a model where each system is formed by one transmitter and many receivers (broadcast system) or many transmitters and a single receiver (multiple access system). In addition, we assume that each system can implement any (single system) capacity achieving communication strategy. That is, in the broadcast case the transmitter communicates data to the multiple receivers using superposition coding, and each receiver decodes and subtracts the messages sent to the weaker users before decoding its own message. In the multiple access case, the receiver performs multi-user detection, that is, it successively decodes the messages and cancels the interference from the different transmitters. However, each system treats the interference received from the other systems as noise.

For the case of a single (degraded) Gaussian broadcast channel, if the background noise is white, any point in the capacity rate region can be achieved by generating the codebook of each user with an i.i.d. Gaussian distribution. When the messages are simultaneously transmitted using superposition coding, the signal transmitted on the channel looks i.i.d. and Gaussian. Therefore, for white background noise there is no loss of optimality for using white Gaussian signals to communicate.

Now consider a single receiver of the broadcast system. This receiver successively decodes the messages sent to the different users, starting with the weakest user, treating the messages of the other users as Gaussian noise. Once a message is decoded, it is subtracted from the received signal, and the next user in increasing level of channel quality is decoded. The process continues until the receiver decodes his own message. At each stage of the decoding process the receiver behaves as a single user receiver. For this single user receiver, we can use the results of Section 13.1 to conclude that the worst type of noise is white and Gaussian. That is, if the background noise distribution is chosen by and adversary that has a power constraint, and whose goal is to minimize the rate at the given receiver, he should choose a white and Gaussian noise distribution. It follows that the worst type of noise for the broadcast system as a whole is white and Gaussian.

A similar argument can be made for a single multiple access system. For white Gaussian background noise, any point in the capacity region of the Gaussian multiple access channel can be achieved by using i.i.d. random Gaussian codebooks at the different transmitters, and using successive decoding at the receiver. Therefore, there is no loss of optimality for the system to use white and Gaussian signals. In any point in space, the superposition of these signals looks white and Gaussian, so the interference created by the system as a whole looks white and Gaussian.

The receiver in the multiple access system performs successive decoding and interference cancellation. At each stage of the decoding process, the receiver decodes a message treating the messages not yet decoded as noise. After decoding the message, it subtracts it from the received signal to reduce the interference seen by the remaining messages. Using the same argument as in the broadcast system case, the worst type of noise for the multi-user detection receiver is white and Gaussian. Therefore, the worst type of noise for the multiple access system is white and Gaussian.

It follows that when many multiple access and broadcast systems share spectrum, white Gaussian signaling forms a Nash Equilibrium of the resulting Gaussian interference game. This is because white Gaussian codebooks are the best response of each system to a white and Gaussian interference. In addition, the strongest punishment that the systems can exert over a misbehaving system is to use white and Gaussian codebooks, so that the interference created onto the system is white and Gaussian. Therefore, Theorems 16 and 17 directly extend to this multi-user systems model.

Chapter 14

Discussion

Our game theoretic analysis showed that in many cases, the fair and efficient operating point can be enforced through punishment strategies. Moreover, the rates that can be obtained with our punishment strategies are essentially the best that one can hope for in a non-cooperative setting. Therefore our results are tight and quantify the best achievable performance with lack of cooperation. The two system example that we presented shows that in most situations the performance loss due to lack of cooperation is small, and vanishes with increasing SNR.

Under the assumption that interference is treated as noise, finding efficient and fair spectrum allocations requires solving a complex optimization problem. We showed that under strong interference, frequency division multiplexing is optimal, and the resulting optimization problem is convex. In the more general case, the problem was shown to be of finite dimension but with non-convex structure. Finding efficient distributed algorithms that can be proved to converge to the optimal solution remains an open problem.

We assumed flat fading channels throughout this third part of the thesis. An interesting problem for future research is to extend our results for a general frequency selective channel model. While the basic idea of enforcing cooperation through credible threats is applicable for the frequency selective case, the choice of punishments and methods to make them credible require further study.

The punishment strategies that we considered, which punish forever any deviation from the desired behavior, are optimal for obtaining the largest achievable rate region in a non-cooperative situation. However, in a practical setting where there may be measurement errors (in the parameters or in the observation of the result of previous stages of the game), punishing forever may be too strict. It is possible to modify the punishment to last a fixed number of stages, where this number is chosen long enough to deter any misbehavior.

Bibliography

[1] R. Ahlswede. The capacity region of a channel with two senders and two receivers. *Annals Probability*, 2(5):805–814, 1974.

[2] R. Ahuja, T. Magnanti, and J. Orlin. *Network Flows: theory, algorithms and applications.* Prentice Hall, Englewood Cliffs, N.J., 1993.

[3] C. R. Baker and I.-F Chao. Information capacity of channels with partially unknown noisepart i: Finite dimensional channels. *SIAM Journal on Applied Mathematics*, 56(3):946–963, June 1996.

