CAPACITY LIMITS

OF BURSTY INTERFERENCE CHANNELS

by

Grace Silvana Villacrés Estrada

in partial fulfillment of the requirements for the degree of

Doctor in Multimedia and Communications

Universidad Carlos III de Madrid

Advisor:

Tobias Koch

April 11, 2019

To my parents

“Only those who will risk going too far
can possibly find out how far one can go.”
T.S. Elliot (1888-1965)

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my PhD advisor, Doktorvater, Dr. Tobias Koch. His knowledge and guidance directed me throughout my research and this dissertation. I feel fortunate to be one of his first PhD students. This, allowed me to enjoy a number of privileges which, I think, a very few Ph.D students can have. Tobi had always the time and patience to listen to me, check rigorously my proofs, teach me how to write papers and give talks. Indeed, it is true that more eyes see more things.

I appreciate his constant encouragement whenever I met a hard problem in research. Tobi always had the words to motivate and guide me, which gave me the confidence to continue with my research. I am indeed grateful to him not only for answering my questions, but also for always taking them seriously, thinking together with me and sharing with me the excitement and enthusiasm about research. He has been an advisor with a true concern for my welfare, an inspiring mentor and a great friend. He is a man who really feels and shares passion for Information Theory. I have learned a lot from him, especially that research means passion, honesty, perseverance and it should be translated to words in papers and talks in a clear, precise and scientific way. For all of the above and many other reasons, I want to say "herzlichen Dank Tobi".

My gratitude further extends to Dr. Gonzalo Vazquez-Vilar for his guidance, all-time availability, for the fruitful discussions and for the kind feedback that he gave to my talks, reports, drafts and papers. I really appreciate his help, patience, generosity and friendship during my Ph.D studies. If one day I have the opportunity to guide students, I will try to keep in mind how much patience and generosity I received from Tobi and Gonzalo.

A huge "Thank you" goes to Ana Hernando for her kindness, complicity and for spending a lot of time to deal with bureaucratic issues. In particular, she helped me at the beginning of my PhD studies when the paperwork was a nightmare. I would like to thank everyone at the Signal Processing Group (GTS), including the faculty, the former and current members. You made my Ph.D journey a really nice and unforgettable experience. The coffee break always was a time to remember how good is to share experiences outside the research time.

I also thank the Ph.D students and post-docs of the Information Theory group, Alex, Vivian, Jithin and Chao, for the stimulating discussions about research and life and for all the experiences that we lived together. I further want to thank Fran Hernando, my great mate in the best cubicle ever, and Gonzalo Ríos, you both were such a good company during our Master and Ph.D time.

Furthermore, I would like to say "thank you!" to Prof. Aydin Sezgin for receiving me at the Digital and Communication (DKS) Group at the Ruhr University Bochum. It was a fruitful experience. I always felt welcome and part of your group. I enjoyed the three months of work under your supervision and guidance. I further thank Soheil Gherekhloo for spending hours with me discussing my project. This academic exchange allowed me to incur in a new interesting topic that now is part of this thesis.

I would also like to profusely thank all my friends, who regardless of time or the place, all of them were just in a click-distance and their presence and friendship were always besides me. It is incalculable how nice it is to have them in my life and know that they are there. Ecuador, Germany and Spain gave me the opportunity to meet them. I am very fortunate to have you as the family I chose.

Thanks to Fundación Dolores Sopeña and its members for giving the opportunity to be part of its project as a volunteer. I just can say that volunteering is one of the most rewarding things one can do. It was a time to go back to the "real non-mathematical" life.

I am always grateful to my parents Blanca and Telmo, for their love, care and support through my life, to my siblings Diana and Thelmo; I can not imagine my life without their presence and encouragement. They are the greatest treasure in my life. I want to thank them for always believing in me when I did not and for reminding me that after a dark night comes a nice morning. Thanks for encouraging me to pursue each new plan, new idea and new dream, no matter how crazy they are. You provided me the tools to dream, and the wings to fly and consequently I could transform such dreams in goals.

Last but not least, I would like to say YOU

Gracias, Danke, Thank you.

PUBLISHED AND SUBMITTED CONTENT

The technical contributions presented in this thesis have been published in the following journals and conference papers:

> Grace Villacrés, Tobias Koch, Aydin Sezgin and Gonzalo Vazquez-Vilar. "Robust Signaling for Bursty Interference". *Entropy*, 20(11):870, 2018.
> Available at `https://doi.org/10.3390/e20110870`.

The whole paper is included in Chapter 4 and in Appendices A and B. The material from this source included in this thesis is not singled out with typographic means and references.

> Grace Villacrés and Tobias Koch. "Wireless Networks of Bounded Capacity". In *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain, July 10-15, 2016, pp. 2584-2588.
> Available at `https://doi.org/10.1109/ISIT.2016.7541766`.

The contribution of this paper is partially included in Chapter 5. The material from this source included in this thesis is not singled out with typographic means and references.

The corresponding preprint versions can be found on Arxiv:

> Grace Villacrés, Tobias Koch, Aydin Sezgin and Gonzalo Vazquez-Vilar. "Robust Signaling for Bursty Interference". *ArXiv preprint*, November, 2018. `arXiv:1809.02022v2`.
> `Available at: https://arxiv.org/abs/1809.02022v2`.

It is wholly included in Chapter 4 and in Appendices A and B. The material from this source included in this thesis is not singled out with typographic means and references.

> Grace Villacrés and Tobias Koch. "Wireless networks of bounded capacity". *ArXiv preprint*, July, 2015. `arXiv:1507.00131`.
> `Available at: https://arxiv.org/abs/1507.00131`.

It is partially included in Chapter 5. The material from this source included in this thesis is not singled out with typographic means and references.

Abstract

This dissertation studies the effects of interference burstiness in the transmission of data in wireless networks. In particular, we investigate the effects of this phenomenon on the largest data rate at which one can communicate with a vanishing small probability of error, i.e., on channel capacity. Specifically, we study the capacity of two different channel models as described in the next sections.

Linear deterministic bursty interference channel

First, we consider a two-user linear deterministic bursty interference channel (IC), where the presence or absence of interference is modeled by a block-independent and identically distributed (IID) Bernoulli process that stays constant for a duration of T consecutive symbols (this is sometimes referred to as a coherence block) and then changes independently to a new interference state. We assume that the channel coefficients of the communication and interference links remain constant during the whole message transmission. For this channel, we consider both its quasi-static setup where the interference state remains constant during the whole transmission of the codeword (which corresponds to the case whether the blocklength N is smaller than T) and its ergodic setup where a codeword spans several coherence blocks. For the quasi-static setup, we follow the seminal works by Khude, Prabhakaran and Viswanath and study the largest sum rate of a coding strategy that provides reliable communication at a basic (or worst-case) rate R and allows an increased (opportunistic) rate ΔR in absence of interference. For the ergodic scenario, we study the largest achievable sum rate as commonly considered in the multi-user information theory literature. We study how (non-causal) knowledge of the interference state, referred to as channel state information (CSI), affects the sum capacity. Specifically, for both scenarios, we derive converse and achievability bounds on the sum capacity for (i) local CSI at the receiver-side only; (ii) when each transmitter and receiver has local CSI, and (iii) global CSI at all nodes, assuming both that interference states are independent of each other and that they are fully correlated. Our bounds allow us to identify regions and conditions where interference burstiness is beneficial and in which scenarios global

CSI improves upon local CSI. Specifically, we show the following:

- Exploiting burstiness: For the quasi-static scenario we have shown that in presence of local CSI, burstiness is only beneficial if the interference region is very weak or weak. In contrast, for global CSI, burstiness is beneficial for all interference regions, except the very strong interference region, where the sum capacity corresponds to that of two parallel channels without interference. For the ergodic scenario, we have shown that, under global CSI, burstiness is beneficial for all interference regions and all possible values of p. For local CSI at the receiver-side only, burstiness is beneficial for all values of p and for very weak and weak interference regions. However, for moderate and strong interference regions, burstiness is only of clear benefit if the interference is present at most half of the time.

- Exploiting CSI: For the quasi-static scenario, local CSI at the transmitter is not beneficial. This is in stark contrast to the ergodic scenario, where local CSI at the transmitter-side is beneficial. Intuitively, in the ergodic scenario the input distributions depend on the realizations of the interference states. Hence, adapting the input distributions to these realizations increases the sum capacity. In contrast, in the quasi-static case, the worst-case scenario (presence of interference) and the best-case scenario (absence of interference) are treated separately. Hence, there is no difference to the case of having local CSI only at the receiver side. Featuring global CSI at all nodes yields an increased sum rate for both the quasi-static and the ergodic scenarios.

The joint treatment of the quasi-static and the ergodic scenarios allows us to thoroughly compare the sum capacities of these two scenarios. While the converse bounds for the quasi-static scenario and local CSI at the receiver-side appeared before in the literature, we present a novel proof based on an information density approach and the Verdú-Han lemma. This approach does not only allow for rigorous yet clear proofs, it also enables more refined analyses of the probabilities of error that worst-case and opportunistic messages can be decoded correctly. For the converse bounds in the ergodic scenario, we use Fano's inequality as the standard approach to derive converse bounds in the multi-user information theory literature.

Bursty noncoherent wireless networks

The linear deterministic model can be viewed as a rough approximation of a fading channel, which has additive and multiplicative noise. The multiplicative noise is referred to as fading. As we have seen in the previous section, the linear deterministic model provides a rough understanding of the effects of interference burstiness on the capacity of the two-user IC. Now, we extend our analysis to a wireless network with a very large number of users and we do not approximate the fading channel by a linear deterministic model. That is, we consider a memoryless flat-fading channel with an infinite number of interferers. We incorporate interference burstiness by an IID Bernoulli process that stays constant during the whole transmission of the codeword.

The channel capacity of wireless networks is often studied under the assumption that the communicating nodes have perfect knowledge of the fading coefficients in the network. However, it is *prima-facie* unclear whether this perfect knowledge of the channel coefficients can actually be obtained in practical systems. For this reason, we study in this dissertation the channel capacity of a noncoherent model where the nodes do not have perfect knowledge of the fading coefficients. More precisely, we assume that the nodes know only the statistics of the channel coefficients but not their realizations. We further assume that the interference state (modeling interference burstiness) is known non-causally at the receiver-side only. To the best of our knowledge, one of the few works that studies the capacity of noncoherent wireless networks (without considering interference burstiness) is by Lozano, Heath, and Andrews. *Inter alia*, Lozano *et al.* show that in the absence of perfect knowledge of the channel coefficients, if the channel inputs are given by the square-root of the transmit power times a power-independent random variable, and if interference is always present (hence, it is non-bursty), then the achievable information rate is bounded in the signal-to-noise ratio (SNR). However, the considered inputs do not necessarily achieve capacity, so one may argue that the information rate is bounded in the SNR because of the suboptimal input distribution. Therefore, in our analysis, we allow the input distribution to change arbitrarily with the SNR. We analyze the asymptotic behavior of the channel capacity in the limit as the SNR tends to infinity. We assume that all nodes (transmitting and interfering) use the same codebook. This implies that each node is transmitting at the same rate, while at the same time it keeps the

analysis tractable. We demonstrate that if the nodes do not cooperate and if the variances of the path gains decay exponentially or slower, then the achievable information rate remains bounded in the SNR, even if the input distribution is allowed to change arbitrarily with the transmit power, irrespective of the interference burstiness. Specifically, for this channel, we show the following:

- The channel capacity is bounded in the SNR. This suggests that noncoherent wireless networks are extremely power inefficient at high SNR.
- Our bound further shows that interference burstiness does not change the behavior of channel capacity. While our upper bound on the channel capacity grows as the channel becomes more bursty, it remains bounded in the SNR. Thus, interference burstiness cannot be exploited to mitigate the power inefficiency at high SNR.

Possible strategies that could mitigate the power inefficiency of noncoherent wireless networks and that have not been explored in this thesis are cooperation between users and improved channel estimation strategies. Indeed, coherent wireless networks, in which users have perfect knowledge of the fading coefficients, have a capacity that grows to infinity with the SNR. Furthermore, for such networks, the most efficient transmission strategies, such as interference alignment, rely on cooperation. Our results suggest that these two strategies may be essential to obtain an unbounded capacity in the SNR.

Keywords: information theory, channel capacity, interference, burstiness, channel with states, interference channel, channel-state information, linear deterministic model, ergodic scenario, quasi-static scenario, exponential decay, converse bounds, achievability bounds, signal-to-noise ratio, wireless networks

Contents

List of Figures

List of Tables

List of Acronyms

BEC	binary erasure channel
BSC	binary symmetric channel
CSI	channel state information
DCT	dominated convergence theorem
GDoF	generalized degrees of freedom
IC	interference channel
IID	independent and identically distributed
INR	interference-to-noise ratio
LDM	linear deterministic model
MAC	multiple access channel
MI	moderate interference
MIMO	multiple-input multiple-output
pdf	probability density function

pmf	probability mass function
RHS	right-hand side
Rx	receiver
SI	strong interference
SISO	single-input single-output
SNR	signal-to-noise ratio
TDMA	time-division multiple access
TIN	treating interference as noise
Tx	transmitter
VSI	very strong interference
VWI	very weak interference
w.p.	with probability
WI	weak interference
wlog	without loss of generality

1 Introduction

1.1 Motivation

The accelerated growth in mobile communications implies a huge increase in the number of users and demands higher data rates. In addition, the design of future wireless networks has to take into account the efficient use of spectrum. However, one of the key limiting factors in the efficient use of spectrum in wireless networks is interference, which also grows with the number of users. For this reason, the effect of interference on the information-theoretical limits of wireless networks is the focus of several works. So far, most of these works do not consider certain physical phenomena such as shadowing, which can make the interference intermittent or bursty. Interference can also be bursty due to the bursty nature of data traffic, distributed medium access control mechanisms, and decentralized networking protocols. Thus, considering that interference is always present is a pessimistic assumption. For this reason, understanding and exploring the effects of interference burstiness is an important point to consider in wireless networks.

In this dissertation, we study the effect of interference burstiness on channel capacity. We capture the burstiness of interference by modeling the wireless network as a channel with states [7], [20, Ch. 7]. The presence/absence of interference is modeled by a block-independent and identically distributed (IID) Bernoulli process as Wang *et al.* have done in [65]. The channel capacity highly depends on the level of knowledge that transmitters (Txs) and receivers (Rxs) have about the interference state. In this thesis, this knowledge is referred to as channel state information (CSI)[1]. We study how knowledge of CSI or lack thereof affects the largest rate at which the users can communicate with very low probability of error when the interference is bursty. We focus our analysis on two kinds of channel models with different limitations.

In the first part, we study the two-user IC as the simplest model to understand the effect of interference in a wireless network. We use the linear deterministic model (LDM) of the IC [5], which is a simplified model, but yields a unified treatment of several aspects previously studied in the literature and gives rise to several new results concerning the effect of CSI on the achievable rates over the bursty IC. We assume that the interference state stays constant for a duration of T symbols (referred to as coherence block) and then changes independently to a new state. We investigate both a quasi-static setup where the interference state remains constant during the whole transmission of the codeword (which corresponds to the case whether the blocklength N is smaller than T) and an ergodic setup where a codeword spans several coherence blocks. For the quasi-static setup, we follow the seminal works by Khude *et al.* [32], [33] and study the largest data rate of a coding strategy that provides reliable communication at a basic (or worst-case) rate R and allows an increased (opportunistic) rate ΔR when there is no interference. For the ergodic setup, we study the largest achievable sum-rate as commonly considered in the multi-user information theory literature. For the two setups, we also derive converse and achievability bounds on capacity, which in many cases are matching. These bounds allow us to study the effect of interference burstiness and level of CSI on capacity.

In the second part, we study the effect of interference burstiness on channel

[1]This CSI is different from the one sometimes considered in the analysis of interference channels (ICs) (see, e.g., [30]), where CSI refers to knowledge of the channel coefficients.

capacity of large wireless networks. The LDM, considered in the first part, can be viewed as a rough characterization of a fading channel which has additive and multiplicative noise, where the latter noise is sometimes referred to as fading. While the LDM provides a rough understanding of the effects of interference burstiness on the capacity of the two-user bursty IC, in the second part we consider a more realistic channel model. Specifically, we model the wireless network as a memoryless flat-fading channel with an infinite number of interferers. We assume that the Tx and Rx are only cognizant of the statistics of the channel coefficients, but they do not know their realizations, which is also known as a noncoherent scenario. We model the interference burstiness by an IID Bernoulli process that stays constant during the whole transmission of the codeword. We assume that the interference states are perfectly known at the Rx side only. For this channel model, we derive an upper bound on the channel capacity. Our results show that the capacity of this channel is bounded in the signal-to-noise ratio (SNR), indicating that noncoherent wireless are extremely power inefficient at high SNR. Since this result holds irrespective of the burstiness, we conclude that interference burstiness cannot mitigate the poor power efficiency. That said, exploiting burstiness in this channel increases the upper bound on capacity compared to the one we can achieve when interference is always present.

Possible strategies that could mitigate the power inefficiency of noncoherent wireless networks and that have not been explored in this thesis are cooperation between users and improved channel estimation strategies. Indeed, coherent wireless networks, in which users have perfect knowledge of the fading coefficients, have a capacity that grows to infinity with the SNR. Furthermore, for such networks, the most efficient transmission strategies, such as interference alignment, rely on cooperation. Our results suggest that these two strategies may be essential to obtain an unbounded capacity in the SNR.

1.2 Outline and Contributions

This dissertation is organized as follows. Chapters 2 and 3 present the background knowledge needed for this thesis. In particular, they introduce channel capacity, channels with state, the interference channel, and the LDM. Chapter 4 addresses the channel capacity for the bursty IC, as modeled by a LDM, and considers its ergodic and its quasi-static setup. Chapter 5 investigates the channel capacity for a network with infinity number of bursty interference signals as modeled by a noncoherent fading channel. Chapter 6 provides a summary and conclusions of the results in this dissertation. Some proofs of the presented results are deferred to the appendix sections. Here, we summarize the main contributions of this thesis:

Chapter 4: Linear deterministic bursty interference channel

In this chapter, we study the sum capacity of the bursty IC as modeled by a LDM. This study is performed for its quasi-static and ergodic setups. We analyze the effects of burstiness on the channel capacity. Specifically, our contributions in this chapter include:

- We perform a joint treatment of the quasi-static and the ergodic model: Previous literature on the bursty IC considers either the quasi-static model or the ergodic model. In contrast, this chapter discusses both models, allowing for a thorough comparison between the two.

- We derive novel achievability and converse bounds: For the ergodic model, the achievability bounds for the case where Tx and Rx know their corresponding interference states, and the achievability and converse bounds for the case when all nodes know all interference states, are novel. In particular, novel achievability strategies are proposed that exploit certain synchronization between the users.

- We provide novel converse proofs for the quasi-static model: In contrast to existing converse bounds, which are based on Fano's inequality, our proofs of the converse bounds for the rates of the worst-case and opportunistic messages are based on an information density approach (more precise, they are based on the Verdú-Han lemma).

- We perform a thorough comparison of the sum capacity of various scenarios: *Inter alia*, the obtained results are used to study the advantage of featuring different levels of CSI, the impact of the burstiness of the interference, and the effect of the correlation between the channel states of both users.

The contribution of this chapter was published in the the following paper [63]:

> Grace Villacrés, Tobias Koch, Aydin Sezgin and Gonzalo Vazquez-Vilar. "Robust Signaling for Bursty Interference". *Entropy*, 20(11):870, 2018.

The corresponding preprint version can be found on Arxiv:

> Grace Villacrés, Tobias Koch, Aydin Sezgin and Gonzalo Vazquez-Vilar. "Robust Signaling for Bursty Interference". *ArXiv preprint*, November, 2018. `arXiv:1809.02022v2.`

Chapter 5: Bursty noncoherent wireless networks

In this chapter, we study the effect of burstiness on channel capacity of wireless network, where two nodes are communicating and an infinite number of nodes are interfering. Specifically, we consider a memoryless flat fading channel with an unbounded number of interferes. Furthermore, we include burstiness in the interference links, where presence or absence is known only at the receiver side and stays constant during the whole transmission. We show that the channel capacity is bounded in the SNR under the assumptions that the interferers do neither cooperate with each other nor with the Tx, and they all use the same codebook. Specifically, our contributions in this chapter include:

- The channel capacity is bounded in the SNR. This suggests that noncoherent wireless networks are extremely power inefficient at high SNR.

- Our bound further shows that interference burstiness does not change the behavior of channel capacity. While our upper bound on the channel capacity grows as the channel becomes more bursty, it remains bounded in the SNR. Thus, interference burstiness cannot be exploited to mitigate the power inefficiency at high SNR.

- Our results suggest that cooperation and better channel estimation strategies may be essential to obtain an unbounded capacity in the SNR.

The contribution of this chapter was partially published in the paper [64]:

Grace Villacrés and Tobias Koch. "Wireless Networks of Bounded Capacity". In *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain, July 10-15, 2016, pp. 2584-2588.

The corresponding preprint version can be found on Arxiv:

Grace Villacrés and Tobias Koch. "Wireless Networks of Bounded Capacity". *ArXiv preprint*, July, 2015. `arXiv:1507.00131.`

1.3 Notation

To differentiate between scalars, vectors, and matrices we use different fonts: scalar random variables and their realizations are denoted by upper and lower case letters, respectively, e.g., B, b; vectors are denoted using bold face, e.g., $\mathbf{X}$, $\mathbf{x}$; random matrices are denoted via a special font, e.g., $\boldsymbol{X}$; and for deterministic matrices we shall use yet another font, e.g., S. For sets we use the calligraphic font, e.g., $\mathcal{S}$. We denote sequences such as $A_{i,1}, \ldots, A_{i,M}$ by A_i^M and sequences such as $E_n, E_{n+1} \ldots, E_M$ by E_n^M . Generic sequences are denoted by $\{A_{i,k}\}$ and $\{E_i\}$. We define $(x)^+$ as $\max\{0, x\}$.

The set $\mathbb{R}$ denotes the set of real numbers, $\mathbb{C}$ denotes the set of complex numbers, $\mathbb{Z}$ denotes the set of integers, $\mathbb{N}$ denotes the set of positive integers.

We use $\mathbb{F}_2$ to denote the binary Galois field and $\oplus$ to denote the modulo 2 addition. We define the $q \times q$ matrix $\mathsf{S}_u \in \mathbb{F}_2^{q\times q}$ as

$$\mathsf{S}_u = \begin{bmatrix} \mathsf{0}_{u\times(q-u)}^T & 0 \\ \mathsf{I}_u & \mathsf{0}_{u\times(q-u)} \end{bmatrix}_{q\times q}$$

with $0_{q-1} \in \mathbb{F}_2^{q-1}$ the all-zero vector and $\mathsf{I}_u \in \mathbb{F}_2^{(n)\times(n)}$ the identity matrix, and we refer to it as down-shift matrix.

Similarly, we define the $q \times q$ matrix $\mathsf{L}_d \in \mathbb{F}_2^{q\times q}$ as

$$\mathsf{L}_d = \begin{bmatrix} 0 & \mathsf{0}_{d\times(q-d)}^T \\ \mathsf{0}_{d\times(q-d)} & \mathsf{I}_d \end{bmatrix}_{q\times q}.$$

The matrix selects the d lowest components of a vector of dimension q.

We shall denote by $H_b(p)$ the entropy of a binary random variable X with probability mass function $(p, 1-p)$ [26, Sec. 3.3] i.e.,

$$H_b(p) \triangleq -p \log p - (1-p) \log(1-p). \tag{1.1}$$

Similarly, we denote by $H_{\text{sum}}(p,q)$ the entropy $H(X \oplus \tilde{X})$ where X and $\tilde{X}$ are two independent binary random variables with probability mass functions (pmfs) $(p, 1-p)$ and $(q, 1-q)$, respectively:

$$H_{\text{sum}}(p,q) \triangleq H_b(p(1-q) + (1-p)q). \tag{1.2}$$

For this function it holds that $H_{\text{sum}}(p,q) = H_{\text{sum}}(1-p,q) = H_{\text{sum}}(p,1-q) = H_{\text{sum}}(1-p,1-q)$.

We denote the floor function by $\lfloor\cdot\rfloor$. Likewise the ceiling function is denoted by $\lceil\cdot\rceil$. Thus $\lfloor a\rfloor$ denotes the largest integer that is smaller than or equal to a and $\lceil a\rceil$ denotes the smallest integer that is greater than or equal to a. We further use $\mathbb{1}(\cdot)$ to denote the indicator function, i.e., $\mathbb{1}(\text{statement})$ is 1 if the statement is true and 0 if it is wrong.

We shall denote the limit superior by $\overline{\lim}$ and the limit inferior by $\underline{\lim}$. Finally, $\underline{A}_1^N$ denotes the liminf in probability of A_1^N. It is defined as the supremum of all reals α for which $\Pr\{A_1^N \leq \alpha\}$ tends to zero as N tends to infinity. Similarly, the limsup in probability of A_1^N is denoted by $\overline{A}_1^N$. It is defined as the infimum of all the reals β for which $\Pr\{A_1^N \geq \beta\}$ tends to zero as N tends to infinity.

The norm-1 or Hamming weight of a length-n vector $\mathbf{x}$ is denoted by $\|x\|_1$ and is defined as $\|\mathbf{x}\|_1 \triangleq |x_1| + |x_2| + \ldots + |x_n|$.

We denote the pmf of X by $P_X(\cdot)$ and the pmf of Y given X by $P_{Y|X}(\cdot|\cdot)$.

2
Channel Capacity

This chapter is devoted to the channel capacity of channels with states. *Inter alia*, channels with states can be used to model communication scenarios with fading and different levels of channel state information (CSI) at transmitter and receiver [7], [20, Ch. 7], [41].

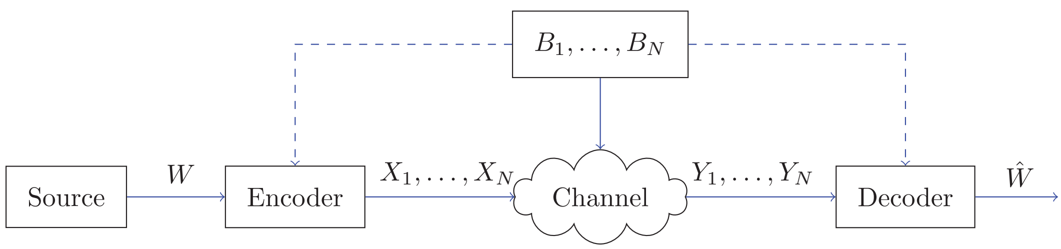


Figure 2.1: Block diagram of a communication system with state [20, Ch. 7].

In particular, we consider the mathematical model of the communication

system depicted in Figure 2.1. In the figure, the channel depends on the state sequence B^N. The message generated by the information-source, denoted by W, is independent of the channel state and is uniformly distributed over the set $\mathcal{W} = \{1, \ldots, |\mathcal{W}|\}$, where $|\mathcal{W}|$ corresponds to the cardinality of $\mathcal{W}$. In this dissertation, we consider the following evolutions of the state sequence B^N:
i) The state sequence remains constant for a duration of T consecutive symbols and then changes independently to a new state (ergodic scenario).
ii) The state sequence remains constant during the whole message transmission (quasi-static scenario).

Transmitter (Tx) or/and receiver (Rx) may have knowledge about B^N. This knowledge can be perfect, partial or fully unknown, as Biglieri *et al.* describe in [7]. Furthermore, the knowledge may be causal or non-causal. In general, we shall use $U^N \in \mathcal{U}^N$ to indicate the level of CSI at the Tx and $V^N \in \mathcal{V}^N$ to indicate the level of CSI at the Rx side. Specifically, in this chapter, we study the perfect non-causal knowledge of the channel state B^N at Tx and/or Rx side, and the perfect causal knowledge at the Tx side.

The **encoder** assigns to the message a length-N sequence $(X_1, \ldots, X_N) \in \mathcal{X}^N$, where N is known as the blocklength and $\mathcal{X}$ denotes the input alphabet of the channel. Mathematically, for non-causal knowledge of U^N at the Tx, the encoder is described by a function $f_N : \mathcal{W} \times \mathcal{U}^N \mapsto \mathcal{X}^N$, such that $X^N = f_N(W, U^N)$. When the channel state is causally known at the Tx, the encoder at time k is described by a function $f_k : \mathcal{W} \times \mathcal{U}^k \mapsto \mathcal{X}$, such that $X_k = f_k(W, U^k)$, $1 \leq k \leq N$. The sequence X^N is then sent over the channel. The channel generates an output sequence $(Y_1, \ldots, Y_N) \in \mathcal{Y}^N$, where $\mathcal{Y}$ denotes the channel output alphabet.

The distribution of the channel outputs Y^N depends on the input sequence and the channel-state sequence, as described by the channel law

$$P_{Y^N|X^N,B^N}(y_1, \ldots, y_N|x_1, \ldots, x_N, b_1, \ldots, b_N). \tag{2.1}$$

Physically, the channel law can be interpreted as the probability that the sequence Y^N appears at the Rx when the Tx sends the sequence X^N and the channel uncertainty is modeled by B^N [3, Ch. 3], [7].

For simplicity, we assume that, given the channel states B^N, the channel is memoryless, i.e.,

$$P_{Y^N|X^N,B^N}(y_1,\dots,y_N|x_1,\dots,x_N,b_1,\dots,b_N) = \prod_{k=1}^{N} P_{Y|X,B}(y_k|x_k,b_k). \quad (2.2)$$

We further assume that the channel is discrete in the sense that $\mathcal{X},\mathcal{Y}$, and $\mathcal{B}$ are finite alphabets. The channel law describes the physical propagation characteristics of the channel model we study. We shall specify in each chapter which channel model we consider.

The **decoder** attempts to guess the transmitted message W based on the sequence of N channel outputs Y^N and the information about B^N given by the CSI V^N, i.e., $(Y_1,\dots,Y_N,V_1,\dots,V_N)$. Mathematically, $\hat{W} = g_N(Y^N,V^N)$ for some mapping $g_N : \mathcal{Y}^N \times \mathcal{V}^N \mapsto \mathcal{W}$, where $\hat{W}$ denotes the decoded message.

Definition 1 (Achievable rate [14, Def. 7.5]) *A rate R is said to be achievable if there exist sequences of mappings $\{f_N, N \in \mathbb{N}\}$ and $\{g_N, N \in \mathbb{N}\}$ such that for each $N \in \mathbb{N}$*

$$R = \frac{\log(|\mathcal{W}|)}{N} \, \frac{\textit{nats/bits}}{\textit{channel use}}$$

and

$$\Pr\{W \neq \hat{W}\} \to 0 \text{ as } N \to \infty. \quad (2.3)$$

Definition 2 (Channel capacity [14, Def. 7.5]) *The capacity C is the supremum over all achievable rates.*

2.1 Capacity of Channels without State

The capacity of a point-to-point memoryless channel was first obtained by Shannon [51]. For the sake of completeness, we provide this capacity in the following theorem.

Theorem 1 *The capacity of a point-to-point memoryless channel is given by*

$$C = \max_{Q} I(X;Y) \quad (2.4)$$

where the optimization is over all distributions of X.

Proof: See [51]. ■

Equation (2.4) reveals that to obtain the channel capacity, one needs to maximize the mutual information $I(X;Y)$ over all input distributions Q of X. A common approach to do this is by computing lower and upper bounds on the channel capacity. Since any choice of input distribution yields a lower bound, obtaining lower bounds on the channel capacity is usually easier than deriving upper bounds. Lapidoth and Moser proposed in [40] a general technique that makes the derivation of upper bounds tractable. Specifically, this approach is based on the following theorem.

Theorem 2 ([40, Th. 5.1]) *Let the channel input and output take value in $\mathcal{X}$ and $\mathcal{Y}$, respectively, let X be of law Q, and let the conditional law of Y, conditioned on X, be given by P. Assume that $\mathcal{X}$ and $\mathcal{Y}$ are separable metric spaces, and assume that for any Borel set $\mathcal{B} \subseteq \mathcal{Y}$ the mapping $x \to P(\mathcal{B}|x)$ from $\mathcal{X}$ to $[0,1]$ is Borel measurable. Then*

$$I(X;Y) \leq \int D(P(\cdot|x)\|R(\cdot))dQ(x), \tag{2.5}$$

where $D(\cdot\|\cdot)$ denotes the relative entropy, i.e.,

$$D(P(\cdot|x)\|R(\cdot)) \triangleq \begin{cases} \int \log \frac{dP(\cdot|x)}{dR(\cdot)} dP(\cdot|x) & \text{if } P(\cdot|x) \ll R(\cdot) \\ +\infty & \text{otherwise} \end{cases} \tag{2.6}$$

and where $R(\cdot)$ is any distribution on $\mathcal{Y}$.

Proof: See [40, Sec. V]. ■

Theorem 2 shows that, for any choice of R, the right-hand side (RHS) of (2.5) yields an upper bound on the mutual information $I(X;Y)$. We will use this technique in Chapter 5.

Remark 1 *Regarding channels that are not memoryless, Verdú and Han [62] showed that the capacity of general channels is given by*

$$C = \sup_{Q} \underline{I}(X^N;Y^N) \tag{2.7}$$

where $\underline{I}(X^N;Y^N)$ denotes the liminf in probability of the normalized information density between input sequence $\mathbf{X}$ and output sequence $\mathbf{Y}$, defined as

$$\frac{1}{N}\mathrm{i}(X^N;Y^N) = \frac{1}{N}\log\frac{P_{Y^N|X^N}(y^N|x^N)}{P_{Y^N}(y^N)}. \tag{2.8}$$

The maximization in (2.7) *is over all input distributions* Q *of* X*. This definition of channel capacity is used for the quasi-static scenario.*

If the channel has memory and behaves ergodically, (2.7) *becomes*

$$C = \lim_{N\to\infty} \sup_{Q^n} \frac{1}{N} I(X^N;Y^N) \tag{2.9}$$

where the supremum is over all input distributions Q^N *of* X^N*. Specifically, Dobrushin [17] showed the validity of* (2.9) *for information stable channels [61, Def. 3].*[1] *Kim [34] showed that* (2.9) *is the capacity of a stationary channel with finite input memory and ergodic noise. However, in general, proving that the expression in* (2.9) *is the capacity is a difficult task.*

For channels with state, its capacity depends further on the level of CSI available at the Tx and/or Rx side. In the following sections, we revise the capacity expressions for the ergodic and quasi-static scenarios under different availabilities of CSI.

2.2 Channels with States: Ergodic Scenario

In the ergodic scenario, the channel state stays constant during the coherence time T and then changes to a new independent state. For simplicity, we assume that $T = 1$, i.e., the state sequence B^N is independent and identically distributed (IID) and $B_k \sim P_B$, where P_B^N denotes any arbitrary distribution of B^N.

2.2.1 No CSI available

We first discuss the case where neither the Tx nor the Rx have access to CSI. Let Q be the input distribution of X. Then, the channel capacity is given by

$$C = \max_Q I(X;Y) \tag{2.10}$$

where the optimization in (2.10) is over all input distributions Q of X and where the channel law is given by

$$P_{Y|X}(y|x) = \sum_{b\in\mathcal{B}} P_{Y|X,B}(y|x,b).$$

[1] Roughly speaking, these channels have the property that the input that maximizes mutual information $I(X^N;Y^N)$ and its corresponding output behave ergodically.

In this case, one may use Theorem 2 to obtain upper bounds on channel capacity.

2.2.2 Perfect CSI available at Receiver

Suppose now that the receiver has access to the channel-state sequence B^N. In this case, the capacity is given by [7, Eq. 3.3.6]

$$C = \max_{Q} I(X; Y|B) \tag{2.11}$$

where the optimization is over all input distributions Q of X. As pointed out by Caire and Shamai [11], this expression can be obtained by treating B^N as an additional channel output. This case is studied in many information-theoretical works that concern, e.g., the capacity of fading channels with side information [24], the bursty-interference channel (IC) [32], [33], [65], or cellular mobile radio networks [46].

2.2.3 Perfect CSI available at Transmitter

When the Tx has access to the CSI (but the Rx has not), the literature focuses on two cases: i) causal CSI and ii) non-causal CSI. The case of non-causal CSI at the Tx (U^N) was considered by Gelfand and Pinsker [23]. In this case, the transmitter knows in advance the entire sequence B^N. This assumption is reasonable, e.g, when one is concerned with storage of encoded information in a computer, like coding in a memory with defective cells [38] or the capacity of computer memory with defects [28]. The channel capacity for this channel is given by [23], [20, Sec. 7.6.1 and 7.6.2]

$$C = \max_{\substack{P_{D|B}, \\ f: X = f(D,B)}} \left(I(D; Y) - I(D; B)\right). \tag{2.12}$$

Here, D is an auxiliary random variable that depends B and has cardinality $|\mathcal{D}| \leq \min\{|\mathcal{X}| \cdot |\mathcal{B}|, |\mathcal{Y}| + |\mathcal{B}| - 1\}$. The maximization is over all mappings $X = f(D, B)$ and all conditional distributions of D given B such that $D \to (X, B) \to Y$ forms a Markov chain.

The capacity of memoryless channels with causal CSI at the Tx was first obtained by Shannon [52] and then generalized by Salehi [48]. In this setting, the transmitter knows at time k the CSI $B_1, \ldots, B_k$. The channel capacity is given in [11], [52] and also in [20, Sec. 7.5]:

$$C = \max_{\substack{P_D,\\ f:X=f(D,B)}} I(D;Y). \tag{2.13}$$

Here, D is an auxiliary random variable independent of B, with cardinality $|\mathcal{D}| \leq \min\{(|\mathcal{X}|-1)|\mathcal{B}|, \mathcal{Y}\}$. The maximization is over all mappings $X = f(D,B)$ and all distributions P_D.

2.2.4 Perfect CSI available at Transmitter and Receiver

We finally discuss the case where CSI is available (causally or non-causally) at both Tx and Rx. The capacity of this channel was first derived by Wolfowitz [67, Th. 4.6.1]

$$C = \max_{P_{X|B}} I(X;Y|B) \tag{2.14}$$

see also [7, Eq. 3.3.7]. The optimization in (2.14) is over all conditional distributions of X given B, which reflects the dependence of the inputs on the channel state. This capacity is achieved when the transmitter adapts its coding scheme, power and data rate to the channel-state variations. This is sometimes referred to as rate adaption in the literature [24]. Goldsmith and Varaiya [24] showed that the optimal power allocation is a time-water-pouring approach, which optimally adapts the power at the transmitter according to the quality of the channel.

2.2.5 Channel Capacity under Different Levels of CSI

To show the effect of CSI on the channel capacity, we compare the results for the cases: i) CSI unavailable at Tx and Rx, ii) non-causal CSI available at Rx side ($V^N = B^N$) iii) non-causal CSI at the Tx side ($U^N = B^N$) and iv) non-causal CSI available at both Tx and Rx sides ($V^N = U^N = B^N$). To this end, let us consider the following example.

Example 1 ([20, Example 7.3], [48]) *Consider a discrete memoryless channel with a discrete memoryless channel state depicted in Figure 2.2. As we can observe, the channel has a ternary channel state $B \in \{0,1,2\}$, whereas the channel input X and the channel output Y are binary. In Figure 2.2 we observe that, if $B = 0$,*

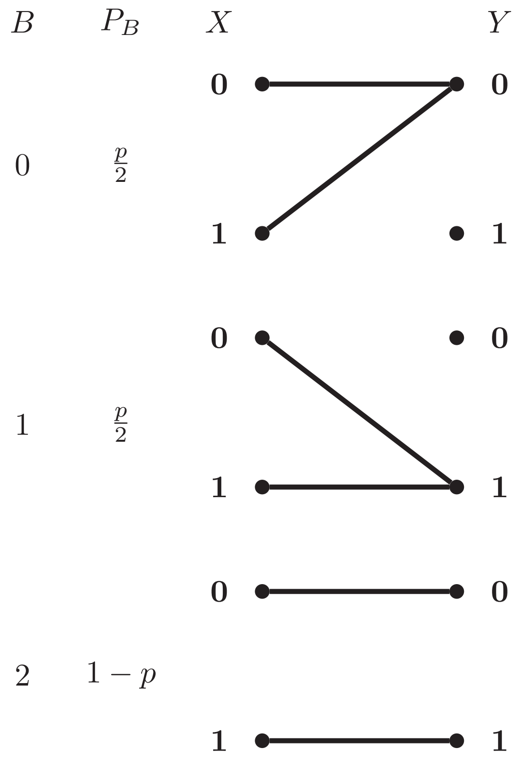


Figure 2.2: Model for a binary memory cell with defects.

then the channel always outputs a 0 independent of its input value; if $B = 1$*, then the channel always outputs a 1 independent of its input value; if* $B = 2$*, then the output has the same value as its input. This channel is a rough model for a binary storage medium with equiprobable one-defects and zero-defects, each with probability (w.p.)*

$$P_B(0) = P_B(1) = \frac{p}{2}, \text{ and} \tag{2.15}$$

$$P_B(2) = 1 - p. \tag{2.16}$$

We present the capacity for the following cases:

i) Non-causal CSI unavailable at both Tx and Rx: In this case

$$P_{Y|X}(1|0) = \frac{p}{2} \tag{2.17}$$

$$P_{Y|X}(0|1) = \frac{p}{2}. \tag{2.18}$$

This corresponds to a binary symmetric channel (BSC) with crossover probability $\frac{p}{2}$*, so the channel capacity is*

$$C = \max_{Q} I(X;Y) = 1 - H_b\left(\tfrac{p}{2}\right) \ \tfrac{bits}{ch.\ use}. \tag{2.19}$$

ii) *Non-causal CSI available at Rx: In this case*

$$P_{Y|B,X}(1|0,0) = \frac{p}{2} \tag{2.20}$$
$$P_{Y|B,X}(0|1,1) = \frac{p}{2}. \tag{2.21}$$

This corresponds to a binary erasure channel (BEC) with erasure probability p*, so the channel capacity is*

$$C = \max_{Q} I(X;Y|V) = 1 - p \ \tfrac{bits}{ch.\ use}. \tag{2.22}$$

iii) *Non-causal CSI available at Tx: Here, we just derive a lower bound on the capacity. Since this lower bound will coincide with the capacity when both Tx and Rx have non-causal CSI, and since CSI at both Tx and Rx cannot be worse than CSI at the Tx only, it follows that this lower bound is also the capacity. To derive a lower bound, we evaluate* (2.12) *for the following distributions. If* $B = 2$*, then we set* $X = D \sim Ber\left(\frac{1}{2}\right)$*. If* $B = 1$ *or* $B = 0$*, then we set* $D = X = B$*. It follows that, for this choice of* $P_{D|B}$ *and* $X = f(D,B)$*, the channel capacity is lower-bounded as*

$$\begin{aligned} C &= \max_{\substack{P_{D|B},\\ f:X=f(D,B)}} \left(I(D;Y) - I(D;B)\right) \\ &\geq H(D|B) - H(D|Y) \\ &= H(D|B) \\ &= 1 - p \end{aligned} \tag{2.23}$$

where the lower bound follows because our choice of $P_{D|B}$ *and* $f(\cdot)$ *may be suboptimal, and the subsequent equality follows because,* $Y = X = B$ *for our choice of* $P_{D|B}$ *and* $f(\cdot)$*.*

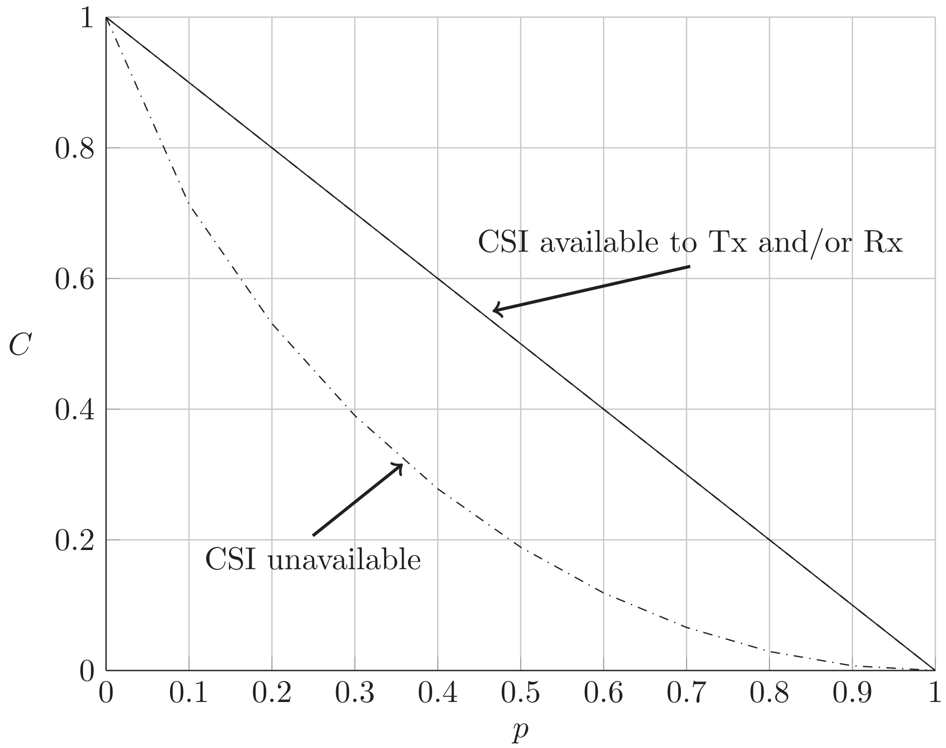


Figure 2.3: Capacity of the channel given in Example 1 as a function of p.

iv) Non-causal CSI available at both Tx and Rx: In this case

$$\begin{aligned} C &= \sum_b P_B(b) \max_{P_{X|B}(\cdot|b)} I(X;Y|B=b) \\ &= \frac{p}{2} I(X;Y|B=0) + \frac{p}{2} I(X;Y|B=1) + (1-p) I(X;Y|B=2) \\ &= 1-p. \end{aligned} \tag{2.24}$$

where we need that $I(X;Y|B=0) = I(X;Y|B=1) = 0$ *and* $I(X;Y|B=2) = 1$.