[4] R. Benzel. The capacity region of a class of discrete additive degraded interference channels. *IEEE Transactions on Information Theory*, 25(2):228–231, March 1979.

[5] D. Bertsimas and J. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, Belmont, Massachusetts, 1997.

[6] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, Cambridge, 2004.

[7] A. B. Carleial. A case where interference does not reduce capacity. *IEEE Transactions on Information Theory*, 21(5):569–570, September 1975.

[8] A. B. Carleial. Interference channels. *IEEE Transactions on Information Theory*, 24(1):60–70, January 1978.

[9] A. B. Carleial. Outer bounds on the capacity of interference channels. *IEEE Transactions on Information Theory*, 29(4):602–606, July 1983.

[10] M. Chiang. Geometric programming for communication systems. *Short monograph in Foundations and Trends in Communications and Information Theory*, 2(1-2):1–154, July 2005.

[11] M. Chiang, C. W. Tan, D. Palomar, D. O'Neill, and D. Julian. Power control by geometric programming. *To appear in IEEE Transactions on Wireless Communications*, 2007.

[12] H. Chong, M. Motani, H. Garg, and H. El Gamal. On the Han-Kobayashi region for the interference channel. *submitted to IEEE Transactions on Information Theory*, August 2006.

[13] S. T. Chung, S. Kim, J. Lee, and J. M. Cioffi. A game-theoretic approach to power allocation in frequency-selective gaussian interference channels. *Proceedings of IEEE International Symposium on Information Theory*, pages 316–316, 29 June - 4 July 2003.

[14] N. Clemens and C. Rose. Intelligent power allocation strategies in an unlicensed spectrum. *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Nov 8-11 2005.

[15] M. H. M. Costa. On the gaussian interference channel. *IEEE Transactions on Information Theory*, 31(5):607–615, September 1985.

[16] M. H. M. Costa and A. A. El Gamal. The capacity region of the discrete memoryless interference channel with strong interference. *IEEE Transactions on Information Theory*, 33(5):710–711, September 1987.

[17] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, November 1991.

[18] H. G. Eggleston. *Convexity*. Cambridge University Press, Cambridge, 1969.

[19] A. A. El Gamal and M. H. M. Costa. The capacity region of a class of deterministic interference channels. *IEEE Transactions on Information Theory*, 28(2):343–346, March 1982.

[20] R. Etkin. Metric multidimensional scaling as a convex optimization problem. Technical report, University of California, Berkeley, California, 2000.

[21] R. Etkin, A. Parekh, and D. Tse. Spectrum sharing for unlicensed bands. *43rd. Allerton Conference on Communication, Control, and Computing*, September 2005.

[22] R. Etkin, A. Parekh, and D. Tse. Spectrum sharing for unlicensed bands. *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, November 2005.

[23] R. Etkin, A. Parekh, and D. Tse. Spectrum sharing for unlicensed bands. *To appear in IEEE Journal on Selected Areas in Commununications, Special Issue (SI) on Adaptive, Spectrum Agile and Cognitive Wirelss Networks*, 2007.

[24] R. Etkin, D. Tse, and H. Wang. Gaussian interference channel capacity to within one bit: the symmetric case. *proceedings of the IEEE Information Theory Workshop*, October 22 - 26, 2006.

[25] C. Fortuin, P. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103, June 1971.

[26] J. Friedman. A non-cooperative equilibrium for supergames. *Review of Economic Studies*, 38(113):1–12, January 1971.

[27] J. Friedman. *Oligopoly and the theory of games*. Amsterdam: North-Holland, 1977.

[28] D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometica*, 54(3):533–554, May 1986.

[29] T. S. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Transactions on Information Theory*, 27(1):49–60, January 1981.

[30] J. Huang, R. Berry, and M. Honig. Distributed interference compensation for wireless networks. *IEEE Journal on Selected Areas in Communications, special issue on "Price-based Access Control and Economics for Communication Networks"*, 24(5):1074–1084, May 2006.

[31] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. Willey, New York, 2000.

[32] S. Karlin and Y. Rinott. Classes of orderings of measures and related correlation inequalities. i. multivariate totally positive distributions. *J. Multivariate Anal.*, 10(4):467–498, 1980.

[33] F. P. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, March 1998.

[34] J. F. C. Kingman. *Poisson Processes*. University Press, Oxford, 1993.

[35] G. Kramer. Outer bounds on the capacity of gaussian interference channels. *IEEE Transactions on Information Theory*, 50(3):581–586, March 2004.

[36] Z. Luo and J. Pang. Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines. *Submitted to the special issue of EURASIP Journal on Applied Signal Processing on Advanced Signal Processing Techniques for Digital Subscriber Lines*.

[37] A. MacKenzie and S. Wicker. Game theory and the design of self-configuring, adaptive wireless networks. *IEEE Communications Magazine*, 39(11):126–131, November 2001.

[38] J. P. Nolan. *Stable Distributions: Models for Heavy Tailed Data*. Preprint, 2004.

[39] O. Popescu, C. Rose, and D. Popescu. Signal space partitioning versus simultaneous signal space partitioning versus simultaneous water filling for mutually interfering systems. *Proceedings 2004 IEEE Global Telecommunications Conference - GLOBECOM 2004*, 5:3128–3132, December 2004.

[40] A. Sahai, N. Hoven, and R. Tandra. Some fundamental limits on cognitive radio. *Allerton Conference on Communication, Control, and Computing*, Oct 2004.

[41] T. Sarkar. Some lower bounds of reliability. Technical Report 124, Stanford University, California, 1969.

[42] I. Sason. On achievable rate regions for the gaussian interference channel. *IEEE Transactions on Information Theory*, 50(6):1345–1356, June 2004.

[43] H. Sato. Two-user communication channels. *IEEE Transactions on Information Theory*, 23(3):295–304, May 1977.

[44] H. Sato. On degraded gaussian two-user channels. *IEEE Transactions on Information Theory*, 24(5):637–640, September 1978.

[45] H. Sato. On the capacity region of a discrete two-user channel for strong interference. *IEEE Transactions on Information Theory*, 24(3):377–379, May 1978.

[46] C. E. Shannon. Two-way communication channels. *Proceedings of the 4th. Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley*, 1:611–644, 1961.

[47] E. Sousa. Performance of a spread spectrum packet radio network link in a poisson field of interferers. *IEEE Transactions on Information Theory*, 38(6):1743–1754, November 1992.

[48] C. W. Tan, D. Palomar, and M. Chiang. Solving nonconvex power control problems: low sir regime and distributed algorithms. *Proceedings 2005 IEEE Global Telecommunications Conference - GLOBECOM 2005*, 6:6, November 2005.

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

[50] E.C. van der Meulen. *Some reflections on the interference channel*. Kluwer, Boston, 1994.

[51] P. Viswanath and D. Tse. Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory*, 48(6):1277–1294, June 2002.

[52] W. Yiu, G. Ginis, and J. Cioffi. Distributed multiuser power control for digital subscriber lines. *IEEE Journal on Selected Areas in Commununications*, 20(5):1105–1115, June 2002.

[53] W. Yu and R. Lui. Dual methods for non-convex spectrum optimization of multi-carrier systems. *IEEE Transactions on Communications*, 54(7):1310–1322, July 2006.

[54] W. Yu, R. Lui, and R. Cendrillon. Dual optimization methods for multiuser orthogonal frequency division multiplex systems. *Proceedings 2004 IEEE Global Telecommunications Conference - GLOBECOM 2004*, 1:225–229, November 29 - December 3, 2004.

Appendix A

Proof of Theorem 15

Theorem 15 *The Gaussian Interference game with M players has an orthogonal Nash equilibrium (i.e. $p_i(f) > 0 \Rightarrow p_j(f) = 0, \forall j \neq i$, for all $f \in [0, W]$) if for all sequences of indices $i_1, i_2, \ldots, i_k$, $2 \leq k \leq M$, $i_s \neq i_t$ for $s \neq t$ the channel gains $\{c_{i,j}\}$ satisfy:*

$$\frac{c_{i_1,i_2} c_{i_2,i_3} \cdots c_{i_k,i_1}}{c_{i_1,i_1} c_{i_2,i_2} \cdots c_{i_k,i_k}} > 1. \tag{A.1}$$

Furthermore, the existence of an orthogonal Nash equilibrium requires that for all sequences of indices $i_1, i_2, \ldots, i_k$, $2 \leq k \leq M$, $i_s \neq i_t$ for $s \neq t$ the channel gains $\{c_{i,j}\}$ satisfy:

$$\frac{c_{i_1,i_2} c_{i_2,i_3} \cdots c_{i_k,i_1}}{c_{i_1,i_1} c_{i_2,i_2} \cdots c_{i_k,i_k}} \geq 1. \tag{A.2}$$

Proof: Every frequency $f \in [0, W]$ must be used by one and only one system. This follows because if a particular frequency remained unused, some system could increase its rate by allocating some power in it, and therefore the corresponding allocations cannot correspond to a Nash equilibrium. Also, if two or more systems allocated power in the same frequency, the corresponding allocation would not be fully orthogonal, which is the assumption here.

Let $\mathcal{W}_m$ be the set of frequencies f in which $p_m(f) > 0$, and $w_m = \int_0^W 1(f \in \mathcal{W}_m) df$ be the bandwidth used by system m. We assume that $w_m > 0$, $m = 1, \ldots, M$ (otherwise we would need infinite power density in some frequencies) and the observation of the previous paragraph implies that $\sum_{m=1}^{M} w_m = W$.

A given power allocation $\{p_m(f)\}$ is a Nash equilibrium of the Gaussian Interference Game (GIG) iff it corresponds to a water-filling power allocation for every user. In the fully orthogonal case, $p_m(f)$ corresponds to a water-filling power allocation iff:

$$p_m(f) = \begin{cases} 1/w_m & f \in \mathcal{W}_m \\ 0 & f \notin \mathcal{W}_m \end{cases}$$

and $c_{i,i}/w_i \leq c_{j,i}/w_j$, $\forall j \neq i$, $i, j = 1, \ldots, M$.