Figure 2.3 shows the capacities obtained for the cases i)-iv) as a function of p. We observe that, by having CSI, we increase the channel capacity compared to the case when CSI is unavailable at both Tx and Rx. We further observe that, in this example, CSI available at the Rx, CSI available at the Tx and CSI available at both Tx and Rx are equivalent. Of course, in general this is not the case.

2.3 Channels with States: Quasi-Static Scenario

In this section, we consider the quasi-static scenario [55, Sec. 5.4.1] of the channel with state. In this scenario, the coherence time is longer than the time needed for the transmission of the whole message and the channel state stays constant during the whole transmission. For the quasi-static scenario we distinguish two scenarios:

- **Composite channel** [18, Def. 2]: A collection of channels modeled as a parameterized sequence of N-dimensional conditional distributions $\{P_{Y^N|X^N,B}(\cdot|\cdot,b),\ b \in \mathcal{B}\}$. The channel law for a given N is determined by the random variable B, chosen according to some channel-state distribution P_B at the beginning of transmission and stays constant for all N.

- **Compound channel** [67, Ch. 4]: A collection of channels modeled as a parameterized sequence of N-dimensional conditional distributions $\{P_{Y^N|X^N,B}(\cdot|\cdot,b),\ b \in \mathcal{B}\}$. In contrast to the composite channel, b is not a random variable, i.e., there is no distribution according to which b is chosen. Instead, encoding and decoding strategies must work for all possible channels in the collection.

The capacity of both, composite and compound channels, is the same. Indeed, Effros *et al.* [18] remarked:

> "In particular, the capacity of the composite channel is a special case of the general channel capacity derived by Verdú and Han. However, the distribution over the collection of channels is not used in this capacity calculation, since the definition of Shannon capacity requires reliable communication for all channels in the collection. Hence, the Verdú–Han (Shannon) capacity of a composite channel will be the same as the Shannon capacity of a compound over the same collection of possible channels $\{b \in \mathcal{B}\}$, regardless of the P_B over the composite channel states."

In the next subsections, we specify the channel capacity of these channels for the cases where an expression is available in the literature.

2.3.1 No CSI available

The capacity in this scenario is based on the *mismatch capacity* [22]. The mismatched capacity is the highest rate at which reliable communication is possible over the channel with a given (possibly suboptimal) decoding rule. In other words, the decoding rule is mismatched to the actual channel.

The capacity of the memoryless compound channel with unavailable CSI at both transmitter and receiver is given by [8], [15], [67, Ch. 4]

$$C = \max_{Q} \min_{b \in \mathcal{B}} I(X;Y) \tag{2.25}$$

where Q corresponds to the input distribution, and the channel law is given by $P_{Y|X,B}(\cdot|\cdot, b)$. Lapidoth and Telatar [42] derived an expression similar to (2.25) for a special class of finite-state channels.

2.3.2 Perfect CSI available at Receiver

Suppose now that the Rx has access to the channel-state sequence B. In this case, the capacity is given by [67, Ch. 4]

$$C = \max_{Q} \min_{b \in \mathcal{B}} I(X;Y|B=b) \tag{2.26}$$

where the optimization is over all input distributions Q of X.

2.3.3 Perfect CSI available at Transmitter

We first consider a channel where the Tx has access to the CSI B. Wolfowitz [67, Ch. 4] has shown that the capacity of these channels corresponds to the capacity of the worst-case channel in B. In particular, the capacity is given by

$$C = \min_{b \in \mathcal{B}} \max_{Q} I(X;Y_b) \tag{2.27}$$

where, conditioned on $X = x$, Y_b has distribution $P_{Y|X,B}(\cdot|x, b)$ and Q corresponds to the input distribution. This capacity might be enhanced compared to the one achieved when no CSI is available at the Tx, since it can adapt its transmission to the current channel.

2.3.4 Perfect CSI available at Transmitter and Receiver

Suppose now that CSI is available at both Tx and Rx. Then, the Tx can adapt the transmission rate to the channel state, hence, it performs rate adaption. In this case, the capacity is given by [67, Ch. 4]

$$C = \max_{P_{X|B}} \min_{b \in \mathcal{B}} I(X;Y|B = b) \tag{2.28}$$

where the optimization is over all conditional distributions of X given B, which reflects the dependence of the inputs on the channel state.

2.3.5 Alternative Metrics

2.3.5.1 Capacity-versus-Outage

As mentioned in Definition 2, the capacity is the maximum rate at which information can be reliably transmitted, i.e., the probability of error (2.3) can be made arbitrarily small by letting the blocklength N tend to infinity. This definition requires that all channels in $\mathcal{B}$ must be treated equally, and a code that performs well on all channels must be designed. Hence, the capacity of a composite or compound channel is typically limited by the **worst channel** in $\mathcal{B}$. Consequently, the capacity may be low even if it is very unlikely that a "bad" channel occurs, [7], [67], because the probability of occurrence of "bad" or "good" channels is not taken into account. In fact, the probability of error may be bounded away from zero for every positive rate $R > 0$ if there is at least one channel in $\mathcal{B}$ that has zero capacity. Hence, the capacity of composite channels can be pessimistic. Taking advantage of the channel-state distribution, it may be possible to allow for errors in rare events. This is the case for the capacity-versus-outage metric.

Consider the composite channel where only the Rx has access to CSI, i.e., $V = B$. As Effros *et al.* [18] pointed out, the capacity-versus-outage metric is applied to cases where a variable rate R is not possible or desirable. Thus, the transmitter sends messages at a fixed R. If the channel is "good" (which happens most of the time) the message is received correctly. However, with some maximal probability q, the channel is "bad" and the decoder declares an outage, in which case the information is not decoded and lost. This approach is referred to as capacity-versus-outage, which we formally introduce next.

Definition 3 (Capacity-versus-outage [18, Def. 5]) *Consider a composite channel with CSI available only at the receiver side, i.e.,* $U = 0$ *and* $V = B$*. An (N,R) code for this channel consists of:*

1. *An independent message* W *uniformly distributed over the message sets* $\mathcal{W} \triangleq \{1, 2, \ldots, 2^{NR}\}$*.*

2. *An encoder:* $f_N : (W) \mapsto X^N$*.*

3. *An outage identification function:* $O : \mathcal{B} \mapsto \{0, 1\}$*.*

4. *A decoder:* $g_N : (Y^N, V^N) \mapsto \hat{W}$*, which only decodes when* $O = 1$*.*

Here, $\hat{W}$ *denotes the decoded message.*

The outage probability corresponds to the probability that the decoder determines that it cannot decode reliably the channel output and declares an outage. Hence, the outage probability is defined as

$$P_{out}^{(N)} \triangleq \Pr\{O = 0\}. \tag{2.29}$$

Correspondingly, the probability of error in non-outage channel states is

$$P_e^{(N)} = \Pr\{W \neq \hat{W} | O = 1\}. \tag{2.30}$$

Definition 4 (Outage-q achievable rate [18, Def. 5]) *A rate* R *is outage-q achievable if there exists a sequence of* (N, R) *codes such that*

$$\lim_{N\to\infty} P_{out}^{(N)} \leq q \tag{2.31}$$

and

$$\lim_{N\to\infty} P_e^{(N)} = 0. \tag{2.32}$$

The capacity-versus-outage C_q *is the supremum over all outage-q achievable rates.*

The following theorem presents the capacity-versus-outage of a composite channel.

Theorem 3 ([18, Th.1]) *The capacity-versus-outage of a composite channel with outage probability q is given by*

$$C_q = \sup_{Q} \underline{I}_q(X^N; Y^N|B) \tag{2.33}$$

where the supremum is over all joint input distributions Q^N of X^N, and $\underline{I}_q(X^N; Y^N|B)$ is the supremum of all αs satisfying

$$\Pr\left\{\frac{1}{N}\log\frac{P_{Y^N|X^N,B}(y^N|x^N,b)}{P_{Y^N|B}(y^N|b)} \leq \alpha\right\} \leq q. \tag{2.34}$$

Proof: The converse bound follows from [62] and the achievability is given in [18, Sec. IV]. ■

Definition 5 (Outage capacity [18]) *The outage capacity is defined as*

$$C_q^O = (1-q)C_q. \tag{2.35}$$

The outage capacity corresponds to the long-term average rate, obtained by sending messages over independent quasi-static composite channels. By the law of large numbers, a fraction $(1-q)$ of the time the Rx can correctly decode the information.

Remark 2 *Sometimes, the capacity-versus-outage C_q is referred to as outage capacity, and the outage capacity C_q^O is referred to as throughput.*

2.3.5.2 Opportunistic Rates

Consider the compound channel where only the Rx has access to CSI, i.e., $V = B$. As mentioned at the beginning of this chapter, we aim of communicating reliably at the highest data rate. For a Tx that wishes to communicate through a compound channel with the Rx at the highest reliable rate, Cover [13] suggested to not only consider the worst-case channel, but also the best one. This idea was generalized by Bergman [6], and is now known as the *broadcast approach.* The broadcast approach was used e.g., by Shamai [53] for a single-input single-output (SISO) Gaussian slowly fading channel.

The broadcast approach allows to deliver information rates which depend on the actual channel realization, when the Tx has no access to CSI. The broadcast

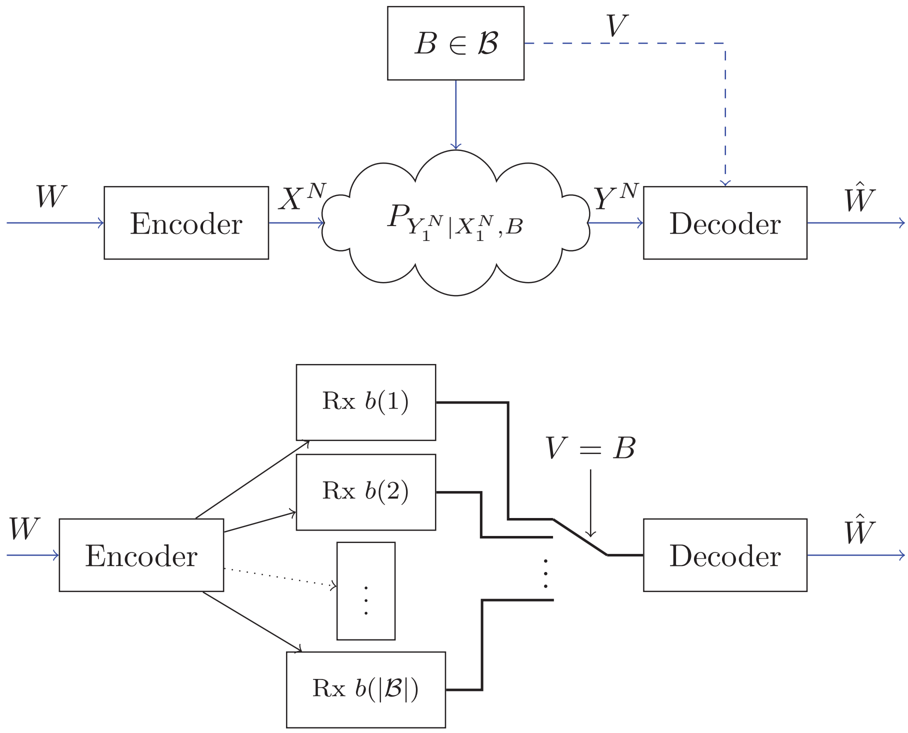


Figure 2.4: Broadcast approach for a channel with state.

approach is depicted in Figure 2.4. Specifically, in the upper part of Figure 2.4 we depict a compound channel with CSI available only at the Rx, in the lower part we represent graphically the broadcast approach for this channel. For the Tx, the compound channel is viewed as a broadcast channel with a given number of virtual Rxs indexed by the channel-state realization $B = b$, where $b \in \mathcal{B}$. The number of virtual Rxs is given by the cardinality of $|\mathcal{B}|$. Then, the encoder uses a broadcast code to encode the message and send it to all virtual Rxs. The Rx, which has access to B, chooses the appropriate decoder based on the current channel-state realization.

Using this approach, Körner and Marton [36] studied the case where the transmitted messages are divided into common and private messages sent over a degraded broadcast channel as in [13]. Common messages are sent to all receivers, private messages are sent only to the stronger receiver. The work by Körner and Marton [36] motivated the idea of **opportunistic codes** introduced by Digavvi and Tse for a quasi-static flat-fading channel [16], where the transmitted messages are divided into messages with high priority (w_H) and low priority (w_L). The messages w_H are recovered reliably for all channel realizations in $\mathcal{B}$, whereas the messages w_L are recovered only when the channel conditions are "good". This allows for reliable communication when the channel is bad and allows transmission at an increased rate when the channel is "good".

Next, we define formally an opportunistic code for the case where CSI is available at the Rx. To this end, we denote the set of opportunistic messages by $\{\Delta W(\cdot)\} \triangleq \{\Delta W(v), v \in \mathcal{V}\}$. Then, we define an opportunistic code as follows.

Definition 6 (Opportunistic code) *An $\big(N, R, \{\Delta R(v), v \in \mathcal{V}\}\big)$ opportunistic code consists of:*

1. *An independent message W uniformly distributed over the message sets $\mathcal{W} \triangleq \{1, 2, \ldots, 2^{NR}\}$.*

2. *An independent set of opportunistic message $\{\Delta W(\cdot)\}$ uniformly distributed over the message set $\Delta\mathcal{W}(v) \triangleq \{1, 2, \ldots, 2^{N\Delta R(V)}\}, v \in \mathcal{V}$.*

3. *An encoder: $f : (W, \{\Delta W(\cdot)\}) \mapsto X^N$.*

4. *A decoder: $g : (Y^N, V) \mapsto (\hat{W}, \Delta\hat{W}(V))$.*

Here, $\hat{W}$ and $\Delta\hat{W}(V)$ denote the decoded message and the decoded opportunistic message, respectively.

Definition 7 (Achievable opportunistic rates) *A rate pair $\big(R, \{\Delta R(\cdot)\}\big)$ is achievable if there exists a sequence of codes $\big(N, R, \{\Delta R(\cdot)\}\big)$ such that*

$$\Pr\big\{\hat{W} \neq W\big\} \to 0 \quad \text{as} \quad N \to \infty \tag{2.36}$$

and

$$\Pr\big\{(\hat{W}, \Delta\hat{W}(V)) \neq (W, \Delta W(V)) | V = v\big\} \to 0 \text{ as } N \to \infty, v \in \mathcal{V}. \tag{2.37}$$

The capacity is the supremum of the set of achievable rate tuples [20, Sec. 6.1].

This opportunism is studied in many fading channel scenarios, including the IC [32], [33]. In this dissertation, we consider opportunistic codes in Chapter 4.

Definition 8 (Average capacity) *Similarly to the outage capacity of the composite channel, for the compound channel, we define the average capacity as*

$$\bar{C} \triangleq \sup_{R,\Delta R} \{R + (1-p)\Delta R\}. \tag{2.38}$$

Intuitively, the average capacity corresponds to the long-term average rate, obtained by sending messages over independent quasi-static compound channels. By the law of large numbers a fraction p of the time the Rx can decode the message W, and a fraction $(1-p)$ of the time it can further decode the opportunistic message ΔW.

3

Interference Channel

As mentioned in Chapter 1, interference is a key limiting factor for the efficient use of the spectrum in modern wireless networks. For this reason, understanding its effects on the reliable communication and how to deal with them is an open research line in Information Theory. The basic model used to better understand these effects is the interference channel (IC).

This chapter introduces the two-user IC depicted in Figure 3.1 [20, Ch. 6]. This channel models the scenario where two independent transmitters (Txs) want to communicate a message W_i, $i = 1, 2$, to two different receivers (Rxs) over a shared channel. In the figure, each message W_i, $i = 1, 2$, is separately encoded into a codeword X_i^N and transmitted over the channel. Rx i, $i = 1, 2$ produces the estimate $\hat{W}_i$ of W_i based on the received signal Y_i^N. Because of the shared medium used for the communication, the signal at each receiver may be affected not only by the noise in the channel, but also by the interference caused by the other transmitted codeword. The maximum data rate at which the users can communicate reliably in such a channel is an interesting problem that has received

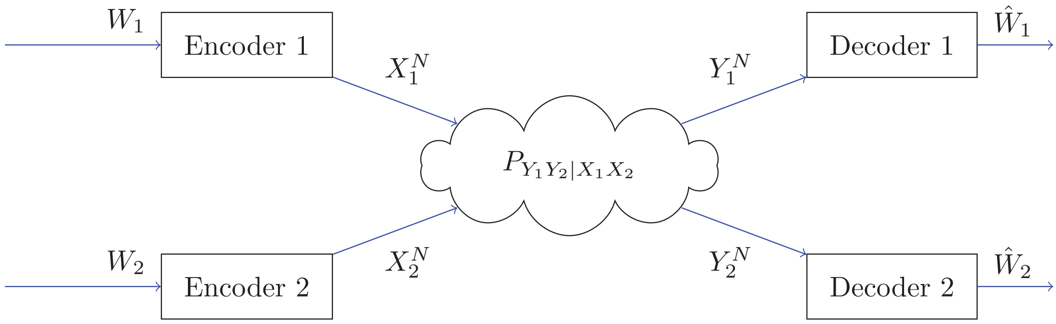


Figure 3.1: Interference channel [20, Ch. 6].

great attention in the information-theory literature; see e.g., [1], [12], [19], [21], [27], [37], [50] and references therein. Despite its vast interest, its channel capacity is still unknown.

3.1 Discrete Memoryless Interference Channel

The discrete memoryless IC $(\mathcal{X}_1 \times \mathcal{X}_2,\ p(y_1, y_2|x_1, x_2),\ \mathcal{Y}_1 \times \mathcal{Y}_2\ )$ [20, Sec. 6.1] is defined by the finite sets $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_1$, $\mathcal{Y}_2$ and the channel transition probability

$$P_{Y_1^N Y_2^N|X_1^N X_2^N}(y_1^N, y_2^N|x_1^N, x_2^N) = \prod_{k=1}^{N} P_{Y_1Y_2|X_1X_2}(y_{1,k}, y_{2,k}|x_{1,k}, x_{2,k}). \tag{3.1}$$

Definition 9 (Code for the discrete memoryless IC) *An (N, R_1, R_2) code for the IC consists of:*

1. *Two independent messages W_1 and W_2 uniformly distributed over the message sets $\mathcal{W}_i \triangleq \{1, 2. \dots, 2^{NR_i}\}$, $i = 1, 2$.*

2. *Two encoders: $f_i : W_i \mapsto X_i^N$, $i = 1, 2$.*

3. *Two decoders: $g_i : Y_i^N \mapsto \hat{W}_i$, $i = 1, 2$.*

Here $\hat{W}_i$ denotes the decoded message.

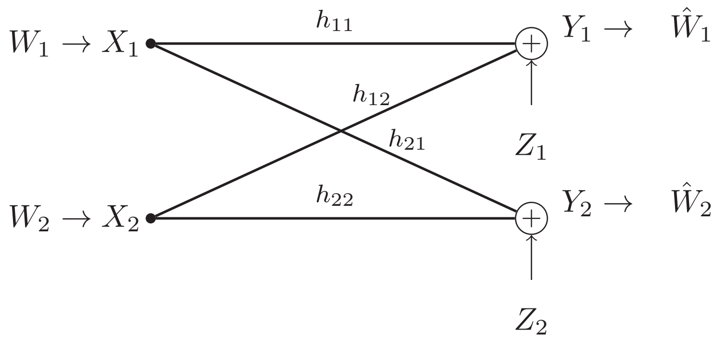


Figure 3.2: Gaussian IC [20, Sec. 6.4].

Definition 10 (Achievable rates [20, Ch. 6]) *A rate pair (R_1, R_2) is achievable if there exists a sequence of (N, R_1, R_2) codes such that*

$$\Pr\{\hat{W}_1 \neq W_1 \cup \hat{W}_2 \neq W_2\} \to 0 \text{ as } N \to \infty. \tag{3.2}$$

Definition 11 (Capacity region [20, Ch. 6]) *The capacity region is the closure of the set of achievable rate pairs.*

The capacity region of the discrete memoryless IC is not known in general. However, when the level of interference is strong or very strong, then the capacity region is known [20, Sec.6.3] Specifically a discrete memoryless IC operates under strong interference level if the following conditions are satisfied [50, Eq. 12]:

$$I(X_1;Y_1|X_2) \leq I(X_1;Y_2|X_2) \tag{3.3}$$

$$I(X_2;Y_2|X_1) \leq I(X_2;Y_1|X_1). \tag{3.4}$$

It operates under very strong interference level if the following conditions are satisfied [50, Eq. 11]:

$$I(X_1;Y_1|X_2) \leq I(X_1;Y_2) \tag{3.5}$$

$$I(X_2;Y_2|X_1) \leq I(X_2;Y_1). \tag{3.6}$$

3.2 Gaussian Interference Channel

As defined in [20, Sec. 6.4], the two-user Gaussian IC (see Figure 3.2) is considered as a simple model for a wireless IC, and hence there are many works related; see e.g., [2], [21], [37] and references therein. We consider a discrete-time channel. The channel outputs at time instant k, corresponding to the channel inputs $X_{1,k}$ and $X_{2,k}$, are given by

$$\begin{aligned} Y_{1,k} &= h_{11}X_{1,k} + h_{12}X_{2,k} + Z_{1,k} \\ Y_{2,k} &= h_{22}X_{2,k} + h_{21}X_{1,k} + Z_{2,k} \end{aligned} \tag{3.7}$$

where h_{ij} is the channel coefficient from Tx j to Rx i, and $Z_1 \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ and $Z_2 \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ are noise components. Here we use the notation $H \sim \mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ to indicate that H is a circularly-symmetric, complex Gaussian random variable of mean μ and variance σ^2. It is assumed that X_1^N and X_2^N are constrained by the average-power constraint

$$\tfrac{1}{N}\mathsf{E}\left[|X_i|^2\right] \leq \mathsf{P}, \; i = 1, 2 \tag{3.8}$$

and we define signal-to-noise ratio (SNR) as

$$\mathrm{SNR}_i = \frac{|h_{ii}|^2\mathsf{P}}{\sigma^2}, \; i = 1, 2 \tag{3.9}$$

and the interference-to-noise ratio (INR) as

$$\mathrm{INR}_i = \frac{|h_{ij}|^2\mathsf{P}}{\sigma^2}, \; i \neq j \text{ and } i, j = 1, 2. \tag{3.10}$$

By assuming without loss of generality (wlog) that $\sigma^2 = 1$, the SNRs are given by $\mathrm{SNR}_1 = |h_{11}|^2\mathsf{P}$ and $\mathrm{SNR}_2 = |h_{22}|^2\mathsf{P}$, and the INRs are given by $\mathrm{INR}_1 = |h_{12}|^2\mathsf{P}$ and $\mathrm{INR}_2 = |h_{21}|^2\mathsf{P}$.

Definition 12 (Code for the Gaussian IC) *An (N, R_1, R_2) code for the Gaussian IC consists of:*

1. *Two independent messages W_1 and W_2 uniformly distributed over the message sets $\mathcal{W}_i \triangleq \{1, 2\ldots, 2^{NR_i}\}$, $i = 1, 2$.*
2. *Two encoders: $f_i : W_i \mapsto X_i^N$, $i = 1, 2$, satisfying the average-power constraint* (3.8)*.*
3. *Two decoders: $g_i : Y_i^N \mapsto \hat{W}_i$, $i = 1, 2$.*

Here, $\hat{W}_i$ denotes the decoded message.

Achievable rates and the capacity region of the Gaussian IC are defined as in Definitions 10 and 11.

3.2.1 Capacity of the Gaussian IC

As mentioned before, the capacity region of the Gaussian IC is not completely characterized in general. However, when the level of interference is strong or very strong, then the capacity region is known [27], [50], [12]. Specifically a Gaussian IC operates under strong interference level if the following conditions are satisfied:

$$|h_{21}| \leq |h_{11}| \tag{3.11}$$
$$|h_{12}| \leq |h_{22}|. \tag{3.12}$$

It operates under very strong interference level if the following conditions are satisfied:

$$|h_{22}| \leq \frac{|h_{12}|}{1+|h_{11}|} \tag{3.13}$$
$$|h_{11}| \leq \frac{|h_{21}|}{1+|h_{22}|}. \tag{3.14}$$

Remark 3 *The conditions* (3.11) *and* (3.12) *are equivalent to the conditions* (3.3) *and* (3.4) *for the discrete memoryless IC. Similarly, the conditions* (3.13) *and* (3.14) *are equivalent to the conditions* (3.5) *and* (3.6) *for the discrete memoryless channel [20, Sec. 6.4.2], [50, Eq. 11, Eq. 12].*

In the following theorems we present the corresponding capacity regions for these cases.

Theorem 4 (Capacity region under strong interference [50]) *The capacity region of the Gaussian IC under strong interference is the union of the set of rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq \log(1+\mathrm{SNR}_1) \tag{3.15}$$
$$R_2 \leq \log(1+\mathrm{SNR}_2) \tag{3.16}$$
$$R_1+R_2 \leq \min\{\log(1+\mathrm{SNR}_1+\mathrm{INR}_1), \log(1+\mathrm{SNR}_2+\mathrm{INR}_2)\}. \tag{3.17}$$

Proof: See [50] and [27]. ■

Theorem 5 (Capacity region under very strong interference [12]) *The capacity region of the Gaussian IC under very strong interference is the union of the set of rate pairs* (R_1, R_2) *satisfying*

$$R_1 \leq \log(1 + \mathrm{SNR}_1) \tag{3.18}$$

$$R_2 \leq \log(1 + \mathrm{SNR}_2). \tag{3.19}$$

Proof: See [12]. ■

Theorem 5 further shows that interference does not affect the capacity when it is very strong. Indeed, the capacity region corresponds to the case of two parallel point-to-point channels.

Some converse and achievability bounds have been derived and proposed for other interference levels using different approaches. We summarize the bounds in the following section.

3.2.1.1 Converse bounds

- **Genie-aided bound**: In this approach, each receiver is provided by additional information, such as the interference signal or a noisy version of the caused interference, that allows it to decode both messages, see e.g., [37], [21].

- **Degraded IC** [37]: In this approach, the IC is transformed to a degraded broadcast channel.

3.2.1.2 Achievability bounds

- **Han-Kobayashi**: The best known achievability strategy for the remaining unsolved cases was proposed by Han and Kobayashi [27]. It combines the ideas of time-sharing and rate-splitting, i.e., dividing the transmitted message into two parts: a common part which can be decoded by both receivers, and a private part, which can be decoded only by the intended receiver.

  Theorem 6 (Han-Kobayashi Achievability Bound) *A rate pair*

(R_1, R_2) is achievable for a discrete memoryless IC if

$$
\begin{aligned}
R_1 &< I(X_1; Y_1 | U_2, T) \\
R_2 &< I(X_2; Y_2 | U_1, T) \\
R_1 + R_2 &< I(X_1, U_2; Y_1 | T) + I(X_2; Y_2 | U_1, U_2, T) \\
R_1 + R_2 &< I(X_2, U_1; Y_2 | T) + I(X_1; Y_1 | U_1, U_2, T) \\
R_1 + R_2 &< I(X_1, U_2; Y_1 | U_1, T) + I(X_2, U_1; Y_2 | U_2, T) \\
2R_1 + R_2 &< I(X_1, U_2; Y_1 | T) + I(X_1; Y_1 | U_1, U_2, T) + I(X_2, U_1; Y_2 | U_2, T) \\
R_1 + 2R_2 &< I(X_2, U_1; Y_2 | T) + I(X_2; Y_2 | U_1, U_2, T) + I(X_1, U_2; Y_1 | U_1, T)
\end{aligned}
\tag{3.20}
$$

for some probability mass function (pmf) $P_T P_{U_1,X_1|T} P_{U_2,X_2|T}$, where $|\mathcal{U}_1| \leq |\mathcal{X}_1| + 4$, $|\mathcal{U}_2| \leq |\mathcal{X}_2| + 4$, and $|\mathcal{T}| \leq 6$.

Proof: See [27] and also [20, Sec. 6.5]. ■

The Han-Kobayashi achievability bound can be extended to the Gaussian IC with average power constraints and for Gaussian codebooks. However, this task is in general very complicated. Etkin *et al.* [21] have shown that a very simple Han–Kobayashi-type scheme can achieve rates within 1 bit/s/Hz of the capacity of the Gaussian IC for all values of the channel parameters.

- **Treating interference as noise (TIN)** [20, Sec. 6.4.1]: The Han-Kobayashi scheme reduces to TIN, if one sets the rate of the common messages to zero, i.e., $U_1 = U_2 = 0$. TIN achieves all rate pairs satisfying

$$R_1 < \log\left(1 + \frac{\mathrm{SNR}_1}{1 + \mathrm{INR}_1}\right) \tag{3.21}$$

$$R_2 < \log\left(1 + \frac{\mathrm{SNR}_2}{1 + \mathrm{INR}_2}\right). \tag{3.22}$$

- **Simultaneous nonunique decoding** [20, Sec. 6.4.1]: The Han-Kobayashi scheme reduces to simultaneous nonunique decoding when $U_1 = X_1$ and $U_2 = X_2$. In this case, all rate pairs (R_1, R_2) are achievable if they satisfy

$$R_1 < \log(1 + \mathrm{SNR}_1) \tag{3.23}$$

$$R_2 < \log(1 + \mathrm{SNR}_2) \tag{3.24}$$

$$R_1 + R_2 < \min\{\log(1 + \mathrm{SNR}_1 + \mathrm{INR}_1), \log(1 + \mathrm{SNR}_2 + \mathrm{INR}_2)\}. \tag{3.25}$$

- **Time division with power control** [20, Sec. 6.4.1]: This approach consists of orthogonalizing the users. Suppose a fraction of time, $\tau \in [0, 1]$, is allocated to Tx_1 for transmission with power $\frac{\mathsf{P}}{\tau}$. The Tx_2 transmits in the remaining $(1-\tau)$ fraction of time with power $\frac{\mathsf{P}}{(1-\tau)}$. With such a scheme, all rate pairs (R_1, R_2) are achievable if they satisfy

$$R_1 \quad < \quad \tau \log\left(1 + \frac{\mathrm{SNR}_1}{\tau}\right) \tag{3.26}$$

$$R_2 \quad < \quad (1-\tau)\log\left(1 + \frac{\mathrm{SNR}_2}{(1-\tau)}\right). \tag{3.27}$$

As mentioned before, Han-Kobayashi is the best known achievability strategy, but it was unclear how close to capacity can such a scheme get and whether there are other strategies that can perform better. Etkin *et al.* [21] demonstrated that the Han-Kobayashi scheme is within 1 bit from a converse bound. Hence, its gap to capacity does never exceed 1 bit. To obtain their result Etkin *et al.* introduced the generalized degrees of freedom (GDoF) as a natural generalization of the degrees of freedom or capacity pre-log of the point-to-point channel [39] to scenarios with multiple users. We formally introduce the GDoF in Section 3.2.2. Before, we introduce the normalized interference as $\alpha \triangleq \frac{\log \mathrm{INR}}{\log \mathrm{SNR}}$. Based on α, we can divide the interference into the following regions (a similar division was used by Jafar and Vishwanath [29]):

- very weak interference (VWI) for $\alpha \leq \frac{1}{2}$,
- weak interference (WI) for $\frac{1}{2} < \alpha \leq \frac{2}{3}$,
- moderate interference (MI) for $\frac{2}{3} < \alpha \leq 1$,
- strong interference (SI) for $1 < \alpha \leq 2$,
- very strong interference (VSI) for $2 < \alpha$.

3.2.2 Generalized Degrees of Freedom

Let us consider a symmetric setup of the Gaussian IC considered in (3.7), i.e., $h_{11} = h_{22} = h_d$, $h_{12} = h_{21} = h_c$. This implies that $\mathrm{SNR}_1 = \mathrm{SNR}_2 = \mathrm{SNR}$ and

$\mathrm{INR}_1 = \mathrm{INR}_2 = \mathrm{INR}$, and the channel model in (3.7) becomes

$$\begin{aligned} Y_{1,k} &= h_d X_{1,k} + h_c X_{2,k} + Z_{1,k} \\ Y_{2,k} &= h_d X_{2,k} + h_c X_{1,k} + Z_{2,k}. \end{aligned} \tag{3.28}$$

The GDoF is given by

$$\mathcal{D}(\alpha) = \lim_{\substack{\mathrm{SNR}\to\infty: \\ \mathrm{INR}=\mathrm{SNR}^{\alpha}}} \frac{C_{\mathrm{sym}}}{\log(1+\mathrm{SNR})} \tag{3.29}$$

where $\alpha = \frac{\log \mathrm{INR}}{\log \mathrm{SNR}}$ and C_{sym} corresponds to the symmetric capacity of the Gaussian IC, which in the symmetric setting is given by the maximum sum rate $R \triangleq R_1 + R_2$.

The GDoF is a useful tool to characterize the channel capacity of the Gaussian IC in the high-SNR regime. Furthermore, coding strategies that achieve the maximum GDoF achieve capacity within a constant number of bits [9], [21]. For the IC, the GDoF as a function of α is given by [21]

$$\mathcal{D}(\alpha) = \begin{cases} 1-\alpha & 0 \le \alpha \le \frac{1}{2} \\ \alpha & \frac{1}{2} \le \alpha \le \frac{2}{3} \\ 1-\frac{\alpha}{2} & \frac{2}{3} \le \alpha \le 1 \\ \frac{\alpha}{2} & 1 \le \alpha \le 2 \\ 1 & 2 \le \alpha. \end{cases}$$

As can be observed from Figure 3.3, $\alpha \mapsto \mathcal{D}(\alpha)$ exhibits a W-shaped curve. This curve is therefore known as the the W-curve.

A useful technique in the characterization of the GDoF is the deterministic approach, which maps a Gaussian network to a deterministic channel, where channel outputs are deterministic functions of their inputs [4], [5], [9].

3.3 Linear Deterministic Model

The general deterministic IC model was first explored by El Gamal and Costa [19], in which a part of the interfering signal is completely invisible to the other link. Then, Avestimehr *et al.* [5] introduced a linear deterministic model (LDM) for wireless relay networks. The main idea of this model is to have a simple model that still captures the key features of a wireless communication channels, but

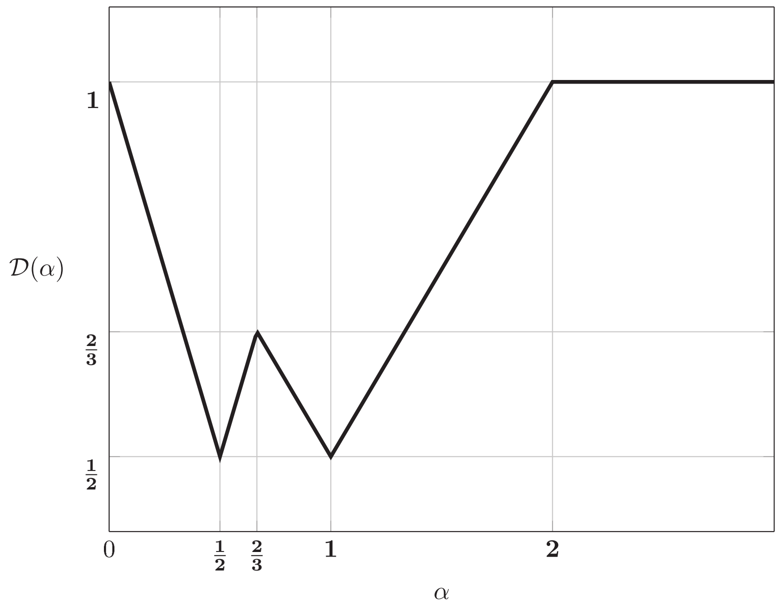


Figure 3.3: W curve.

simplifies the analysis by eliminating the randomness of noise from the setup. In this model, the Txs send bit vectors, and depending on the channel strength, a certain number of bits will be received at the Rxs. The model is an approximation of the Gaussian IC under the following assumptions:

- the operation regime is at high-SNR, where the signal power is larger than the noise power.

- at a given receiver, the attained signals from different transmitters are received at different power levels.

Avestimehr *et al.* [5] motivate and explain their idea for the point-to-point channel, the broadcast channel, the multiple access channel (MAC) and relay networks. In order to introduce the LDM of the IC we are going to study in Chapter 4, we first motivate the LDM on a simple point-to-point channel.

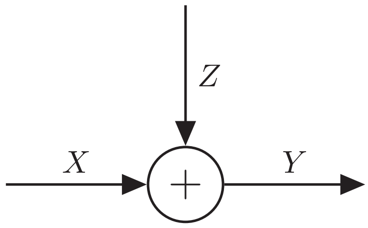


Figure 3.4: Point-to-point channel.

3.3.1 Linear Deterministic Point-to-point Channel

We consider the Gaussian point-to-point channel depicted in Figure 3.4. For the sake of simplicity, we assume that the channel output is given by

$$Y = X + Z \tag{3.30}$$

where X denotes the real-valued channel input with power constraint $\mathsf{E}\left[X^2\right] = 1$, and Z corresponds to the additive Gaussian noise $Z \sim \mathcal{N}(0,1)$. The corresponding LDM is presented in Figure 3.5. The real-valued channel input X is expanded as a binary vector, which is interpreted as a succession of bits at different signal levels. For example, in Figure 3.5 we have $X = \{b_1, b_2, b_3\}$, where $b_i,\ i = 1,2,3$ are bits. The most significant bit (b_1) coincides with the highest signal level, the least significant bit (b_3) with the lowest signal level. In the deterministic model, the noise is modeled as a truncation of the signal, and bits below the noise level are discarded. For this example, the Rx can see only the n most significant bits of $\mathbf{X}$ without any noise and the rest are not seen at all. Mathematically, the channel output of the LDM of the point-to-point channel is given by

$$\mathbf{Y} = \mathsf{S}_n \mathbf{X} \tag{3.31}$$

where S is a $q \times q$ down-shift matrix as defined in Section 1.3 and with the correspondence $n = \log \mathrm{SNR}$. Thus, there is a correspondence between n and SNR in dB scale. The capacity of the channel (3.31), which is $C = n$, is an approximation of the capacity of the Gaussian point-to-point channel at high SNR [14, Ch. 9].

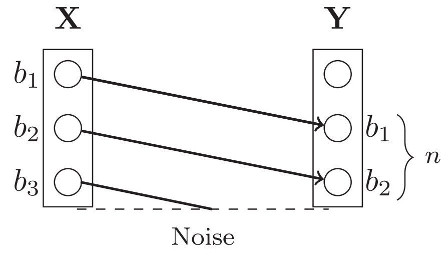


Figure 3.5: LDM for the point-to-point channel.

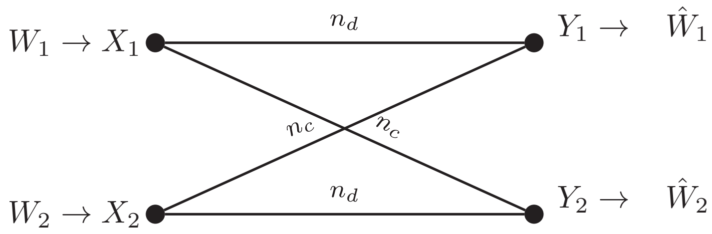


Figure 3.6: LDM for the IC.

3.3.2 Linear Deterministic IC

Bresler and Tse [9] proposed the LDM of the Gaussian IC. For simplicity, we shall focus here on the symmetric case. The same model will also be used in Chapter 4. The LDM of the IC is depicted in Figure 3.6, where the channel outputs at time instant k are given by [9]

$$\mathbf{Y}_{1,k} = \mathsf{S}_{n_d}\mathbf{X}_{1,k} \oplus \mathsf{S}_{n_c}\mathbf{X}_{2,k} \tag{3.32}$$

$$\mathbf{Y}_{2,k} = \mathsf{S}_{n_d}\mathbf{X}_{2,k} \oplus \mathsf{S}_{n_c}\mathbf{X}_{1,k}. \tag{3.33}$$

The signal strengths or power of the channel coefficients for a normalized power constraint $\mathsf{E}\left[|X_i|^2\right] = 1,\ i = 1, 2$, are given by

$$n_d = \lfloor \log_2 |h_d|^2 \rfloor = \lfloor \log_2 \mathrm{SNR} \rfloor \tag{3.34}$$

$$n_c = \lfloor \log_2 |h_c|^2 \rfloor = \lfloor \log_2 \mathrm{INR} \rfloor. \tag{3.35}$$

Let $q \triangleq \max\{n_d, n_c\}$. In (3.32) and (3.33), the channel inputs are the binary vectors $\mathbf{X}_{i,k} \in \mathbb{F}_2^{q\times 1}$ and the channel outputs are $\mathbf{Y}_{i,k} \in \mathbb{F}_2^{q\times 1}$, $i = 1, 2$. The operations of (3.32) and (3.33) are illustrated in Figure 3.7. In the figure, we observe the LDM of the IC for one time instant.[1] Both Txs use the same transmission strategy. In this case, $n_d = 5$ and $n_c = 3$ and $q = \max(5, 3) = 5$. Each Tx_i, $i = 1, 2$ sends a vector of 5 bits which is received by the corresponding Rx_i. Furthermore, each Tx interferes the communication, i.e., Tx_2 interferes to Rx_1 and Tx_1 interferes to Rx_2. The number of bits that interfere to the intended signal is given by the first n_c most significant bits in $\mathbf{X}_i$, $i = 1, 2$. For both signals (communication and interfering signals), the down-shift matrix S (defined in Section 1.3) is the matrix that mathematically shifts the input vectors $\mathbf{X}_i$, $i = 1, 2$ down according to n_d and n_c. Specifically, the down-shift matrix S_{n_d}, in this example 5×5, shifts $\mathbf{X}_i$, $i = 1, 2$, down by $q - n_d$ elements, which in this case corresponds to zero-shifted elements. The down-shift matrix S_{n_c} shifts $\mathbf{X}_j$, $j = 1, 2$, down by $q - n_c$ elements, which in this case corresponds to two-shifted elements. Then, at the Rx side, the signals are received at different levels (see Figure 3.7). At the Rxs, all received bits at the same signal level will be added using a modulo-2 sum.