This set of inequalities can be written as:

$$w_j - \frac{c_{j,i}}{c_{i,i}} w_i \leq 0 \ ; \ j \neq i \ ; \ i, j = 1, \ldots, M \tag{A.3}$$

which constitutes a system of $M(M-1)$ inequalities. In addition to these inequalities, the vector $\mathbf{w} = (w_1, \ldots, w_M)^T$ must satisfy $w_m > 0$ for $m = 1, \ldots, M$.

We first prove that (A.2) is a necessary condition for the existence of a $\mathbf{w} > \mathbf{0}$ (we use '>' to denote component-wise inequality) that satisfies (A.3). Let $\mathbf{w} > \mathbf{0}$ be such that (A.3) is satisfied, and consider a particular sequence of indices $i_1, i_2, \ldots, i_k$, $2 \leq k \leq M$, $i_s \neq i_t$ for $s \neq t$. We need only consider a subset of the inequalities (A.3), which for convenience we rewrite here as:

$$w_{i_1} \leq \tfrac{c_{i_1,i_2}}{c_{i_2,i_2}} w_{i_2} \quad w_{i_2} \leq \tfrac{c_{i_2,i_3}}{c_{i_3,i_3}} w_{i_3} \quad \cdots \quad w_{i_{k-1}} \leq \tfrac{c_{i_{k-1},i_k}}{c_{i_k,i_k}} w_{i_k} \quad w_{i_k} \leq \tfrac{c_{i_k,i_1}}{c_{i_1,i_1}} w_{i_1}$$

By repeatedly replacing the w_i in the right hand side of each inequality with the right hand side of the following inequality we obtain:

$$w_{i_1} \leq \frac{c_{i_1,i_2}}{c_{i_2,i_2}} \cdot \frac{c_{i_2,i_3}}{c_{i_3,i_3}} \cdots \frac{c_{i_{k-1},i_k}}{c_{i_k,i_k}} \cdot \frac{c_{i_k,i_1}}{c_{i_1,i_1}} w_{i_1}$$

and noting that $w_{i_1} > 0$ the inequality (A.2) follows. Since this is true for any choice of the index sequence we have proved the necessity of the conditions (A.2).

The conditions (A.3) are necessary and sufficient for the existence of an orthogonal Nash equilibrium. If the inequality in (A.3) is replaced by strict inequality we get only a set of sufficient conditions for the existence of an orthogonal equilibrium. We now express this set of sufficient conditions in matrix form.

Let $\mathbf{A} \in \mathcal{R}^{M(M-1)\times M}$ be the matrix whose rows, indexed by the pair (i,j), $i,j = 1,\ldots,M$, $j \neq i$ have 1 in the jth column, $-c_{j,i}/c_{i,i}$ in the ith column and 0 in the other columns. Then, a sufficient condition on for the existence of an orthogonal Nash equilibrium is:

$$\begin{bmatrix} \mathbf{A} \\ -\mathbf{I}_M \end{bmatrix} \mathbf{w} < \mathbf{0} \tag{A.4}$$

where $\mathbf{I}_M$ is the $M \times M$ identity matrix, $\mathbf{w} = (w_1, \ldots, w_M)^T$, and $\mathbf{0} \in \mathcal{R}^M$ is the all zeros vector.

Note that if $\mathbf{w}$ satisfies (A.4) then $\tilde{\mathbf{w}} = (W/\sum_{m=1}^{M} w_m)\mathbf{w}$ also satisfies it, and $\sum_{m=1}^{M} \tilde{w}_m = W$. So it follows that there exists an orthogonal Nash equilibrium in the GIG if there exists $\mathbf{w} \in \mathcal{R}^M$ that satisfies (A.4).

To prove the sufficiency of the conditions (A.1) we need to prove that if (A.1) is satisfied for every choice of the index sequence, then there exists a $\mathbf{w}$ that satisfies (A.4). For this we use the theorem of alternatives [6] to write the dual statement to (A.4):

$$\exists \mathbf{w} \in \mathcal{R}^M : \begin{bmatrix} \mathbf{A} \\ -\mathbf{I}_M \end{bmatrix} \mathbf{w} < \mathbf{0} \quad \Leftrightarrow \quad \nexists \lambda \in \mathcal{R}^{M^2} : \lambda \geq 0, \lambda \neq \mathbf{0}, \left[\mathbf{A}^T - \mathbf{I}_M\right] \lambda = \mathbf{0} \tag{A.5}$$