In the figure, we also see the matrix L_d (defined in Section 1.3) of dimension 5×5 ($\mathsf{L}_{\mathsf{n_c}}\mathbf{Y}_1$ in the figure). This matrix is not directly in the channel model but is used in the proofs of Chapter 4. This matrix selects the d lowest components of a vector of dimension q, i.e., $d = n_c = 3$ in this case. The normalized interference level α for the LDM is $\alpha \triangleq \frac{n_d}{n_c}$.

3.3.3 Channel Capacity

The channel capacity of the linear deterministic IC was obtained by Bresler and Tse [9] and by El Gamal and Costa [19]. For the sake of completeness, we next present these results specialized to the symmetric setup.

Theorem 7 (Channel Capacity of the linear deterministic IC) *The sum capacity of the two-user linear deterministic IC is equal to the union of the set of all sum rates $R \triangleq R_1 + R_2$ satisfying*

$$R \leq 2n_d \tag{3.36}$$

$$R \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{3.37}$$

$$R \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{3.38}$$

[1]For the sake of simplicity, we omit the temporal index.

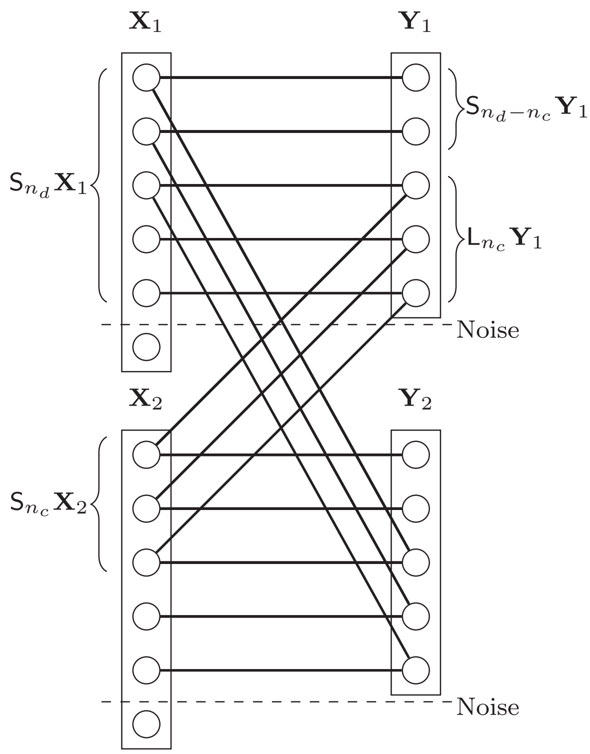


Figure 3.7: LDM of the IC.

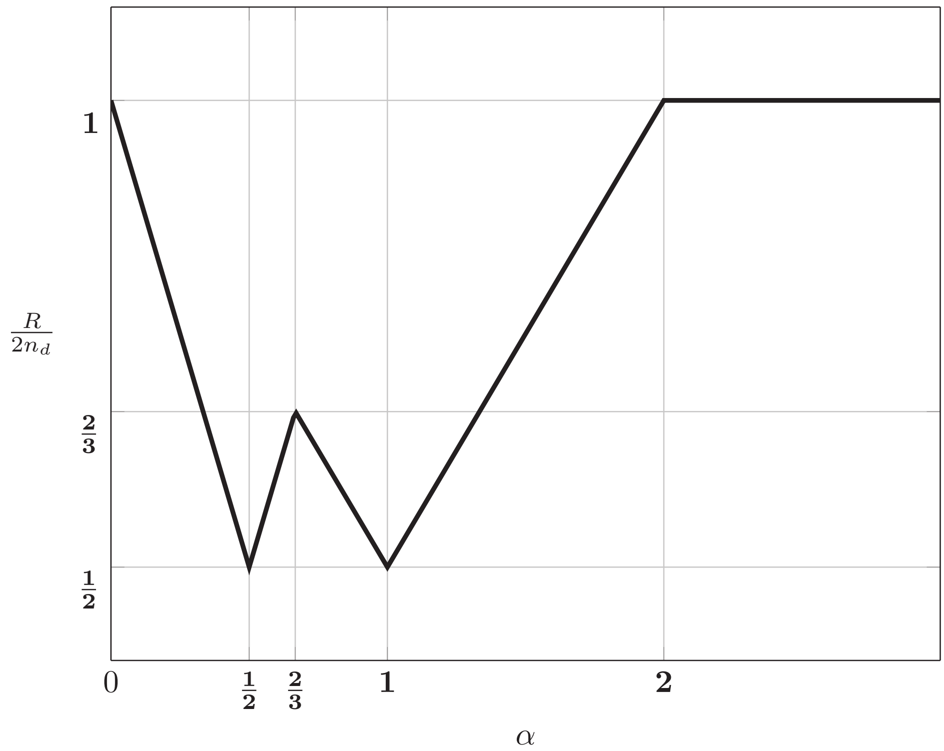


Figure 3.8: Capacity normalized by $2n_d$ of the linear deterministic IC.

Proof: The proof is given in [19, Sec. II]. For the achievability bounds, El Gamal and Costa [19, Th. 1] use the Han-Kobayashi scheme [27] for a general IC. Bresler and Tse [9, Section 4] use a specific Han-Kobayashi strategy for the special case of the LDM. Jafar and Vishwanath [29] present an alternative achievability scheme for the K-user deterministic IC, which particularized for the two-user IC will be referenced in this thesis. For the sake of completeness, we present the proofs of converse and achievability bounds of Theorem 7 in Sections 3.3.4 and 3.3.5, respectively. ■

Note that the bound (3.36) is only tight in the VSI region or in absence of interference.

In Figure 3.8, we plot the channel capacity normalized by $2n_d$ of the linear deterministic IC. We observe that the normalized capacity coincides with the W-curve of the Gaussian IC in Figure 3.3. Hence, the LDM provides insights on the Gaussian IC at high SNR. Furthermore, Bresler and Tse [9] demonstrated that the linear deterministic IC uniformly approximates the Gaussian channel within a constant number of bits. This result is presented in the following theorem.

Theorem 8 ([9, Th. 1]) *The capacity of the two-user Gaussian IC with signal and interference to noise ratios* SNR *and* INR *is within 42 bits per user of the capacity of the linear deterministic IC with gains* $n_d = \lfloor \log_2 \mathrm{SNR} \rfloor$, $n_c = \lfloor \log_2 \mathrm{INR} \rfloor$.

Proof: See [9]. ■

3.3.4 Converse Bounds for the Linear Deterministic IC

In this section we present the proofs of the converse bounds in Theorem 7.

3.3.4.1 Proof of (3.36)

This bound corresponds to the bound derived in [49]. We begin by applying Fano's inequality [14, Th. 2.10.1]

$$\begin{aligned} N(R_1 - \epsilon_{1N}) &\le I(W_1; \mathbf{Y}_1^N | W_2) \\ &= H(\mathbf{Y}_1^N | W_2) - H(\mathbf{Y}_1^N | W_1, W_2) \\ &\le H(\mathsf{S}_{n_d} \mathbf{X}_1^N) \end{aligned} \tag{3.39}$$

where $\epsilon_{1N} \to 0$ as $N \to \infty$. Here, the first step follows by giving W_2 as extra information and because W_1 and W_2 are independent of each other. The last step follows because W_2 determines $\mathbf{X}_2^N$, so we can subtract its contribution from Y_1^N. The entropy can be upper-bounded as

$$H(\mathsf{S}_{n_d} \mathbf{X}_1^N) \le N n_d. \tag{3.40}$$

By symmetry, we obtain the same bound for the other user. By combining the results for both users, dividing the result by N and taking the limit as $N \to \infty$, we prove (3.36).

3.3.4.2 Proof of (3.37)

This bound was derived in [19]. It was also obtained independently by Kramer [37]. We begin by applying Fano's inequality to obtain

$$N(R_1 - \epsilon_{1N}) \le I(W_1; \mathbf{Y}_1^N | W_2)$$

$$
\begin{aligned}
&= H(\mathbf{Y}_1^N|W_2) - H(\mathbf{Y}_1^N|W_1, W_2) \\
&= H(\mathsf{S}_{n_d}\mathbf{X}_1^N) \qquad (3.41)
\end{aligned}
$$

where $\epsilon_{1N} \to 0$ as $N \to \infty$. Here, the first step follows by giving W_2 as extra information and because W_1 and W_2 are independent of each other. The last step follows because W_2 determines $\mathbf{X}_2^N$, so we can subtract its contribution from Y_1^N. Likewise we have

$$
\begin{aligned}
N(R_2 - \epsilon_{2N}) &\leq I(W_2; \mathbf{Y}_2^N) \\
&= H(\mathbf{Y}_2^N) - H(\mathbf{Y}_2^N|W_2) \\
&= H(\mathbf{Y}_2^N) - H(\mathsf{S}_{n_c}\mathbf{X}_1^N) \qquad (3.42)
\end{aligned}
$$

where $\epsilon_{2N} \to 0$ as $N \to \infty$. The last step follows because W_2 determines $\mathbf{X}_2^N$, so we can subtract its contribution from Y_2^N. Combining (3.41) and (3.42) yields

$$
\begin{aligned}
N(R_1 + R_2 - \epsilon_{1N} - \epsilon_{2N}) &\leq H(\mathsf{S}_{n_d}\mathbf{X}_1^N) + H(Y_2^N) - H(\mathsf{S}_{n_c}\mathbf{X}_1^N) \\
&\leq H(\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N) + H(\mathbf{Y}_2^N) \qquad (3.43)
\end{aligned}
$$

where the last step follows because $H(F) - H(G) \leq H(F|G)$ for any two random variables F and G. The entropies in (3.43) can be upper-bounded as

$$
\begin{aligned}
H(\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N) &\leq N(n_d - n_c)^+ \qquad (3.44) \\
H(\mathbf{Y}_2^N) &\leq N\max(n_d, n_c). \qquad (3.45)
\end{aligned}
$$

By dividing (3.43) by N and taking the limit as $N \to \infty$, we obtain (3.37).

3.3.4.3 Proof of (3.38)

This bound is proved by using the genie-aided approach suggested by Etkin *et al.* [21]. Specifically, the bound follows by giving the extra information $(\mathsf{S}_{n_c}\mathbf{X}_1^N)$ to Rx_1. By Fano's inequality, we have

$$
\begin{aligned}
N(R_1 - \epsilon_{1N}) &\leq I(W_1; \mathbf{Y}_1^N) \\
&\leq I(W_1; \mathbf{Y}_1^N, \mathsf{S}_{n_c}\mathbf{X}_1^N) \\
&= I(W_1; \mathsf{S}_{n_c}\mathbf{X}_1^N) + I(W_1; \mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N) \\
&= H(\mathsf{S}_{n_c}\mathbf{X}_1^N) + H(\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N) - H(\mathbf{Y}_1^N|W_1, \mathsf{S}_{n_c}\mathbf{X}_1^N) \\
&= H(\mathsf{S}_{n_c}\mathbf{X}_1^N) + H(\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N) - H(\mathsf{S}_{n_c}\mathbf{X}_2^N) \qquad (3.46)
\end{aligned}
$$

where $\epsilon_{1N} \to 0$ as $N \to \infty$. Analogously, by giving the extra information $(\mathsf{S}_{n_c}\mathbf{X}_2^K)$ to Rx_2, we obtain

$$N(R_2 - \epsilon_{2N}) \leq H(\mathsf{S}_{n_c}\mathbf{X}_2^N) + H(\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N)H(\mathsf{S}_{n_c}\mathbf{X}_1^N) \tag{3.47}$$

where $\epsilon_{2N} \to 0$ as $N \to \infty$. Thus, (3.46) and (3.47) yield

$$N(R_1 + R_2 - \epsilon_{1N} - \epsilon_{2N}) \leq H(\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N) + H(\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N). \tag{3.48}$$

The individual entropies can be upper-bounded as

$$\begin{aligned} H(\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N) &= H(\mathsf{S}_{n_d}\mathbf{X}_1^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N|\mathsf{S}_{n_c}\mathbf{X}_1^N) \\ &\leq N\max\{(n_d - n_c)^+, n_c\} \qquad (3.49) \\ H(\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N) &= H(\mathsf{S}_{n_d}\mathbf{X}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_2^N) \\ &\leq N\max\{(n_d - n_c)^+, n_c\}. \qquad (3.50) \end{aligned}$$

By dividing (3.48) by N and taking the limit as $N \to \infty$, we obtain (3.38).

3.3.5 Achievability Bounds for the Linear Deterministic IC

In this section, we present the proofs of the achievability bounds in Theorem 7. We specialize the schemes presented by Jafar and Viswanath in [29, Sec. C] for the symmetric K-user IC to the two-user IC. For all interference regions, we assume that both Txs use the same transmission strategy.

3.3.5.1 Very Weak Interference

The symbols transmitted by both Txs (normalized by n_d) are depicted in Figure 3.9. Specifically, both Txs use uncoded transmission. Figure 3.10 depicts the signal levels of the transmitted signals (normalized by n_d) as observed at Rx_1, when it is affected by interference. At the Rx side we observe that the most significant $n_d(1-\alpha)$ levels are received without interference. Block $\boxed{\mathrm{B}}$ is affected by interference and is treated as an erasure [57]. We thus obtain the individual rate

$$R_1 = (n_d - n_c)\frac{\text{bits}}{\text{ch. use}}. \tag{3.51}$$

User 2 achieves the same rate. Thus, the sum rate R is given by

$$R = 2(n_d - n_c)^+ \ \frac{\text{bits}}{\text{sub-channel use}}. \tag{3.52}$$

This coding scheme achieves (3.38) evaluated for VWI.

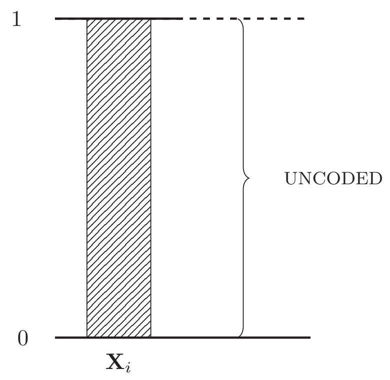


Figure 3.9: Transmitted symbols (VWI).

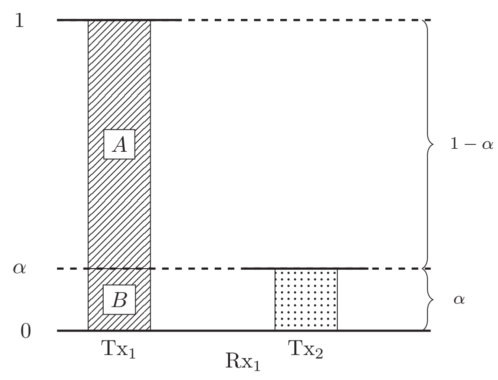


Figure 3.10: Signal levels at Rx_1 (VWI).

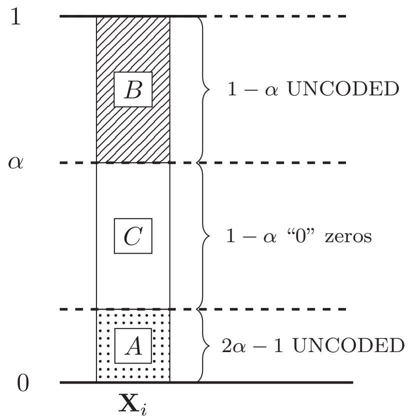


Figure 3.11: Transmitted symbols (WI).

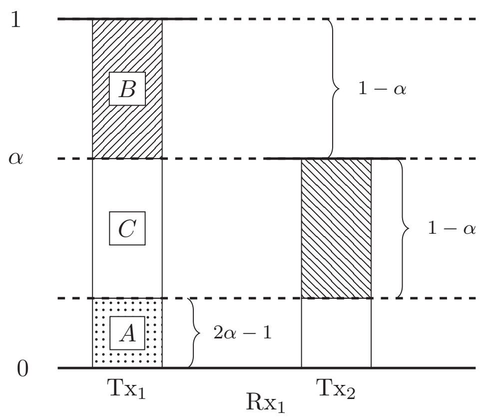


Figure 3.12: Signal levels at Rx_1 (WI).

3.3.5.2 Weak Interference

The symbols transmitted by both Txs (normalized by n_d) are depicted in Figure 3.11. Specifically, we transmit a block of $n_d(1-\alpha)$ bits in the most significant levels. In the subsequent levels, we transmit a block of $n_d(1-\alpha)$ zeros. Finally, in the least significant levels, we transmit a block of $n_d(2\alpha-1)$ bits. Figure 3.12 depicts the normalized signal levels of the transmitted signal as observed by Rx_1. At the Rx side, we observe that the most significant $n_d(1-\alpha)$ (block $\boxed{\text{B}}$), and the least $n_d(2\alpha-1)$ levels (block $\boxed{\text{A}}$) are received without interference. Thus, we achieve the rate

$$\begin{aligned} R_1 &= n_d - n_c + 2n_c - n_d \\ &= n_c. \end{aligned} \tag{3.53}$$

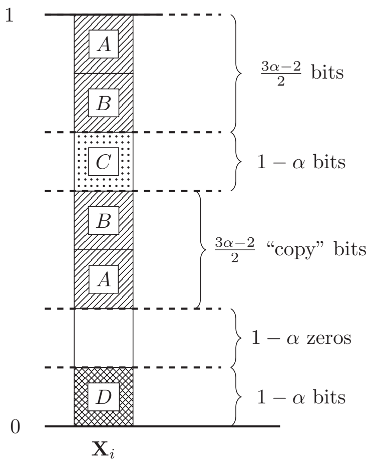


Figure 3.13: Transmitted symbols (MI).

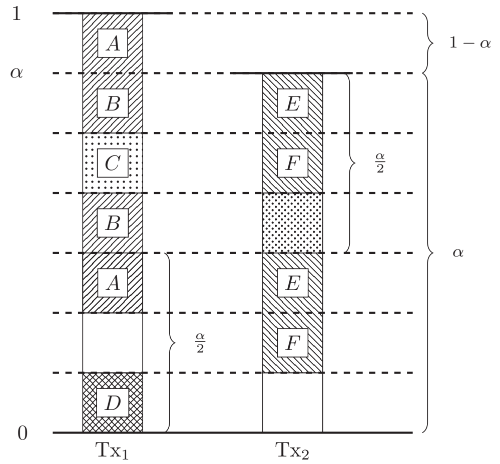


Figure 3.14: Signal levels at Rx_1 (MI).

User 2 achieves the same rate. Thus, the sum rate R is given by

$$R = 2n_c \; \frac{\text{bits}}{\text{sub-channel use}}. \tag{3.54}$$

Observe that this rate coincides with the upper bound (3.38), evaluated for WI.

3.3.5.3 Moderate Interference

The symbols transmitted by both Txs (normalized by n_d) are depicted in Figure 3.13. Specifically, we transmit in the most significant levels a block of $n_d\left(\frac{3\alpha}{2}\right)$ bits (we divided this block into two sub-blocks $\boxed{\text{A}}$ and $\boxed{\text{B}}$). In the subsequent levels, we transmit a block of $n_d(1-\alpha)$ bits. Next, we transmit a copy of the first block but in a reverse order, i.e., first $\boxed{\text{B}}$ and then $\boxed{\text{A}}$. Then, we transmit a block of $n_d(1-\alpha)$ zeros. Finally, we transmit a block of $n_d(1-\alpha)$ bits in the least significant levels. Figure 3.14 depicts the normalized signal levels of the transmitted signals as observed by Rx_1. Note that in this interference region the interfering signal must also be decoded. At the Rx side, we have the following procedure:

1. The block $\boxed{\text{A}}$ (at the first $n_d(1-\alpha)$ most-significant bits) and the block $\boxed{\text{D}}$ can be decoded interference free. Likewise, the block $\boxed{\text{F}}$ of interfering bits can be decoded interference free.

2. By subtracting the bits in block $\boxed{\text{F}}$, we can decode block $\boxed{\text{C}}$ interference free.

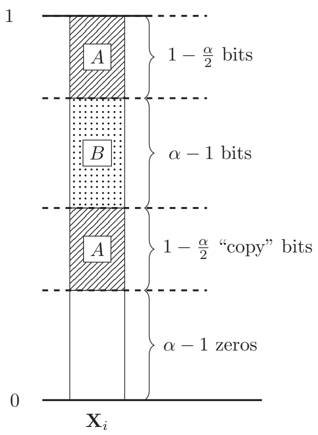


Figure 3.15: Transmitted symbols (SI).

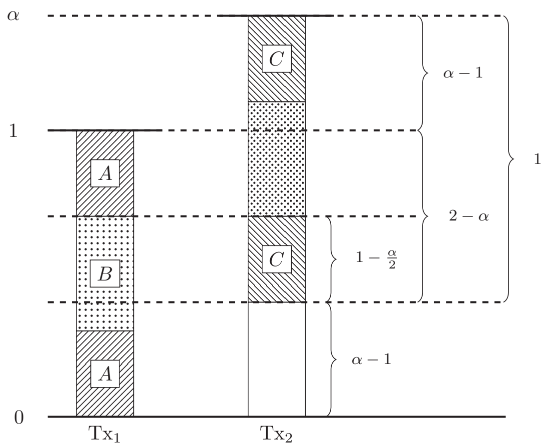


Figure 3.16: Signal levels at Rx_1 (SI).

3. By subtracting the bits in block $\boxed{\text{A}}$, we can decode block $\boxed{\text{E}}$ interference free.

4. By subtracting the bits in block $\boxed{\text{E}}$, we can decode the bits in block $\boxed{\text{B}}$ interference free.

This process is also valid at the other Rx, so the sum rate R achieved with this scheme is

$$\begin{aligned} R &= 2\left(\frac{3n_c - 2n_d}{2} + 2n_d - 2n_c\right) \\ &= 2n_d - n_c \ \frac{\text{bits}}{\text{sub-channel use}}. \end{aligned} \tag{3.55}$$

Observe that this rate coincides with (3.37) evaluated for MI.

3.3.5.4 Strong Interference

The symbols transmitted by both Txs (normalized by n_d) are depicted in Figure 3.15. Specifically, we transmit a block of $n_d\left(1-\frac{\alpha}{2}\right)$ bits in the most significant levels. In the subsequent levels, a block of $n_d(1-\alpha)$ bits is transmitted. Next, we transmit a block with a copy of the first block. Finally, we transmit a block of $n_d(\alpha-1)$ zeros in the least significant levels. The transmitted signal levels normalized by n_d, as observed by Rx_1, are depicted in Figure 3.16. As in the MI region, for SI the interfering signal must also be decoded. At the Rx side, we have the following procedure:

1. The block $\boxed{\mathrm{A}}$ (at the least-significant bits) and part of $\boxed{\mathrm{B}}$ can be decoded interference free. Likewise, the block $\boxed{C}$ of interfering bits can also be decoded interference free.

2. By subtracting the bits in block $\boxed{C}$, we can decode the rest of block $\boxed{B}$ interference free.

This process is also valid at the other Rx, so the sum rate R achieved using this scheme is

$$\begin{aligned} R &= 2\left(n_c - n_d + n_d - \frac{n_c}{2}\right) \\ &= n_c \, \frac{\text{bits}}{\text{sub-channel use}}. \end{aligned} \tag{3.56}$$

Observe that this rate coincides with (3.37) evaluated for SI.

4

Linear Deterministic Bursty Interference Channel

4.1 Introduction

Interference is a key limiting factor for the efficient use of the spectrum in modern wireless networks. It is, therefore, not surprising that the interference channel (IC) has been studied extensively in the past; see, e.g., [20, Ch. 6] and references therein. Most of the information-theoretic work developed for the IC assumes that interference is always present. However, certain physical phenomena, such as shadowing, can make the presence of interference intermittent or bursty. Interference can also be bursty due to the bursty nature of data traffic, distributed medium access control mechanisms, and decentralized networking protocols. For this reason, there has been an increasing interest in understanding and exploring the effects of burstiness of interference.

Seminal works in this area were performed by Khude *et al.* in [32] for the Gaussian channel and in [33] by using a model which corresponds to an approximation to the two-user Gaussian IC. They tried to harness the burstiness of the interference by taking advantage of the time instants when the interference is not present to send opportunistic data. Specifically, [32], [33] considered a channel model where the interference state stays constant during the transmission of the entire codeword, which corresponds to a quasi-static channel. Motivated by the idea of degraded message sets by Körner and Marton [36], Khude *et al.* studied the largest rate of a coding strategy that provides reliable communication at a basic rate R and allows an increased (opportunistic) rate $R + \Delta R$ when there is no interference. The idea of opportunism was also used by Diggavi and Tse [16] for the quasi-static flat fading channel and, recently, by Yi and Sun [69] for the K-user IC with states.

Wang *et al.* [65] modeled the presence of interference using an independent and identically distributed (IID) Bernoulli process that indicates whether interference is present or not, which corresponds to an ergodic channel. They further assume that the interference links are fully correlated. Wang *et al.* mainly studied the effect of causal feedback under this model, but also presented converse bounds for the non-feedback case. Mishra *et al.* considered the generalization of this model to multicarrier systems, modeled as parallel two-user bursty ICs, for the feedback [45] and non-feedback case [44].

The bursty IC is related to the binary fading IC, for which the four channel coefficients are in the binary field $\{0, 1\}$ according to some Bernoulli distribution.[1] Vahid *et al.* [56], [57], [58], [59], [60] studied the capacity region of the binary fading IC. Specifically, [58], [60] study the capacity region of the binary fading IC when the transmitters do not have access to the channel coefficients, and [59] study the capacity region when the transmitters have access to the past channel coefficients. Vahid and Calderbank additionally study the effect on the capacity region when certain correlation is available to all nodes as side information [56].

The focus of the works by Khude *et al.* [33] and Wang *et al.* [65] was on the linear deterministic model (LDM), which was first introduced by Avestimehr [5],

[1] Note, however, that neither of the two models is a special case of the other. While a zero channel coefficient of the cross link corresponds to intermittence of interference, the bursty IC allows for non-binary signals. Conversely, in contrast to the binary fading IC, the direct links in the bursty IC cannot be zero, since only the interference can be intermittent.

but falls within the class of more general deterministic channels whose capacity was obtained by El Gamal and Costa in [19]. The LDM maps the Gaussian IC to a channel whose outputs are deterministic functions of their inputs. Bresler and Tse demonstrated in [9] that the generalized degrees of freedom (first-order capacity approximation) of the two-user Gaussian IC coincides with the normalized capacity of the corresponding deterministic channel. The LDM thus offers insights on the Gaussian IC.

4.1.1 Contributions

In this chapter, we consider the LDM of a bursty IC. We study how interference burstiness and the knowledge of the interference states (throughout referred to as channel state information (CSI)) affects the capacity of this channel. We point out that this CSI is different from the one sometimes considered in the analysis of ICs (see, e.g., [30]), where CSI refers to knowledge of the channel coefficients. (In this regard, we assume that all transmitters and receivers have access to the channel coefficients.) For the sake of compactness, we focus on non-causal CSI and leave other CSI scenarios, such as causal or delayed CSI, for future work.

We consider the following cases: (i) only the receivers know the corresponding interference state (local CSIR); (ii) transmitters and receivers know their corresponding interference states (local CSIRT); and (iii) both transmitters and receivers know all interference states (global CSIRT). For each CSI level we consider both (i) the quasi-static channel and (ii) the ergodic channel. Specifically, in the quasi-static channel the interference is present or absent during the whole message transmission and we harness the realizations when the channel experiences better conditions (no presence of interference) to send extra messages. In the ergodic channel the presence/absence of interference is modeled as a Bernoulli random variable which determines the interference state. The interference state stays constant for a certain coherence time T and then changes independently to a new state. This model includes the IID model by Wang *et al.* as a special case, but also allows for scenarios where the interference state changes more slowly.[2] The proposed analysis is performed for the two extreme cases where the states of each

[2]Note, however, that when the receivers know the interference state (as we shall assume in this work), then the capacity of this model becomes independent of T and coincides with that of the IID model.

of the interfering links are independent, and where states of the interfering links are fully correlated. Hence we unify the scenarios already treated in the literature [33], [32], [65]. Nevertheless, some of our presented results can be extended to consider an arbitrary correlation between the interfering states. The works by Vahid and Calderbank [56] and Yeh and Wang [68] characterize the capacity region of the two-user binary IC and the multiple-input multiple-output (MIMO) X-channel, respectively. While [56], [68] consider a general spatial correlation between communication and interfering links, they do not consider the correlation between interfering links.

Our analysis shows that, for both the quasi-static and ergodic channels, for all interference regions except the very strong interference region, global CSIRT outperforms local CSIR/CSIRT. This result does not depend on the correlation between the states of the interfering links. For local CSIR/CSIRT and the quasi-static scenario, the burstiness of the channel is of benefit only in the very weak and weak interference regions. For the ergodic case and local CSIR, interference burstiness is only of clear benefit if the interference is either weak or very weak, or if it is present at most half of the time. This is in contrast to local CSIRT, where interference burstiness is beneficial in all interference regions.

Specific contributions of this chapter include:

- A joint treatment of the quasi-static and the ergodic model: Previous literature on the bursty IC considers either the quasi-static model or the ergodic model. Furthermore, due to space constraints, the proofs of some of the existing results were either omitted or contain little details. In contrast, this chapter discusses both models, allowing for a thorough comparison between the two.

- Novel achievability and converse bounds: For the ergodic model, the achievability bounds for local CSIRT, and the achievability and converse bounds for global CSIRT, are novel. In particular, novel achievability strategies are proposed that exploit certain synchronization between the users.

- Novel converse proofs for the quasi-static model: In contrast to existing converse bounds, which are based on Fano's inequality, our proofs of the converse bounds for the rates of the worst-case and opportunistic messages are based on an information density approach (more precise, they are based on the Verdú-Han lemma). This approach does not only allow for rigorous yet clear proofs, but it would also enable a more refined analysis of the probabilities that worst-case and opportunistic messages can be decoded correctly.

- A thorough comparison of the sum capacity of various scenarios: *Inter alia*, the obtained results are used to study the advantage of featuring different levels of CSI, the impact of the burstiness of the interference, and the effect of the correlation between the channel states of both users.

4.2 Channel Model

Our analysis is based on the LDM, introduced by Avestimehr *et al.* [5] for some relay network. This model is, on the one hand, simple to analyze and, on the other hand, captures the essential structure of the Gaussian channel in the high signal-to-noise ratio regime.

We consider a bursty IC where i) the interference state remains constant during the whole transmission of the codeword of length N (quasi-static setup) or ii) the interference state remains constant for a duration of T consecutive symbols and then changes independently to a new state (ergodic setup). For one coherence block, the two-user bursty IC is depicted in Figure 4.1, where n_d and n_c are the channel gains of the direct and cross links, respectively. We assume that n_d and n_c are known to both the transmitter and receiver and remain constant during the whole transmission of the codeword. For simplicity, we shall assume that n_d and n_c are equal for both users. Nevertheless, most of our results generalize to the asymmetric case. More precisely, all converse and achievability bounds generalize to the asymmetric case, while the direct generalization of the proposed achievability schemes may be loose in some asymmetric regions.

For the k-th block, the input-output relation of the channel is given by

$$\boldsymbol{Y}_{1,k} = \mathsf{S}_{n_d}\boldsymbol{X}_{1,k} \oplus B_{1,k}\mathsf{S}_{n_c}\boldsymbol{X}_{2,k} \quad (4.1)$$

$$\boldsymbol{Y}_{2,k} = \mathsf{S}_{n_d}\boldsymbol{X}_{2,k} \oplus B_{2,k}\mathsf{S}_{n_c}\boldsymbol{X}_{1,k}. \quad (4.2)$$

Let $q \triangleq \max\{n_d, n_c\}$. In (4.1) and (4.2), $\boldsymbol{X}_{i,k} \in \mathbb{F}_2^{q\times T}$ and $\boldsymbol{Y}_{i,k} \in \mathbb{F}_2^{q\times T}$, $i = 1, 2$. The interference states $B_{i,k}$, $i = 1, 2$, $k = 1, \ldots, K$, are sequences of IID Bernoulli random variables with activation probability p.

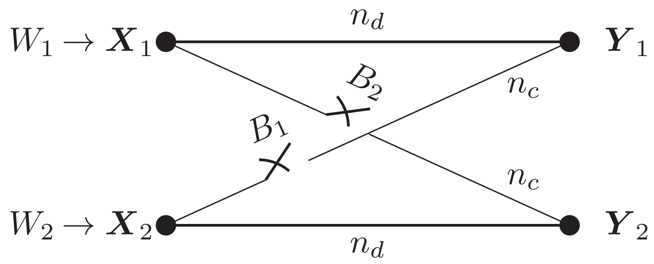


Figure 4.1: Channel model of the bursty interference channel.

Regarding the sequences B_1^K and B_2^K, we consider two cases: (i) B_1^K and B_2^K are independent of each other and (ii) B_1^K and B_2^K are fully correlated sequences, i.e., $B_1^K = B_2^K$. For both cases we assume that the sequences are independent of the messages W_1 and W_2.

4.2.1 Quasi-Static Channel

The channel defined in (4.1) and (4.2) may experience a slowly-varying change on the interference state. In this case, the duration of each of the transmitted codewords of length $N = KT$ is smaller than the coherence time T of the channel and the interference state stays constant over the duration of each codeword, i.e., $K = 1$, $T = N$. In the wireless communications literature such a channel is usually referred to as a quasi-static channel [55, Sec. 5.4.1]. In this scenario, the rate pair of achievable rates (R_1, R_2) is dominated by the worst case, which corresponds to the presence of interference at both receivers. However, in absence of interference, it is possible to communicate at a higher date rate, so planning a system for the worst case may be too pessimistic. Assuming that the receivers have access to the interference states, the transmitters could send opportunistic messages that are decoded only if the interference is absent, in addition to the regular messages that are decoded irrespective of the interference state. We make the notion of opportunistic messages and rates precise in the subsequent paragraphs.

Let $U_{i,k}$ indicate the level of CSI available at the transmitter side in coherence block k, and let $V_{i,k}$ indicate the level of CSI at the receiver side in coherence block k:

1. local CSIR: $\quad U_{i,k} = \emptyset$ and $V_{i,k} = B_{i,k},\ i = 1, 2,\ k = 1, \ldots, K,$

2. local CSIRT: $U_{i,k} = V_{i,k} = B_{i,k},\ i = 1, 2,\ k = 1, \ldots, K,$

3. global CSIRT: $U_{i,k} = V_{i,k} = (B_{1,k}, B_{2,k}),\ i = 1, 2,\ k = 1, \ldots, K.$

We define the set of opportunistic messages according to the level of CSI at the receiver as $\{\Delta W_i(\cdot)\} \triangleq \{\Delta W_i(v_i), v_i \in \mathcal{V}_i\}$, where $\mathcal{V}_i$ denotes the set of possible interference states V_i. Specifically,

1. for local CSIR:
   $\{\Delta W_i(\cdot)\} = \{\Delta W_i(1), \Delta W_i(0)\},\ i = 1, 2,$

2. for local CSIRT:
   $\{\Delta W_i(\cdot)\} = \{\Delta W_i(1), \Delta W_i(0)\},\ i = 1, 2,$

3. for global CSIRT:
   $\{\Delta W_i(\cdot)\} = \{\Delta W_i(00), \Delta W_i(01), \Delta W_i(10), \Delta W_i(11)\},\ i = 1, 2.$

Then, we define an opportunistic code as follows.

Definition 13 (Opportunistic code for the bursty IC) *An $\big(N, R_1, R_2, \{\Delta R_1(\cdot)\}, \{\Delta R_2(\cdot)\}\big)$ opportunistic code for the bursty IC consists of:*

1. *Two independent messages W_1 and W_2 uniformly distributed over the message sets $\mathcal{W}_i \triangleq \{1, 2, \ldots, 2^{NR_i}\},\ i = 1, 2.$*

2. *Two independent sets of opportunistic messages $\{\Delta W_1(\cdot)\}$ and $\{\Delta W_2(\cdot)\}$ uniformly distributed over the message sets $\Delta\mathcal{W}_i(v_i) \triangleq \{1, 2, \ldots, 2^{N\Delta R_i(v_i)}\}, v_i \in \mathcal{V}_i,\ \ i = 1, 2.$*

3. *Two encoders: $f_i : (W_i, \{\Delta W_i(\cdot)\}, U_i) \mapsto \boldsymbol{X}_i,\ i = 1, 2,.$*

4. *Two decoders: $g_i : (\boldsymbol{Y}_i, V_i) \mapsto (\hat{W}_i, \Delta\hat{W}_i(V_i)),\ i = 1, 2.$*

Here, $\hat{W}_i$ and $\Delta\hat{W}_i(V_i)$ denote the decoded message and the decoded opportunistic message, respectively. We set $\Delta R_i(1) = 0$, $i = 1, 2$ (for local CSIR/CSIRT) and $\Delta R_i(11) = 0$ (for global CSIRT).

To better distinguish the rates (R_1, R_2) from the opportunistic rates $\{\Delta R_i(\cdot)\}$, $i = 1, 2$, we shall refer to (R_1, R_2) as worst-case rates, because the corresponding messages can be decoded even if the channel is in its worst state (see also Definition 14).

Definition 14 (Achievable opportunistic rates) *A rate tuple* $\big(R_1, R_2, \{\Delta R_1(\cdot)\}, \{\Delta R_2(\cdot)\}\big)$ *is achievable if there exists a sequence of codes* $\big(N, R_1, R_2, \{\Delta R_1(\cdot)\}, \{\Delta R_2(\cdot)\}\big)$ *such that*

$$\Pr\big\{\hat{W}_1 \neq W_1 \cup \hat{W}_2 \neq W_2\big\} \to 0 \quad as \quad N \to \infty \tag{4.3}$$

and

$$\Pr\big\{(\hat{W}_1, \Delta\hat{W}_1(V_1)) \neq (W_1, \Delta W_1(V_1)) | V_1 = v_1\big\} \to 0 \text{ as } N \to \infty, \; v_1 \in \mathcal{V}_1, \tag{4.4}$$

$$\Pr\big\{(\hat{W}_2, \Delta\hat{W}_2(V_2)) \neq (W_2, \Delta W_2(V_2)) | V_2 = v_2\big\} \to 0 \text{ as } N \to \infty, \; v_2 \in \mathcal{V}_2. \tag{4.5}$$

The capacity region is the closure of the set of achievable rate tuples [20, Sec. 6.1]. We define the worst-case sum rate as $R \triangleq R_1 + R_2$ *and the opportunistic sum rate as* $\Delta R(V_1, V_2) \triangleq \Delta R_1(V_1) + \Delta R_2(V_2)$*. The worst-case sum capacity* C *is the supremum of all achievable worst-case sum rates, the opportunistic sum capacity* $\Delta C(V_1, V_2)$ *is the supremum of all opportunistic sum rates, and the total sum capacity is defined as* $C + \Delta C(V_1, V_2)$*. Note that the opportunistic sum capacity depends on the worst-case sum rate.*

Remark 4 *The worst-case sum rate and opportunistic sum rates in the quasi-static setting depend only on the collection of possible interference states: for independent interference states we have* $\mathbf{B} \in \{00, 01, 10, 11\}$*, and for fully correlated interference states we have* $\mathbf{B} \in \{00, 11\}$*. In principle, our proof techniques could also be applied to analyze other collections of interference states.*

Remark 5 *In the CSIRT setting the transmitters have access to the interference state. Therefore, in this setting the messages are strictly speaking not opportunistic. Instead, transmitters can adapt their rate based on the state of the interference links, which is sometimes referred to as rate adaptation in the literature.*

4.2.2 Ergodic Channel

In this setup, we shall restrict ourselves to codes whose blocklength N is an integer multiple of the coherence time T. A codeword of length $N = KT$ thus spans K independent channel realizations.

Definition 15 (Code for the bursty IC) *A* $\big(K, T, R_1, R_2\big)$ *code for the bursty IC consists of:*

1. *Two independent messages* W_1 *and* W_2 *uniformly distributed over the message sets* $\mathcal{W}_i \triangleq \{1, 2, \ldots, 2^{KTR_i}\},\ i = 1, 2.$

2. *Two encoders:* $f_i : (W_i, U_i^K) \mapsto \boldsymbol{X}_i^K,\ i = 1, 2.$

3. *Two decoders:* $g_i : (\mathbf{Y}_i^K, V_i^K) \mapsto \hat{W}_i,\ i = 1, 2.$

Here, $\hat{W}_i$ *denotes the decoded message, and* U_i^K *and* V_i^K *indicate the level of CSI at the transmitter and receiver side, respectively, which are defined as for the quasi-static channel in Section 4.2.1.*

Definition 16 (Ergodic achievable rates) *A rate pair* (R_1, R_2) *is achievable for a fixed* T *if there exists a sequence of codes* $\big(K, T, R_1, R_2\big)$ *(parameterized by* K*) such that*

$$\Pr\big\{\hat{W}_1 \neq W_1 \cup \hat{W}_2 \neq W_2\big\} \to 0 \quad as \quad K \to \infty. \tag{4.6}$$

The capacity region is the closure of the set of achievable rate pairs. We define the sum rate as $R \triangleq R_1 + R_2$*, the sum capacity* C *is the supremum of all achievable sum rates.*

4.2.3 The Sum Capacities of the Non-Bursty and the Quasi-Static Bursty IC

When the activation probability p is 1, we recover in both the ergodic and quasi-static scenarios the deterministic IC. For a general deterministic IC the capacity region was obtained in [19, Th. 1] and then by Bresler and Tse in [9] for a specific deterministic IC. For completeness, we recall the sum capacity region for the deterministic non-bursty IC in the following theorem (it was already presented in Chapter 3).

Theorem 9 *The sum capacity region of the two-user deterministic IC is equal* *to the union of the set of all sum rates R satisfying*

$$R \leq 2n_d \tag{4.7}$$

$$R \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{4.8}$$

$$R \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{4.9}$$

Proof: Sections 3.3.4 and 3.3.5. ∎

We can achieve the sum rates (4.7) and (4.9) over the quasi-static channel by treating the bursty IC as a non-bursty IC. The following theorem demonstrates that this is the largest achievable worst-case sum rate irrespective of the availability of CSI and the correlation between B_1 and B_2.

Theorem 10 (Sum capacity for the quasi-static bursty IC) *For $0 \leq p \leq 1$, the worst-case sum capacity of the bursty IC is equal to the supremum of the set of sum rates R satisfying*

- *For $p = 0$,*

$$R \leq 2n_d. \tag{4.10}$$

- *For $0 < p \leq 1$*

$$R \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{4.11}$$

$$R \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{4.12}$$

Proof: The converse bounds are proved in Appendix A.1.1. Achievability follows directly from Theorem 9 by treating the bursty IC as a non-bursty IC. ∎

Theorem 10 shows that the worst-case sum capacity does not depend on the level of CSI available at the transmitter and receiver side. However, this is not the case for the opportunistic rates as we will see in the next sections.

Remark 6 *In principle, one could reduce the worst-case rates in order to increase the opportunistic rates. However, it turns out that such a strategy is not beneficial in terms of total rates $R_i + \Delta R_i(V_i)$, $i = 1, 2$. In other words, setting $\Delta R_i(1) = 0$, $i = 1, 2$ (for local CSIR/CSIRT) and $\Delta R_i(11) = 0$ (for global CSIRT), as we have done in Definition 14, incurs no loss in total rate. Furthermore, in most cases it is preferable to maximize the worst-case rate, since it can be guaranteed irrespective of the interference state.*

4.3 Local CSIR

For the quasi-static and ergodic setups, described in Sections 4.2.1 and 4.2.2, respectively, we derive converse and achievability bounds for the independent and fully correlated scenarios when the interference state is only available at the receiver side.

4.3.1 Quasi-Static Channel

4.3.1.1 Independent Case

We present converse and achievability bounds for local CSIR when B_1 and B_2 are independent. The converse bounds are derived for local CSIRT, hence they also apply to this case. Since converse and achievability bounds coincide, this implies that local CSI at the transmitter is not beneficial in the quasi-static setup.

Theorem 11 (Opportunistic sum capacity for local CSIR/CSIRT) *Assume that B_1 and B_2 are independent of each other. For $0 < p < 1$, the opportunistic sum capacity region is the union of the set of rate tuples $(R, \{\Delta R_1(b_1) + \Delta R_2(b_2), b_i \in \{0,1\}\})$, where $\Delta R_1(1) = \Delta R_2(1) = 0$, and R, $\Delta R_1(0)$ and $\Delta R_2(0)$ satisfy (4.10)–(4.12) and*

$$R + \Delta R_1(0) + \Delta R_2(0) \leq 2n_d \tag{4.13}$$

$$R + \Delta R_1(0) \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{4.14}$$

$$R + \Delta R_2(0) \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{4.15}$$

Proof: The converse bounds are proved in Appendix A.1.2 and the achievability bounds are proved in Appendix A.1.3. ■

Remark 7 *The converse bounds in Theorem 11 coincide with those in [33, Th. 2.1], particularized for the symmetric setting. Theorem 11, however, is proven for local CSIRT, which is not considered in the model from [33]. The proof included in Appendix A.1.2 is based on an information density approach and provides a unified framework for treating local CSIR, local CSIRT and global CSIRT, as will be shown in Section 4.5.*

As discussed in Remark 6, one could reduce the worst-case sum rate R and increase the opportunistic rates $\Delta R(V_1, V_2)$. However, in the case of one-shot

Table 4.1: Opportunistic sum capacity for local CSIR when the worst-case sum rate is maximized.

Rates	**VWI**	**WI**	**MI**	**SI**
C	$2(n_d - n_c)$	$2n_c$	$2n_d - n_c$	n_c
$\Delta C(00)$	$2n_c$	$2(2n_d - 3n_c)$	0	0
$\Delta C(01)/\Delta C(10)$	n_c	$2n_d - 3n_c$	0	0

transmission[3] this is not desirable, since the worst-case sum rate is the only rate that can be guaranteed irrespective of the interference state. Thus, one is typically interested in the opportunistic sum capacity when the worst-case rate R is maximized. For this case, the results of Theorem 11 are summarized in Table 4.1 for the very weak interference (VWI), weak interference (WI), moderate interference (MI) and strong interference (SI) regions. Observe that converse and achievability bounds coincide. Further observe that opportunistic messages can only be transmitted reliably for VWI or WI. In the other interference regions, the opportunistic sum capacity is zero.

4.3.1.2 Fully Correlated Case

Assume now that the sequences B_1 and B_2 are fully correlated ($B_1 = B_2$). For local CSIR, the correlation between B_1 and B_2 has no influence on the opportunistic sum capacity region. Indeed, in this case the channel inputs are independent of (B_1, B_2) and the opportunistic sum capacity region of the quasi-static bursty IC depends on (B_1, B_2) only via the marginal distributions of B_i, $i = 1, 2$. Hence, it follows that Theorem 11 as well as Table 4.1 apply also to the fully correlated case and local CSIR scenario. For completeness, a proof of the converse part is given in Appendix A.1.4. The achievability part is included in Appendix A.1.3.

[3] With one-shot transmission we refer to the case where we transmit one codeword of length N over the quasi-static channel. This is in contrast to the case discussed, e.g., in Section 4.3.3, where we are interested in transmitting many codewords, each over N channel uses of independent quasi-static channels.

4.3.2 Ergodic Channel

4.3.2.1 Independent Case

For the case where the sequences B_1^K and B_2^K are independent of each other, we have the following theorems.

Theorem 12 (Converse bounds for local CSIR) *Assume that B_1^K and B_2^K are independent of each other. The sum rate R for the bursty IC is upper-bounded by*

$$R \leq 2\frac{1-p}{1+p}n_d + 2\frac{p}{1+p}\left[(n_d-n_c)^+ + \max(n_d,n_c)\right] \tag{4.16}$$

and

$$R \leq \begin{cases} 2(1-2p)n_d + 2p\big[(n_d-n_c)^+ + \max(n_d,n_c)\big] & p \leq \frac{1}{2}, \\ 2(1-p)\left[(n_d-n_c)^+ + \max(n_d,n_c)\right] & \\ \quad + 2(2p-1)\left[\max\{(n_d-n_c)^+, n_c\}\right] & p > \frac{1}{2}. \end{cases} \tag{4.17}$$

Proof: Bound (4.16) coincides with [65, Eq. (3)]. Specifically, [65, Eq. (3)] derives (4.16) for the considered channel model with $T=1$ and feedback. The proof for this bound under local CSIRT (without feedback) is given in Appendix A.2.1. Bound (4.17) coincides with [66, Lemma A.1]. Specifically, [66, Lemma A.1] derives (4.17) for the model considered with $T=1$. The proof of [66, Lemma A.1] directly generalizes to arbitrary T. ■

Theorem 13 (Achievability bounds for local CSIR) *Assume that B_1^K and B_2^K are independent of each other. The following sum rate R is achievable over the bursty IC:*

$$R = \begin{cases} 2(1-2p)n_d + 2p\big[(n_d-n_c)^+ + \max(n_d,n_c)\big], & p \leq \frac{1}{2}, \\ \min\{(n_d-n_c)^+ + \max(n_d,n_c), & \\ \ 2(1-p)\left[(n_d-n_c)^+ + \max(n_d,n_c)\right] & \\ \quad + 2(2p-1)\left[\max\{(n_d-n_c)^+, n_c\}\right]\}, & p > \frac{1}{2}. \end{cases} \tag{4.18}$$

Proof: The achievability scheme for VWI for all values of p, and for WI and MI when $0 \leq p \leq \frac{1}{2}$, is described in Appendix A.2.2.1. The achievability scheme for WI and $\frac{1}{2} < p \leq 1$ is described in Appendix A.2.2.2. The scheme for SI and $0 \leq p \leq \frac{1}{2}$ is summarized in Appendix A.2.2.3. For MI and SI when $\frac{1}{2} < p \leq 1$, the achievability bound in the theorem corresponds to the one of the non-bursty IC [29]. This also implies that in this sub-region we do not exploit the burstiness of the IC. ■

Remark 8 *The achievability schemes presented in Theorem 13 are similar to those described in [58], [60]. They achieve the capacity region by applying point-to-point erasure codes with appropriate rates at each transmitter and using either treating-interference-as-erasure or interference-decoding at each receiver. Specifically, we apply treating-interference-as-erasure in the VWI region and for all values of p, and for all interference regions, except very strong interference (VSI), and $p \leq \frac{1}{2}$. Interference-decoding at each receiver is applied in the MI and SI regions for $p > \frac{1}{2}$.*

Remark 9 *Wang et al. claim in [66, Lemma A.1] that the converse bound (18) is tight for $0 \leq p \leq \frac{1}{2}$ without providing an achievability bound. Instead, they refer to Khude et al. [33] for the inner bound which, alas, does not apply to the ergodic setup. While it is possible to adapt the achievability schemes considered in [33] to prove* (4.18)*, a number of steps are required. For completeness, we include the achievability schemes for the ergodic setup and $0 \leq p \leq \frac{1}{2}$ in Appendix A.2.2.1.*

Table 4.2 summarizes the results of Theorems 12 and 13. We write the sum capacities in bold face when the converse and achievability bounds match. In Table 4.2, we define

$$\mathbf{C}_{\text{LMI}} \triangleq \min\left\{2[2(n_d - n_c) + p(3n_c - 2n_d)],\right.$$
$$\left. 2\left[\frac{1-p}{1+p}n_d + \frac{p}{1+p}(2n_d - n_c)\right]\right\} \tag{4.19}$$

$$\mathbf{C}_{\text{LSI}} \triangleq \min\left\{2pn_c, 2\left[\frac{1-p}{1+p}n_d + \frac{p}{1+p}n_c\right]\right\} \tag{4.20}$$

where "L" stands for "local CSIR".

Table 4.2: Sum capacity for local CSIR.

Regions	$p \leq \frac{1}{2}$	$p > \frac{1}{2}$
VWI	$\mathbf{2(n_d - pn_c)}$	$\mathbf{2(n_d - pn_c)}$
WI	$\mathbf{2(n_d - pn_c)}$	$\mathbf{4(n_d - n_c) + 2p(3n_c - 2n_d)}$
MI	$\mathbf{2(n_d - pn_c)}$	$2n_d - n_c \leq R \leq \mathbf{C}_{\text{LMI}}$
SI	$\mathbf{2(1-2p)n_d + 2pn_c}$	$n_c \leq R \leq \mathbf{C}_{\text{LSI}}$

4.3.2.2 Fully Correlated Case

For local CSIR, the dependence between B_1^K and B_2^K has no influence on the capacity region. Indeed, in this case the channel inputs are independent of (B_1^K, B_2^K) and decoder i has only access to $B_{i,k}$ and $(\mathsf{S}_{n_d}\boldsymbol{X}_{i,k} \oplus B_{i,k}\mathsf{S}_{n_c}\boldsymbol{X}_{j,k})$, $k = 1, \ldots, K$, $j = 3 - i$ and $i = 1, 2$. Furthermore, $\Pr\{\hat{W}_1 \neq W_1 \cup \hat{W}_2 \neq W_2\}$ vanishes as $K \to \infty$ if, and only if, $\Pr\{\hat{W}_i \neq W_i\}$, $i = 1, 2$, vanishes as $K \to \infty$. Since $\Pr(\hat{W}_i \neq W_i)$ depends only on B_i^K, the capacity region of the bursty IC depends on (B_1^K, B_2^K) only via the marginal distributions of B_1^K and B_2^K. Hence, Theorems 12 and 13 as well as Table 4.2 apply also to the case where $B_1^K = B_2^K$. This is consistent with the observation by Sato [49] that "the capacity region is the same for all two-user channels that have the same marginal probabilities."

4.3.3 Quasi-Static vs. Ergodic Setup

In general, the sum capacities of the quasi-static and ergodic channels cannot be compared, because in the former case we have a set of sum capacities (worst case and opportunistic), whereas in the latter case only one is defined. To allow for a comparison, we introduce for the quasi-static channel the average sum capacity as

$$\bar{C} \triangleq \sup_{(R, \Delta R_1(0), \Delta R_2(0))} \{R + (1-p)(\Delta R_1(0) + \Delta R_2(0))\} \tag{4.21}$$

where the suprema is over all tuples $(R, \Delta R_1(0), \Delta R_2(0))$ that satisfy (4.10)–(4.15). Intuitively, the average rate corresponds to the case where we send many messages over independent quasi-static fading channels. By the law of large numbers, a fraction of p transmissions will be affected by interference, the remaining transmissions will be interference-free. Table 4.3 summarizes the average sum capacity for the different interference regions.

Table 4.3: Average sum capacities for local CSIR.

Regions	$p \leq \frac{1}{2}$	$p > \frac{1}{2}$
VWI	$2(n_d - pn_c)$	$2(n_d - pn_c)$
WI	$2(n_d - pn_c)$	$4(n_d - n_c) + 2p(3n_c - 2n_d)$
MI	$2(n_d - pn_c)$	$2n_d - n_c$
SI	$2(1 - 2p)n_d + 2pn_c$	n_c

By comparing Tables 4.2 and 4.3, we can observe that for $p \leq \frac{1}{2}$ and all interference regions, and for $p > \frac{1}{2}$ and VWI/WI, the average sum capacity in the quasi-static setup coincides with the sum capacity in the ergodic setup. For $p > \frac{1}{2}$, and MI/SI (where converse and achievability bounds do not coincide), the average sum capacities in the quasi-static setup coincide with the achievability bounds of the ergodic setup.

4.4 Local CSIRT

For the quasi-static and ergodic setups, we present converse and achievability bounds when transmitters and receivers have access to their corresponding interference states. We shall only consider the independent case here, because when $B_1^K = B_2^K$ local CSIRT coincides with global CSIRT, which will be discussed in Section 4.5.

4.4.1 Quasi-Static Channel

For the quasi-static channel, the converse and achievability bounds were already presented in Theorem 11 in Section 4.3.1.1. Indeed, the converse bounds were derived for local CSIRT, whereas the achievability bounds in that theorem were derived for local CSIR. Since these bounds coincide for all interference regions and all probabilities of $0 < p < 1$ it follows that, for the quasi-static channel, availability of local CSI at the transmitter in addition to local CSI at the receiver is not beneficial. The converse and achievability bounds are then given in Theorem 11.

4.4.2 Ergodic Channel

The converse bound (4.16) presented in Theorem 12 was derived for local CSIRT, so it applies to the case at hand. We next present achievability bounds for this setup that improve upon those for CSIR. The aim of these bounds is to provide computable expressions showing that local CSIRT outperforms local CSIR in the whole range of the α parameter. While the particular achievability schemes are sometimes involved, the intuition behind these schemes can be explained with the following toy example.

Example: *Let us assume that $n_d = n_c = T = 1$, and suppose that at time k the transmitters send the bits $(B_{1,k}, B_{2,k}) \in \{0,1\}^2$. If there is no interference, then receiver i receives $X_{i,k}$. If there is interference, then receiver i receives $X_{1,k} \oplus X_{2,k}$. Consequently, the channel flips $X_{1,k}$ if $B_{1,k} = X_{2,k} = 1$, and it flips $X_{2,k}$ if $B_{2,k} = X_{1,k} = 1$. It follows that each transmitter-receiver pair experiences a binary symmetric channel (BSC) with a given crossover probability that depends on p and on the probabilities that (X_1, X_2) are one. Specifically, let*

$$P_{X_1|B_1}(X_1 = 1|B_1 = 0) \triangleq p_1 \tag{4.22}$$

$$P_{X_1|B_1}(X_1 = 1|B_1 = 1) \triangleq p_2 \tag{4.23}$$

$$P_{X_2|B_2}(X_2 = 1|B_2 = 0) \triangleq q_1 \tag{4.24}$$

$$P_{X_2|B_2}(X_2 = 1|B_2 = 1) \triangleq q_2 \tag{4.25}$$

and define $p_3 \triangleq (1-p)p_1 + pp_2$ and $q_3 \triangleq (1-p)q_1 + pq_2$, which are the crossover probabilities of the BSCs experienced by receivers 1 and 2, respectively, when they are affected by interference. By drawing for each user two codebooks (one for $B_{i,k} = 0$ and one for $B_{i,k} = 1$) IID at random according to the probabilities p_1, p_2, q_1, and q_2, and by following a random-coding argument, it can be shown that this scheme achieves the sum rate

$$\begin{aligned} R = (1-p)[H_b(p_1) + H_b(q_1)] + p[H_{\text{sum}}(p_2, q_3) - H_b(q_3)] \\ + p[H_{\text{sum}}(q_2, p_3) - H_b(p_3)]. \end{aligned} \tag{4.26}$$

This expression holds for any set of parameters (p_1, p_2, q_1, q_2), and the largest sum rate achieved by this scheme is obtained by maximizing over $(p_1, p_2, q_1, q_2) \in \left[0, \frac{1}{2}\right]^4$.

In the following, we present the achievable sum rates that can be obtained by generalizing the above achievability scheme to general n_d and n_c. The achievability schemes that achieve these rates are presented in Appendix A.3. The largest achievable sum rates can then be obtained by numerically maximizing over the parameters $(p_1, p_2, q_1, q_2, \ldots)$ (which depend on the interference region).

1. For the VWI region, we achieve the sum rate

$$R = 2(n_d - pn_c). \tag{4.27}$$

2. For the WI region, we can achieve for any $(p_1, p_2, q_1, q_2) \in \left[0, \frac{1}{2}\right]^4$

$$\begin{aligned} R_1 &= (n_d - n_c) + (1-p)[(n_d - n_c) + (2n_c - n_d)H_b(p_1)] \\ &\quad + p(2n_c - n_d)(1 - H_b(q_3)) \end{aligned} \tag{4.28}$$

$$\begin{aligned} R_2 &= (n_d - n_c) + (1-p)[(n_d - n_c) + (2n_c - n_d)H_b(q_1)] \\ &\quad + p(2n_c - n_d)(1 - H_b(p_3)) \end{aligned} \tag{4.29}$$

where $p_3 = (1-p)p_1 + pp_2$ and $q_3 = (1-p)q_1 + pq_2$.

3. To present the achievable rates for MI, we need to divide the region into the following four subregions:

(a) For $\frac{2}{3} \le \alpha \le \frac{3}{4}$, we can achieve for any $(p_1, p_2, \tilde{p}_1, \tilde{p}_2, \hat{p}_1, q_1, q_2, \tilde{q}_1, \tilde{q}_2, \hat{q}_1) \in \left[0, \frac{1}{2}\right]^{10}$ and $(\eta_1, \gamma_1) \in \left[\frac{1}{2}, 1\right]^2$

$$\begin{aligned} R_1 =& (n_d - n_c) \\ &+ (1-p)\left[\left(\tfrac{3n_c - 2n_d}{2}\right)(H_b(\eta_1) + H_b(\hat{p}_1) + H_b(p_1)) \right.\\ &\qquad \left. + \left(\tfrac{4n_d - 5n_c}{2}\right) H_b(\tilde{p}_1) + (n_d - n_c)\right] \\ &+ p\Big[\left(\tfrac{3n_c - 2n_d}{2}\right)\Big(1 + H_{\mathrm{sum}}(p_2, \tilde{\gamma}) - H_b(\tilde{\gamma}) + H_{\mathrm{sum}}(\tilde{p}_2, q_3) \\ &\qquad - H_b(q_3) - H_b(\hat{q}_3)\Big) \\ &\quad + \left(\tfrac{4n_d - 5n_c}{2}\right)(1 - H_b(\tilde{q}_3))\Big] \end{aligned} \tag{4.30}$$

where $q_3 = (1-p)q_1 + pq_2$, $\tilde{q}_3 = (1-p)\tilde{q}_1 + p\tilde{q}_2$, $\hat{q}_3 = (1-p)\hat{q}_1$, and

$\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$\begin{aligned} R_2 =&(n_d - n_c) \\ &+ (1-p)\left[\left(\tfrac{3n_c-2n_d}{2}\right)\left(H_b(\gamma_1) + H_b(\hat{q}_1) + H_b(q_1)\right)\right. \\ &\qquad\left. + \left(\tfrac{4n_d-5n_c}{2}\right) H_b(\tilde{q}_1) + (n_d - n_c)\right] \\ &+ p\Big[\left(\tfrac{3n_c-2n_d}{2}\right)\Big(1 + H_{\text{sum}}(q_2,\tilde{\eta}) - H_b(\tilde{\eta}) + H_{\text{sum}}(\tilde{q}_2, p_3) \\ &\qquad - H_b(p_3) - H_b(\hat{p}_3)\Big) \\ &\quad + \left(\tfrac{4n_d-5n_c}{2}\right)\left(1 - H_b(\tilde{p}_3)\right)\Big] \end{aligned} \tag{4.31}$$

where $p_3 = (1-p)p_1 + pp_2$, $\tilde{p}_3 = (1-p)\tilde{p}_1 + p\tilde{p}_2$, $\hat{p}_3 = (1-p)\hat{p}_1$, and $\tilde{\eta} = p + \eta_1(1-p)$.

Remark 10 *After combining* (4.30) *and* (4.31), η_1 *and* γ_1 *appear only through the functions* $H_b(\eta_1) - H_b(p + \eta_1(1-p))$ *and* $H_b(\gamma_1) - H_b(p + \gamma_1(1-p))$*, respectively. Hence,* η_1 *and* γ_1 *can be optimized separately from the remaining terms.*

(b) For $\frac{3}{4} \leq \alpha \leq \frac{4}{5}$, we can achieve for any $(p_1, p_2, \tilde{p}_1, \tilde{p}_2, \hat{p}_1, q_1, q_2, \tilde{q}_1, \tilde{q}_2, \hat{q}_1) \in \left[0, \frac{1}{2}\right]^{10}$ and $(\eta_1, \gamma_1) \in \left[\frac{1}{2}, 1\right]^2$

$$\begin{aligned} R_1 =&(n_d - n_c) \\ &+ (1-p)\Big[\left(\tfrac{3n_c-2n_d}{2}\right)\left(H_b(p_1) + H_b(\eta_1) + H_b(\hat{p}_1)\right) \\ &\qquad + \left(\tfrac{4n_d-5n_c}{2}\right)H_b(\tilde{p}_1) + (n_d - n_c)\Big] \\ &+ p\Big[\left(\tfrac{3n_c-2n_d}{2}\right)\left(H_{\text{sum}}(p_2,\tilde{\gamma}) - H_b(\tilde{\gamma}) + 1 - H_b(\hat{q}_3)\right) \\ &\quad + \left(\tfrac{4n_d-5n_c}{2}\right)\left(H_{\text{sum}}(\tilde{p}_2, q_3) - H_b(q_3) + 1 - H_b(\tilde{q}_3)\right)\Big] \end{aligned} \tag{4.32}$$

where $q_3 = (1-p)q_1 + pq_2$, $\tilde{q}_3 = (1-p)\tilde{q}_1 + p\tilde{q}_2$, $\hat{q}_3 = (1-p)\hat{q}_1$, and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$\begin{aligned} R_2 =&(n_d - n_c) \\ &+ (1-p)\Big[\left(\tfrac{3n_c-2n_d}{2}\right)\left(H_b(q_1) + H_b(\gamma_1) + H_b(\hat{q}_1)\right) \\ &\qquad + \left(\tfrac{4n_d-5n_c}{2}\right)H_b(\tilde{q}_1) + (n_d - n_c)\Big] \\ &+ p\Big[\left(\tfrac{3n_c-2n_d}{2}\right)\left(H_{\text{sum}}(q_2,\tilde{\eta}) - H_b(\tilde{\eta}) + 1 - H_b(\hat{p}_3)\right) \\ &\quad + \left(\tfrac{4n_d-5n_c}{2}\right)\left(H_{\text{sum}}(\tilde{q}_2, p_3) - H_b(p_3) + 1 - H_b(\tilde{p}_3)\right)\Big] \end{aligned} \tag{4.33}$$

where $p_3 = (1-p)p_1 + pp_2$, $\tilde{p}_3 = (1-p)\tilde{p}_1 + p\tilde{p}_2$, $\hat{p}_3 = (1-p)\hat{p}_1$, and $\tilde{\eta} = p + \eta_1(1-p)$. Remark 10 also applies to the parameters η_1 and γ_1 in (4.32) and (4.33).

(c) For $\frac{4}{5} \le \alpha \le \frac{6}{7}$, we can achieve for any $(p_1, p_2, \hat{p}_1, q_1, q_2, \hat{q}_1) \in \left[0, \frac{1}{2}\right]^6$ and $(\eta_1, \eta', \gamma_1, \gamma') \in \left[\frac{1}{2}, 1\right]^4$

$$\begin{aligned} R_1 =& (n_d - n_c) \\ &+ (1-p)\Big[\left(\tfrac{5n_c - 4n_d}{2}\right)(1 + H_b(\eta')) \\ &\qquad + (n_d - n_c)\left(1 + H_b(p_1) + H_b(\eta_1) + H_b(\hat{p}_1)\right)\Big] \\ &+ p\Big[\left(\tfrac{5n_c - 4n_d}{2}\right)\Big(1 - H_b(\tilde{\gamma}) + H_{\text{sum}}(p_2, \gamma') - H_b(\gamma') \\ &\qquad + H_{\text{sum}}(\eta'(1-\tilde{\gamma}) + (1-\eta')\tilde{\gamma}, q_3) - H_b(q_3)\Big) \\ &\quad + \left(\tfrac{6n_d - 7n_c}{2}\right)\Big(H_{\text{sum}}(p_2, \tilde{\gamma}) - H_b(\tilde{\gamma})\Big) \\ &\quad + (n_d - n_c)(1 - H_b(\hat{q}_3))\Big] \end{aligned} \tag{4.34}$$

where $q_3 = (1-p)q_1 + pq_2$, $\hat{q}_3 = (1-p)\hat{q}_1$, and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$\begin{aligned} R_2 =& (n_d - n_c) \\ &+ (1-p)\Big[\left(\tfrac{5n_c - 4n_d}{2}\right)(1 + H_b(\gamma')) \\ &\qquad + (n_d - n_c)\Big(1 + H_b(q_1) + H_b(\gamma_1) + H_b(\hat{q}_1)\Big)\Big] \\ &+ p\Big[\left(\tfrac{5n_c - 4n_d}{2}\right)\Big(1 - H_b(\tilde{\eta}) + H_{\text{sum}}(q_2, \eta') - H_b(\eta') \\ &\qquad + H_{\text{sum}}(\gamma'(1-\tilde{\eta}) + (1-\gamma')\tilde{\eta}, p_3) - H_b(p_3)\Big) \\ &\quad + \left(\tfrac{6n_d - 7n_c}{2}\right)\Big(H_{\text{sum}}(q_2, \tilde{\eta}) - H_b(\tilde{\eta})\Big) \\ &\quad + (n_d - n_c)(1 - H_b(\hat{p}_3))\Big] \end{aligned} \tag{4.35}$$

where $p_3 = (1-p)p_1 + pp_2$, $\hat{p}_3 = (1-p)\hat{p}_1$, and $\tilde{\eta} = p + \eta_1(1-p)$.

(d) For $\frac{6}{7} \le \alpha \le 1$ we can achieve for any $(p_1, p_2, \hat{p}_1, q_1, q_2, \hat{q}_1) \in \left[0, \frac{1}{2}\right]^6$

and $(\eta_1, \eta', \gamma_1, \gamma') \in \left[\frac{1}{2}, 1\right]^4$

$$
\begin{aligned}
R_1 =& (n_d - n_c) \\
&+ (1-p)\Big[(6n_c - 5n_d)H_b(p_1) \\
&\qquad + (n_d - n_c)\left(2 + H_b(\eta_1) + H_b(\eta') + H_b(\hat{p}_1)\right)\Big] \\
&+ p\Big[(n_d - n_c)\Big(2 - H_b(\tilde{\gamma}) - H_b(\hat{q}_3) \\
&\qquad + H_{\text{sum}}(\eta'(1-\tilde{\gamma}) + (1-\eta')\tilde{\gamma}, q_3) - H_b(q_3)\Big) \\
&\quad + (n_d - n_c)\left(H_{\text{sum}}(p_2, \gamma') - H_b(\gamma')\right) \\
&\quad + (7n_c - 6n_d)\left(H_{\text{sum}}(p_2, q_3) - H_b(q_3)\right)\Big]
\end{aligned}
\tag{4.36}
$$

where $q_3 = (1-p)q_1 + pq_2$, $\hat{q}_3 = (1-p)\hat{q}_1$, and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$
\begin{aligned}
R_2 =& (n_d - n_c) \\
&+ (1-p)\Big[(6n_c - 5n_d)H_b(q_1) \\
&\qquad + (n_d - n_c)\left(2 + H_b(\gamma_1) + H_b(\gamma') + H_b(\hat{q}_1)\right)\Big] \\
&+ p\Big[(n_d - n_c)\Big(2 - H_b(\tilde{\eta}) - H_b(\hat{p}_3) \\
&\qquad + H_{\text{sum}}(\gamma'(1-\tilde{\eta}) + (1-\gamma')\tilde{\eta}, p_3) - H_b(p_3)\Big) \\
&\quad + (n_d - n_c)\left(H_{\text{sum}}(q_2, \eta') - H_b(\eta')\right) \\
&\quad + (7n_c - 6n_d)\left(H_{\text{sum}}(q_2, p_3) - H_b(p_3)\right)\Big]
\end{aligned}
\tag{4.37}
$$

where $p_3 = (1-p)p_1 + pp_2$, $\hat{p}_3 = (1-p)\hat{p}_1$, and $\tilde{\eta} = p + \eta_1(1-p)$.

4. To present the achievable rates for SI, we divide the region into the following four subregions:

   (a) For $1 \leq \alpha \leq \frac{6}{5}$, we can achieve for any $(p_1, p_2, q_1, q_2) \in \left[0, \frac{1}{2}\right]^4$ and

$(\eta_1, \eta', \gamma_1, \gamma') \in \left[\frac{1}{2}, 1\right]^4$

$$\begin{aligned}
R_1 =&(n_c - n_d) \\
&+ (1-p)\Big[(5n_d - 4n_c)H_b(p_1) \\
&\qquad\qquad + (n_c - n_d)\left(1 + H_b(\eta_1) + H_b(\eta')\right)\Big] \\
&+ p\Big[(n_c - n_d)\Big(1 - H_b(\tilde{\gamma}) \\
&\qquad\qquad + H_{\text{sum}}(\eta'(1-\tilde{\gamma}) + (1-\eta')\tilde{\gamma}, q_3) - H_b(q_3)\Big) \\
&\qquad + (n_c - n_d)\left(H_{\text{sum}}(p_2, \gamma') - H_b(\gamma')\right) \\
&\qquad + (6n_d - 5n_c)\left(H_{\text{sum}}(p_2, q_3) - H_b(q_3)\right)\Big]
\end{aligned} \tag{4.38}$$

where $q_3 = (1-p)q_1 + pq_2$ and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$\begin{aligned}
R_2 =&(n_c - n_d) \\
&+ (1-p)\big[(5n_d - 4n_c)H_b(q_1) \\
&\qquad\qquad + (n_c - n_d)\left(1 + H_b(\gamma_1) + H_b(\gamma')\right)\big] \\
&+ p\Big[(n_c - n_d)\Big(1 - H_b(\tilde{\eta}) \\
&\qquad\qquad + H_{\text{sum}}(\gamma'(1-\tilde{\eta}) + (1-\gamma')\tilde{\eta}, q_3) - H_b(p_3)\Big) \\
&\qquad + (n_c - n_d)\left(H_{\text{sum}}(q_2, \eta') - H_b(\eta')\right) \\
&\qquad + (6n_d - 5n_c)\left(H_{\text{sum}}(q_2, p_3) - H_b(p_3)\right)\big]
\end{aligned} \tag{4.39}$$

where $p_3 = (1-p)p_1 + pp_2$ and $\tilde{\eta} = p + \eta_1(1-p)$.

(b) For $\frac{6}{5} \leq \alpha \leq \frac{4}{3}$, we can achieve for any $(p_1, p_2, q_1, q_2) \in \left[0, \frac{1}{2}\right]^4$ and $(\eta_1, \gamma_1) \in \left[\frac{1}{2}, 1\right]^2$

$$\begin{aligned}
R_1 = &\left(2n_d - \tfrac{3n_c}{2}\right) \\
&+ (1-p)\Big[\left(2n_d - \tfrac{3n_c}{2}\right)H_b(\eta_1) \\
&\qquad\qquad + 2(n_c - n_d) + (3n_d - 2n_c)H_b(p_1)\Big] \\
&+ p\Big[(n_c - n_d)\left(1 - H_b(q_3)\right) + \left(2n_d - \tfrac{3n_c}{2}\right)\left(1 - H_b(\tilde{\gamma})\right) \\
&\qquad + \left(\tfrac{5n_c}{2} - 3n_d\right)\Big]
\end{aligned} \tag{4.40}$$

where $q_3 = (1-p)q_1 + pq_2$, and $\tilde{\gamma} = p + \gamma_1(1-p)$, and

$$\begin{aligned} R_2 = &\left(2n_d - \tfrac{3n_c}{2}\right) \\ &+ (1-p)\Big[\left(2n_d - \tfrac{3n_c}{2}\right) H_b(\gamma_1) \\ &\qquad + 2(n_c - n_d) + (3n_d - 2n_c)H_b(q_1)\Big] \\ &+ p\Big[(n_c - n_d)\left(1 - H_b(p_3)\right) + \left(2n_d - \tfrac{3n_c}{2}\right)\left(1 - H_b(\tilde{\eta})\right) \\ &\qquad + \left(\tfrac{5n_c}{2} - 3n_d\right)\Big] \end{aligned} \tag{4.41}$$

where $p_3 = (1-p)p_1 + pp_2$, and $\tilde{\eta} = p + \eta_1(1-p)$. Remark 10 also applies to the parameters η_1 and γ_1 in (4.40) and (4.41).

(c) For $\frac{4}{3} \le \alpha \le \frac{3}{2}$, we can achieve for any $(p_1, p_2, q_1, q_2) \in \left[0, \frac{1}{2}\right]^4$ and $(\eta_1, \gamma_1) \in \left[\frac{1}{2}, 1\right]^2$,

$$\begin{aligned} R_1 = &(n_d - \tfrac{n_c}{2}) \\ &+ (1-p)\Big[(3n_d - 2n_c)(1 + H_b(p_1)) \\ &\qquad + \left(\tfrac{3n_c}{2} - 2n_d\right)\left(1 + H_b(\eta_1)\right)\Big] \\ &+ p\Big[(3n_d - 2n_c)(1 - H_b(q_3)) + (\tfrac{3n_c}{2} - 2n_d)(1 - H_b(\tilde{\gamma})\Big] \end{aligned} \tag{4.42}$$

$$\begin{aligned} R_2 = &(n_d - \tfrac{n_c}{2}) \\ &+ (1-p)\Big[(3n_d - 2n_c)(1 + H_b(q_1)) \\ &\qquad + \left(\tfrac{3n_c}{2} - 2n_d\right)\left(1 + H_b(\gamma_1)\right)\Big] \\ &+ p\left[(3n_d - 2n_c)(1 - H_b(p_3)) + (\tfrac{3n_c}{2} - 2n_d)(1 - H_b(\tilde{\eta})\right] \end{aligned} \tag{4.43}$$

where $q_3 = (1-p)q_1 + pq_2$, $\tilde{\gamma} = p + \gamma_1(1-p)$, $p_3 = (1-p)p_1 + pp_2$ and $\tilde{\eta} = p + \eta_1(1-p)$. Remark 10 also applies to the parameters η_1 and γ_1 in (4.42) and (4.43).

(d) For $\frac{3}{2} \le \alpha \le 2$, we can achieve for any $\eta_1, \gamma_1 \in \left[\frac{1}{2}, 1\right]$

$$\begin{aligned} R_1 = &(n_c - n_d) + (1-p)\left[(n_d - \tfrac{n_c}{2})(1 + H_b(\eta_1))\right] \\ &+ p(n_d - \tfrac{n_c}{2})(1 - H_b(\tilde{\gamma})) \end{aligned} \tag{4.44}$$

$$\begin{aligned} R_2 = &\left(n_c - n_d\right) + (1-p)\left[(n_d - \tfrac{n_c}{2})(1 + H_b(\gamma_1))\right] \\ &+ p\left(n_d - \tfrac{n_c}{2}\right)\left(1 - H_b(\tilde{\eta})\right) \end{aligned} \tag{4.45}$$

where $\tilde{\gamma} = p + \gamma_1(1-p)$ and $\tilde{\eta} = p + \eta_1(1-p)$. Remark 10 also applies to the parameters η_1 and γ_1 in (4.44) and (4.45).

In each region, we optimize numerically over the set of parameters, exploiting in some cases that there is symmetry (except for $\alpha = 1$) between the corresponding parameters of both users.

4.4.3 Local CSIRT vs. Local CSIR

To evaluate the effect of exploiting local CSI at the transmitter side, we plot in Figures 4.2–4.4 the converse and achievability bounds for local CSIR and local CSIRT. For each interference region, we choose one value of α. We omit the VWI region because in this region both local CSIR and local CSIRT coincide. We observe that for all interference regions, except in the VWI region, local CSIRT outperforms local CSIR. We further observe that the largest improvement is obtained for $p = \frac{1}{2}$. This is not surprising, since in this case the uncertainty about the interference states is the largest.

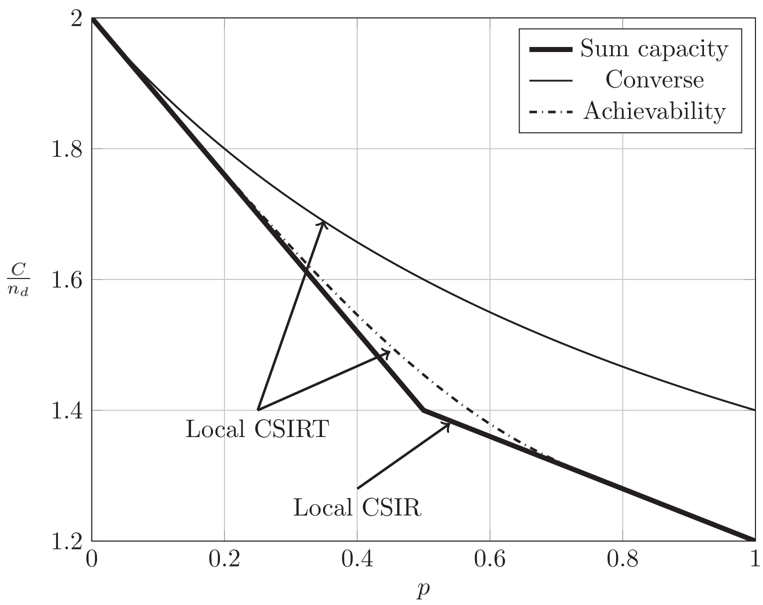


Figure 4.2: Local CSIRT vs. local CSIR for $\alpha = \frac{3}{5}$ (WI).

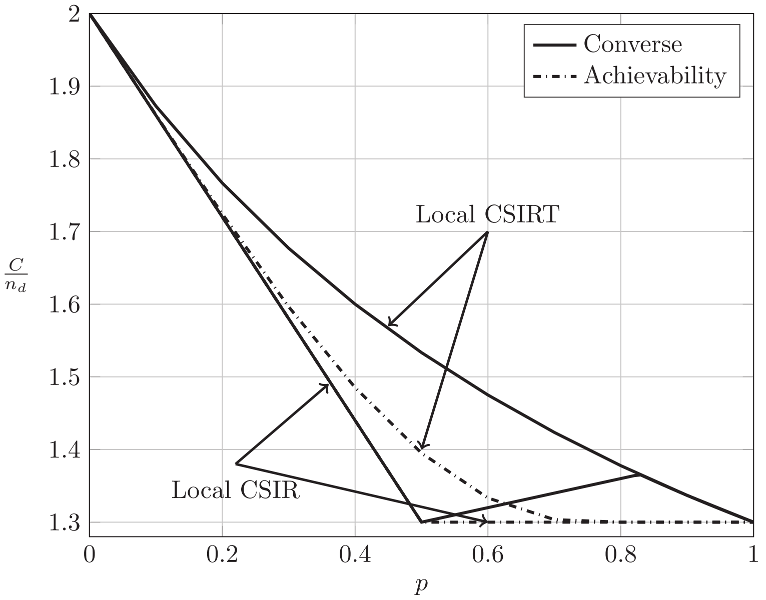


Figure 4.3: Local CSIRT vs. local CSIR for $\alpha = \frac{7}{10}$ (MI).

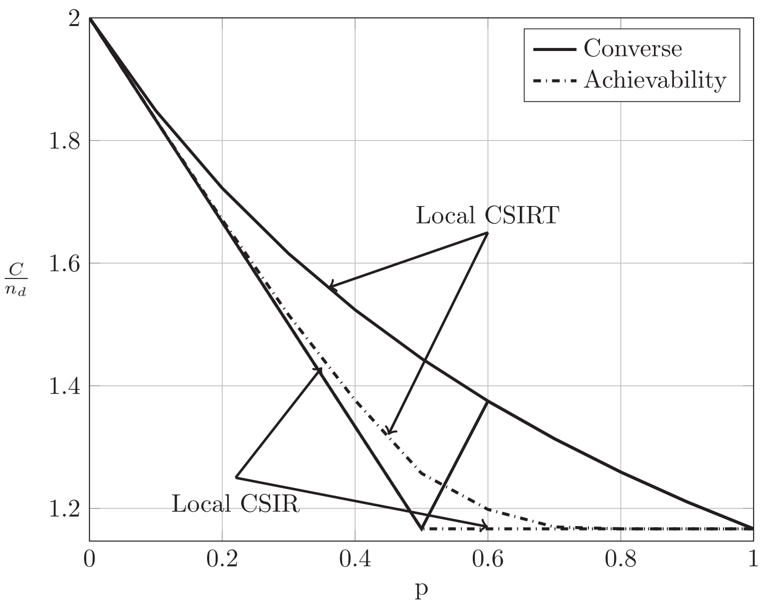


Figure 4.4: Local CSIRT vs. local CSIR for $\alpha = \frac{7}{6}$ (SI).

4.4.4 Quasi-Static vs. Ergodic Setup

As observed in the previous subsection, for the ergodic setup local CSIRT outperforms local CSIR in all interference regions (except VWI). In contrast, the opportunistic rates achievable in the quasi-static setup for local CSIRT coincide with those achievable for local CSIR. In other words, the availability of local CSI at the transmitter is only beneficial in the ergodic setup but not in the quasi-static one. This remains to be true even if we consider the average sum capacity rather than the sum rate region. Intuitively, in the coherent setup, the achievable rates depend on the input distributions of X_1^K and X_2^K, and adapting these distributions to the interference state yields a rate gain. In contrast, in the quasi-static setup, we treat the two interference states separately: the worst-case rates are designed for the worst case (where both receivers experience interference), and the opportunistic rates are designed for the best case (where the corresponding receiver is interference-free).

Given that the opportunistic rate region $(R, \Delta R(V_1, V_2))$ is not enhanced by the availability of local CSI at the transmitter, it follows directly that the same is true for the average sum capacity, defined in (4.21). Note, however, that it is unclear whether (4.21) corresponds to the best strategy to transmit several messages over independent uses of a quasi-static channel when the transmitters have access to local CSI. Indeed, in this case transmitter i may choose the values for R_i and $\Delta R_i(0)$ as a function of the interference state B_i, potentially giving rise to a larger average sum capacity. Yet, the set of achievable rate pairs $(R_i, \Delta R_i(0))$ depends on the choice of $(R_j, \Delta R_j(0))$ of transmitter $j \neq i$, which transmitter i may not deduce since it has no access to the other transmitter's CSI. How the transmitters should adapt their rates to the interference state remains therefore an open question.

4.5 Global CSIRT

We next present converse and achievability bounds for global CSIRT. In this scenario, the transmitters may agree on a specific coding scheme that depends on the realization of (B_1^K, B_2^K). This allows for a more elaborated cooperation between the transmitters and strictly increases the sum capacity compared to the local CSIR/CSIRT scenarios.

4.5.1 Quasi-Static Channel

In the quasi-static scenario with global CSIRT, the messages are, strictly speaking, not opportunistic. Instead, transmitters can choose the message depending on the true state of the interference links, so the strategy is perhaps better described as *rate adaptation*. Nevertheless, the definitions of worst-case sum rate and opportunistic sum rate in Section 4.2.1 still apply in this case. To keep notation consistent, we use the definition of "opportunism" also for global CSIRT.

4.5.1.1 Independent Case

Assume first that the sequences B_1 and B_2 are independent of each other.

Theorem 14 (Opportunistic sum capacity for global CSIRT) *Assume that B_1 and B_2 are independent of each other. For $0 < p < 1$, the opportunistic sum capacity region is the union of the set of rate tuples $(R, \Delta R(00), \Delta R(01), \Delta R(10))$ satisfying (4.10)–(4.12) and*

$$R + \Delta R(00) \leq 2n_d \tag{4.46}$$

$$R + \Delta R(01) \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{4.47}$$

$$R + \Delta R(10) \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{4.48}$$

Proof: The converse bounds are proved in Appendix A.1.4.1. The achievability bounds are achieved by the following achievability scheme: For $\mathbf{B} = [0, 0]$ we use all the n_d sub-channels of both parallel channels. For $\mathbf{B} = [0, 1]$ and $\mathbf{B} = [1, 0]$ and the VWI/WI regions, we use all n_d sub-channels and the receivers decode them only if they are not affected by interference. For the MI/SI regions, we treat the bursty IC as a non-bursty IC and use the achievability schemes of the IC proposed in [29]. The details can be found in Appendix A.1.4.2. ■

Remark 11 *The proofs of Theorems 11 and 14 merely require that the joint distribution $p_{b_1 b_2} \triangleq \Pr\{\mathbf{B} = [b_1, b_2]\}$ satisfies $p_{00} < 1$, $p_{01} > 0$, $p_{10} > 0$ and $p_{11} > 0$. Thus, these theorems also apply to the case where B_1 and B_2 are dependent, as long as they are not fully correlated.*

Table 4.4 summarizes the results of Theorem 14. Observe that for VWI and WI opportunistic messages can be transmitted reliably at a positive rate, while for MI and SI this is only the case if both links are interference-free.