To prove that (A.1) implies that there is no λ that satisfies the three conditions in the right of (A.5) we associate these three conditions to the feasibility of a non-zero flow in a generalized network flow problem. For this we let $\mathbf{x} = (\lambda_1, \cdots, \lambda_{M(M-1)})^T$ and $\mathbf{b} = (\lambda_{M(M-1)+1}, \cdots, \lambda_{M^2})^T$ and rewrite the conditions on the right of (A.5) as:

$$\nexists (\mathbf{x}, \mathbf{b}) : \mathbf{x} \geq 0, \mathbf{b} \geq 0, (\mathbf{x}, \mathbf{b})^T \neq \mathbf{0}, \mathbf{A}^T\mathbf{x} = \mathbf{b}$$

We associate the equation $\mathbf{A}^T\mathbf{x} = \mathbf{b}$ to a mass balance equation in a generalized network. Consider a directed graph $\mathcal{G}$ with M nodes labeled $1, \ldots, M$ and $M(M-1)$ arcs. Each column of $\mathbf{A}^T$ has two nonzero entries, one of which is 1 (in the jth row) the other being $-c_{j,i}/c_{i,i}$ (in the ith row).

We use each such column to represent an arc with origin in node j and end in node i, with arc gain $\mu_{j,i} = c_{j,i}/c_{i,i}$. The variable x_k represents the flow that enters the arc defined by the kth column of $\mathbf{A}^T$. In a generalized network the flow that exits a given arc may be larger or smaller than the flow that enters the arc depending on the arc gain μ. The parameter b_m represents a flow that enters node m from the exterior. If $b_m > 0$ there is a source in node m that generates flow, and if $b_m < 0$ there is a sink in node m that consumes flow. Row m of $\mathbf{A}^T\mathbf{x} = \mathbf{b}$ indicates that the sum of the flows that leave node m minus the sum of the flows that enter node m (multiplied by the appropriate arc gains) must be equal to the flow generated in the node. Note that in our case the special structure of the matrix $\mathbf{A}^T$ results in a fully connected network, with two arcs (one in each direction) connecting every pair of nodes (see Figure A.1).

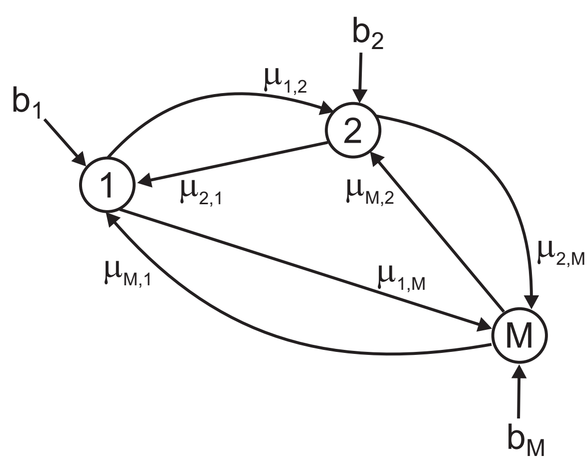


Figure A.1: Fully connected network with M nodes.

In a generalized network flow problem each arc (i, j) has an associated cost $z_{i,j}$ per unit of flow that enters the arc. Also the flow of each arc may be upper and lower bounded by some constants. The generalized network flow problem can be formulated as a linear program:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{z}^T\mathbf{x} \\ \text{Subject to:} \quad & \mathbf{A}^T\mathbf{x} = \mathbf{b} \\ & \mathbf{lb} \leq \mathbf{x} \leq \mathbf{ub} \end{aligned}$$

where $\mathbf{z}$ is the vector of arc costs, $\mathbf{lb}$ are the lower bounds on the arc flows and $\mathbf{ub}$ are the upper bounds on the arc flows.

For our purposes we set $\mathbf{lb} = \mathbf{0}$, $\mathbf{z} = -\mathbf{1}$, and do not use upper bound constraints (uncapacited network). With these choices the linear program is in standard form. If the optimal value is ∞ we have that the problem is infeasible, and if it is 0 the only feasible solution is $\mathbf{x} = \mathbf{0}$. We need to prove that if (A.1) holds, then the optimization problem is infeasible for $\mathbf{b} \neq \mathbf{0}$, and is feasible with optimal value 0 for $\mathbf{b} = \mathbf{0}$.

A set of M linearly independent columns of $\mathbf{A}^T$ is called a basis of the linear program. The corresponding entries of $\mathbf{x}$ are called basic variables. A vector $\mathbf{x}$ that satisfies the inequality and equality constraints of the linear program such that the non-basic variables satisfy the inequality constraints with equality is called a basic feasible solution.

The constraints of the linear program define a polyhedron. By Corollary 2.2 of [5] every non-empty polyhedron in standard form has at least one basic feasible solution. Theorem 2.8 of [5] states that if the polyhedron defined by the constraints of the linear program has at least one basic feasible solution, then either the optimal value is $-\infty$ or there exists a basic feasible solution that is optimal. It follows that for a linear program in standard form either:

- The problem is infeasible.
- The optimal value is $-\infty$.
- There is a basic feasible solution that is optimal.

Therefore, assuming that we are in the third case, to solve a linear program we can restrict our attention to basic feasible solutions. The simplex algorithm uses this fact to solve linear programs efficiently.