Table 4.4: Opportunistic sum capacity for global CSIRT when the worst-case sum rate is maximized and B_1 and B_2 are independent.

Rates	**VWI**	**WI**	**MI**	**SI**
C	$2(n_d - n_c)$	$2n_c$	$2n_d - n_c$	n_c
$\Delta C(00)$	$2n_c$	$2(n_d - n_c)$	n_c	$2n_d - n_c$
$\Delta C(01)/\Delta C(10)$	n_c	$2n_d - 3n_c$	0	0

4.5.1.2 Fully Correlated Case

Next, we consider the case in which the interference states are fully correlated. In this scenario, local CSIRT coincides with global CSIRT.

Theorem 15 (Opportunistic sum capacity for global CSIRT) *Assume that B_1 and B_2 are fully correlated. For $0 \leq p < 1$, the opportunistic sum capacity region is the union of the set of rate pairs $(R, \Delta R(00))$ satisfying* (4.10)–(4.12) *and*

$$R + \Delta R(00) \leq 2n_d. \tag{4.49}$$

Proof: For the converse bound, we note that the analysis in Appendix A.1.4.1 applies directly to the case where the states B_1 and B_2 are fully correlated, with the only difference that there are only two possible cases $\mathbf{B} = [0, 0]$ and $\mathbf{B} = [1, 1]$. The result follows then from (A.59), (A.60) and (A.62). For the achievability bound, we use an achievability scheme where the opportunistic messages are only decoded in absence of interference at the intended receiver. In this case, we have two parallel interference-free channels, for which the optimal strategy consists of transmitting uncoded bits in the n_d sub-channels. ■

Table 4.5 summarizes the results of Theorem 15. Observe that the worst-case sum capacity C and the opportunistic sum capacity $\Delta C(00)$ when the channel is interference-free do not depend on the correlation between B_1 and B_2. The only difference between the independent and fully correlated case is that the interference states $[0, 1]$ and $[1, 0]$ are impossible if $B_1 = B_2$.

Table 4.5: Opportunistic sum capacity for global CSIRT when the worst-case sum rate is maximized and B_1 and B_2 are fully correlated.

Rates	VWI	WI	MI	SI
C	$2(n_d - n_c)$	$2n_c$	$2n_d - n_c$	n_c
$\Delta C(00)$	$2n_c$	$2(n_d - n_c)$	n_c	$2n_d - n_c$

4.5.2 Ergodic Channel

4.5.2.1 Independent Case

When the sequences B_1^K and B_2^K are independent of each other, we have the following theorems.

Theorem 16 (Converse bounds for global CSIRT) *Assume that B_1^K and B_2^K are independent of each other. The sum rate R for the bursty IC is upper-bounded by*

$$R \leq 2(1-p)n_d + p\left[(n_d - n_c)^+ + \max(n_d, n_c)\right] \tag{4.50}$$

and

$$\begin{aligned} R \leq 2\big[&p(1-p)\{(n_d - n_c)^+ + \max(n_d, n_c)\} \\ &+ (1-p)^2 n_d + p^2 \max\{(n_d - n_c)^+, n_c\}\big]. \end{aligned} \tag{4.51}$$

Proof: The proof of (4.50) follows along similar lines as (4.16) but noting that, for global CSIRT, $\boldsymbol{X}_i^K$ depends on both B_1^K and B_2^K. The proof of (4.51) is based on pairing the interference states according the four possible combinations of $(B_{1,k}, B_{2,k})$. See Appendix A.2.3 for details. ■

Remark 12 *The proof of Theorem 16 can be extended to consider an arbitrary joint distribution $p_{b_1 b_2} \triangleq \Pr\{\mathbf{B}_k = [b_1, b_2]\}$. In this case (4.50) is replaced by*

$$R \leq 2(p_{00} + p_{01})n_d + (p_{10} + p_{11})\left[(n_d - n_c)^+ + \max(n_d, n_c)\right]$$
$$R \leq 2(p_{00} + p_{10})n_d + (p_{01} + p_{11})\left[(n_d - n_c)^+ + \max(n_d, n_c)\right]$$

and (4.51) becomes

$$R \leq (p_{01} + p_{10})[(n_d - n_c)^+ + \max(n_d, n_c)] + 2\left[p_{00} n_d + p_{11} \max\{(n_d - n_c)^+, n_c\}\right].$$

Theorem 17 (Achievability bounds for global CSIRT) *Assume that B_1^K and B_2^K are independent of each other. The following sum rates R are achievable over the bursty IC:*

$$\begin{aligned} R &= 2\big[p(1-p)(2n_d - n_c) + (1-p)^2 n_d \\ &\qquad + p^2 \max\{(n_d - n_c)^+, n_c\}\big], && \textit{(VWI/WI)} \quad (4.52) \\ R &= 4n_d p_{\min} + 2n_d(1-p)^2 + \big(2n_d - n_c\big)\big(2p - p^2 - 3p_{\min}\big), && \textit{(MI)} \quad (4.53) \\ R &= 2(n_d + n_c)p_{\min} + 2n_d(1-p)^2 + n_c\big(2p - p^2 - 3p_{\min}\big), && \textit{(SI)} \quad (4.54) \end{aligned}$$

where $p_{\min} \triangleq \min(p^2, p(1-p))$.

Proof: The sum rate (4.52) is achieved by using the optimal scheme for the non-bursty IC when any of the two receivers is affected by interference [29], and by using uncoded transmission when there is no interference. The sum rates (4.53) and (4.54) are novel. See Appendix A.2.4 for details. ∎

Remark 13 *In contrast to the local CSIR scenario, the achievability schemes presented in Theorem 17 differ noticeably from those in [59] for the binary IC. Indeed, while both works exploit global CSIRT to enable cooperation between users, [59] assumes that only delayed CSI is present. The achievability schemes presented in Theorem 17 thus cannot be applied directly to the model considered in [59].*

Table 4.6 summarizes the results of Theorems 16 and 17. We write the sum capacity in bold face when converse and achievability bounds coincide. In Table 4.6, we define

$$\mathbf{C}_{\text{GMI}} \triangleq \min\left\{2n_d - pn_c, 2\left[(1-p^2) - (1-2p)\alpha p\right]\right\} \qquad (4.55)$$

$$\mathbf{C}_{\text{GSI}} \triangleq \min\left[n_c p + 2(1-p)n_d, 2n_d(1-p)^2 + 2n_c p\right] \qquad (4.56)$$

where "G" stands for "global CSIRT".

4.5.2.2 Fully Correlated Case

We next discuss the case where the sequences B_1^K and B_2^K are fully correlated, i.e., $B_1^K = B_2^K$.

Table 4.6: Bounds on the sum capacity C for global CSIRT when B_1^K and B_2^K are independent.

Regions	Achievability	Converse
VWI	$\mathbf{2(n_d - pn_c)}$	
WI	$\mathbf{2[(1-p^2)n_d + (1-2p)pn_c]}$	
MI	$4n_d p_{\min} + 2n_d(1-p)^2 + (2n_d - n_c)(2p - p^2 - 3p_{\min})$	$\mathbf{C}_{\mathrm{GMI}}$
SI	$2(n_d + n_c)p_{\min} + 2n_d(1-p)^2 + n_c(2p - p^2 - 3p_{\min})$	$\mathbf{C}_{\mathrm{GSI}}$

Theorem 18 (Converse bounds for global CSIRT) *Assume that B_1^K and B_2^K are fully correlated. The sum rate R for the bursty IC is upper-bounded by*

$$R \leq 2(1-p)n_d + p\{(n_d - n_c)^+ + \max(n_d, n_c)\} \tag{4.57}$$

$$R \leq 2\left[(1-p)n_d + p\max\{(n_d - n_c)^+, n_c\}\right]. \tag{4.58}$$

Proof: The proof of (4.57) follows similar steps as in Appendix A.2.3.1 but considering $B_1^K = B_2^K = B^K$. The proof of (4.58) is given in Appendix A.2.5. See also Remark 12. ∎

Theorem 19 (Achievability bounds for global CSIRT) *Assume that B_1^K and B_2^K are fully correlated. The following sum rates R are achievable over the bursty IC:*

$$R = 2\left[(1-p)n_d + p\max\{(n_d - n_c)^+, n_c\}\right], \quad \textit{(VWI/WI)} \tag{4.59}$$

$$R = 2(1-p)n_d + p\{(n_d - n_c)^+ + \max(n_d, n_c)\}, \quad \textit{(MI/SI).} \tag{4.60}$$

Proof: The sum rates (4.59) and (4.60) are achieved by using the optimal scheme for the non-bursty IC when the two receivers are affected by interference [29], and by using uncoded transmission in absence of interference. ∎

Table 4.7 summarizes the results of Theorems 18 and 19. For global CSIRT and fully correlated B_1^K and B_2^K, converse and achievability bounds coincide. Thus, (4.59) and (4.60) indicate the sum capacity.

Table 4.7: Bounds on the sum capacity C for global CSIRT when B_1^K and B_2^K are fully correlated.

Regions	Bounds
VWI	$2(n_d - pn_c)$
WI	$2[(1-p)n_d + pn_c]$
MI	$2(1-p)n_d + p(2n_d + n_c)$
SI	$2(1-p)n_d + p(n_c)$

4.5.3 Quasi-Static vs. Ergodic Setup

Similar to the average sum capacity for local CSIR defined in Section 4.3.3, we define the average sum capacity for global CSIRT when B_1 and B_2 are independent as

$$\begin{aligned} \bar{C} =& p^2 \sup_R \{R\} + p(1-p) \sup_{(R, \Delta R(01))} \{R + \Delta R(01)\} \\ &+ p(1-p) \sup_{(R, \Delta R(10))} \{R + \Delta R(10)\} \\ &+ (1-p)^2 \sup_{(R, \Delta R(00))} \{R + \Delta R(00)\} \end{aligned} \tag{4.61}$$

where the suprema are over all rate tuples $(R, \Delta R(00), \Delta R(01), \Delta R(10))$ that satisfy Theorems 10 and 14. The intuition behind (4.61) is the same as that behind (4.21) for local CSIR, but with global CSIRT the transmitters can adapt their rates $(R_i, \Delta R_i(V_i))$ to the interference state. For example, the first term on the right-hand side (RHS) of (4.61) corresponds to the interference state $[1, 1]$, in which case we transmit at total sum rate R; the second term corresponds to the interference state $[0, 1]$, in which case we transmit at total sum rate $R + \Delta R(01)$; and so on.

Table 4.8 summarizes the average sum capacity for the different interference regions. The average sum capacities for VWI and WI coincide with the sum capacities in the ergodic setup (see Table 4.6). In contrast, for MI and SI, the average sum capacities are smaller than the sum capacities in the ergodic setup.

Similarly, in the fully correlated case, we define the average sum capacity as

$$\bar{C} \triangleq p \sup_R \{R\} + (1-p) \sup_{(R, \Delta R(00))} \{(R + \Delta R(00))\} \tag{4.62}$$

Table 4.8: Average sum capacity when B_1 and B_2 are independent.

Regions	Bounds
VWI	$2(n_d - pn_c)$
WI	$2[(1-p^2)n_d + (1-2p)pn_c]$
MI	$2n_d - pn_c(2-p)$
SI	$2n_d(1-p)^2 + pn_c(2-p)$

Table 4.9: Average sum capacity when B_1 and B_2 are fully correlated.

Regions	Bounds
VWI	$2(n_d - pn_c)$
WI	$2[(1-p)n_d + pn_c]$
MI	$2(1-p)n_d + p(2n_d + n_c)$
SI	$2(1-p)n_d + p(n_c)$

where the suprema are over all rate pairs $(R, \Delta R(00))$ that satisfy Theorems 10 and 15. The corresponding results are summarized in Table 4.9.

We observe that the average sum capacities coincide with the sum capacities of the ergodic setup.

4.6 Exploiting CSI

In this section, we study how the level of CSI affects the sum rate in the quasi-static and ergodic setups.

For the quasi-static channel, Figures 4.5 and 4.6 show the total sum capacity presented in Theorems 11, 14 and 15. Specifically, we plot the normalized total sum capacity $\frac{C+\Delta C}{n_d}$ versus α, comparing scenarios of local CSIR/CSIRT and global CSIRT. We analyze separately the cases $\mathbf{B} = [0,0]$ and $\mathbf{B} = [0,1]$. For the case where $\mathbf{B} = [0,0]$ and global CSIRT, the total sum capacity is $2n_d$ for all interference regions. For $\mathbf{B} = [0,0]$ and local CSIR/CSIRT, the total sum capacity is $2n_d$ for VWI and VSI, but is strictly smaller in the remaining interference regions. Hence, in these regions global CSIRT outperforms local CSIR/CSIRT. For the case where $\mathbf{B} = [0,1]$, the total sum capacity is equal to $(n_d - n_c)^+ + \max(n_d, n_c)$

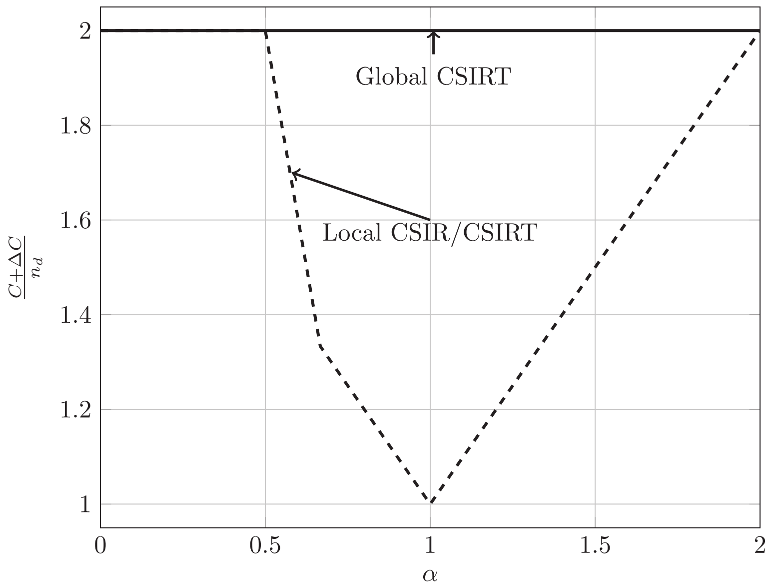


Figure 4.5: Total sum capacity for $\mathbf{B} = [0, 0]$, for local CSIR/CSIRT and global CSIRT.

irrespective of the level of CSI.

We further observe that the opportunistic-capacity region for local CSIRT is equal to that for local CSIR. Thus, local CSI at the transmitter is not beneficial. As we shall see later, this is in stark contrast to the ergodic setup, where local CSI at the transmitter-side is beneficial. Intuitively, in the ergodic case the input distributions of $\boldsymbol{X}_1^K$ and $\boldsymbol{X}_2^K$ depend on the realizations of B_1^K and B_2^K, respectively. Hence, adapting the input distributions to these realizations increases the sum capacity. In contrast, in the quasi-static case, the worst-case scenario (presence of interference) and the best-case scenario (absence of interference) are treated separately. Hence, there is no difference to the case of local CSIR.

For the ergodic setup, Figures 4.7–4.10 show the converse and achievability bounds presented in Theorems 12, 13, 16 and 17. We further include the results on local CSIRT presented in Section 4.4. Specifically, we plot the normalized sum capacity $\frac{C}{n_d}$ versus the probability of presence of interference p, comparing scenarios of local CSIR, local CSIRT and global CSIRT when B_1^K and B_2^K are independent of each other. The shadowed areas correspond to the regions where achievability and converse bounds do not coincide.

Figure 4.7 reveals that in the VWI region the sum capacity is equal to $2(n_d - pn_c)$, irrespective of the availability of CSI (see Figure 4.7). Thus, in

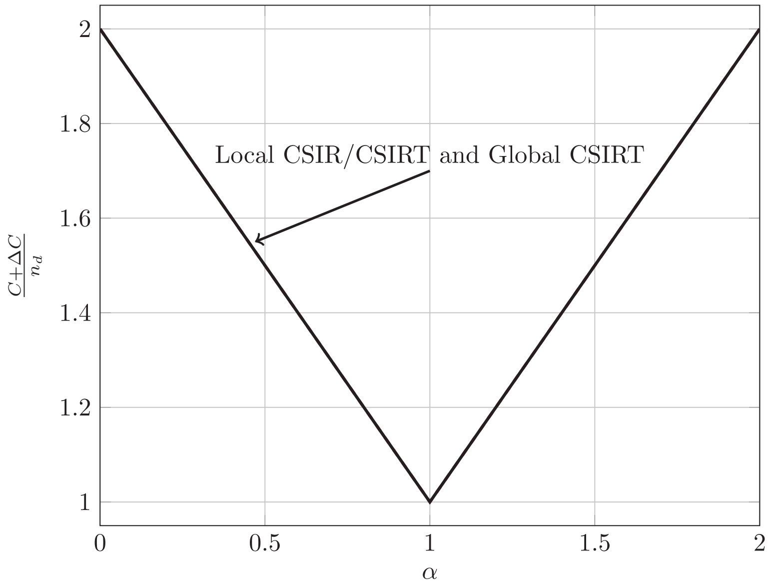


Figure 4.6: Total sum capacity for $\mathbf{B} = [0, 1]$, for local CSIR/CSIRT and global CSIRT.

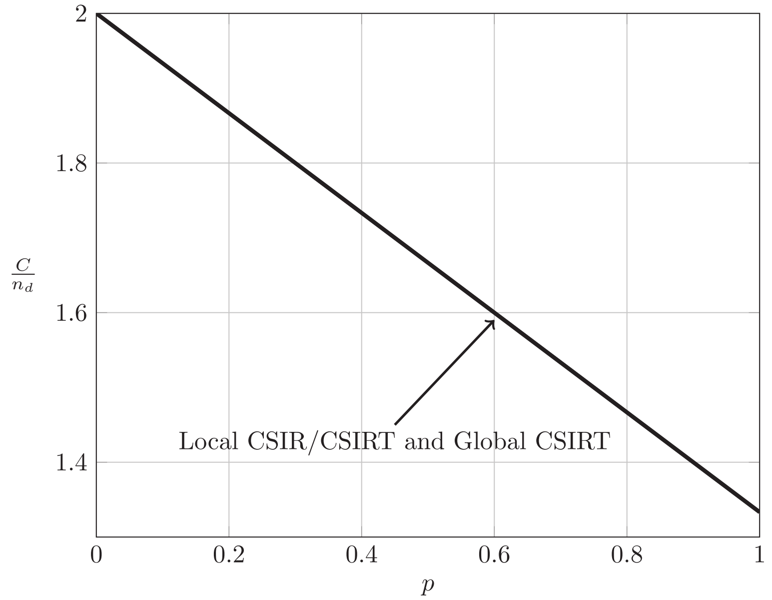


Figure 4.7: Sum capacity for local CSIR/CSIRT and global CSIRT when B_1^K and B_2^K are independent and $\alpha = \frac{1}{3}$ (VWI).

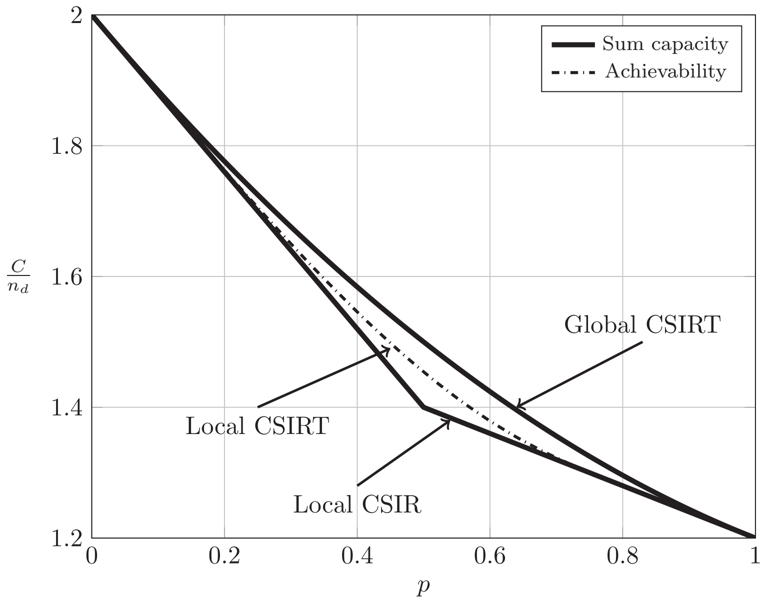


Figure 4.8: Sum capacity for local CSIR/CSIRT and global CSIRT when B_1^K and B_2^K are independent and $\alpha = \frac{3}{5}$ (WI).

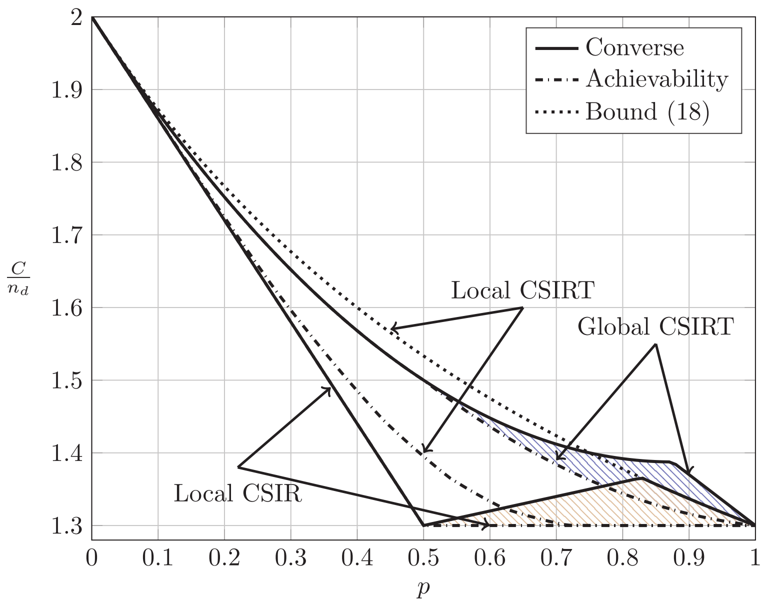


Figure 4.9: Sum capacity for local CSIR/CSIRT and global CSIRT when B_1^K and B_2^K are independent and $\alpha = \frac{7}{10}$ (MI).

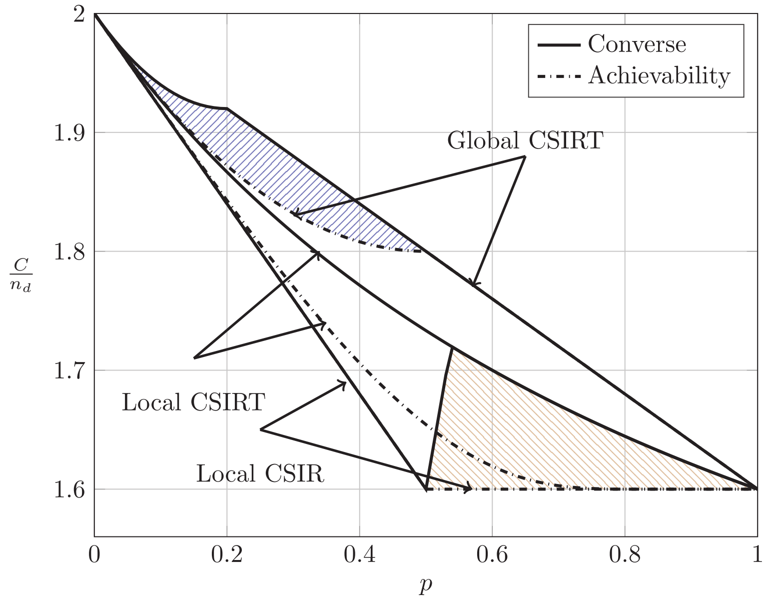


Figure 4.10: Sum capacity for local CSIR/CSIRT and global CSIRT when B_1^K and B_2^K are independent and $\alpha = \frac{8}{5}$ (SI).

this region access to global CSIRT is not beneficial compared to the local CSIR scenario. In the VSI region, the sum capacity of the non-bursty IC is equal to $2n_d$, which is that of two parallel channels without interference [5, Sec. II-A]. Therefore, burstiness of the interference (and hence CSI) does not affect the sum capacity.

In the WI region, shown in Figure 4.8, the converse and achievability bounds for local CSIR and global CSIRT coincide and it is apparent that global CSIRT outperforms local CSIR. In theMI and SI regions, the converse and achievability bounds only coincide for certain regions of p. Nevertheless, Figures 4.9 and 4.10 show that, in almost all cases, global CSIRT outperforms local CSIR. (For the case presented in Figure 4.9 $\left(\alpha = \frac{7}{10}\right)$, we also present the local CSIRT converse bound (4.16), although it is looser for some values of p, with respect to the one depicted for global CSIRT.) Local CSIRT outperforms local CSIR in all interference regions (except VWI). We stress again the fact that this was not the case in the quasi-static scenario, where both coincide.

We next consider the case where B_1^K and B_2^K are fully correlated. For this scenario, [65], [66] studied the effect of perfect feedback on the bursty IC. For comparison, the non-bursty IC with feedback was studied by Suh *et al.* in [54], where it was demonstrated that the gain of feedback becomes arbitrarily large for certain interference regions (VWI and WI) when the signal-to-noise-ratio increases. This gain corresponds to a better resource utilization and thereby a better resource sharing between users. Specifically, [65], [66] (bursty IC) and [54] (non-bursty IC) assume that noiseless, delayed feedback is available from receiver i to transmitter i ($i = 1, 2$). For the symmetric setup treated in this chapter, [65, Th. 3.2] or [66, Th. 3.2] showed the following:

Theorem 20 (Capacity for the bursty IC with feedback [65], [66]) *The sum capacity of the bursty IC with noiseless, delayed feedback is given by*

$$C = \begin{cases} 2n_d - 2\frac{p}{1+p}n_c, & \alpha \leq 1, \\ 2\frac{1-p}{1+p}n_d + 2\frac{p}{1+p}n_c & 1 < \alpha \leq 2, \\ 2(1-p)n_d + pn_c, & 2 < \alpha. \end{cases} \tag{4.63}$$

Proof: See [65, Sec. IV and V], [66, Sec. IV and V, Appendices A, C, D]. ■

Observe that (4.63) for $\alpha \leq 2$ coincides with (4.16). This implies that local CSIRT can never outperform delayed feedback. Intuitively, feedback contains not only information about the channel state, but also about the previous symbols transmitted by the other transmitter, which can be exploited to establish a certain cooperation between the transmitters. Figures 4.11–4.14 show the bounds on the normalized sum capacity, $\frac{C}{n_d}$, comparing the scenarios of local CSIR versus global CSIRT when the interference states are fully correlated, i.e., $B_1^K = B_2^K$. They further show the sum capacity for the case where the transmitters have noiseless delayed feedback [65]. The shadowed areas correspond to the regions where achievability and converse bounds do not coincide.

Figure 4.11 reveals that feedback in the VWI region outperforms the non-feedback case, irrespective of the availability of CSI. Wang *et al.* [65] have further shown that feedback also outperforms the non-feedback case in the VSI region. The order between global CSIRT and the feedback scheme is not obvious. There are regions where global CSIRT outperforms the feedback scheme and vice versa. Indeed, on the one hand, feedback contains information about the previous interference states and previous symbols transmitted by the other transmitter, permitting the resolution of collisions in previous transmissions. On the other hand, global CSIRT provides *non-causal* information about the interference states,

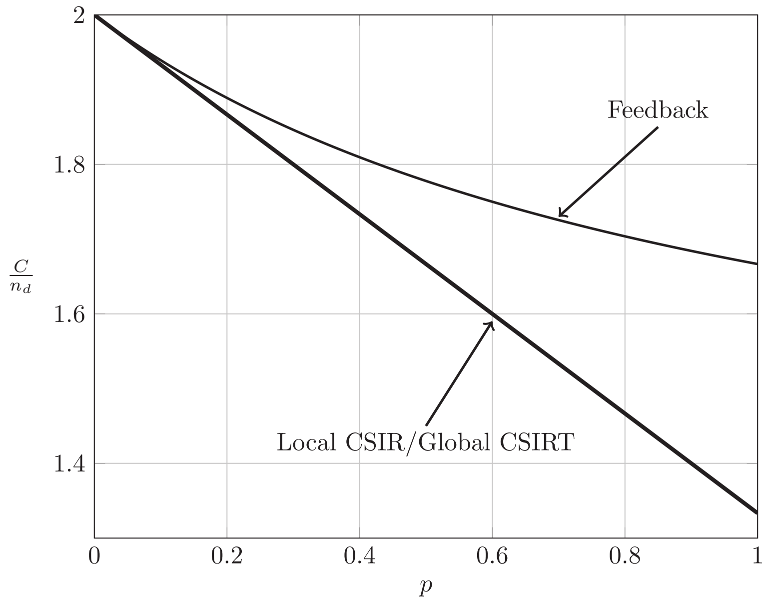


Figure 4.11: Sum capacity for local CSIR and global CSIRT when $B_1^K = B_2^K$ and $\alpha = \frac{1}{3}$ (VWI).

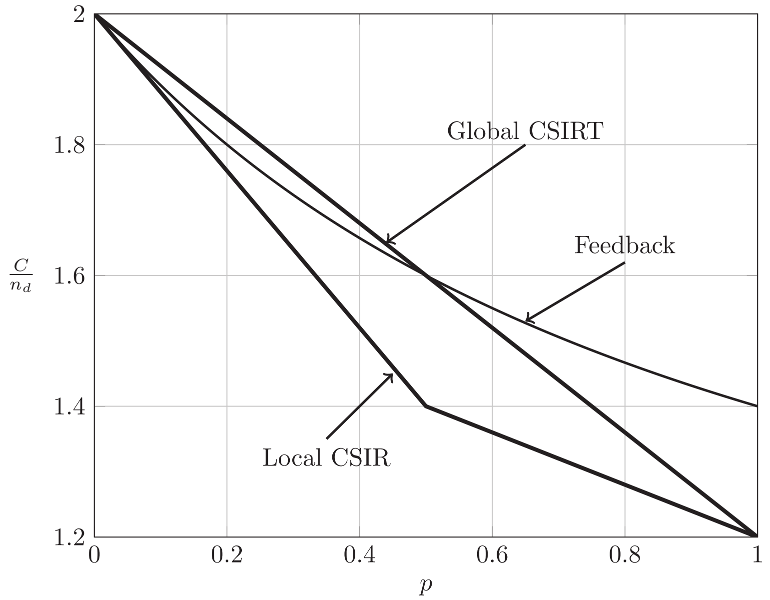


Figure 4.12: Sum capacity for local CSIR and global CSIRT when $B_1^K = B_2^K$ and $\alpha = \frac{3}{5}$ (WI).

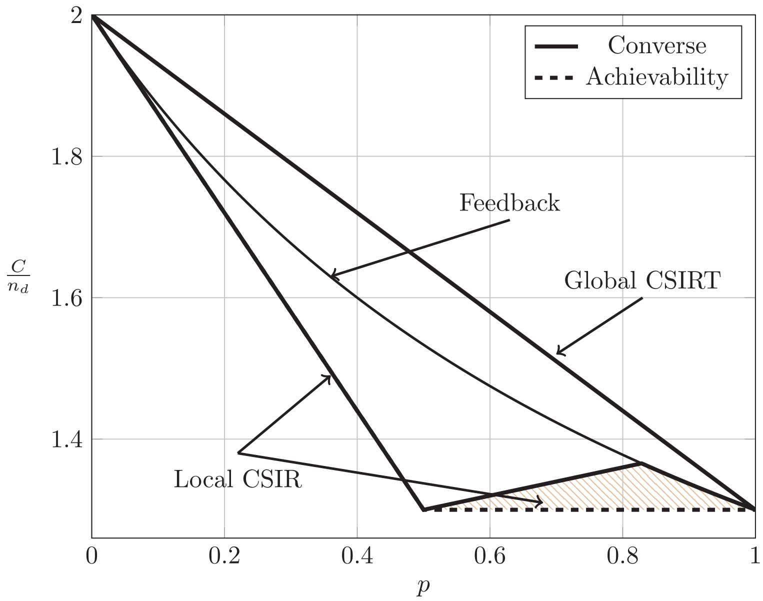


Figure 4.13: Sum capacity for local CSIR and global CSIRT when $B_1^K = B_2^K$ and $\alpha = \frac{7}{10}$ (MI).

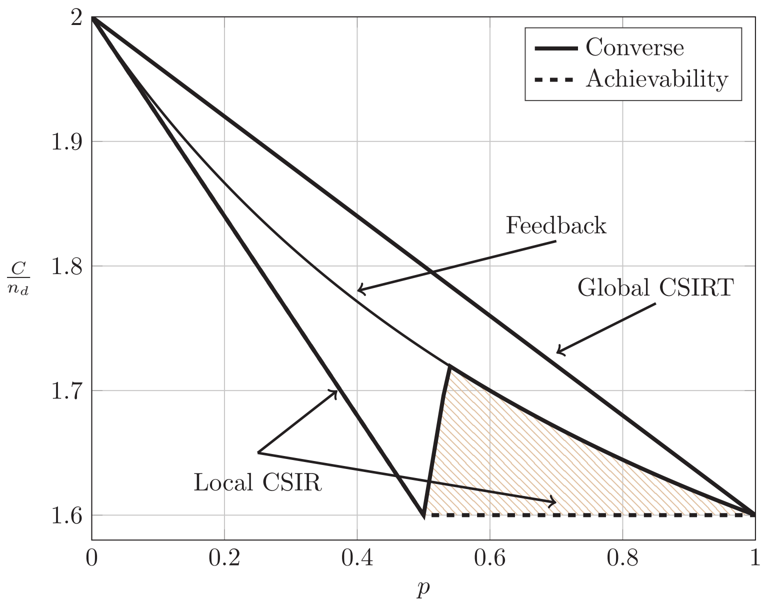


Figure 4.14: Sum capacity for local CSIR and global CSIRT when $B_1^K = B_2^K$ and $\alpha = \frac{8}{5}$ (SI).

allowing a better adaptation of the transmission strategy to the interference burstiness.

4.7 Exploiting Interference Burstiness

To better illustrate the benefits of interference burstiness, we show the normalized sum capacity as a function of α, in order to appreciate all the interference regions. In the non-bursty IC ($p = 1$), this curve corresponds to the well-known W-curve obtained by Etkin *et al.* in [21]. We next study how burstiness affects this curve in the different considered scenarios.

In the quasi-static setup, burstiness can be exploited by sending opportunistic messages. We consider the total sum capacity for the case where the worst-case rate R is maximized. For local CSIR/CSIRT, Theorem 11 suggests that the use of an opportunistic code is only beneficial if the interference region is VWI or WI. For other interference regions there is no benefit. In contrast, for global CSIRT an opportunistic code is beneficial for all interference regions (except for VSI where the sum capacity corresponds to that of two parallel channels without interference).

Figures 4.15 and 4.16 illustrate these observations. Specifically, in Figures 4.15 and 4.16 we show the normalized total sum capacity achieved under local CSIR/CSIRT and global CSIRT when the interference states are independent. We observe that, for local CSIR, the opportunistic rates $\Delta R_1(0)$ and $\Delta R_2(0)$, are only positive in the VWI and WI regions. In these regions, if only one of the receivers is affected by interference the sum capacity is given by the worst-case rate R plus one opportunistic rate of the user which is not affected by interference. In absence of interference at both receivers, both receivers can decode opportunistic messages. Hence, the total sum capacity is equal to $C + \Delta C_1(0) + \Delta C_2(0)$. For global CSIRT we can observe that, when only one of the receivers is affected by interference, we achieve the same total sum capacity as in the local CSIR/CSIRT. However, in absence of interference at both receivers, we achieve the trivial upper bound corresponding to two parallel channels. The fully correlated scenario can be considered as a subset of the independent scenario. Indeed, for the case $\mathbf{B} = [0, 0]$ and $\mathbf{B} = [1, 1]$ we obtain the same total sum capacity as for the independent scenario. The main difference is that in the fully correlated scenario the interference

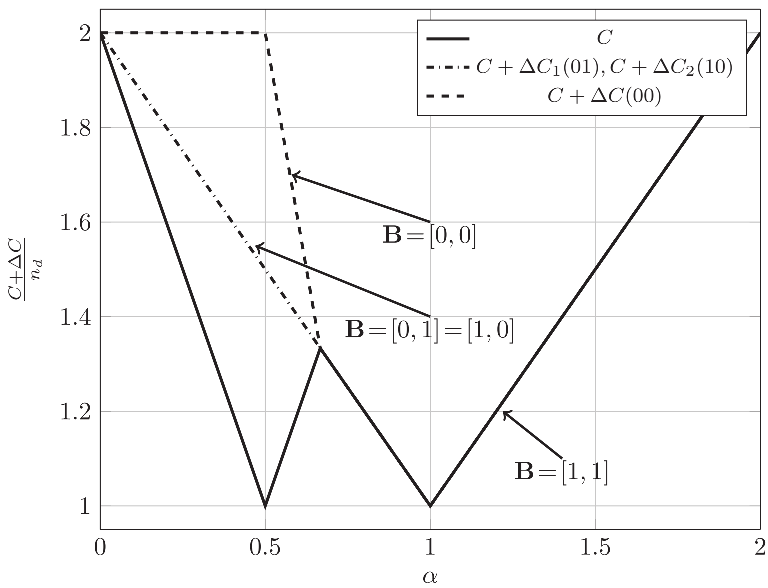


Figure 4.15: Normalized total sum capacity $\frac{C+\Delta C}{n_d}$ as a function of α for local CSIR/CSIRT when B_1 and B_2 are independent.

states $\mathbf{B} = [0, 1]$ and $\mathbf{B} = [1, 0]$ are impossible.

For the ergodic case, Figures 4.17 and 4.18 show the bounds on the normalized sum capacity, $\frac{C}{n_d}$, as a function of α when B_1^K and B_2^K are independent. The shadowed areas correspond to the regions where achievability and converse bounds do not coincide. We further show the W-curve. Observe that for $p \leq \frac{1}{2}$ the sum capacity as a function of α forms a V-curve instead of the W-curve. Further observe how the sum capacity approaches the W-curve as p tends to one.

In Figure 4.19 we show the bounds on the normalized sum capacity, $\frac{C}{n_d}$, as a function of α for global CSIRT when B_1^K and B_2^K are fully correlated. (For local CSIR the sum capacity is not affected by the correlation between B_1^K and B_2^K, so the curve for $\frac{C}{n_d}$ as a function of α coincides with the one obtained in Figure 4.17.) We observe that, for all values of $p > 0$, the sum capacity forms a W-curve similar to the W-curve for $p = 1$. This is the case because, when both interference states are fully correlated, the bursty IC is a combination of an IC and two parallel channels.

We observe that for global CSIRT the burstiness of the interference is beneficial for all interference regions and all values of p. For local CSIR, burstiness is beneficial for all values of p for VWI and WI. However, for MI and SI, burstiness is

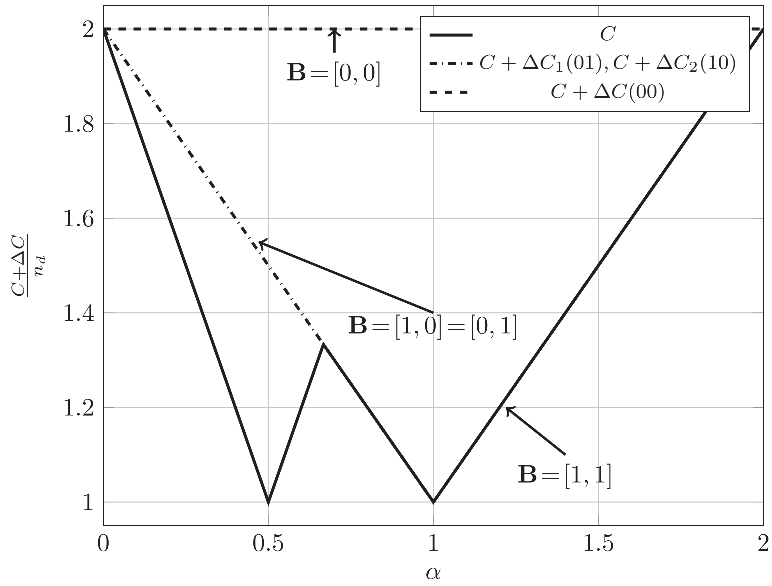


Figure 4.16: Normalized total sum capacity $\frac{C+\Delta C}{n_d}$ as a function of α for global CSIRT when B_1 and B_2 are independent.

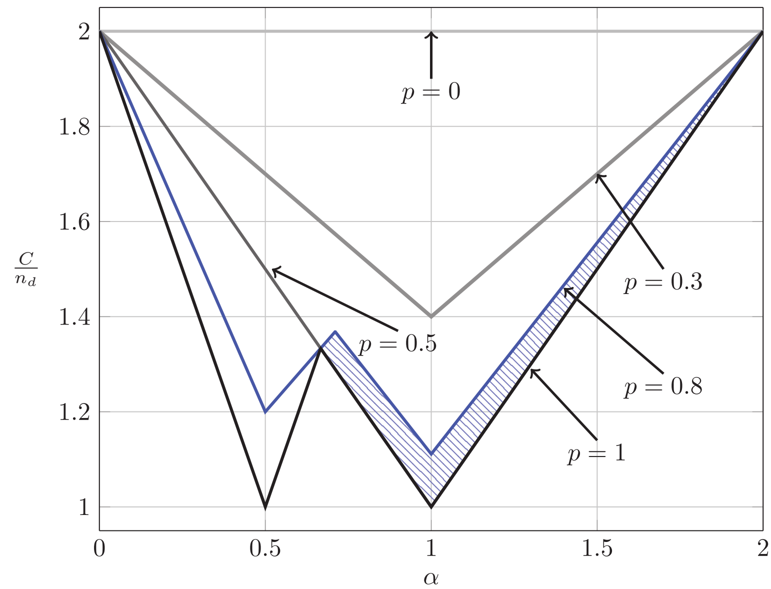


Figure 4.17: Normalized sum capacity $\frac{C}{n_d}$ as a function of α for local CSIR/CSIRT when B_1^K and B_2^K are independent.

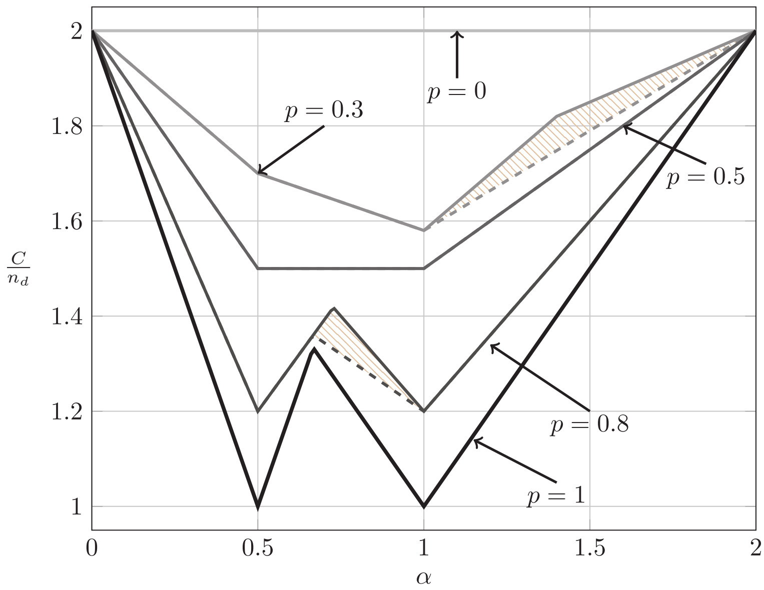


Figure 4.18: Normalized sum capacity $\frac{C}{n_d}$ as a function of α for global CSIRT when B_1^K and B_2^K are independent.

only of clear benefit for $p \leq \frac{1}{2}$. It is yet unclear whether burstiness is also beneficial in these interference regions when $p > \frac{1}{2}$. To shed some light on this question, note that evaluating the converse bound in [66, Lemma A.1], which yields (4.19), for inputs $\boldsymbol{X}_1^K$ and $\boldsymbol{X}_2^K$ that are temporally independent, we recover the achievability bound (4.18). Since for MI/SI and $p \geq \frac{1}{2}$ this bound coincides with the rates achievable over the non-bursty IC, this implies that an achievability scheme can only exploit the burstiness of the interference in this regime if it introduces some temporal correlation (this observation is also revealed by considering the average sum capacity for the quasi-static case). In fact, for global CSIRT the achievability schemes proposed in Theorem 17 for MI and SI copy the same bits over several coherence blocks, i.e., they exhibit a temporal correlation, which cannot be achieved using temporally independent distributions. However, the temporal pattern of these bits requires knowledge of both interference states, so this approach cannot be adapted to the cases of local CSIR/CSIRT. In contrast, for global CSIRT in the fully correlated case where converse and achievability bounds coincide, it is not necessary to introduce temporal memory. This scenario is simpler, since in this case the channel exhibits only two channel states, a non-bursty IC and two parallel channels.

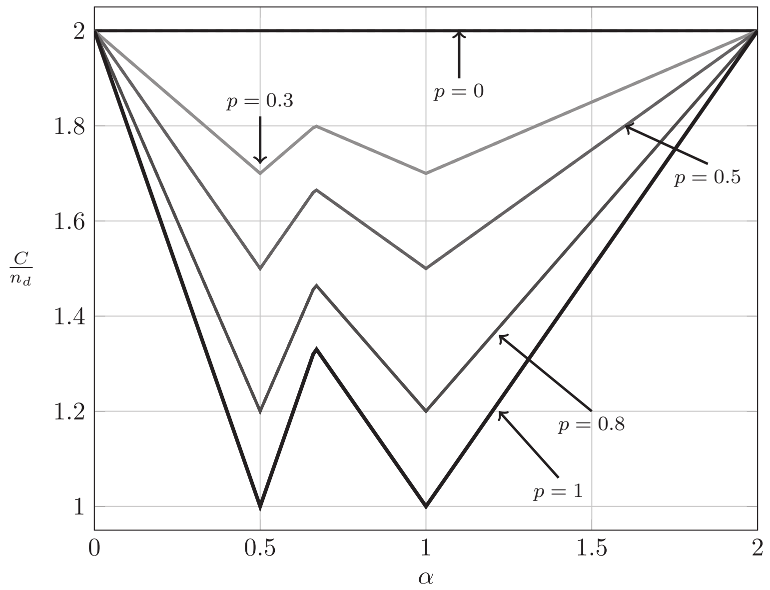


Figure 4.19: Normalized sum capacity $\frac{C}{n_d}$ as a function of α for global CSIRT when $B_1^K = B_2^K$.

4.8 Summary and Conclusions

In this thesis, we considered a two-user bursty IC in which the presence/absence of interference is modeled by a block-IID Bernoulli process while the power of the direct and cross links remains constant during the whole transmission. This scenario corresponds, e.g., to a slow-fading scenario in which all the nodes can track the channel gains of the different links, but where the interfering links are affected by intermittent occlusions due to some physical process. While this model may appear over-simplified, it yields a unified treatment of several aspects previously studied in the literature and gives rise to several new results on the effect of the CSI in the achievable rates over the bursty IC. Our channel model encompasses both the quasi-static scenario studied in [16], [33] and the ergodic scenario (see, e.g., [65], [59]). While the model recovers several cases studied in the literature, it also presents scenarios which have not been previously analyzed. This is the case, for example, for the ergodic setup with local and global CSIRT. Our analysis in these scenarios does not yield matching upper and lower bounds for all interference and burstiness levels. Yet, examining the obtained results, we observe that the best strategies in these scenarios often require elaborated coding strategies for both users that feature memory across different interference. This fact probably explains why no previous results exist in these scenarios. Furthermore, several of our proposed achievability schemes require complex correlation among signal levels. Thus, while the LDM in general provides insights on the Gaussian IC, the proposed schemes may actually be difficult to convert to the Gaussian case.

In the quasi-static scenario, the highest sum rate R that can be achieved is limited by the worst realization of the channel and thus coincides with that of the (non-bursty) IC. We can however transmit at an increased (opportunistic) sum rate $R + \Delta R$ when there is no interference at any of the interfering links. For the ergodic setup, we showed that an increased rate can be obtained when local CSI is present at both transmitter and receiver, compared to that obtained when CSI is only available at the receiver side. This is in contrast to the quasi-static scenario, where the achievable rates for local CSIR and local CSIRT coincide. Featuring global CSIRT at all nodes yields an increased sum rate for both the quasi-static and the ergodic scenarios. In the quasi-static channel, global CSI yields increased opportunistic rates in all the regions except in the very strong interference region, which is equivalent to having two parallel channels with no interference.

Both in the quasi-static and ergodic scenarios, global CSI exploits interference burstiness for all interference regions (except for very strong interference), irrespective of the level of burstiness. When local CSI is available only at the receiver side, interference burstiness is of clear benefit if the interference is either weak or very weak, or if the channel is ergodic and interference is present at most half of the time. When local CSI is available at each transmitter and receiver and the channel is ergodic, interference burstiness is beneficial in all interference regions except in the very weak and very strong interference regions.

In order to compare the achievable rates of the quasi-static and ergodic setup, one can define the average sum rate of the quasi-static setup for local CSIR/CSIRT as $R+(1-p)(\Delta R_1(0)+\Delta R_2(0))$, with a similar definition for the average sum rate for global CSIRT. The average sum rate corresponds to a scenario where several codewords are transmitted over independent quasi-static bursty ICs. This, in turn, could be the case if a codeword spans several coherence blocks, but no coding is performed over these blocks. This is in contrast to the ergodic setup where coding is typically performed over different coherence blocks. By the law of large numbers, roughly a fraction of p codewords experiences interference, the remaining codewords are transmitted free of interference. Consequently, an opportunistic transmission strategy achieves the rate $pR + (1-p)(R + \Delta R_1(0) + \Delta R_2(0))$, which corresponds to the average sum rate. Our results demonstrate that, for local CSIR, the average sum capacity, obtained by maximizing the average sum rate over all achievable rate pairs $(R, \Delta R_1(0) + \Delta R_2(0))$, coincides with the achievable rates in the ergodic setup for all interference regions. In contrast, for local CSIRT, the average sum capacity is strictly smaller than the sum capacity in the ergodic setup. For global CSIRT, average sum capacity and sum capacity coincide for all interference regions when the interference states are fully correlated, and they coincide for VWI and WI when the interference states are independent. For global CSIRT, MI/SI, and independent interference states, the average sum capacity is smaller than the sum capacity in the ergodic setup. In general, the average sum capacity defined for the quasi-static setup never exceeds the sum capacity in the ergodic setup. This is perhaps not surprising if we recall that the average sum capacity corresponds to the case where no coding is performed over coherence blocks. Interestingly, the average sum capacity is not always achieved by maximizing the worst-case rate. For small values of p, it is beneficial to reduce the worst-case rate in order to achieve a larger opportunistic rate.

In our work we considered both the case where the interference states of the two users are independent and the case where the interference states are fully correlated. In both ergodic and quasi-static setups, the results for local CSIR are independent of the correlation between interference states. For other CSI levels, dependence between the interference states helps in all interference regions except very weak and very strong interference regions.

5

Bursty Noncoherent Wireless Networks

5.1 Introduction

The information-theoretical limits of wireless networks have mostly been studied under the assumption that the nodes have perfect knowledge of the fading coefficients. We made the same assumption in Chapter 4. For the fully-connected wireless interference channel (IC), it has been shown that if the nodes in the network have perfect knowledge of the fading coefficients then, irrespective of the number of users in the network, each user can achieve 1/2 degrees of freedom by using a transmission strategy called *interference alignment* [10]. However, it is *prima facie* unclear whether perfect knowledge of the fading coefficients can actually be obtained in practical systems.

In this chapter, we analyze the channel capacity of wireless networks when the nodes do not have this knowledge (noncoherent setting) and the interference is bursty. This chapter is along the lines of the work by Lozano, Heath, and Andrews [43], which demonstrates that in the absence of perfect knowledge of the channel coefficients realizations, and under some simplifying assumptions, the channel capacity for wireless networks is bounded in the signal-to-noise ratio (SNR). Specifically, the main results in [43] are based on the analysis of a block-fading channel that models the channel within a cluster and takes out-of-cluster interference into account. *Inter alia*, [43] considers a fully cooperative system, where all transmitters (Txs) and all receivers (Rxs) cooperate, resulting effectively in multiple-input multiple-output (MIMO) transmission. It is assumed that the number of Txs is greater than the number of time instants L over which the block-fading channel stays constant. This precludes an accurate channel estimation. For this scenario, Lozano *et al.* study the maximum achievable rate when the time-k channel input is of the form $\sqrt{\mathrm{SNR}}U_k$, where the distribution of U_k does not depend on the SNR. They demonstrate that, in the absence of perfect knowledge of the channel coefficients realizations, this achievable rate is bounded in the SNR, hence the transmission over such networks is highly power-inefficient.

However, one may argue that the rates achievable with inputs of the form $\sqrt{\mathrm{SNR}}U_k$ are bounded in the SNR because of the suboptimal input distribution. In fact, it has been demonstrated by Lapidoth and Moser [40, Th. 4.3] that for a memoryless channel and noncoherent setting, such inputs give rise to a bounded information rate also in the point-to-point case. In other words, for noncoherent, point-to-point, memoryless fading channels a more elaborate dependence between the input distribution and the SNR is necessary in order to achieve an unbounded information rate.[1] Since the block-fading channel specializes to the memoryless fading channel when $L = 1$, the observation that inputs of the form $\sqrt{\mathrm{SNR}}U_k$ yield a bounded information rate may perhaps not be surprising.

In this chapter, we explore whether the capacity of noncoherent wireless networks is bounded in the SNR if we allow the input distribution to change arbitrarily with the SNR and the interference is bursty. In contrast to the analysis by Lozano *et al.* [43], we assume interference burstiness and that the nodes do not cooperate. We further consider a memoryless flat-fading channel with an infinite

[1]However, in contrast to the case of perfect CSI, in its absence the capacity only grows double-logarithmically with the SNR [40, Th. 4.2].

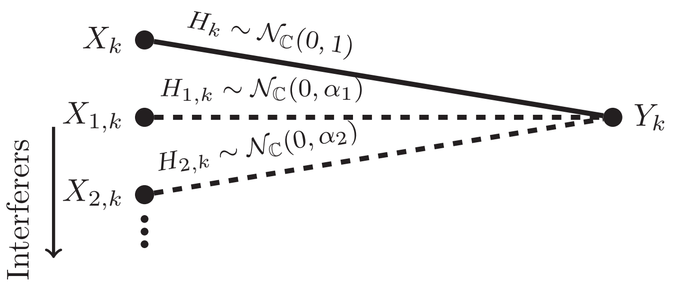


Figure 5.1: Channel model.

number of interferers. The locations of these interferers enter the channel model through the variance of the fading coefficients corresponding to the paths between the interferers and the intended receiver. Without loss of generality, we order the interferers with respect to the variances of the corresponding fading coefficients: the fading coefficient of the first interferer has the largest variance, denoted by α_1, the fading coefficient of the second interferer has the second-largest variance, denoted by α_2, and so on. We model the presence/absence of the corresponding interference links, as in Chapter 4, by an independent and identically distributed (IID) Bernoulli process. We consider a noncoherent scenario where Tx and Rx are cognizant of the statistics of the fading coefficients, but are ignorant of their realization. We further assume that the Rx knows perfectly the interference states. We demonstrate that the result by Lozano *et al.* continues to hold even if the input distribution is allowed to change arbitrarily with the SNR and even if we assume interference burstiness, provided that the variances $\{\alpha_\ell\}$ decay at most exponentially and all nodes use the same codebook.

5.2 Channel Model

A network consists of a number of users that are communicating with each other. For simplicity, we assume that the set of transmitting nodes and the set of receiving nodes are disjoint, and that they do not cooperate.

Since a characterization of all achievable rates in the network is unfeasible when the number of nodes is large, it is common to study the *sum-rate capacity* of the network. However, it is *prima facie* unclear whether a transmission strategy that achieves the sum-rate capacity is also practical. Indeed, it may well be that the optimal transmission strategy consists of turning off all but one of the transmitting nodes, thereby minimizing the interference. Such a strategy allows only one node to transmit its message and is probably not desirable in practice. In fact, practical constraints may demand that each node is offered roughly the same transmission rate. In order to enforce such a solution, one could study the sum-rate capacity of the network under the constraint that all transmitting nodes transmit at the same rate, but obtaining an expression for such a capacity seems again unfeasible. Alternatively, one may consider more elaborate rate allocation strategies, such as the proportional fair strategy [31], but these may also be difficult to analyze.

In this chapter, we simplify the original problem as follows: Firstly, we consider the case where one transmitting node communicates with one receiving node and the interfering nodes emit symbols that interfere with this communication. To model a large network, we assume that there are infinitely many interfering nodes. The presence of interference is modeled using an IID Bernoulli process B, i.e., $B \sim Ber(p)$, that indicates whether the interference links are present or not. We further assume that the interference states remain constant during the whole transmission and are known by the Rx. As performance measure we consider the capacity of the channel between the transmitting and receiving node. Secondly, to avoid transmission strategies for which the interfering nodes are turned off (which would, in fact, maximize the capacity), we assume that all nodes (transmitting and interfering) use the same codebook. This implies that each node is transmitting at the same rate, while at the same time it keeps the analysis tractable.

Note that the above simplifications permit a mathematical analysis of the channel capacity of the network, but they preclude strategies such as time-division multiple access (TDMA), where the nodes do not use the same codebook, but communicate nevertheless at the same rate.

We model the channel between the transmitting and receiving node by a discrete-time memoryless flat-fading channel whose complex-valued output Y_k at time $k \in \mathbb{Z}$ (where $\mathbb{Z}$ denotes the set of integers) corresponding to the time-k channel input X_k and the time-k interfering symbols $X_{\ell,k}$, $\ell = 1, 2, \ldots$ is given by

$$Y_k = H_k X_k + \sum_{\ell=1}^{\infty} B_\ell H_{\ell,k} X_{\ell,k} + Z_k. \tag{5.1}$$

In (5.1), Z_k models the time-k additive noise; H_k denotes the time-k fading coefficient of the channel between the Tx and Rx; $H_{\ell,k}$, $\ell = 1, 2, \dots$ denotes the time-k fading coefficient of the link between the ℓ-th interfering node and the receiver, and B_ℓ denotes the state of the ℓ interfering link; see Figure 5.1. We assume that $\{Z_k,\, k \in \mathbb{Z}\}$, $\{H_k,\, k \in \mathbb{Z}\}$, $\{H_{\ell,k},\, k \in \mathbb{Z}\}$, and $\{B_\ell,\ \ell = 1, 2, \dots\}$ are independent sequences of IID complex random variables. We further assume that $Z_k \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$, $H_k \sim \mathcal{N}_{\mathbb{C}}(0, 1)$, $H_{\ell,k} \sim \mathcal{N}_{\mathbb{C}}(0, \alpha_\ell)$ for some $\alpha_\ell > 0$, and $B_\ell \sim Ber(p)$. The variance α_ℓ is related to the path loss between interferer ℓ and the receiving node. Furthermore, we make the following assumptions:

A1) We consider a noncoherent scenario where Tx and Rx are cognizant of the statistics of the fading coefficients, but are ignorant of their realization. However, the states of the interfering links, i.e., $\{B_\ell,\ \ell = 1, 2, \dots\}$, are known at the Rx side.

A2) We assume that the interferers do neither cooperate with each other nor with the Tx and they all use the same codebook. Hence, $\{X_k,\, k \in \mathbb{Z}\}$ and $\{X_{\ell,k},\, k \in \mathbb{Z}\}$, $\ell = 1, 2, \dots$ are independent and follow the same input distribution Q^N.

A3) Without loss of generality (wlog), we assume that the interfering nodes are ordered according to the variances of the corresponding fading coefficients, i.e., $\alpha_\ell \geq \alpha_{\ell'}$ for any $\ell < \ell'$. We further assume that there exists a $0 < \rho < 1$ such that

$$\frac{\alpha_{\ell+1}}{\alpha_\ell} \geq \rho, \quad \ell = 1, 2, \dots \tag{5.2}$$

We finally assume that $\sum_{\ell=1}^{\infty} \alpha_\ell < \infty$. We believe that the assumption (5.2) is reasonably mild. For example, suppose that the path loss grows polynomially with the distance. Thus, (5.2) implies that the distance from the interferers to the receiver decays at most exponentially.

5.3 Channel Capacity and Main Result

We notice that $I(X_1^N; Y_1^N, B_1^\infty) = \lim_{L\to\infty} I(X_1^N; Y_1^N, B_1^L)$.

We define the capacity of the above channel (5.1) as[2]

$$C(\mathsf{P}) \triangleq \varlimsup_{N\to\infty} \frac{1}{N} \sup_{Q^N} \varlimsup_{L\to\infty} I(X_1^N; Y_1^N, B_1^L) \tag{5.3}$$

where we assume that the sequences X_1^N and $X_{\ell,1}^N$, $\ell = 1, 2, \ldots$ are independent, and that each such sequence has distribution Q^N. Furthermore, the input sequences are independent of the interference state B_1^L. The channel interference state is known at the receiver and it is considered as an additional channel output [11]. The supremum in (5.3) is over all N-dimensional probability distributions Q^N satisfying

$$\frac{1}{N} \sum_{k=1}^{N} \int |x_k|^2 dQ^N(x_1^N) \leq \mathsf{P}. \tag{5.4}$$

The SNR is defined as

$$\mathrm{SNR} \triangleq \frac{\mathsf{P}}{\sigma^2}. \tag{5.5}$$

By Fano's inequality [14, Sec. 7.9], any encoding and decoding scheme with a rate above $C(\mathsf{P})$ has a decoding error probability that is bounded away from zero as n tends to infinity. By demonstrating that $C(\mathsf{P})$ is bounded in P, we therefore demonstrate that there exists no encoding and decoding scheme that has a rate that tends to infinity as $\mathsf{P} \to \infty$ and for which the decoding error probability vanishes as N tends to infinity.

Our main result is the following.

Theorem 21 (Upper bound for the bursty case) *Consider the channel model introduced in Section 5.2. For every $\mathsf{P} > 0$ and $B_\ell \sim Ber(p)$, $\ell = 1, 2, \ldots$ and $0 \leq p \leq 1$, the channel capacity is upper-bounded by*

$$C(\mathsf{P}) \leq \frac{1-p}{p} \log\left(\rho^{-\frac{3}{2}}\right) + \frac{1}{2}\log(\eta_{\max}) + \log(1+\eta_{\max}) + \log\frac{\pi}{e} \tag{5.6}$$

where $\eta_{\max}$ is defined as

$$\eta_{\max} \triangleq \max\left(\frac{1}{\alpha_1}, \frac{1}{\rho}\right). \tag{5.7}$$

[2] The logarithms used in this chapter are natural logarithms. The capacity has thus the dimension "nats per channel use".

Proof: See Section 5.4. ■

For the non-bursty case, i.e., $p = 1$ the upper bound (5.6) becomes

$$C(\mathsf{P}) \leq \frac{1}{2}\log\eta_{\max} + \log\left(1+\eta_{\max}\right) + \log\frac{\pi}{e} \tag{5.8}$$

We presented this result in [64, Section V].

Remark 14 *The upper bound* (5.6) *depends on* η_{max}*, which in turn depends on* ρ *given by* (5.2)*. One may wonder whether ordering the interfering nodes differently (i.e., not according to the value of* α_ℓ*) would give rise to a larger* ρ *satisfying* (5.2) *and therefore to a tighter upper bound on* $C(\mathsf{P})$*. However, this is not the case. It can be shown that the ordering used in this chapter yields the largest* ρ*, see Appendix C.*

5.4 Proof of Theorem 21

To obtain (5.6), we begin by deriving an upper bound on the mutual information $I(X_1^N; Y_1^N, B_1^L)$ as follows.

$$\begin{aligned} I(X_1^N; Y_1^N, B_1^L) &= I(X_1^N; Y_1^N | B_1^L) \\ &= h(Y_1^N | B_1^L) - h(Y_1^N | X_1^N, B_1^L) \\ &\leq h(Y_1^N | B_1^L) - h(Y_1^N | X_1^N, H_1^N, B_1^L) \\ &= h(Y_1^N | B_1^L) - h(Y_1^N - H_1^N X_1^N | B_1^L) \end{aligned} \tag{5.9}$$

where the first step follows because X_1^N and B_1^L are independent. The inequality follows because conditioning reduces entropy.

For $\mathbf{B} = \mathbf{b} = [b_1, \ldots, b_L]$, we define the random variables

$$\begin{aligned} Y_k(\mathbf{b}) &= H_k X_k + \sum_{\ell=1}^{L} b_\ell H_{\ell,k} X_{\ell,k} + \sum_{\ell=L+1}^{\infty} B_\ell H_{\ell,k} X_{\ell,k} + Z_k \\ &= \sum_{\ell=0}^{L} b_\ell H_{\ell,k} X_{\ell,k} + \sum_{\ell=L+1}^{\infty} B_\ell H_{\ell,k} X_{\ell,k} + Z_k \end{aligned} \tag{5.10}$$

where, for compactness, we let $b_0 = 1$, and $X_{0,k} = X_k$, and

$$\hat{Y}_k(\mathbf{b}) \triangleq Y_k(\mathbf{b}) - H_k X_k = \sum_{\ell=1}^{L} b_\ell H_{\ell,k} X_{\ell,k} + \sum_{\ell=L+1}^{\infty} B_\ell H_{\ell,k} X_{\ell,k} + Z_k. \tag{5.11}$$

Using these definitions, we rewrite (5.9) as

$$I(X_1^N; Y_1^N, B_1^L) \leq \sum_{\mathbf{b}\in\mathcal{B}_L} \Pr\{\mathbf{B}=\mathbf{b}\} h\big(Y_1^N(\mathbf{b})\big) - \sum_{\tilde{\mathbf{b}}\in\mathcal{B}_L} \Pr\{\mathbf{B}=\tilde{\mathbf{b}}\} h(\hat{Y}_1^N(\tilde{\mathbf{b}})). \quad (5.12)$$

where $\mathcal{B}_L \triangleq \{0,1\}^L$ denotes the set of all binary sequences of length L. We consider a partition of $\mathcal{B}_L$ based on the position of the leading 1 in each sequence. In particular, for $m = 1, \ldots, L+1$, we define

$$\mathcal{B}_L(m) = \begin{cases} \big\{\mathbf{b} : b_1^m = [0_1^{m-1}, 1]\big\}, & 1 \leq m \leq L, \\ \{0_1^L\}, & m = L+1. \end{cases} \quad (5.13)$$

In words, $\mathcal{B}_L(m)$ is the set of all sequences of length L whose leading 1 is in the m-th position. The sets $\mathcal{B}_L(m)$, $m = 1, \ldots, L+1$ are disjoint and define a partition of $\mathcal{B}_L$, i.e., $\mathcal{B}_L(m) \cap \mathcal{B}_L(m') = \emptyset$, $m \neq m'$, and $\bigcup_{m=1}^{L+1} \mathcal{B}_L(m) = \mathcal{B}_L$.

To upper-bound (5.12), we will pair the sequences $\mathbf{b}$ and $\tilde{\mathbf{b}}$ according to the mapping described in the next proposition.

Proposition 1 *There exists a one-to-one and onto mapping*

$$f_L : \mathcal{B}_L \quad \rightarrow \quad \mathcal{B}_L,\ \mathbf{b} \longmapsto \tilde{\mathbf{b}}, \quad (5.14)$$

such that $\mathbf{b} \in \mathcal{B}_L \mapsto \tilde{\mathbf{b}} = [0_1^{m-1}, 1, b_1^{L-m}] \in \mathcal{B}_L(m)$, *and* $\|\mathbf{b}\|_1 = \|\tilde{\mathbf{b}}\|_1$.

Proof: We consider the Algorithm 1, which is described below

Data: Binary sequence $\mathbf{b}$ of length L with Hamming weight $\|\mathbf{b}\|_1 = W$
Result: Binary sequence $\tilde{\mathbf{b}} = [0_1^{m-1}, 1, b_1^{L-m}]$ of length L and Hamming weight $\|\tilde{\mathbf{b}}\|_1 = W$.
if $\mathbf{b} = 0_1^L$ **then**
  $\tilde{\mathbf{b}} = \mathbf{b}$
else
  $i \leftarrow$ take the position of the right-most 1 in sequence $\mathbf{b}$
  $m \leftarrow L - i + 1$ (length of b_i^L)
  $\tilde{\mathbf{b}} = [0_1^{m-1}, 1, b_1^{L-m}]$
end

Algorithm 1: Mapping between binary sequences $\mathbf{b}$ and $\tilde{\mathbf{b}}$.

It generates a vector $\tilde{\mathbf{b}}$ such that the conditions in the proposition hold. In particular, the output of Algorithm 1 satisfies $\tilde{\mathbf{b}} = [0_1^{m-1}, 1, b_1^{L-m}]$ by construction. The number of ones in the sequences $\mathbf{b}$ and $\tilde{\mathbf{b}}$ is the same, since the proposed algorithm just reorders the different positions of the original vector to generate the output vector. Finally, note that the Algorithm 2 recovers the original sequence $\mathbf{b}$ from the corresponding $\tilde{\mathbf{b}}$ for any $\mathbf{b} \in \mathcal{B}_L$.

Data: Binary sequence $\tilde{\mathbf{b}}$ of length L with Hamming weight $\|\tilde{\mathbf{b}}\|_1 = W$
Result: Binary sequence $\mathbf{b} = [b_{m+1}^L, 1, 0_1^{m-1}]$ of length L and Hamming weight $\|\mathbf{b}\|_1 = W$.
if $\tilde{\mathbf{b}} = 0_1^L$ **then**
  $\mathbf{b} = \tilde{\mathbf{b}}$
else
  $i \leftarrow$ take the position of the left-most 1 in sequence $\mathbf{b}$
  $m \leftarrow i$: (length of b_1^i)
  $\mathbf{b} = [\tilde{b}_{m+1}^L, 1, 0_1^{m-1}]$
end

Algorithm 2: Mapping between binary sequences $\tilde{\mathbf{b}}$ and $\mathbf{b}$.

Thus, Algorithms 1 and 2 show that the correspondence of both sequences $\mathbf{b}$ and $\tilde{\mathbf{b}}$ is one-to-one and onto. ∎

We apply the mapping described in Proposition 1 to pair up $\mathbf{b}$ and $\tilde{\mathbf{b}}$ in (5.12). In particular, since $\|\mathbf{b}\|_1 = \|\tilde{\mathbf{b}}\|_1$ it follows that $\Pr\{\mathbf{B} = \mathbf{b}\} = \Pr\{\mathbf{B} = \tilde{\mathbf{b}}\}$ and (5.12) becomes

$$\begin{aligned} I(X_1^N; Y_1^N, B_1^L) &\\ &\leq \sum_{\mathbf{b}} \Pr\{\mathbf{B} = \mathbf{b}\} \left[h(Y_1^N(\mathbf{b})) - h(\hat{Y}_1^N(f_L(\mathbf{b}))) \right] \\ &= \sum_{m=1}^{L} \sum_{\mathbf{b}: f_L(\mathbf{b}) \in \mathcal{B}_L(m)} \Pr\{\mathbf{B} = \mathbf{b}\} \left[h(Y_1^N(\mathbf{b})) - h(\hat{Y}_1^N(f_L(\mathbf{b}))) \right]. \end{aligned} \tag{5.15}$$

We now focus on the bracketed term of (5.15), to this end we first introduce Lemma 1.

Lemma 1 *Let f and g be arbitrary probability density function (pdf). If $-\int f(x) \log f(x) dx$ is finite, then $-\int f(x) \log g(x) dx$ exists and*

$$-\int f(x) \log f(x) \mathrm{d}x \leq -\int f(x) \log g(x) \mathrm{d}x. \tag{5.16}$$

Proof: See [3, Lemma 8.3.1]. Inequality (5.16) is a consequence of the nonnegativity of the relative entropy between f and g. ■

Based on Lemma 1, we obtain the following upper bound on (5.15).

$$\begin{aligned}
&h(Y_1^N(\mathbf{b})) - h(\hat{Y}_1^N(f_L(\mathbf{b}))) \\
&\quad \leq N\left[(m-1)\log\left(\rho^{-\frac{3}{2}}\right) + \frac{1}{2}\log(\eta_{\max}) + \log(1+\eta_{\max}) + \log\frac{\pi}{e}\right] \\
&\qquad + N\sum_{\ell=L-m+1}^{L}\alpha_\ell \mathsf{P}(b_\ell+p)\left(\frac{2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right) \\
&\qquad + N\sum_{\ell=L+1}^{\infty}\alpha_\ell p\mathsf{P}\left(\frac{4}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right) \qquad (5.17)
\end{aligned}$$

where $\eta_{\max}$ is defined in (5.7).

Next, we prove (5.17). To this end, we first define the random variable

$$\begin{aligned}
\tilde{Y}_k(\mathbf{b},m) &= \tilde{H}_{m,k}X_{0,k} + \sum_{\ell=1}^{L-m} b_\ell \tilde{H}_{\ell+m,k}X_{\ell,k} + \sum_{\ell=L-m+1}^{\infty}\tilde{B}_\ell\tilde{H}_{\ell+m,k}X_{\ell,k} + \tilde{Z}_k \\
&= \sum_{\ell=0}^{L-m} b_\ell \tilde{H}_{\ell+m,k}X_{\ell,k} + \sum_{\ell=L-m+1}^{\infty}\tilde{B}_\ell\tilde{H}_{\ell+m,k}X_{\ell,k} + \tilde{Z}_k \qquad (5.18)
\end{aligned}$$

where for every $\ell = 1,2,\dots$ the fading coefficients $\{\tilde{H}_{\ell,k}, k\in\mathbb{N}\}$ have the same distribution as $\{H_{\ell,k}, k\in\mathbb{N}\}$ in (5.11) but are independent of $\{H_{\ell,k},\ k\in\mathbb{N}\}$. Likewise, the additive noise terms $\{\tilde{Z}_k, k\in\mathbb{N}\}$ have the same distribution as $\{Z_k, k\in\mathbb{N}\}$ but are independent of $\{Z_k, k\in\mathbb{N}\}$. Since $X_{0,1}^N$ and $X_{\ell,1}^N$, $\ell=1,2,\dots$ have the same distribution by assumption A2), we conclude that $\tilde{Y}_k(\mathbf{b},m)$ and $\hat{Y}_k(\tilde{\mathbf{b}})$ have the same distribution for $\tilde{\mathbf{b}}\in\mathcal{B}_L(m)$. Hence,

$$\begin{aligned}
&h(Y_1^N(\mathbf{b})) - h(\hat{Y}_1^N(f_L(\mathbf{b}))) \\
&\quad = h\big(Y_1^N(\mathbf{b})\big) - h(\tilde{Y}_1^N(\mathbf{b},m)) \\
&\quad = h\big(Y_1^N(\mathbf{b})|\tilde{Y}_1^N(\mathbf{b},m)\big) - h(\tilde{Y}_1^N(\mathbf{b},m)|Y_1^N(\mathbf{b})) \\
&\quad \leq \sum_{k=1}^{N}[h(Y_k(\mathbf{b})|\tilde{Y}_k(\mathbf{b},m)) - h(\tilde{Y}_k(\mathbf{b},m)|\tilde{Y}_1^{k-1}(\mathbf{b},m), Y_1^N(\mathbf{b}))] \qquad (5.19)
\end{aligned}$$

where in the second step we use the identity $h(A)-h(B) = h(A|B)-h(B|A)$. The last step follows by applying the chain rule for entropy and because conditioning reduces entropy. To find an upper bound on (5.19), we first upper-bound the conditional differential entropy $h(Y_k(\mathbf{b})|\tilde{Y}_k(\mathbf{b},m))$ by applying Lemma 1. Let $f_{Y_k|\tilde{Y}_k}$ denote the true conditional pdf of $Y_k(\mathbf{b})$ given $\tilde{Y}_k(\mathbf{b},m)$. Lemma 1 allows us to upper-bound the conditional differential entropy of $Y_k(\mathbf{b})$ given $\tilde{Y}_k(\mathbf{b},m)$ by replacing $f_{Y_k|\tilde{Y}_k}$ by an auxiliary pdf $g_{Y_k|\tilde{Y}_k}$. For any given $\tilde{Y}_k(\mathbf{b},m)=\tilde{y}_k$, we choose

$$g_{Y_k|\tilde{Y}_k}(y_k|\tilde{y}_k)=\frac{\sqrt{\beta}}{\pi^2|y_k|}\frac{1}{1+\beta|y_k|^2}, \quad y_k\in\mathbb{C} \tag{5.20}$$

with $\beta=1/|\tilde{y}_k|^2$. This is the density of a circularly-symmetric complex random variable whose magnitude is Cauchy distributed. A similar pdf has been used by Koch and Lapidoth [35] to obtain their result for frequency-selective fading channels and in [64] to obtain the upper bound on the capacity of the non-bursty case of the channel studied in this chapter.

Using (5.20) in (5.16), we obtain that

$$\begin{aligned} h(Y_k(\mathbf{b})|\tilde{Y}_k(\mathbf{b},m)) \le{}& \frac{1}{2}\mathsf{E}\left[\log|Y_k(\mathbf{b})|^2\right]+2\log\pi+\frac{1}{2}\mathsf{E}\left[\log|\tilde{Y}_k(\mathbf{b},m)|^2\right]\\ &+\mathsf{E}\left[\log\left(1+\frac{|Y_k(\mathbf{b})|^2}{|\tilde{Y}_k(\mathbf{b},m)|^2}\right)\right] \end{aligned} \tag{5.21}$$

Next, we consider the second term in (5.19). By conditioning on $\{X_{\ell,k}\}_{\ell=1}^{\infty}$ and $\{\tilde{B}_\ell\}_{\ell=L-m+1}^{\infty}$, the random variable $\tilde{Y}_k(\mathbf{b},m)$ is independent of $(\tilde{Y}_1^{k-1}(\mathbf{b},m),Y_1^N)$ and has a Gaussian distribution. Hence,

$$\begin{aligned} &h(\tilde{Y}_k(\mathbf{b},m)|\tilde{Y}_1^{k-1}(\mathbf{b},m),Y_1^N)\\ &\ge h(\tilde{Y}_k(\mathbf{b},m)|\{X_{\ell,k}\}_{\ell=1}^{\infty},\{\tilde{B}_\ell\}_{\ell=L-m+1}^{\infty})\\ &=\log(\pi e)\\ &\quad+\mathsf{E}\left[\log\left(\sum_{\ell=0}^{L-m}b_\ell\alpha_{\ell+m}|X_{\ell,k}|^2+\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_{\ell+m}|X_{\ell,k}|^2+\sigma^2\right)\right]. \end{aligned} \tag{5.22}$$

Using (5.21) and (5.22) in (5.19) we have that

$$\begin{aligned} &h\big(Y_1^N(\mathbf{b})\big)-h(\tilde{Y}_1^N(\mathbf{b},m))\\ &\le\frac{1}{2}\mathsf{E}\left[\log|Y_k(\mathbf{b})|^2\right]-\frac{1}{2}\mathsf{E}\left[\log|\tilde{Y}_k(\mathbf{b},m)|^2\right]+\log\frac{\pi}{e}\\ &\quad+\mathsf{E}\left[\log\left(|Y_k(\mathbf{b})|^2+|\tilde{Y}_k(\mathbf{b},m)|^2\right)\right] \end{aligned}$$

$$-\,\mathsf{E}\left[\log\left(\sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m}|X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_{\ell+m}|X_{\ell,k}|^2 + \sigma^2\right)\right]. \quad (5.23)$$

For the fourth term in (5.23), Jensen's inequality yields

$$\begin{aligned}
&\mathsf{E}\left[\log\left(|Y_k(\mathbf{b})|^2 + |\tilde{Y}_k(\mathbf{b},m))|^2\right)\right] \\
&\le \mathsf{E}\left[\log\left(\mathsf{E}\left[|Y_k(\mathbf{b})|^2 + |\tilde{Y}_k(\mathbf{b},m)|^2 \,\middle|\, \{X_{\ell,k}\}_{\ell=0}^{\infty}, \{B_\ell, \tilde{B}_\ell\}_{\ell=L-m+1}^{\infty}\right]\right)\right] \\
&= \mathsf{E}\left[\log\left(\sum_{\ell=0}^{L-m} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{L} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty} |B_\ell|^2 \alpha_\ell |X_{\ell,k}|^2 + \sigma^2\right.\right. \\
&\qquad\left.\left. + \sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m}|X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_{\ell+m}|X_{\ell,k}|^2 + \sigma^2\right)\right] \quad (5.24)
\end{aligned}$$

where in the last step we used (5.10) and (5.18), and we define $\alpha_0 \triangleq 1, b_0 = 1$ and $X_{0,k} = X_k$ to stress the equivalence between the two summations. Next, in (5.24) we add and subtract $\sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_\ell |X_{\ell,k}|^2$ to obtain

$$\begin{aligned}
&\mathsf{E}\left[\log\left(|Y_k(\mathbf{b})|^2 + |\tilde{Y}_k(\mathbf{b},m))|^2\right)\right] \\
&\qquad \le \mathsf{E}\left[\log\left(\sum_{\ell=0}^{L-m} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_\ell |X_{\ell,k}|^2 + \sigma^2\right.\right. \\
&\qquad\qquad + \sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m}|X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_{\ell+m}|X_{\ell,k}|^2 + \sigma^2 \\
&\qquad\qquad + \sum_{\ell=L-m+1}^{L} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty} |B_\ell|^2 \alpha_\ell |X_{\ell,k}|^2 \\
&\qquad\qquad \left.\left. - \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_\ell |X_{\ell,k}|^2\right)\right]. \quad (5.25)
\end{aligned}$$

We next recall assumption (5.2), namely that

$$\frac{\alpha_{\ell+1}}{\alpha_\ell} \ge \rho, \quad \ell = 1, 2, \ldots \quad (5.26)$$

for some $0 < \rho < 1$. Since $\alpha_0 = 1$ and the condition $\frac{\alpha_{\ell+1}}{\alpha_\ell} \ge \rho$ may not hold for $\ell = 0$, we define

$$\eta_{\max} \triangleq \max\left(\frac{1}{\alpha_1}, \frac{1}{\rho}\right). \quad (5.27)$$

It then follows that $\alpha_\ell \leq \eta_{\max}\alpha_{\ell+1}$ for $\ell = 0, 1, \ldots$, and using (5.26) iteratively we conclude that

$$\alpha_\ell \leq \frac{\eta_{\max}}{\rho^{m-1}}\alpha_{\ell+m}, \quad \ell = 0, 1, \ldots \tag{5.28}$$

Applying (5.28) to the first two terms in (5.25) gives

$$\begin{aligned}
&\sum_{\ell=0}^{L-m} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_\ell |X_{\ell,k}|^2 \\
&\qquad \leq \frac{\eta_{\max}}{\rho^{m-1}} \left(\sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m} |X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_{\ell+m} |X_{\ell,k}|^2 \right).
\end{aligned} \tag{5.29}$$

Then, using (5.29) and that $\frac{\eta_{\max}}{\rho^{m-1}} \geq 1$, in (5.24), we obtain

$$\begin{aligned}
&\mathsf{E}\left[\log\left(|Y_k(\mathbf{b})|^2 + |\tilde{Y}_k(\mathbf{b}, m))|^2\right)\right] \\
&\leq \mathsf{E}\left[\log\left(\left(1 + \frac{\eta_{\max}}{\rho^{m-1}}\right)\left(\sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m}|X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_{\ell+m}|X_{\ell,k}|^2 + \sigma^2\right)\right.\right. \\
&\qquad\qquad + \sum_{\ell=L-m+1}^{L} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty} |B_\ell|^2 \alpha_\ell |X_{\ell,k}|^2 \\
&\qquad\qquad \left.\left. - \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_\ell |X_{\ell,k}|^2 \right)\right].
\end{aligned} \tag{5.30}$$

By applying the identity $\log(A+B) = \log A + \log\left(1 + \frac{B}{A}\right)$, by upper-bounding $\log\left(1 + \frac{B}{A}\right)$ by $\frac{B}{A}$, where in our case we have

$$A \triangleq \left(1 + \frac{\eta_{\max}}{\rho^{m-1}}\right)\left(\sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m}|X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_{\ell+m}|X_{\ell,k}|^2 + \sigma^2\right)$$

and

$$B \triangleq \sum_{\ell=L-m+1}^{L} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty} |B_\ell|^2 \alpha_\ell |X_{\ell,k}|^2 - \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_\ell |X_{\ell,k}|^2$$

and by defining

$$D \triangleq \sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m}|X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_{\ell+m}|X_{\ell,k}|^2 + \sigma^2$$

in (5.30), we obtain

$$
\begin{aligned}
&\mathsf{E}\left[\log\left(|Y_k(\mathbf{b})|^2+|\tilde{Y}_k(\mathbf{b},m))|^2\right)\right]\\
&\leq \mathsf{E}\Bigg[\log\Bigg(\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)\Bigg(\sum_{\ell=0}^{L-m} b_\ell\alpha_{\ell+m}|X_{\ell,k}|^2\\
&\qquad\qquad\qquad\qquad + \sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_{\ell+m}|X_{\ell,k}|^2+\sigma^2\Bigg)\Bigg)\Bigg]\\
&\quad + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{L} b_\ell\alpha_\ell|X_{\ell,k}|^2+\sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)D}\right]\\
&\quad - \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)D}\right]\\
&\leq \mathsf{E}\Bigg[\log\Bigg(\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)\Bigg(\sum_{\ell=0}^{L-m} b_\ell\alpha_{\ell+m}|X_{\ell,k}|^2\\
&\qquad\qquad\qquad\qquad + \sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_{\ell+m}|X_{\ell,k}|^2+\sigma^2\Bigg)\Bigg)\Bigg]\\
&\quad + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{L} b_\ell\alpha_\ell|X_{\ell,k}|^2+\sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)D}\right]\\
&\quad + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)D}\right]\\
&\leq \mathsf{E}\Bigg[\log\Bigg(\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right)\Bigg(\sum_{\ell=0}^{L-m} b_\ell\alpha_{\ell+m}|X_{\ell,k}|^2\\
&\qquad\qquad\qquad\qquad + \sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_{\ell+m}|X_{\ell,k}|^2+\sigma^2\Bigg)\Bigg)\Bigg]\\
&\quad + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{L} b_\ell\alpha_\ell|X_{\ell,k}|^2+\sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right]\\
&\quad + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right]
\end{aligned}
\tag{5.31}
$$

where the second step follows by changing the sign of third expected value, and the last step follows because we lower-bound $D \geq \sigma^2$ and $1 + \frac{\eta_{\max}}{\rho^{m-1}} \geq \frac{\eta_{\max}}{\rho^{m-1}}$.