We need to introduce some definitions which are extracted from [2]:

Definition 2 *Let P be a path (not necessarily directed) from node s to node t. Let $\bar{P}$ and $\underline{P}$ denote* *the sets of forward and backward arcs in P. We define the path multiplier $\mu(P)$ of the path P as follows:*

$$\mu(P) = \frac{\prod_{(i,j)\in\bar{P}} \mu_{i,j}}{\prod_{(i,j)\in\underline{P}} \mu_{i,j}}$$

Property 1 *If we send 1 unit of flow from node s to another node t along a path P, then $\mu(P)$ units become available at node t.*

Definition 3 *Let W be a cycle (not necessarily directed) from a specified node s to itself whose orientation has already been defined. Let $\bar{W}$ and $\underline{W}$ denote the sets of forward and backward arcs in the cycle. With respect to the cycle's orientation, we define its cycle multiplier $\mu(W)$ as follows:*

$$\mu(W) = \frac{\prod_{(i,j)\in\bar{W}} \mu_{i,j}}{\prod_{(i,j)\in\underline{W}} \mu_{i,j}}$$

Sending flow along a cycle is the same as sending flow along a path except that the flow comes back to itself. Property 1 implies that if we send 1 unit of flow along the cycle W starting from node s, then $\mu(W)$ units return to this node. If $\mu(W) > 1$, we create an excess at node s; in this case we refer to the cycle W as a *gainy cycle*. If $\mu(W) < 1$, we create a deficit at node s; in this case we refer to the cycle W as a *lossy cycle*. If $\mu(W) = 1$, the flow around this cycle conserves mass balance at all its nodes; we refer to any such cycle W as a *breakeven cycle*.

Definition 4 *A subgraph of $\mathcal{G}$ is an augmented tree if it is connected and it contains exactly one cycle (see Figure A.2). A good augmented tree is an augmented tree whose cycle is either lossy or gainy but not breakeven.*

Definition 5 *An augmented forest is a collection of node-disjoint augmented trees that span all the nodes in $\mathcal{G}$. A good augmented forest is an augmented forest whose components are good augmented trees.*

Definition 6 *Suppose that the sets* **F**, **L** *and* **U** *define a partition of the arc set of $\mathcal{G}$ and that* **F** *is a good augmented forest. We refer to the arcs in* **L** *as nonaugmented-forest arcs at their lower*

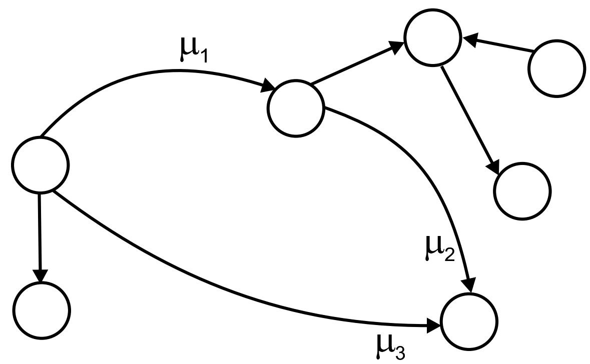


Figure A.2: Example of an augmented tree.

bounds, and the arcs in $\mathbf{U}$ *as nonaugmented-forest arcs at their upper bounds. We refer to the triple* $(\mathbf{F}, \mathbf{U}, \mathbf{L})$ *as a good augmented forest structure.*

Theorem 15.8 of [2] and the remark that follows it state that each good augmented forest structure defines a basis of the generalized network flow problem and to each augmented forest structure we can associate a basis of the linear program. Therefore in the search for a solution to the generalized network flow problem we can restrict our attention to good augmented forest structures. This fact is the essence of the generalized network simplex algorithm.

In our case we don't have upper bound constraints on $\mathbf{x}$ and therefore $\mathbf{U} = \emptyset$. The set $\mathbf{L}$ defines the set of components of the basic point $\mathbf{x}$ with zero value. Let $\mathbf{A}_F^T$ be the matrix obtained by extracting from $\mathbf{A}^T$ the columns corresponding to the arcs in $\mathbf{F}$, and $\mathbf{x}_F$ the corresponding arc flows. The fact that the good augmented forest structure defines a basis for the generalized network flow problem implies that the equation $\mathbf{A}_F^T \mathbf{x}_F = \mathbf{b}$ has a unique solution. If this solution satisfies the inequality constraints $\mathbf{x}_F \geq 0$ the corresponding basic point is feasible.

The above matrix equation can be decomposed into many matrix equations, one for each component of the augmented forest structure. Each of these equations has a unique solution. We will now show that if $\mathbf{b} \geq 0$ with $\mathbf{b} \neq \mathbf{0}$ then the basic point is infeasible, and that if $\mathbf{b} = 0$ then $\mathbf{x}_F = 0$.