Combining (5.31) and (5.23) yields

$$
\begin{aligned}
&h(Y_1^N(\mathbf{b})) - h(\tilde{Y}_1^N(\mathbf{b},m)) \\
&\quad \leq \sum_{k=1}^{N} \Bigg[\frac{1}{2}\mathsf{E}\Big[\log|Y_k(\mathbf{b})|^2\Big] - \frac{1}{2}\mathsf{E}\Big[\log|\tilde{Y}_k(\mathbf{b},m)|^2\Big] + \log\left(\frac{\eta_{\max}}{\rho^{m-1}}\right) + \log\frac{\pi}{e} \\
&\qquad\qquad + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{L} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty} |B_\ell|^2 \alpha_\ell |X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right] \\
&\qquad\qquad + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_\ell |X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right]\Bigg]. && (5.32)
\end{aligned}
$$

To upper-bound the first two terms on the right-hand side (RHS) of (5.32), we note that, conditioned on $X_{\ell,k} = x_\ell$, $\ell = 0, 1, \ldots$, both $|Y_k(\mathbf{b})|^2$ and $|\tilde{Y}_k(\mathbf{b},m)|^2$ have an exponential distribution with mean

$$\sum_{\ell=0}^{L} b_\ell \alpha_\ell |x_\ell|^2 + \sum_{L+1}^{\infty} |b_\ell|^2 \alpha_\ell |x_\ell|^2 + \sigma^2$$

and

$$\sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m} |x_\ell|^2 + \sum_{L-m+1}^{\infty} |\tilde{b}_\ell|^2 \alpha_\ell |x_\ell|^2 + \sigma^2,$$

respectively. Consequently, by [25, p. 571, 4.331.1]

$$
\begin{aligned}
&\mathsf{E}\Big[\log|Y_k(\mathbf{b})|^2\Big] \\
&= \mathsf{E}\left[\log\left(\sum_{\ell=0}^{L} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty} |B_\ell|^2 \alpha_\ell |X_{\ell,k}|^2 + \sigma^2\right)\right] - \gamma, && (5.33) \\
&\mathsf{E}\Big[\log|\tilde{Y}_k(\mathbf{b},m)|^2\Big] \\
&= \mathsf{E}\left[\log\left(\sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m} |X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_{\ell+m} |X_{\ell,k}|^2 + \sigma^2\right)\right] - \gamma, && (5.34)
\end{aligned}
$$

where $\gamma \approx 0.577$ denotes the Euler-Mascheroni constant. It follows that

$$
\begin{aligned}
&\mathsf{E}\Big[\log|Y_k(\mathbf{b})|^2\Big] - \mathsf{E}\Big[\log|\tilde{Y}_k(\mathbf{b},m)|^2\Big] \\
&= \mathsf{E}\left[\log\frac{\sum_{\ell=0}^{L} b_\ell \alpha_\ell |X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty} |B_\ell|^2 \alpha_\ell |X_{\ell,k}|^2 + \sigma^2}{\sum_{\ell=0}^{L-m} b_\ell \alpha_{\ell+m} |X_{\ell,k}|^2 + \sum_{\ell=L-m+1}^{\infty} |\tilde{B}_\ell|^2 \alpha_{\ell+m} |X_{\ell,k}|^2 + \sigma^2}\right]. && (5.35)
\end{aligned}
$$

To upper-bound (5.35) we add and subtract $\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2$. We further recall that $D=\sum_{\ell=0}^{L-m}b_\ell\alpha_{\ell+m}|X_{\ell,k}|^2+\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_{\ell+m}|X_{\ell,k}|^2+\sigma^2$. This yields

$$\begin{aligned}
&\mathsf{E}\left[\log|Y_k(\mathbf{b})|^2\right]-\mathsf{E}\left[\log|\tilde{Y}_k(\mathbf{b},m)|^2\right]\\
&\quad=\mathsf{E}\left[\log\left(\frac{\sum_{\ell=0}^{L-m}b_\ell\alpha_\ell|X_{\ell,k}|^2+\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2+\sigma^2}{D}\right.\right.\\
&\qquad\qquad+\frac{\sum_{\ell=L-m+1}^{L}b_\ell\alpha_\ell|X_{\ell,k}|^2+\sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{D}\\
&\qquad\qquad\left.\left.-\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{D}\right)\right].
\end{aligned}\tag{5.36}$$

Then, by using (5.29) and that $\frac{\eta_{\max}}{\rho^{m-1}}\geq 1$, we obtain

$$\begin{aligned}
&\mathsf{E}\left[\log|Y_k(\mathbf{b})|^2\right]-\mathsf{E}\left[\log|\tilde{Y}_k(\mathbf{b},m)|^2\right]\\
&\leq\mathsf{E}\left[\log\left(\frac{\eta_{\max}}{\rho^{m-1}}+\frac{\sum_{\ell=L-m+1}^{L}b_\ell\alpha_\ell|X_{\ell,k}|^2+\sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{D}\right.\right.\\
&\qquad\left.\left.-\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{D}\right)\right].
\end{aligned}\tag{5.37}$$

By applying the identity $\log(A+B)=\log A+\log\left(1+\frac{B}{A}\right)$ and by upper-bounding $\log\left(1+\frac{B}{A}\right)$ by $\frac{B}{A}$, where in this case

$$A\triangleq\frac{\eta_{\max}}{\rho^{m-1}}$$

and

$$\begin{aligned}
B\ \triangleq\ &\frac{\sum_{\ell=L-m+1}^{L}b_\ell\alpha_\ell|X_{\ell,k}|^2+\sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{D}\\
&-\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{D}
\end{aligned}$$

this can be upper-bounded as

$$\begin{aligned}
&\mathsf{E}\left[\log|Y_k(\mathbf{b})|^2\right]-\mathsf{E}\left[\log|\tilde{Y}_k(\mathbf{b},m)|^2\right]\\
&\leq\log\left(\frac{\eta_{\max}}{\rho^{m-1}}\right)+\mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{L}b_\ell\alpha_\ell|X_{\ell,k}|^2+\sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}D}\right]
\end{aligned}$$

$$
\begin{aligned}
&- \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}D}\right] \\
&\leq \log\left(\frac{\eta_{\max}}{\rho^{m-1}}\right) + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{L} b_\ell\alpha_\ell|X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}D}\right] \\
&\quad + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}D}\right] \\
&\leq \log\left(\frac{\eta_{\max}}{\rho^{m-1}}\right) + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{L} b_\ell\alpha_\ell|X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right] \\
&\quad + \mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right]
\end{aligned} \tag{5.38}
$$

where the second step follows because by changing the sign of third expected value, and the last step follows because we lower-bound the denominator by $\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2$.

Combining (5.38) with (5.32) yields

$$
\begin{aligned}
&h\big(Y_1^N(\mathbf{b})\big) - h(\tilde{Y}_1^N(\mathbf{b},m)) \\
&\quad\leq \sum_{k=1}^{N}\Bigg[\frac{1}{2}\log\left(\frac{\eta_{\max}}{\rho^{m-1}}\right) + \log\left(1+\frac{\eta_{\max}}{\rho^{m-1}}\right) + \log\frac{\pi}{e} \\
&\qquad\qquad + 2\mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{L} b_\ell\alpha_\ell|X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right] \\
&\qquad\qquad + 2\mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right]\Bigg].
\end{aligned} \tag{5.39}
$$

It remains to analyze the last two terms, i.e.,

$$
\begin{aligned}
S &= 2\sum_{k=1}^{N}\mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{L} b_\ell\alpha_\ell|X_{\ell,k}|^2 + \sum_{\ell=L+1}^{\infty}|B_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right] \\
&\quad + 2\sum_{k=1}^{N}\mathsf{E}\left[\frac{\sum_{\ell=L-m+1}^{\infty}|\tilde{B}_\ell|^2\alpha_\ell|X_{\ell,k}|^2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right].
\end{aligned} \tag{5.40}
$$

By the linearity of expectation, the power constraint (5.4), and $\mathsf{E}\left[|B_\ell|^2\right] = \mathsf{E}\left[|\tilde{B}_\ell|^2\right] = p$, S can be upper-bounded as

$$
S \leq 2N\frac{\sum_{\ell=L-m+1}^{L} b_\ell\alpha_\ell\mathsf{P} + \sum_{\ell=L+1}^{\infty} p\alpha_\ell\mathsf{P}}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}
$$

$$+\,2N\frac{\sum_{\ell=L-m+1}^{L} p\alpha_\ell \mathsf{P} + \sum_{\ell=L+1}^{\infty} p\alpha_\ell \mathsf{P}}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}. \tag{5.41}$$

After simplifying the terms, we obtain

$$S \le N\sum_{\ell=L-m+1}^{L}\alpha_\ell \mathsf{P}(b_\ell+p)\left(\frac{2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right) + N\sum_{\ell=L+1}^{\infty}\alpha_\ell p\mathsf{P}\left(\frac{4}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right). \tag{5.42}$$

from which we obtain the upper bound

$$\begin{aligned}
&h(Y_1^N(\mathbf{b})) - h(\tilde{Y}_1^N(\mathbf{b},m))\\
&\le N\left[(m-1)\log\left(\rho^{-\frac{3}{2}}\right) + \frac{1}{2}\log(\eta_{\max}) + \log(1+\eta_{\max}) + \log\frac{\pi}{e}\right]\\
&\quad + N\sum_{\ell=L-m+1}^{L}\alpha_\ell \mathsf{P}(b_\ell+p)\left(\frac{2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right)\\
&\quad + N\sum_{\ell=L+1}^{\infty}\alpha_\ell p\mathsf{P}\left(\frac{4}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right)
\end{aligned} \tag{5.43}$$

where we used that $\log\bigl(1+\frac{\eta_{\max}}{\rho^{m-1}}\bigr) = \log\bigl(\frac{\rho^{m-1}+\eta_{\max}}{\rho^{m-1}}\bigr) \le \log\bigl(\frac{1+\eta_{\max}}{\rho^{m-1}}\bigr)$. This proves (5.17).

Back to (5.15), by using the upper bound (5.17), we obtain for any pair $(\mathbf{b},\tilde{\mathbf{b}})$ satisfying Proposition 1

$$\begin{aligned}
&\frac{1}{N}I(X_1^N;Y_1^N,B_1^L)\\
&\le \sum_{m=1}^{L}\sum_{\mathbf{b}:f_L(\mathbf{b})\in\mathcal{B}_L(m)} \Pr\{\mathbf{B}=\mathbf{b}\}\Biggl[(m-1)\log\left(\rho^{-\frac{3}{2}}\right) + \frac{1}{2}\log(\eta_{\max})\\
&\qquad + \log(1+\eta_{\max}) + \log\frac{\pi}{e}\\
&\qquad + \sum_{\ell=L-m+1}^{L}\alpha_\ell \mathsf{P}(b_\ell+p)\left(\frac{2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right)\\
&\qquad + \sum_{\ell=L+1}^{\infty}\alpha_\ell p\mathsf{P}\left(\frac{4}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2}\right)\Biggr].
\end{aligned} \tag{5.44}$$

We next note that $\sum_{\mathbf{b}:f_L(\mathbf{b})\in\mathcal{B}_L(m)}\Pr\{\mathbf{B}=\mathbf{b}\}$ corresponds to the probability of $\mathbf{B}$ being a sequence with $m-1$ leading zeros followed by a one, so

$$\sum_{\mathbf{b}:f_L(\mathbf{b})\in\mathcal{B}_L(m)}\Pr\{\mathbf{B}=\tilde{\mathbf{b}}\} = p(1-p)^{m-1}.$$

Consequently,

$$
\begin{aligned}
&\frac{1}{N} I(X_1^N; Y_1^N, B_1^L) \\
&\leq \sum_{m=1}^{L} p(1-p)^{m-1} \left[(m-1) \log\left(\rho^{-\frac{3}{2}}\right) + \frac{1}{2}\log(\eta_{\max}) + \log(1+\eta_{\max}) + \log\frac{\pi}{e} \right] \\
&\quad + \sum_{m=1}^{L} p(1-p)^{m-1} \left[\sum_{\ell=L-m+1}^{L} \alpha_\ell \mathsf{P}(b_\ell - p) \left(\frac{2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2} \right) \right. \\
&\qquad\qquad \left. + \sum_{\ell=L+1}^{\infty} \alpha_\ell p \mathsf{P}\left(\frac{4}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2} \right) \right]. \qquad (5.45)
\end{aligned}
$$

To complete this bound, we have to analyze the behavior of (5.45) when L tends to infinity. To this end, we analyze separately the terms in (5.45). For the first term in (5.45),

$$
\sum_{m=1}^{L} p(1-p)^{m-1} \left[(m-1) \log\left(\rho^{-\frac{3}{2}}\right) + \frac{1}{2}\log(\eta_{\max}) + \log(1+\eta_{\max}) + \log\frac{\pi}{e} \right].
$$

We observe that inside the brackets only the first term depends on L and when $L \to \infty$, we have

$$
\sum_{m=1}^{\infty} p(1-p)^{m-1} = \sum_{t=0}^{\infty} p(1-p)^t = 1
$$

where it follows by [25, p. 8, 0.231.1], and

$$
\sum_{m=1}^{\infty} p(1-p)^{m-1}(m-1) = \sum_{t=0}^{\infty} p(1-p)^t t = \frac{1-p}{p}
$$

where it follows by [25, p. 8, 0.231.2]. Then, when $L \to \infty$, the first term in (5.45) yields

$$
\frac{(1-p)}{p} \log\left(\rho^{-\frac{3}{2}}\right) + \frac{1}{2}\log(\eta_{\max}) + \log(1+\eta_{\max}) + \log\frac{\pi}{e}. \qquad (5.46)
$$

Next, we show that the second term in (5.45) vanishes as L tends to infinity. The second term in (5.45) is given by

$$
\begin{aligned}
&\sum_{m=1}^{L} p(1-p)^{m-1} \left[\sum_{\ell=L-m+1}^{L} \alpha_\ell \mathsf{P}(b_\ell + p) \left(\frac{2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2} \right) \right. \\
&\qquad\qquad \left. + \sum_{\ell=L+1}^{\infty} \alpha_\ell p \mathsf{P}\left(\frac{4}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2} \right) \right]. \qquad (5.47)
\end{aligned}
$$

By reordering terms in (5.47), we have for any arbitrary $\delta \geq 0$,

$$
\begin{aligned}
&\frac{2\mathsf{P}}{\eta_{\max}\sigma^2}\sum_{m=1}^{L} p[\rho(1-p)]^{m-1}\left[\sum_{\ell=L-m+1}^{L}\alpha_\ell(b_\ell+p)+2\sum_{\ell=L+1}^{\infty}\alpha_\ell p\right] \\
&\quad \leq \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}\left(\sum_{m=0}^{L-\delta-1} p[\rho(1-p)]^{m}\sum_{\ell=L-m+1}^{L}\alpha_\ell + \sum_{m=L-\delta}^{L-1} p[\rho(1-p)]^{m}\sum_{\ell=1}^{\infty}\alpha_\ell\right) \\
&\qquad + \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}\sum_{m=0}^{L-1} p[\rho(1-p)]^{m}\sum_{\ell=L+1}^{\infty}\alpha_\ell p \\
&\quad = \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}\sum_{m=0}^{L-\delta-1} p[\rho(1-p)]^{m}\sum_{\ell=L-m+1}^{L}\alpha_\ell \\
&\qquad + \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}[\rho(1-p)]^{L-\delta}\frac{1-[\rho(1-p)]^{\delta}}{1-\rho(1-p)}\sum_{\ell=1}^{\infty}\alpha_\ell \\
&\qquad + \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}\sum_{m=0}^{L-1} p[\rho(1-p)]^{m}\sum_{\ell=L+1}^{\infty}\alpha_\ell p \\
&\quad \leq \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}\sum_{m=0}^{\infty} p[\rho(1-p)]^{m}\sum_{\ell=\delta+2}^{\infty}\alpha_\ell \\
&\qquad + \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}[\rho(1-p)]^{L-\delta}\frac{1-[\rho(1-p)]^{\delta}}{1-\rho(1-p)}\sum_{\ell=1}^{\infty}\alpha_\ell \\
&\qquad + \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}\sum_{m=0}^{\infty} p[\rho(1-p)]^{m}\sum_{\ell=L+1}^{\infty}\alpha_\ell p \\
&\quad = \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}\frac{p}{1-\rho(1-p)}\sum_{\ell=\delta+2}^{\infty}\alpha_\ell \\
&\qquad + \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}[\rho(1-p)]^{L-\delta}\frac{1-[\rho(1-p)]^{\delta}}{1-\rho(1-p)}\sum_{\ell=1}^{\infty}\alpha_\ell \\
&\qquad + \frac{4\mathsf{P}}{\eta_{\max}\sigma^2}\frac{p^2}{1-\rho(1-p)}\sum_{\ell=L+1}^{\infty}\alpha_\ell
\end{aligned} \tag{5.48}
$$

where the first step follows because we divided the first sum (over m) into two sums, because we upper-bound $b_\ell + p$ by 2 and because we sum the α_ℓ over more terms. The second step follows by computing the sum over m of the second term. The third step follows because we increase the number of terms to be added in the first and last terms. The last step follows by computing the infinite sum over m of the first term.

As $L \to \infty$, the second and third terms on the RHS of (5.48) vanish, since $\rho(1-p) < 1$ and $\sum_{\ell=1}^{\infty} \alpha_\ell < \infty$ by assumption. Consequently,

$$\begin{aligned}
\overline{\lim_{L\to\infty}} \sum_{m=1}^{L} p(1-p)^{m-1} \Bigg[\sum_{\ell=L-m+1}^{L} & \alpha_\ell \mathsf{P}(b_\ell + p) \left(\frac{2}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2} \right) \\
& + \sum_{\ell=L+1}^{\infty} \alpha_\ell p \mathsf{P} \left(\frac{4}{\frac{\eta_{\max}}{\rho^{m-1}}\sigma^2} \right) \Bigg] \\
\leq \frac{4\mathsf{P}}{\eta_{\max}\sigma^2} \frac{p}{1-\rho(1-p)} \sum_{\ell=\delta+2}^{\infty} \alpha_\ell & \qquad (5.49)
\end{aligned}$$

for any arbitrary $\delta \geq 0$. Next, by letting $\delta \to \infty$, we obtain that (5.49) vanishes, and consequently (5.47) also vanishes.

Finally, using (5.46) and that (5.47) vanishes as $L \to \infty$, the limit of (5.45) as $L \to \infty$ becomes

$$\begin{aligned}
& \overline{\lim_{L\to\infty}} \frac{1}{N} I(X_1^N; Y_1^N, B_1^L) \\
& \quad \leq \frac{1-p}{p} \log\left(\rho^{-\frac{3}{2}}\right) + \frac{1}{2}\log(\eta_{\max}) + \log(1+\eta_{\max}) + \log\frac{\pi}{e}. \qquad (5.50)
\end{aligned}$$

Since the RHS of (5.50) neither depends on the input distribution nor on N, it follows from (5.3) that it is also an upper bound on the capacity $C(\mathsf{P})$. This proves Theorem 21.

5.5 Conclusions

Lozano, Heath, and Andrews demonstrated that, in the absence of perfect knowledge of the realizations of the channel coefficients, the information rate achievable over wireless networks with inputs of the form $\sqrt{\text{SNR}}U_k$ is bounded in the SNR [43]. In our analysis we incorporated the possibility that the interference is intermittent and allow the channel inputs to change arbitrarily with the SNR. Our work is thus more general than [43] in the sense that we consider interference burstiness and we optimize over all possible input distributions, but it is less general in the sense that we do not allow the nodes to cooperate and we require all nodes to use the same codebook. Since in the presence of perfect knowledge of the realizations of the channel coefficients, the most efficient transmission strategies, such as interference alignment, rely on cooperation between the users, the former constraint seems particularly restrictive. It is yet unknown whether cooperative strategies can achieve rates that are unbounded in the SNR when a noncoherent setting is assumed.

6 Summary and Conclusion

In this dissertation we studied the effect of interference burstiness in communication channels. In particular, we investigated the impact of these phenomena on channel capacity.

On the one hand, we studied the channel capacity of the bursty interference channel (IC) as modeled by a linear deterministic model (LDM). For this simplified model with only two users, we obtained a complete characterization of the channel capacity under different scenarios of channel state information (CSI) that provides an understanding of the effect of interference burstiness on the data transmission. On the other hand, we extended our analysis to a more realistic wireless network with an infinite number of users, as modeled by a noncoherent fading channel. For this case, we performed an asymptotic analysis on the channel capacity. Specifically, we showed that the capacity of this channel is bounded in the signal-to-noise ratio (SNR), suggesting that such networks are highly power inefficient. We next present a more detailed summary of our main contributions.

Linear deterministic bursty interference channel

The effect of burstiness was studied in **Chapter 4** for the two-user linear deterministic bursty IC. We modeled the presence/absence of interference by a block-independent and identically distributed (IID) Bernoulli process. While the LDM may appear over-simplified, it allowed us to unify several aspects previously studied in the literature and gives rise to several new results on the effects of CSI in the capacity of the bursty IC. We studied both the quasi-static and ergodic setups. In the quasi-static setup, the highest sum rate R is limited by the worst-case channel and coincides with that of the (non-bursty) IC. However, by using opportunistic codes, we can transmit at a higher sum rate $R + \Delta R$ when the channel realization is good, i.e., when at least one receiver (Rx) is not affected by interference. For this setup, we derived matching converse and achievability bounds. For the ergodic setup, we also derived converse and achievability bounds on the channel capacity, that are matching in most cases. The main conclusions of Chapter 4 are as follows:

- For the quasi-static setup, the achievable rates for local CSIR and local CSIRT coincide. However, for the ergodic setup, having local CSIRT yields an increased data rate compared to the one achieved by only having CSI at the Rx side. Featuring global CSI at all nodes yields a sum rate that outperforms the rates achieved by having only local CSI for both quasi-static and ergodic setups.

- Regarding burstiness, global CSI exploits interference for all regions (except for very strong interference) in both quasi-static and ergodic setups. When local CSI is available only at the Rx side, interference burstiness is of clear benefit if the interference is either weak or very weak, or if the channel is ergodic and interference is present at most half of the time. Furthermore, when the channel is ergodic and local CSI is available at each transmitter and receiver, interference burstiness is beneficial in all interference regions except the very weak and very strong interference regions.

- We compared the quasi-static and ergodic setups by using the average sum capacity metric. We demonstrated that, for local CSIR, the average sum capacity coincides with the achievable rates in the ergodic setup for all interference regions. In contrast, for local CSIRT, the average sum capacity is strictly smaller than the sum capacity in the ergodic setup. For global CSIRT, average sum capacity and sum capacity coincide for all interference regions when the interference states are fully correlated, and they coincide in the very weak interference (VWI) and weak interference (WI) regions when the interference states are independent. For global CSIRT, moderate interference (MI) and strong interference (SI), and independent interference states, the average sum capacity is smaller than the sum capacity in the ergodic setup.

Bursty noncoherent wireless networks

In **Chapter 5**, we studied a wireless network with an infinite number of users, modeled by a noncoherent fading channel. To make the analysis tractable, we considered a channel where two nodes are communicating and infinite number of nodes interfere the communication. We modeled again the presence/absence of interference as a Bernoulli process, with activation probability p. We assumed a noncoherent setting where the realizations of the channel coefficients are not available to transmitter (Tx) and Rx, and where they do not perform channel estimation to obtain information on the fading coefficients. We assumed, however, that the Rx knows the interference states of the interfering links, and that the states remain constant during the whole codeword transmission. This study is similar to the one performed by Lozano, Heath and Andrews [43] for non-bursty wireless networks. Indeed, Lozano *et al.* demonstrated that, when Tx and Rx have no knowledge of the realizations of the channel coefficients, the capacity of wireless networks is bounded in the SNR, under the assumption that channel inputs are of the form $\sqrt{\mathrm{SNR}}U_k$. In our analysis, we incorporated the effect of interference burstiness and the possibility that the channel inputs change arbitrarily with the SNR. The main conclusions of this chapter are thus as follows:

- We demonstrated that the result obtained by Lozano, Heath and Andrews [43] continues to hold even if the channel inputs are allowed to change arbitrarily with the SNR, and even if the interference is bursty, provided that the variances of the path gains decay at most exponentially. Since this last assumption is very mild, this suggests that noncoherent wireless networks are highly power inefficient.

- Our bound further showed that interference burstiness does not change the behavior of channel capacity. While our upper bound on the channel capacity grows as the channel becomes more bursty, it remains bounded in the SNR. Thus, interference burstiness cannot be exploited to mitigate the power inefficiency at high SNR.

- Possible strategies that could mitigate the power inefficiency of noncoherent wireless networks and that have not been explored in this thesis are cooperation between users and improved channel estimation strategies. Indeed, coherent wireless networks, in which users have perfect knowledge of the fading coefficients, have a capacity that grows to infinity with the SNR. Furthermore, for such networks, the most efficient transmission strategies, such as interference alignment, rely on cooperation. Our results suggested that these two strategies may be essential to obtain an unbounded capacity in the SNR.

“If at first you don’t succeed (first you don’t succeed) / Dust yourself off, and try again”

Aaliyah. Song: Try Again

A

Appendix to Chapter 4

A.1 Proofs for the Quasi-Static Setup

We define $p_{\mathbf{b}} = \Pr\{\mathbf{B} = \mathbf{b}\}$. Clearly, when B_1, B_2 are independent, we have $p_{00} = (1-p)^2$, $p_{11} = p^2$ and $p_{01} = p_{10} = p(1-p)$, and when B_1, B_2 are fully correlated $p_{00} = 1-p$, $p_{11} = p$ and $p_{01} = p_{10} = 0$.

The converse bounds in the quasi-static case are based on an information density approach [62]. In particular, we define the information densities for the bursty interference channel (IC)

$$\mathrm{i}_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b}) \triangleq \mathrm{i}_{\mathbf{X}_1^N \mathbf{Y}_1^N|\mathbf{B}}(\mathbf{x}_1^N; \mathbf{y}_1^N|\mathbf{b}) = \log \frac{P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathbf{y}_1^N|\mathbf{x}_1^N, \mathbf{b})}{P_{\mathbf{Y}_1^N|\mathbf{B}}(\mathbf{y}_1^N|\mathbf{b})} \tag{A.1}$$

$$\mathrm{i}_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b}) \triangleq \mathrm{i}_{\mathbf{X}_2^N \mathbf{Y}_2^N|\mathbf{B}}(\mathbf{x}_2^N; \mathbf{y}_2^N|\mathbf{b}) = \log \frac{P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_2^N|\mathbf{x}_2^N, \mathbf{b})}{P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{y}_2^N|\mathbf{b})}. \tag{A.2}$$

Here and throughout the appendices, we use the notations $\mathbf{X}_i^N = \boldsymbol{X}_i$, $\mathbf{x}_i^N = \boldsymbol{x}_i$, $\mathbf{Y}_i^N = \boldsymbol{Y}_i$, and $\mathbf{y}_i^N = \boldsymbol{y}_i$ to highlight the fact that, in the quasi-static setting, we transmit N symbols in one coherence block.

We further consider the individual error events

$$\mathcal{E}_i(\Gamma_i) \quad \triangleq \quad \left\{\frac{1}{N}\mathrm{i}_i(\mathbf{x}_i^N, \mathbf{y}_i^N, \mathbf{b}) \leq \Gamma_i\right\}, \; i = 1, 2 \tag{A.3}$$

and the joint error event

$$\mathcal{E}_{12}(\Gamma) \quad \triangleq \quad \left\{\frac{1}{N}\left(\mathrm{i}_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b}) + \mathrm{i}_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b})\right) \leq \Gamma\right\}. \tag{A.4}$$

The proofs of the converse results are based on the following lemmas.

Lemma 2 (Verdú-Han lemma) *Every (N, R, P_e) code over a channel $P_{\mathbf{Y}^N|\mathbf{X}^N}$ satisfies*

$$P_e \geq \Pr\left\{\frac{1}{N}\mathrm{i}_{\mathbf{X}^N\mathbf{Y}^N}(\mathbf{X}^N; \mathbf{Y}^N) \leq R - \gamma\right\} - e^{-\gamma N} \tag{A.5}$$

for every $\gamma > 0$, where X^N places probability mass $\frac{1}{2^{NR}}$ on each codeword and $\mathrm{i}_{\mathbf{X}^N\mathbf{Y}^N}(\mathbf{X}^N; \mathbf{Y}^N) \triangleq \log \frac{P_{\mathbf{Y}^N|\mathbf{X}^N}(\mathbf{y}^N|\mathbf{x}^N)}{P_{\mathbf{Y}^N}(\mathbf{y}^N)}$.

Proof: See [62, (Th. 4)]. ■

Lemma 3 *Suppose that $\Pr\{\mathcal{E}_{12}(\Gamma)\big|\mathbf{B} = \mathbf{b}\} \to 0$ as $N \to \infty$. Then, for each pair $\mathbf{b} \in \{0, 1\}^2$, the threshold Γ must satisfy the following conditions:*

- *For $\mathbf{B} = [0, 0]$, Γ satisfies*

$$\Gamma \quad \leq \quad 2n_d. \tag{A.6}$$

- *For $\mathbf{B} = [0, 1]$ and $\mathbf{B} = [1, 0]$, Γ satisfies (A.6) and*

$$\Gamma \quad \leq \quad (n_d - n_c)^+ + \max(n_d, n_c). \tag{A.7}$$

- *For $\mathbf{B} = [1, 1]$, Γ satisfies (A.6) and (A.7), and*

$$\Gamma \quad \leq \quad 2\max\{(n_d - n_c)^+, n_c\}. \tag{A.8}$$

Proof: See Appendix B. ■

A.1.1 Proof of Theorem 10

In this section we prove the IC converse bounds for $p > 0$. This proof assumes global CSIRT, hence the resulting bounds also apply to local CSIR and local CSIRT. Let $P_e^{(N)} = \Pr\{(\hat{W}_1 \neq W_1 \cup \hat{W}_2 \neq W_2)\}$, and let us denote by $P_{e_1}^{(N)}$ and $P_{e_2}^{(N)}$ the error probabilities at decoders one and two, respectively:

$$P_{e1}^{(N)} \triangleq \Pr\{\hat{W}_1 \neq W_1\}, \tag{A.9}$$

$$P_{e2}^{(N)} \triangleq \Pr\{\hat{W}_2 \neq W_2\}. \tag{A.10}$$

Clearly, the error probabilities $P_e^{(N)}$, $P_{e1}^{(N)}$ and $P_{e2}^{(N)}$ are related by the following sets of inequalities

$$\max\Big(P_{e1}^{(N)}, P_{e2}^{(N)}\Big) \leq P_e^{(N)} \leq P_{e1}^{(N)} + P_{e2}^{(N)} \leq 2\max\Big(P_{e1}^{(N)}, P_{e2}^{(N)}\Big). \tag{A.11}$$

Using these inequalities we conclude that

$$P_e^{(N)} \geq \frac{1}{2}\Big(P_{e1}^{(N)} + P_{e2}^{(N)}\Big). \tag{A.12}$$

We now rewrite (A.9) and (A.10) as

$$P_{e1}^{(N)} = \sum_{\mathbf{b}} p_{\mathbf{b}} \Pr\{\hat{W}_1 \neq W_1 | \mathbf{B} = \mathbf{b}\}, \tag{A.13}$$

$$P_{e2}^{(N)} = \sum_{\mathbf{b}} p_{\mathbf{b}} \Pr\{\hat{W}_2 \neq W_2 | \mathbf{B} = \mathbf{b}\} \tag{A.14}$$

and apply the Verdú-Han lemma (Lemma 2) to each of the probability terms $\Pr\{\hat{W}_i \neq W_i | \mathbf{B} = \mathbf{b}\}$, $i = 1, 2$, in (A.13) and (A.14). This yields

$$\begin{aligned}&\Pr\{\hat{W}_1 \neq W_1 | \mathbf{B} = \mathbf{b}\}\\ &\qquad \geq \Pr\Big\{\frac{1}{N}\mathrm{i}_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b}) \leq R_1 - \gamma_1 | \mathbf{B} = \mathbf{b}\Big\} - e^{-\gamma_1 N},\end{aligned} \tag{A.15}$$

$$\begin{aligned}&\Pr\{\hat{W}_2 \neq W_2 | \mathbf{B} = \mathbf{b}\}\\ &\qquad \geq \Pr\Big\{\frac{1}{N}\mathrm{i}_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b}) \leq R_2 - \gamma_2 | \mathbf{B} = \mathbf{b}\Big\} - e^{-\gamma_2 N}.\end{aligned} \tag{A.16}$$

We set $\Gamma_i = R_i - \gamma_i$ and $\Gamma = \Gamma_1 + \Gamma_2 = R - \gamma_1 - \gamma_2$. Then, using the definition of $\mathcal{E}_i$ in (A.3), we can write (A.15) and (A.16) as

$$\Pr\{\hat{W}_1 \neq W_1 | \mathbf{B} = \mathbf{b}\} \geq \Pr\{\mathcal{E}_1(\Gamma_1) | \mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N}, \tag{A.17}$$

$$\Pr\{\hat{W}_2 \neq W_2 | \mathbf{B} = \mathbf{b}\} \geq \Pr\{\mathcal{E}_2(\Gamma_2) | \mathbf{B} = \mathbf{b}\} - e^{-\gamma_2 N}. \tag{A.18}$$

Comparing the joint error event $\mathcal{E}_{12}(\Gamma)$ in (A.4) with $\mathcal{E}_1(\Gamma_1)$ and $\mathcal{E}_2(\Gamma_2)$ in (A.3), it can be shown that

$$\mathcal{E}_1(\Gamma_1) \cap \mathcal{E}_2(\Gamma_2) \quad \subseteq \quad \mathcal{E}_{12}(\Gamma), \tag{A.19}$$

$$\mathcal{E}_1^c(\Gamma_1) \cap \mathcal{E}_2^c(\Gamma_2) \quad \subseteq \quad \mathcal{E}_{12}^c(\Gamma) \;\Rightarrow\; \mathcal{E}_{12}(\Gamma) \subseteq \mathcal{E}_1(\Gamma_1) \cup \mathcal{E}_2(\Gamma_2). \tag{A.20}$$

Using (A.20) and the union bound, we thus obtain

$$\begin{aligned}\Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B}=\mathbf{b}\} &\le \Pr\{\mathcal{E}_1(\Gamma_1) \cup \mathcal{E}_2(\Gamma_2)|\mathbf{B}=\mathbf{b}\} \\ &\le \Pr\{\mathcal{E}_1(\Gamma_1)|\mathbf{B}=\mathbf{b}\} + \Pr\{\mathcal{E}_2(\Gamma_2)|\mathbf{B}=\mathbf{b}\}.\end{aligned} \tag{A.21}$$

Combining this result with (A.12), (A.17) and (A.18) gives

$$\begin{aligned}&P_e^{(N)} \\ &\ge \frac{1}{2}\left(P_{e1}^{(N)} + P_{e2}^{(N)}\right) \\ &\ge \frac{1}{2}\sum_{\mathbf{b}} p_{\mathbf{b}}\big(\Pr\{\mathcal{E}_1(\Gamma_1)|\mathbf{B}=\mathbf{b}\} + \Pr\{\mathcal{E}_2(\Gamma_2)|\mathbf{B}=\mathbf{b}\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}\big) \\ &\ge \frac{1}{2}\sum_{\mathbf{b}} p_{\mathbf{b}}\big(\Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B}=\mathbf{b}\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}\big).\end{aligned} \tag{A.22}$$

The remainder of this section is devoted to an analysis of $\Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B}=\mathbf{b}\}$. Indeed, by (A.22) we have for any $\gamma_1, \gamma_2 > 0$ that

$$\lim_{N\to\infty} P_e^{(N)} \ge \lim_{N\to\infty} \frac{1}{2}\left[p_{11}\epsilon_{11} + p_{00}\epsilon_{00} + p_{10}\epsilon_{10} + p_{01}\epsilon_{01}\right], \tag{A.23}$$

where $\epsilon_{\mathbf{b}} \triangleq \Pr\{\mathcal{E}_{12}(\Gamma)\big|\mathbf{B}=\mathbf{b}\}$. When $p > 0$, the probability p_{11} is strictly positive both when (B_1, B_2) are independent and when they are fully correlated. Since $p_{\mathbf{b}}$ does not depend on N, it follows that the only way that $\lim_{N\to\infty} P_e^{(N)} = 0$ is that $\epsilon_{11} \to 0$ as $N \to \infty$. The conditions on R under which this happens are summarized in Lemma 3. Specifically, recalling that $\Gamma = R - (\gamma_1 + \gamma_2)$, we obtain from Lemma 3 that $P_e^{(N)} \to 0$ only if

$$R - (\gamma_1 + \gamma_2) \quad \le \quad 2n_d \tag{A.24}$$

$$R - (\gamma_1 + \gamma_2) \quad \le \quad (n_d - n_c)^+ + \max(n_d, n_c) \tag{A.25}$$

$$R - (\gamma_1 + \gamma_2) \quad \le \quad 2\max\{(n_d - n_c)^+, n_c\}. \tag{A.26}$$

Since $\gamma_1, \gamma_2 > 0$ are arbitrary, we obtain the converse bounds (4.11) and (4.12) in Theorem 10 from (A.24)–(A.26) upon letting $N \to \infty$ and then $\gamma_1 \to 0$ and $\gamma_2 \to 0$.

When $p = 0$, the only positive probability is p_{00}. A necessary condition for $\lim_{N\to\infty} P_e^{(N)} = 0$ is that $\epsilon_{00} \to 0$ as $N \to \infty$. By following the same approach as for the case $p > 0$, we obtain the converse bound (4.10) in Theorem 10.

A.1.2 Converse Proof of Theorem 11

In this section, we analyze the opportunistic rate $\Delta R_1(b_1) + \Delta R_2(b_2),\ b_i \in \{0, 1\}$ for local CSIRT and independent B_1 and B_2. Let us denote by $\hat{P}^{(n)}_{e_{1(b_1)}}$ and $\hat{P}^{(n)}_{e_{2(b_2)}}$ the error probabilities at decoders one and two, defined in (4.4) and (4.5), i.e.,

$$\hat{P}^{(n)}_{e_{1(b_1)}} \triangleq \Pr\{(\hat{W}_1, \Delta\hat{W}_1(B_1)) \neq (W_1, \Delta W_1(B_1))|B_1 = b_1\}, \tag{A.27}$$

$$\hat{P}^{(n)}_{e_{2(b_2)}} \triangleq \Pr\{(\hat{W}_2, \Delta\hat{W}_2(B_2)) \neq (W_2, \Delta W_2(B_2))|B_2 = b_2\}. \tag{A.28}$$

where $b_1 \in \{0, 1\}$ in (A.27) and $b_2 \in \{0, 1\}$ in (A.28).

Before we apply the Verdú-Han lemma, we have to deal with the fact that (A.27) and (A.28) are conditioned on two different variables but we need to analyze the probability of error jointly. To solve this problem, we expand the probability of error (A.27) as

$$\hat{P}^{(n)}_{e_{1(b_1)}} = \sum_{b_2=0,1} \Pr\{B_2 = b_2\} \Pr\big\{(\hat{W}_1, \Delta\hat{W}_1(B_1)) \neq (W_1, \Delta W_1(B_1))\big|\mathbf{B} = \mathbf{b}\big\}. \tag{A.29}$$

Since, by assumption, $\Pr\{B_2 = b_2\} \in (0, 1)$, it follows that

$$\Pr\big\{(\hat{W}_1, \Delta\hat{W}_1(B_1)) \neq (W_1, \Delta W_1(B_1))|B_1 = b_1\big\} \to 0 \text{ as } N \to \infty$$

if, and only if,

$$\Pr\big\{(\hat{W}_1, \Delta\hat{W}_1(B_1)) \neq (W_1, \Delta W_1(B_1))|\mathbf{B} = \mathbf{b}\big\} \to 0,\ b_2 \in \{0, 1\} \text{ as } N \to \infty. \tag{A.30}$$

We shall lower-bound (A.29) by considering only one of the two terms in the sum. Proceeding analogously for the second user and applying the Verdú-Han lemma (Lemma 8), we obtain

$$\hat{P}^{(n)}_{e_{1(b_1)}} \geq \left(\Pr\Big\{\frac{1}{N}\mathrm{i}_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b}) \leq R_1 + \Delta R_1(B_1) - \gamma_1|\mathbf{B} = \mathbf{b}\Big\} - e^{-\gamma_1 N}\right) \Pr\{B_2 = b_2\} \tag{A.31}$$

where $\quad b_2 = 0, 1,$

$$\hat{P}^{(n)}_{e_{2(b_2)}} \geq \left(\Pr\Big\{\tfrac{1}{N}\mathrm{i}_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b}) \leq R_2 + \Delta R_2(B_2) - \gamma_2|\mathbf{B} = \mathbf{b}\Big\} - e^{-\gamma_2 N}\right) \Pr\{B_1 = b_1\}. \tag{A.32}$$

where $b_1 = 0, 1$. Let $\Gamma_i = R_i + \Delta R_i - \gamma_i$, $i = 1, 2$ and $\Gamma = R + \Delta R_1(B_1) + \Delta R_2(B_2) - (\gamma_1 + \gamma_2)$. Then, (A.31) and (A.32) can be written as

$$\hat{P}^{(n)}_{e_{1(b_1)}} \geq \left(\Pr\{\mathcal{E}_1(\Gamma_1)|\mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N}\right) \Pr\{B_2 = b_2\}, \tag{A.33}$$
$$\hat{P}^{(n)}_{e_{2(b_2)}} \geq \left(\Pr\{\mathcal{E}_2(\Gamma_2)|\mathbf{B} = \mathbf{b}\} - e^{-\gamma_2 N}\right) \Pr\{B_1 = b_1\}. \tag{A.34}$$

where $b_2 = 0, 1$ in (A.33) and $b_1 = 0, 1$ in (A.34).

Proceeding analogously as in (A.19)–(A.22), and using that $\Pr\{B_i = b_i\} \geq \min\{p, 1 - p\}$, we obtain

$$\begin{aligned} &\hat{P}^{(n)}_{e_{1(b_1)}} + \hat{P}^{(n)}_{e_{2(b_2)}} \\ &\geq \left(\Pr\{\mathcal{E}_1(\Gamma_1)|\mathbf{B} = \mathbf{b}\} + \Pr\{\mathcal{E}_2(\Gamma_2)|\mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}\right) \min\{p, 1 - p\} \qquad \text{(A.35)} \\ &\geq \left(\Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N} - e^{\gamma_2 N}\right) \min\{p, 1 - p\}. \end{aligned}$$

Since $\gamma_1, \gamma_2 > 0$, the left-hand side (LHS) of (A.35) only tends to zero as $N \to \infty$ if $\Pr(\mathcal{E}_{12}(\Gamma)|\mathbf{B} = \mathbf{b}) \to 0$ as $N \to \infty$. It thus follows from Lemma 3 that $\hat{P}^{(n)}_{e_{1(b_1)}} + \hat{P}^{(n)}_{e_{2(b_2)}} \to 0$ as $N \to \infty$ only if conditions (A.6)-(A.8) are satisfied. Letting $\gamma_1 \to 0$ and $\gamma_2 \to 0$ then gives the following constraints:

- For $\mathbf{B} = [1, 1]$

$$R_1 + \Delta R_1(1) + R_2 + \Delta R_2(1) \leq 2n_d \tag{A.36}$$
$$R_1 + \Delta R_1(1) + R_2 + \Delta R_2(1) \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{A.37}$$
$$R_1 + \Delta R_1(1) + R_2 + \Delta R_2(1) \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{A.38}$$

- For $\mathbf{B} = [0, 0]$,

$$R_1 + \Delta R_1(0) + R_2 + \Delta R_2(0) \leq 2n_d. \tag{A.39}$$

- For $\mathbf{B} = [0, 1]$, using that $\Delta R_2(1) = 0$,

$$R_1 + \Delta R_1(0) + R_2 \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A.40}$$

- For $\mathbf{B} = [1, 0]$, using that $\Delta R_1(1) = 0$,

$$R_1 + R_2 + \Delta R_2(0) \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A.41}$$

The constraints (A.39)–(A.41) yield (4.13)–(4.15). This proves Theorem 11.

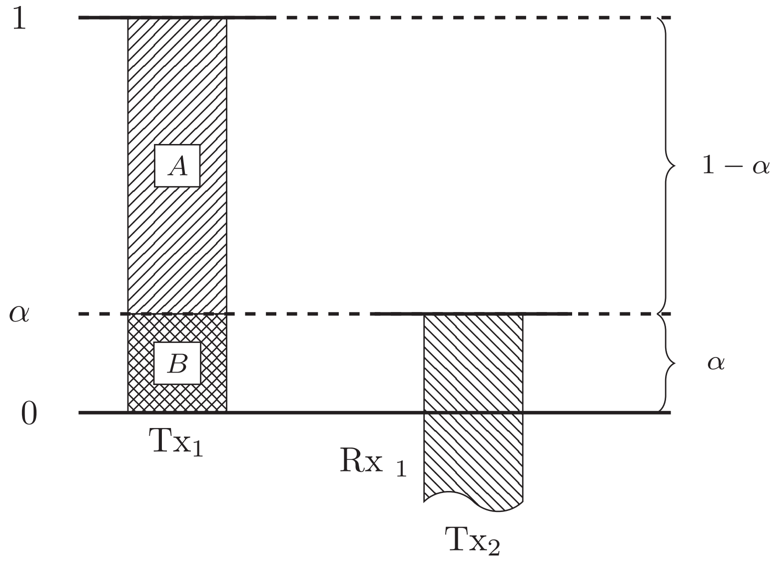


Figure A.1: Normalized signal levels at Rx_1 for $\alpha \leq \frac{1}{2}$.