Consider any good augmented forest of $\mathbf{F}$. Since $\mathbf{b} \geq 0$ the mass balance equation implies that all flows in the non-cycle arcs must be flowing toward the cycle nodes. Therefore each cycle-node receives a non-negative amount of flow from all non-cycle arcs. If all cycle arcs have the same orientation, condition (A.1) implies that the cycle of the augmented forest is gainy. The mass balance equation can only be satisfied by non-negative flows if all the flows in the cycle are zero and all flows arriving at cycle-nodes from non-cycle arcs are also zero.

If the arcs in the cycle have different orientations, then there is at least 1 node in the cycle that has two cycle arcs entering it. Since the amount of flow entering the node from non-cycle arcs is non-negative, the only way in which mass balance can be satisfied with non-negative flows is by having all flows entering the node equal to zero. This results from the fact that flows entering the node don't have a way of exiting it. Having a flow in a cycle arc equal to zero breaks the cycle and forces all flows in cycle arcs to be zero too (the augmented forest becomes a tree with no sinks, so any non-zero flow cannot leave the network and mass balance cannot hold). As before, it follows that mass balance can only be satisfied by non-negative flows if all the flows in the cycle are zero, and all flows arriving at cycle-nodes from non-cycle arcs are also zero.

To have all flows equal to zero we need $\mathbf{b} = \mathbf{0}$. Since this is true for all augmented forests in $\mathbf{F}$, the statement that we want to prove follows.

Since the choice of the good augmented forest structure is arbitrary, it follows that for $\mathbf{b} = 0$ all basic feasible solutions are $\mathbf{0}$, and for $\mathbf{b} \neq 0$, $\mathbf{b} \geq \mathbf{0}$ there are no basic feasible solutions.

Since our optimization problem is in standard form, the non-existence of basic feasible solutions for $\mathbf{b} \neq \mathbf{0}$ implies that the polyhedron defined by the constraints of the linear program is empty. In other words, the problem is infeasible for $\mathbf{b} \neq \mathbf{0}$.

To finish the proof of the theorem, we need to show that for $\mathbf{b} = \mathbf{0}$ the only feasible point is $\mathbf{x} = \mathbf{0}$. In this way we would prove that for $\mathbf{b} = \mathbf{0}$ the optimal value of the linear program is $\mathbf{x} = \mathbf{0}$.

To get a contradiction, assume that there is a feasible point $\hat{\mathbf{x}} \neq \mathbf{0}$. Then $\hat{\mathbf{x}} \geq 0$ with $\hat{x}_k > 0$ for some k. Let $\mathbf{a}_k$ be the kth column of $\mathbf{A}^T$, $\tilde{\mathbf{A}}$ be the matrix obtained by removing the kth column of

$\mathbf{A}^T$, and $\tilde{\mathbf{x}}$ be the vector obtained by removing the kth component of $\hat{\mathbf{x}}$. The feasibility of $\hat{\mathbf{x}}$ implies that $\mathbf{A}^T\hat{\mathbf{x}} = \mathbf{0}$. Then we can write $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = -\hat{x}_k\mathbf{a}_k$ with $\tilde{\mathbf{x}} \geq 0$. Note that $\mathbf{a}_k$ has exactly two non-zero entries, one being 1 and the other being $-\mu_{j,i} = -c_{j,i}/c_{i,i}$, where we assumed that the arc defined by $\mathbf{a}_k$ starts in node j and ends in node i. With this transformation we have a new generalized network similar to the original network but with arc k removed, and a source of value $\hat{x}_k\mu_{j,i}$ at node i and a sink of value $\hat{x}_k$ at node j.

Consider a good augmented forest structure $\mathbf{F}$ of this modified network. There are two cases to consider depending on whether the nodes i and j are in the same or different components of $\mathbf{F}$.

If the nodes i and j are in different components of $\mathbf{F}$ then the good augmented forest that contains node i has a non-zero source and no sinks. As was shown before, there is no flow that satisfies mass balance for this good augmented forest and therefore $\mathbf{F}$ is infeasible.

If on the other hand the nodes i and j are in the same component of $\mathbf{F}$, the source at node j and the sink at node i can be replaced by an arc starting at node j and ending at node i with arc gain $\mu_{j,i}$ and with an amount of flow $\hat{x}_k > 0$. With this transformation the resulting component has no sources or sinks. Note that by introducing the extra arc we have generated a second cycle and possibly a third cycle, so the resulting subgraph is not a good augmented tree anymore. Note that the arc (j, i) is part of at least one of these cycles. The parts of the subgraph other than the cycles and the nodes and arcs that connect them are trees with no sources or sinks and the flow going through them must be zero. We can therefore just focus on the cycles and the nodes and arcs that connect them. There are two cases to consider:

- Two cycles are created which share at most one vertex (see Figure A.3). If the orientations of the arcs in any of the cycles do not follow the same direction then the flow in the corresponding cycle must be zero (flows cannot be negative), the cycle is broken, and we are left with a good augmented tree with no sources or sinks. The only feasible flow for this graph is the all zeros flow, which contradicts the fact that arc (j, i) has a strictly positive flow.

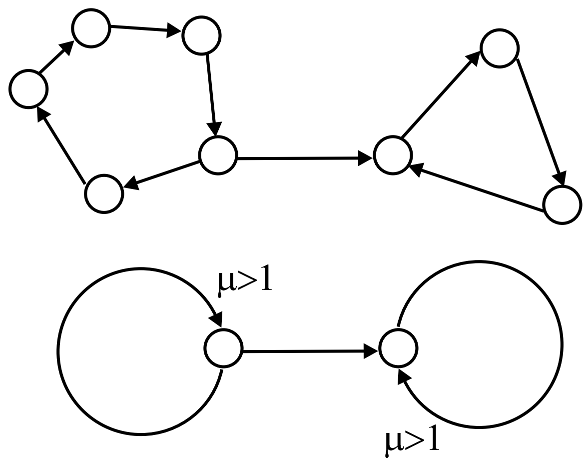


Figure A.3: Example where the addition of the arc (j, i) creates 1 additional cycle. The lower figure represents the 2 cycles with the corresponding cycle multipliers, and the connection between them.

Assuming that the orientation of the arcs in both cycles follow the same direction, we start computing the flows that circulate in the cycle that contains the arc (j, i). Since the cycle is gainy, there is a non-negative amount of flow that is generated, and that must leave the cycle toward the other cycle. The other cycle receives a non-negative amount of flow, and since it is gainy it cannot consume it. Therefore there is no feasible flow that satisfies mass balance.

- Three cycles are created which share 2 nodes, p and q (see Figure A.4).

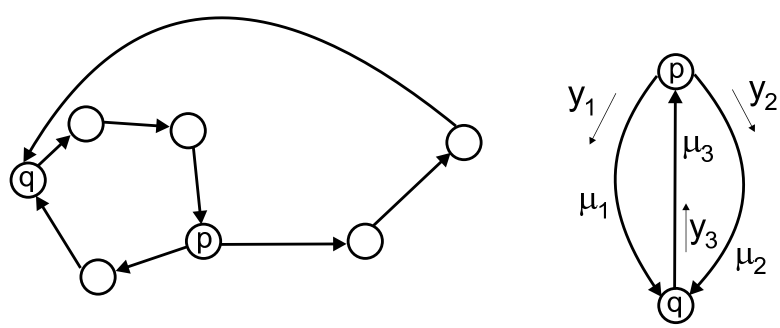


Figure A.4: Example where the addition of the arc (j, i) creates 2 additional cycles. The figure on the right represents the paths that form the 3 cycles, with their orientations, path multipliers and flows.

If the orientation of the arcs in any of the three paths that connect p and q are not the same, then the flow in the corresponding path must be zero, one of the cycles is broken and we are left with a good augmented tree with no sinks or sources. The only feasible flow for this augmented tree is the all zeros flow, which contradicts the fact that arc (j, i) has a strictly positive flow. Assuming that the 3 paths have arcs with the same orientation, then the three of them cannot be pointing from p to q (or q to p) simultaneously, as there would be no incoming flow into node p. This would force the 3 outgoing flows to be zero. We would then have all flows equal to zero contradicting the fact that arc (j,i) has a strictly positive flow. Then we must have two outgoing paths and one incoming path at node p (or q). Let μ_1 and μ_2 be the path gains of the two outgoing paths, and let y_1 and y_2 be the corresponding flows. Also let μ_3 be the path gain of the remaining path, and let y_3 be the corresponding flow. By considering the mass balance equations we have:

$$\begin{aligned} y_3 &= \mu_1 y_1 + \mu_2 y_2 \\ y_1 + y_2 &= \mu_3 y_3 \end{aligned}$$

Therefore,

$$\begin{aligned} \mu_3 \left(\mu_1 y_1 + \mu_2 y_2\right) &= y_1 + y_2 \\ y_1 \left(\mu_3 \mu_1 - 1\right) + y_2 \left(\mu_3 \mu_2 - 1\right) &= 0 \end{aligned}$$

The conditions (A.1) imply that $\mu_3\mu_2 - 1 > 0$ and $\mu_3\mu_2 - 1 > 0$ which with the above equations and the non-negativity of feasible flows imply $y_1 = y_2 = y_3 = 0$, contradicting the fact that the flow in the arc (j,i) is strictly positive.

We see that in both cases the good augmented forest structure $\mathbf{F}$ is infeasible. Since the choice of $\mathbf{F}$ was arbitrary, all basic solutions are infeasible. The constraints on the flows of the modified network are given in standard form, so the absence of basic feasible solutions implies that the feasible region is empty. This proves that there is no such $\tilde{\mathbf{x}}$ such that $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = -\tilde{x}_k \mathbf{a}_k$ with $\hat{x}_k > 0$, and we have a contradiction.

Having proved that $\mathbf{x} = \mathbf{0}$ is the only feasible point of the linear program for $\mathbf{b} = \mathbf{0}$, it follows that the optimal value is 0. The sufficiency of the conditions (A.1) follows. ■