A.1.3 Achievability Proof of Theorem 11

In this section, we present the achievability bounds in Theorem 11 for the regions in which it is possible to transmit opportunistic messages, namely the very weak interference (VWI) and weak interference (WI) regions. The presented bounds are valid for local CSIR and local CSIRT.

A.1.3.1 Very Weak Interference

Transmitter 1 (transmitter $(\mathrm{Tx})_1$) and transmitter 2 (Tx_2) transmit in the most significant levels a block of $n_d(1-\alpha)$ bits, and they transmit in the least significant levels a block of $n_d\alpha$ bits. The same construction is used for both transmitters. Figure A.1 depicts the signal levels of the transmitted signals (normalized by n_d) as observed at receiver 1 (receiver $(\mathrm{Rx})_1$), when it is affected by interference. At the receiver side, we have the following procedure:

- In presence of interference: decode block $\boxed{A}$ in the desired signal which is interference free, and treat the block $\boxed{B}$ as noise. We thus obtain the individual rate

$$R_1 = (n_d - n_c)^+ \ \frac{\text{bits}}{\text{sub-channel use}}. \tag{A.42}$$

- In absence of interference: decode blocks $\boxed{A}$ and $\boxed{B}$. We thus obtain the individual rate

$$R_1 + \Delta R_1(0) = n_d \ \frac{\text{bits}}{\text{sub-channel use}}. \tag{A.43}$$

where $\Delta R_1(0) = n_c \frac{\text{bits}}{\text{sub-channel use}}$ corresponds to the opportunistic rate.

The bounds (A.42) and (A.43) coincide with the bounds of user 2. In order to obtain the possible sum rates according to the interference states, we combine (A.42) (which corresponds to $B_1 = 1$) and (A.43) (which corresponds to $B_1 = 0$) to obtain the converse bounds (4.13)–(4.14).

A.1.3.2 Weak Interference

The symbol transmitted by Tx_1 (normalized by n_d) is depicted in Figure A.2a. Specifically, we transmit in the most significant levels a block of $n_d(1-\alpha)$ bits. In the subsequent levels we transmit a block of $n_d(2\alpha - 1)$ zeros, followed by $n_d(2-3\alpha)$ opportunistic bits. Finally, in the least significant levels, we transmit a block of $n_d(2\alpha - 1)$ bits. The same construction is used for both transmitters.

Figure A.2b depicts the normalized signal levels of the transmitted signals as observed by Rx_1. At the receiver side, we have the following procedure:

- In presence of interference: The channel pushes the interference level by $n_d - n_c$ bits. Thus, the least significant $2n_c - n_d$ bits of the desired signal (block $\boxed{A}$) align with the zeros of the interference signal and can be decoded free from interference. Since $(n_d - n_c) \leq n_c$, the most significant $n_d - n_c$ bits (block $\boxed{B}$) are also free from interference. Thus, we achieve the rate

$$\begin{aligned} R_1 &= n_d - n_c + 2n_c - n_d \\ &= n_c \ \frac{\text{bits}}{\text{sub-channel use}}. \end{aligned} \tag{A.44}$$

- In absence of interference: The bits in blocks $\boxed{A}$, $\boxed{B}$, and $\boxed{D}$ can be decoded free from interference. Thus, we achieve the rate

$$\begin{aligned} R_1 + \Delta R_1(0) &= n_d - n_c + 2n_c - n_d + 2n_d - 3n_c \\ &= 2(n_d - n_c) \ \frac{\text{bits}}{\text{sub-channel use}} \end{aligned} \tag{A.45}$$

where $\Delta R_1(0) = 2n_d - 3n_c \frac{\text{bits}}{\text{sub-channel use}}$ corresponds to the opportunistic rate.

By symmetry, the bounds (A.44) and (A.45) also apply for the achievable rates of user 2. In order to obtain the possible sum rates according to the interference states, we combine (A.44) (which corresponds to $B_1 = 1$) and (A.45) (which corresponds to $B_1 = 0$) to obtain the achievability bounds in Theorem 11.

A.1.4 Converse Proof of Theorem 11 when $B_1 = B_2$

The proof of the converse bound (4.13) for local CSIR when $B_1 = B_2$ is similar to the proof when B_1 and B_2 are independent; see Appendix A.1.2. However, to prove the converse bound (4.14) for the case where $B_1 = B_2$ we cannot simply reproduce the steps for the independent case. The reason is that, in the correlated case, we only have the interference states $[0, 0]$ and $[1, 1]$, but the derivation of (4.14) for the independent case follows from the analysis of the states $\mathbf{B} = [0, 1]$ and $\mathbf{B} = [1, 0]$ (see (A.40) and (A.41) in Appendix A.1.2). To sidestep this problem, we follow a slightly different approach. Specifically, we combine the error probability of user 1 when $\mathbf{B} = [0, 0]$ with that of user 2 when $\mathbf{B} = [1, 1]$. This approach yields a tighter converse bound compared to the one obtained by simply considering $\mathbf{B} = [0, 0]$ in both probabilities.

Consider $\hat{P}^{(n)}_{e_{1(b_1)}}$ and $\hat{P}^{(n)}_{e_{2(b_2)}}$ defined in (A.27) and (A.28). Applying the Verdú-Han lemma (Lemma 2) with $\Gamma_1 = R_1 + \Delta R_1(0) - \gamma_1$ and $\Gamma_2 = R_2 - \gamma_2$, and using (A.29), we obtain the lower bounds

$$\hat{P}^{(n)}_{e_{1(0)}} \geq \left(\Pr\Big\{\mathcal{E}_1(\Gamma_1)|\mathbf{B} = [0, 0]\Big\} - e^{-\gamma_1 N}\right) \Pr\{B_2 = 0\} \tag{A.46}$$

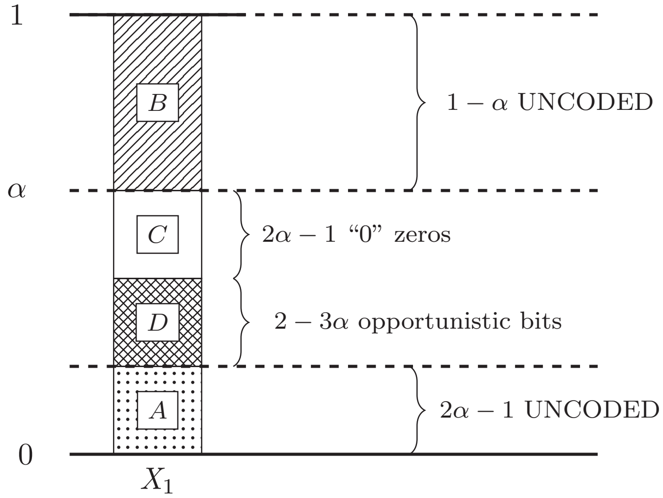
$$\hat{P}^{(n)}_{e_{2(1)}} \geq \left(\Pr\Big\{\mathcal{E}_2(\Gamma_2)|\mathbf{B} = [1, 1]\Big\} - e^{-\gamma_2 N}\right) \Pr\{B_1 = 1\}. \tag{A.47}$$

Note that compared to the derivation in Section A.1.2, the two error events $\mathcal{E}_1(\Gamma_1)$ and $\mathcal{E}_2(\Gamma_2)$ are conditioned on different interference states. In order to derive a joint error event for $\mathcal{E}_1(\Gamma_1)$ and $\mathcal{E}_2(\Gamma_2)$, we use the next lemma.

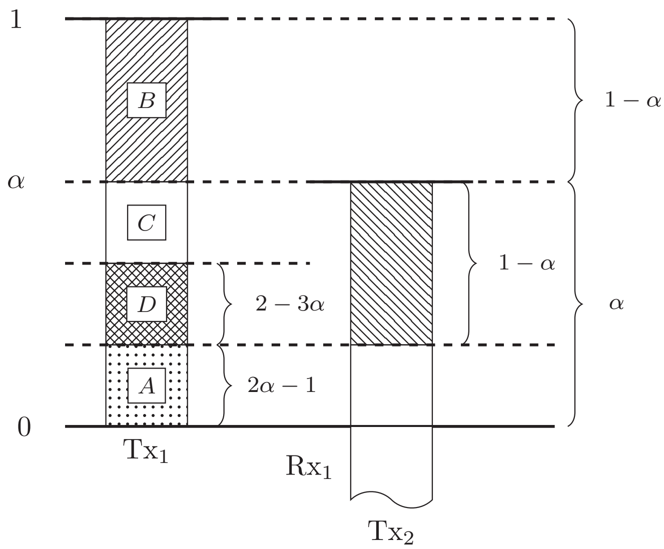
Lemma 4 *For local CSIR, the information density* i_i, $i = 1, 2$ *depends only on* $(\mathbf{x}_i^N, \mathbf{y}_i^N)$ *and the corresponding state* b_i, *i.e.,*

$$\mathrm{i}_1(\mathbf{x}_1^N; \mathbf{y}_1^N, [b_1, 0]) = \mathrm{i}_1(\mathbf{x}_1^N; \mathbf{y}_1^N, [b_1, 1]) \triangleq \mathrm{i}_1(\mathbf{x}_1^N, \mathbf{y}_1^N, b_1) \tag{A.48}$$

$$\mathrm{i}_2(\mathbf{x}_2^N; \mathbf{y}_2^N, [0, b_2]) = \mathrm{i}_2(\mathbf{x}_2^N; \mathbf{y}_2^N, [1, b_2]) \triangleq \mathrm{i}_2(\mathbf{x}_2^N, \mathbf{y}_2^N, b_2). \tag{A.49}$$



(a)



(b)

Figure A.2: (**a**) Normalized transmitted symbol at Tx_1; (**b**) Normalized signal levels at Rx_1.

Proof: We prove (A.48) for user 1. By the definition of the information density (A.1), it follows that

$$\mathrm{i}_1(\mathbf{x}_1^N, \mathbf{y}_1^N, [b_1, b_2]) = \log \frac{P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathbf{y}_1^N|\mathbf{x}_1^N, [b_1, b_2])}{P_{\mathbf{Y}_1^N|\mathbf{B}}(\mathbf{y}_1^N|[b_1, b_2])} \tag{A.50}$$

Evaluating i_1 for $\mathbf{B} = [0, b_2]$, $b_2 = 0, 1$ and $\mathbf{B} = [1, b_2]$, $b_2 = 0, 1$ we obtain that both cases are independent of b_2. The identity (A.48) can be proven in the same way. ■

We next analyze the probability terms in (A.46) and (A.47). It follows from (A.48) in Lemma 4 that $\mathrm{i}_1(\mathbf{x}_1^N, \mathbf{y}_1^N, [0, b_2])$ is independent of b_2. Consequently,

$$\begin{aligned}
\Pr\{\mathcal{E}_1(\Gamma_1)\big|\mathbf{B} = [0,0]\} &= \mathsf{E}\left[\mathbb{1}\left\{\frac{1}{N}\mathrm{i}_1(\mathbf{X}_1^N, \mathbf{Y}_1^N, [0,0]) \le \Gamma_1\right\}\right] \\
&= \mathsf{E}\left[\mathbb{1}\left\{\frac{1}{N}\mathrm{i}_1(\mathbf{X}_1^N, \mathbf{Y}_1^N, [0,1]) \le \Gamma_1\right\}\right] \\
&= \Pr\{\mathcal{E}_1(\Gamma_1)\big|\mathbf{B} = [0,1]\}.
\end{aligned} \tag{A.51}$$

Analogously, using (A.49) in (A.47), we obtain

$$\Pr\{\mathcal{E}_2(\Gamma_2)\big|\mathbf{B} = [1,1]\} = \Pr\{\mathcal{E}_2(\Gamma_2)\big|\mathbf{B} = [0,1]\}. \tag{A.52}$$

Adding (A.46) and (A.47), using (A.51) and (A.52), and lower-bounding $\Pr\{B_1 = 1\}$ and $\Pr\{B_2 = 0\}$ by $\min\{p, 1-p\}$, we obtain

$$\begin{aligned}
&\hat{P}_{e_{1(0)}}^{(n)} + \hat{P}_{e_{2(1)}}^{(n)} \\
&\ge \left(\Pr\{\mathcal{E}_1(\Gamma_1)|\mathbf{B} = [0,1]\} + \Pr\{\mathcal{E}_2(\Gamma_2)|\mathbf{B} = [0,1]\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}\right)\min\{p, 1-p\} \\
&\ge \Big(\Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B} = [0,1]\} - e^{-\gamma_1 N} - e^{\gamma_2 N}\Big)\min\{p, 1-p\}
\end{aligned} \tag{A.53}$$

where $\Gamma = \Gamma_1 + \Gamma_2$. We next apply Lemma 3 with $\Gamma = R + \Delta R_1(0) + \Delta R_2(0) - (\gamma_1 + \gamma_2)$. Since $\min\{p, 1-p\}$ is strictly positive for $0 < p < 1$, and since $-e^{-\gamma_1 N} - e^{-\gamma_2 N} \to 0$ as $N \to \infty$ for any fixed $\gamma_1, \gamma_2 > 0$, a necessary condition for (A.53) going to zero is that $\Pr\{\mathcal{E}_{12}(\Gamma)|\mathbf{B} = [0,1]\} \to 0$ as $N \to \infty$. This is the case if, and only if, (A.7) in Lemma 3 is fulfilled. Since $\gamma_1, \gamma_2 > 0$ are arbitrary, we conclude the proof by letting $\gamma_1 \to 0$ and $\gamma_2 \to 0$ and using that $\Delta R_2(1) = 0$ to obtain

$$R_1 + \Delta R_1(0) + R_2 \quad \le \quad (n_d - n_c)^+ + \max(n_d, n_c). \tag{A.54}$$

Given the symmetry of the problem, a bound on $\Delta R_2(0)$ follows by swapping the roles of users 1 and 2, yielding in this case

$$R_1 + R_2 + \Delta R_2(0) \quad \leq \quad (n_d - n_c)^+ + \max(n_d, n_c). \tag{A.55}$$

Finally, combining (A.54) and (A.55), we obtain the bound (4.14) in Theorem 11 for the fully correlated scenario.

A.1.4.1 Converse Proof of Theorem 14

In this section, we analyze the opportunistic rates $\{\Delta R(\mathbf{b}), \mathbf{b} \in \{0,1\}^2\}$ for global CSIRT and independent B_1 and B_2. Let us denote by $\hat{P}^{(n)}_{e_{1(\mathbf{b})}}$ and $\hat{P}^{(n)}_{e_{2(\mathbf{b})}}$ the error probabilities at decoders 1 and 2, defined in (4.4) and (4.5), namely,

$$\hat{P}^{(n)}_{e_{1(\mathbf{b})}} \quad \triangleq \quad \Pr\{(\hat{W}_1, \{\Delta\hat{W}_1(\mathbf{B})\}) \neq (W_1, \{\Delta W_1(\mathbf{B})\}) | \mathbf{B} = \mathbf{b}\}, \tag{A.56}$$
$$\hat{P}^{(n)}_{e_{2(\mathbf{b})}} \quad \triangleq \quad \Pr\{(\hat{W}_2, \{\Delta\hat{W}_2(\mathbf{B})\}) \neq (W_2, \{\Delta W_2(\mathbf{B})\}) | \mathbf{B} = \mathbf{b}\}. \tag{A.57}$$

where $\mathbf{b} \in \{0,1\}^2$.

We shall follow analogous steps as in Section A.1.2 and set $\Gamma_i = R_i + \Delta R_i(\mathbf{B}) - \gamma_i$, $i = 1, 2$, and $\Gamma = R + \Delta R(\mathbf{B}) - (\gamma_1 + \gamma_2)$. Proceeding analogously as in (A.19)–(A.21), we obtain

$$\hat{P}^{(n)}_{e_{1(\mathbf{b})}} + \hat{P}^{(n)}_{e_{2(\mathbf{b})}} \quad \geq \quad \Pr\{\mathcal{E}_{12}(\Gamma) | \mathbf{B} = \mathbf{b}\} - e^{-\gamma_1 N} - e^{-\gamma_2 N}. \tag{A.58}$$

By invoking Lemma 3 for fixed (but arbitrary) $\gamma_1, \gamma_2 > 0$, and letting then $\gamma_1 \to 0$ and $\gamma_2 \to 0$, we obtain that the right-hand side (RHS) of (A.58) vanishes as $N \to \infty$ only if the following constraints are satisfied:

- For $\mathbf{B} = [1, 1]$,

$$R_1 + \Delta R_1(11) + R_2 + \Delta R_2(11) \leq 2n_d \tag{A.59}$$
$$R_1 + \Delta R_1(11) + R_2 + \Delta R_2(11) \leq (n_d - n_c)^+ + \max(n_d, n_c) \tag{A.60}$$
$$R_1 + \Delta R_1(11) + R_2 + \Delta R_2(11) \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{A.61}$$

- For $\mathbf{B} = [0, 0]$,

$$R_1 + \Delta R_1(00) + R_2 + \Delta R_2(00) \leq 2n_d. \tag{A.62}$$

- For $\mathbf{B} = [0, 1]$,

$$R_1 + \Delta R_1(01) + R_2 + \Delta R_2(01) \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A.63}$$

- For $\mathbf{B} = [1, 0]$,

$$R_1 + \Delta R_1(10) + R_2 + \Delta R_2(10) \leq (n_d - n_c)^+ + \max(n_d, n_c). \tag{A.64}$$

This proves the converse bounds in Theorem 14.

A.1.4.2 Achievability Proof of Theorem 14

In this section, we present the achievability schemes for global CSIRT when B_1 and B_2 are independent. In contrast to the local CSIR/CSIRT case, we can adapt our transmission strategy to the interference states.

When $\mathbf{B} = [0, 0]$, the capacity-achieving scheme consists of sending uncoded bits in all n_d level. We thus achieve the sum rate $R + \Delta R(00) = 2n_d \frac{\text{bits}}{\text{sub-channel use}}$.

When $\mathbf{B} = [0, 1]$ or $\mathbf{B} = [1, 0]$, the achievability schemes coincide with the schemes described in Section A.1.3. In this case, we can only send opportunistic messages when we have VWI or WI.

A.1.4.3 Very Weak Interference

Consider the achievability scheme depicted in Figure A.1. By (A.42) and (A.43),

$$R_1 + \Delta R_1(01) = R_2 + \Delta R_2(10) = n_d \frac{\text{bits}}{\text{sub-ch.use}} \tag{A.65}$$

$$R_1 + \Delta R_1(10) = R_2 + \Delta R_2(01) = n_d - n_c \frac{\text{bits}}{\text{sub-ch.use}}. \tag{A.66}$$

This proves the achievability bounds in Theorem 14 for VWI.

A.1.4.4 Weak Interference

Consider the achievability scheme depicted in Figure A.2a. By (A.44) and (A.45),

$$R_1 + \Delta R_1(01) = R_2 + \Delta R_2(10) = 2(n_d - n_c) \frac{\text{bits}}{\text{sub-ch.use}} \tag{A.67}$$

$$R_1 + \Delta R_1(10) = R_2 + \Delta R_1(01) = n_c \frac{\text{bits}}{\text{sub-ch.use}}. \tag{A.68}$$

Combining (A.67) and (A.68), we obtain the achievability bounds in Theorem 14 for WI.

A.2 Proofs for the Ergodic Setup

A.2.1 Proof of (4.16) in Theorem 12

The bound (4.16) coincides with [65, Th. 3.1]. However, [65, Th. 3.1] derives (4.16) for the considered channel model with $T = 1$ and feedback. In this section we show that (4.16) also holds for general T in the no-feedback case. We follow along the lines of the proof of [65, Thm 3.1]. We begin by applying Fano's inequality to obtain

$$
\begin{aligned}
N(R_1 - \epsilon_{1K}) \leq & I(W_1; \boldsymbol{Y}_1^K \big| B_1^K) \\
= & \sum_{k=1}^{K} \left[H(\boldsymbol{Y}_{1,k} \big| \boldsymbol{Y}_1^{k-1}, B_1^K) - H(\boldsymbol{Y}_{1,k} \big| W_1, \boldsymbol{Y}_1^{k-1}, B_1^K) \right] \\
\stackrel{(a)}{=} & \sum_{k=1}^{K} \Big[H(\boldsymbol{Y}_{1,k} \big| \boldsymbol{Y}_1^{k-1}, B_{1,k}, B_1^{k-1}, B_{1,k+1}^K) \\
& \quad - H(B_{1,k}\mathsf{S}_{n_c}\boldsymbol{X}_{2,k} \big| \{B_{1,\ell}\mathsf{S}_{n_c}\boldsymbol{X}_{2,\ell}\}_{\ell=1}^{k-1}, W_1, B_1^K) \Big] \\
= & \sum_{k=1}^{K} \Big[(1-p) H(\boldsymbol{Y}_{1,k} \big| \boldsymbol{Y}_1^{k-1}, B_{1,k} = 0, B_1^{k-1}, B_{1,k+1}^K) \\
& \quad + pH(\boldsymbol{Y}_{1,k} \big| \boldsymbol{Y}_1^{k-1}, B_{1,k} = 1, B_1^{k-1}, B_{1,k+1}^K) \\
& \quad - pH(\mathsf{S}_{n_c}\boldsymbol{X}_{2,k} \big| \{B_{1,\ell}\mathsf{S}_{n_c}\boldsymbol{X}_{2,\ell}\}_{\ell=1}^{k-1}, W_1, B_{1,k} = 1, B_{1,k+1}^K, B_1^{k-1}) \Big] \\
\stackrel{(b)}{\leq} & \sum_{k=1}^{K} \Big[(1-p) H(\mathsf{S}_{n_d}\boldsymbol{X}_{1,k} | B_{1,k} = 0) + pH(\boldsymbol{Y}_{1,k} | B_{1,k} = 1) \\
& \quad - pH(\mathsf{S}_{n_c}\boldsymbol{X}_{2,k} \big| \{B_{1,\ell}\mathsf{S}_{n_c}\boldsymbol{X}_{2,\ell}\}_{\ell=1}^{k-1}, B_1^{k-1}) \Big]
\end{aligned}
\tag{A.69}
$$

where $\epsilon_{1K} \to 0$ as $K \to \infty$. Here, (a) follows because (W_1, B_1^K) determine $\boldsymbol{X}_1^K$, so we can subtract the contribution of $\boldsymbol{X}_1^K$ in the second entropy and by evaluating the entropy for different interference states. Step (b) follows because $(B_1^{k-1}, \mathbf{X}_2^k)$ are independent of $(B_{1,k}^K, W_1)$ (which in turn follows because $\boldsymbol{X}_2^K$ only depends on (B_2^K, W_2), which is independent of (B_1^K, W_1)) and because conditioning reduces entropy.

Likewise, we have

$$
\begin{aligned}
N(R_2 - \epsilon_{2K}) &\leq I(W_2; \boldsymbol{Y}_2^K \big| B_2^K) \\
&\overset{(a)}{\leq} I(W_2; \boldsymbol{Y}_1^K, \boldsymbol{Y}_2^K \big| W_1, B_1^K, B_2^K) \\
&= H(\boldsymbol{Y}_1^K, \boldsymbol{Y}_2^K \big| W_1, B_1^K, B_2^K) \\
&= \sum_{k=1}^{K} H(\boldsymbol{Y}_{1,k}, \boldsymbol{Y}_{2,k} \big| W_1, B_1^K, B_2^K, \boldsymbol{Y}_1^{k-1}, \boldsymbol{Y}_2^{k-1}) \\
&\overset{(b)}{\leq} \sum_{k=1}^{K} H(\mathsf{S}_{n_c}\boldsymbol{X}_{2,k}, \mathsf{S}_{n_d}\boldsymbol{X}_{2,k} \big| W_1, B_1^K, \{B_{1,\ell}\mathsf{S}_{n_c}\boldsymbol{X}_{2,\ell}\}_{\ell=1}^{k-1}) \\
&\overset{(c)}{\leq} \sum_{k=1}^{K} \Big[H(\mathsf{S}_{n_c}\boldsymbol{X}_{2,k} \big| \{B_{1,\ell}\mathsf{S}_{n_c}\boldsymbol{X}_{2,\ell}\}_{\ell=1}^{k-1}, B_1^{k-1}) \\
&\qquad + H(\mathsf{S}_{n_d}\boldsymbol{X}_{2,k} \big| \mathsf{S}_{n_c}\boldsymbol{X}_{2,k}) \Big]
\end{aligned}
\tag{A.70}
$$

where $\epsilon_{2K} \to 0$ as $K \to \infty$. Here, (a) follows because W_2, W_1 and B_1^K are independent. Step (b) follows because (W_1, B_1^K) determines $\boldsymbol{X}_1^K$, so we can subtract its contribution from $(\boldsymbol{Y}_{1,k}, \boldsymbol{Y}_{2,k})$, because $\boldsymbol{Y}_{1,k} \oplus \mathsf{S}_{n_d}\boldsymbol{X}_{1,k} = B_{1,k}\mathsf{S}_{n_c}\mathbf{X}_{2,k}$ has a lower entropy than $\mathsf{S}_{n_c}\mathbf{X}_{2,k}$, and because conditioning reduces entropy. Step (c) follows by the chain rule, and because conditioning reduces entropy.

Combining (A.69) and (A.70) yields

$$
\begin{aligned}
&N(R_1 + pR_2) - N(\epsilon_{1K} + p\epsilon_{2K}) \\
&\qquad \leq \sum_{k=1}^{K} \Big[(1-p) H(\mathsf{S}_{n_d}\boldsymbol{X}_{1,k} | B_{1,k} = 0) \\
&\qquad\qquad + pH(\boldsymbol{Y}_{1,k} | B_{1,k} = 1) + pH(\mathsf{S}_{n_d}\boldsymbol{X}_{2,k} | \mathsf{S}_{n_c}\boldsymbol{X}_{2,k}) \Big].
\end{aligned}
\tag{A.71}
$$

By maximizing the individual entropies in (A.71) over all input distributions, dividing both sides of (A.71) by $N = KT$, and by letting then K tend to infinity, we obtain that

$$
R_1 + pR_2 \leq (1-p)n_d + p[(n_d - n_c)^+ + \max(n_d, n_c)]. \tag{A.72}
$$

By symmetry, the same bound also holds for $R_2 + pR_1$. Thus, by averaging over the two cases, it follows that (A.72) is also an upper bound on $(R_1 + R_2)(1+p)/2$. The final result (4.16) follows by dividing (A.72) by $\frac{1+p}{2}$.

A.2.2 Achievability Proof of Theorem 13

In this section, we describe the achievability schemes that yield the rates presented in Theorem 13 for local CSIR. The bursty IC described in Section 4.2 is treated here as a set of n_d parallel sub-channels. We begin by considering VWI, WI and moderate interference (MI) regions and then we consider the strong interference (SI) region.

A.2.2.1 Scheme 1 (VWI; WI, MI for $0 \le p \le \frac{1}{2}$)

The achievability scheme is illustrated in Figure A.3a. In the figure, we present the normalized received signal at Rx_1, i.e., we represent graphically the time-k channel output $\boldsymbol{Y}_{1,k}$ given by (4.1), where the signal level from Tx_1 corresponds to $\mathsf{S}_{n_d}\boldsymbol{X}_{1,k}$ and the signal level from Tx_2 corresponds to $\mathsf{S}_{n_c}\boldsymbol{X}_{2,k}$, both normalized by n_d. In our scheme, the upper $n_d - n_c$ sub-channels (block $\boxed{A}$ in the figure) carry uncoded data (rate 1 bits/sub-channel use), while in the lower n_c channels (block $\boxed{B}$ in the figure) a capacity-achieving code of blocklength $N = KT$ for a binary erasure channel (BEC) with erasure probability p is used (with asymptotic rate $1-p$ bits/sub-channel use) [14, Sec. 7.1.5]. Block $\boxed{A}$ is received free of interference and can be directly decoded at the receiver. Block $\boxed{B}$ is affected by interference with probability (w.p.) p. Since the fading state $B_{i,k}$ is known to the i-th receiver, interfered slots are treated as erasures. Consequently, when K tends to infinity, user i achieves the rate $R_i = (n_d - n_c) + (1-p)n_c$. The sum rate R is thus given by

$$R = 2(n_d - pn_c), \quad n_d \ge n_c. \tag{A.73}$$

This scheme is tight for VWI and for WI and MI when $p \le \frac{1}{2}$.

A.2.2.2 Scheme 2 (WI, $\frac{1}{2} < p \le 1$)

We next consider the achievability scheme illustrated in Figure A.3b. In blocks $\boxed{A}$ and $\boxed{B}$ uncoded data is transmitted (rate 1 bits/sub-channel use), block $\boxed{C}$ carries the deterministic all-zeros sequence (rate 0 bit/sub-channel use) and in block $\boxed{D}$ a capacity-achieving code for the BEC (with asymptotic rate $1-p$ bits/sub-channel use) is used. As in Scheme 1, blocks $\boxed{A}$ and $\boxed{B}$ can be decoded without interference, and block $\boxed{D}$ is decoded by treating interfered symbols as erasures. The rate

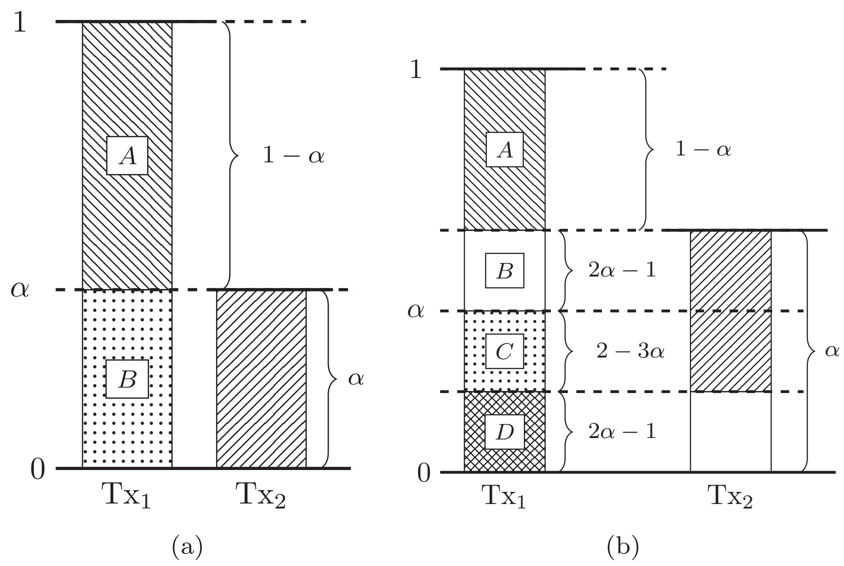


Figure A.3: Normalized signal levels at Rx_1. (**a**) VWI; WI; MI, $p \leq \frac{1}{2}$; (**b**) WI, $p > \frac{1}{2}$.

achieved by this scheme at user i is $R_i = (n_d - n_c) + (2n_c - n_d) + (1-p)(2n_d - 3n_c)$, so

$$R = 4(n_d - n_c) + p(6n_c - 4n_d), \quad \tfrac{2n_d}{3} \geq n_c \geq \tfrac{n_d}{2}. \tag{A.74}$$

A.2.2.3 Scheme 3 (SI, $0 \leq p \leq \frac{1}{2}$)

We use an achievability scheme similar to Scheme 1. Now, the upper $2n_d - n_c$ sub-channels carry a capacity-achieving code for a BEC with erasure probability p, and the lower $n_c - n_d$ sub-channels carry uncoded data. Consequently, when K tends to infinity, user i achieves the rate $R_i = (n_c - n_d) + (1-p)(2n_d - n_c)$. The sum rate $R = R_1 + R_2$ is thus given by

$$R = 2(1-2p)n_d + 2pn_c, \quad 2n_d \geq n_c \geq n_d. \tag{A.75}$$

This proves Theorem 13.

A.2.3 Proof of Theorem 16

In this section, we prove the converse bounds for global CSIRT and independent B_1^K and B_2^K.

A.2.3.1 Converse Bound (4.50) for Global CSIRT

By Fano's inequality, we have

$$
\begin{aligned}
N(R_1-\epsilon_{1K}) &\le I(W_1;\mathbf{Y}_1^K|\mathbf{B}^K)\\
&\overset{(a)}{=}\sum_{k=1}^{K}\left[H(\mathbf{Y}_{1,k}|\mathbf{Y}_1^{k-1},\mathbf{B}^K)-H(B_{1,k}\mathsf{S}_{n_c}\boldsymbol{X}_{2,k}|W_1,\mathbf{Y}_1^{k-1},\mathbf{B}^K)\right]\\
&=\sum_{k=1}^{K}\Big[(1-p)H(\mathbf{Y}_{1,k}|\mathbf{Y}_1^{k-1},B_{1,k}=0,B_1^{k-1},B_{1,k+1}^K,B_2^K)\\
&\qquad\quad+pH(\mathbf{Y}_{1,k}|\mathbf{Y}_1^{k-1},B_{1,k}=1,B_1^{k-1},B_{1,k+1}^K,B_2^K)\\
&\qquad\quad-pH(\mathsf{S}_{n_c}\boldsymbol{X}_{2,k}|W_1,\mathbf{Y}_1^{k-1},B_{1,k}=1,B_1^{k-1},B_{1,k+1}^K,B_2^K)\Big]\\
&\le\sum_{k=1}^{K}\Big[(1-p)H(\mathsf{S}_{n_d}\boldsymbol{X}_{1,k}|B_{1,k}=0)+pH(\mathbf{Y}_{1,k}|B_{1,k}=1)\\
&\qquad\quad-pH(\mathsf{S}_{n_c}\boldsymbol{X}_{2,k}|W_1,\mathbf{Y}_1^{k-1},B_{1,k}=1,B_1^{k-1},B_{1,k+1}^K,B_2^K)\Big]
\end{aligned}
\tag{A.76}
$$

where $\epsilon_{1K}\to 0$ as $K\to\infty$. Here, (a) follows because $(W_1,\mathbf{B}^K)$ determines $\boldsymbol{X}_{1,k}$, so we can subtract its contribution from the second entropy. Likewise,

$$
\begin{aligned}
N(R_2-\epsilon_{2K}) \le& I(W_2;\mathbf{Y}_2^K|\mathbf{B}^K)\\
\overset{(a)}{\le}& I(W_2;\mathbf{Y}_1^K,\mathbf{Y}_2^K|W_1,\mathbf{B}^K)\\
=&\sum_{k=1}^{K}H(\mathbf{Y}_{1,k},\mathbf{Y}_{2,k}|W_1,\mathbf{Y}_1^{k-1},\mathbf{Y}_2^{k-1},\mathbf{B}^K)\\
\overset{(b)}{\le}&\sum_{k=1}^{K}H(B_{1,k}\mathsf{S}_{n_c}\boldsymbol{X}_{2,k},\mathsf{S}_{n_d}\boldsymbol{X}_{2,k}|W_1,\mathbf{Y}_1^{k-1},\mathbf{B}^K)\\
\overset{(c)}{\le}&\sum_{k=1}^{K}\Big[(1-p)H(\mathsf{S}_{n_d}\boldsymbol{X}_{2,k}|,B_{1,k}=0)\\
&\qquad+pH(\mathsf{S}_{n_c}\boldsymbol{X}_{2,k}|W_1,\mathbf{Y}_1^{k-1},B_{1,k}=1,B_1^{k-1},B_{1,k+1}^K,B_2^K)\\
&\qquad+pH(\mathsf{S}_{n_d}\boldsymbol{X}_{2,k}|\mathsf{S}_{n_c}\boldsymbol{X}_{2,k},B_{1,k}=1)\Big]
\end{aligned}
\tag{A.77}
$$

where $\epsilon_{2K}\to 0$ as $K\to\infty$. Here, step (a) follows because W_2 and (W_1,B_1^K) are independent. Step (b) follows because $(W_1,\mathbf{B}^K)$ determines $\boldsymbol{X}_{1,k}$, so we can subtract its contribution from $\boldsymbol{Y}_{1,k}$ and $\boldsymbol{Y}_{2,k}$, and because conditioning reduces entropy. Step (c) follows by evaluating the entropies for different interference states and because conditioning reduces entropy. Combining (A.76) and (A.77) yields

$$
\begin{aligned}
&N(R_1 + R_2) - N(\epsilon_{1K} + \epsilon_{2K}) \\
&\leq \sum_{k=1}^{K} \left[(1-p) \left(H(\mathsf{S}_{n_d} \boldsymbol{X}_{1,k} | B_{1,k} = 0) + H(\mathsf{S}_{n_d} \boldsymbol{X}_{2,k} | B_{1,k} = 0) \right) \right. \\
&\qquad \left. + pH(\boldsymbol{Y}_{1,k} | B_{1,k} = 1) + pH(\mathsf{S}_{n_d} \boldsymbol{X}_{2,k} | \mathsf{S}_{n_c} \boldsymbol{X}_{2,k}, B_{1,k} = 1) \right].
\end{aligned} \tag{A.78}
$$

By maximizing the entropies in (A.78) over all input distributions, dividing by $N = KT$, and letting K tend to infinity, we obtain that

$$
R \leq 2(1-p)n_d + p \max(n_d, n_c) + p(n_d - n_c)^+ \tag{A.79}
$$

which is (4.50).

A.2.3.2 Converse Bound (4.51) for Global CSIRT

Let $\mathbf{b}^K$ denote the realizations of the interference states $\mathbf{B}^K$. We label the set of time indices where the pair $(b_{1,k}, b_{2,k})$ takes the value (0,1) by A; (1,1) by B; (1,0) by C; and (0,0) by D. We denote the length of each of these states by $j_{\mathsf{A}}, j_{\mathsf{B}}, j_{\mathsf{C}}$ and j_{D}, respectively. For example,

$$
\mathsf{A} \triangleq \{i = 1, \dots, K : \mathbf{b}_k = [1,1]\}
$$

and

$$
j_{\mathsf{A}} = \sum_{k=1}^{K} \mathbb{1}\{\mathbf{B} = [1,1]\}.
$$

These states are schematically shown in Figure A.4, where shaded areas correspond to $b_i = 1$.

For global CSIRT, $(\boldsymbol{X}_1^K, \boldsymbol{X}_2^K)$ may depend on $\mathbf{B}^K = \mathbf{b}^K$. We shall denote by $\boldsymbol{X}_i^{\mathsf{A}}, \boldsymbol{X}_i^{\mathsf{B}}, \boldsymbol{X}_i^{\mathsf{C}}$ and $\boldsymbol{X}_i^{\mathsf{D}}$ the $\boldsymbol{X}_{1,k}$'s with indices in $\mathsf{A}, \mathsf{B}, \mathsf{C}$ and D. For example, $\boldsymbol{X}_i^{\mathsf{A}} = \{\boldsymbol{X}_{i,k} : k \in \mathsf{A}\}$. At time k, the interference states $\mathbf{B}_k = \mathbf{b}_k$ can be in one of the 4 possible cases, as depicted in Figure A.4. The converse bound (4.51) is

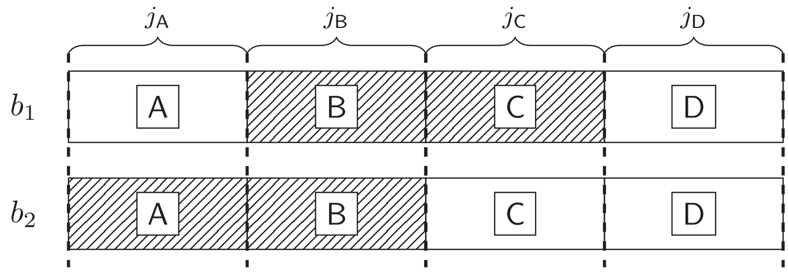


Figure A.4: Possible interference states.

proved as follows. We begin by applying Fano's inequality to obtain

$$\begin{aligned}
&N(R_1+R_2) - N(\epsilon_{1K}+\epsilon_{2K})\\
&\leq I(W_1;\boldsymbol{Y}_1^K|\mathbf{B}^K) + I(W_2;\boldsymbol{Y}_2^K|\mathbf{B}^K)\\
&= \sum_{\mathbf{b}\in\{0,1\}^K} \mathcal{P}(\mathbf{B}=\mathbf{b}^K)\left[I(W_1;\boldsymbol{Y}_1^K|\mathbf{B}^K=\mathbf{b}^K) + I(W_2;\boldsymbol{Y}_2^K|\mathbf{B}^K=\mathbf{b}^K)\right]
\end{aligned} \tag{A.80}$$

where $\epsilon_{1K}\to 0$ and $\epsilon_{2K}\to 0$ as $N\to\infty$. For every $\mathbf{b}^K$, we have

$$\begin{aligned}
&I(W_1;\boldsymbol{Y}_1^K|\mathbf{B}^K=\mathbf{b}^K) + I(W_2;\boldsymbol{Y}_2^K|\mathbf{B}^K=\mathbf{b}^K)\\
&= H(\boldsymbol{Y}_1^K|\mathbf{B}^K=\mathbf{b}^K) - H(\boldsymbol{Y}_1^K|W_1,\mathbf{B}^K=\mathbf{b}^K)\\
&\quad + H(\boldsymbol{Y}_2^K|\mathbf{B}^K=\mathbf{b}^K) - H(\boldsymbol{Y}_2^K|W_2,\mathbf{B}^K=\mathbf{b}^K)\\
&\overset{(a)}{=} H(\boldsymbol{Y}_1^\mathsf{C}|\mathbf{B}^K=\mathbf{b}^K) + H(\boldsymbol{Y}_1^\mathsf{A},\boldsymbol{Y}_1^\mathsf{B}|\boldsymbol{Y}_1^\mathsf{C},\mathbf{B}^K=\mathbf{b}^K)\\
&\quad + H(\boldsymbol{Y}_1^\mathsf{D}|\boldsymbol{Y}_1^\mathsf{A},\boldsymbol{Y}_1^\mathsf{B},\boldsymbol{Y}_1^\mathsf{C},\mathbf{B}^K=\mathbf{b}^K)\\
&\quad - H(\mathsf{S}_{n_c}\boldsymbol{X}_2^\mathsf{B},\mathsf{S}_{n_c}\boldsymbol{X}_2^\mathsf{C}|\mathbf{B}^K=\mathbf{b}^K)\\
&\quad + H(\boldsymbol{Y}_2^\mathsf{A}|\mathbf{B}^K=\mathbf{b}^K) + H(\boldsymbol{Y}_2^\mathsf{B},\boldsymbol{Y}_2^\mathsf{C}|\boldsymbol{Y}_2^\mathsf{A},\mathbf{B}^K=\mathbf{b}^K)\\
&\quad + H(\boldsymbol{Y}_2^\mathsf{D}|\boldsymbol{Y}_2^\mathsf{A},\boldsymbol{Y}_2^\mathsf{B},\boldsymbol{Y}_2^\mathsf{C},\mathbf{B}^K=\mathbf{b}^K)\\
&\quad - H(\mathsf{S}_{n_c}\boldsymbol{X}_1^\mathsf{A},\mathsf{S}_{n_c}\boldsymbol{X}_1^\mathsf{B}|\mathbf{B}^K=\mathbf{b}^K)\\
&\overset{(b)}{\leq} H(\boldsymbol{Y}_1^\mathsf{C}|\mathbf{B}^K=\mathbf{b}^K) + H(\boldsymbol{Y}_1^\mathsf{A},\boldsymbol{Y}_1^\mathsf{B}|\mathbf{B}^K=\mathbf{b}^K) + H(\boldsymbol{Y}_1^\mathsf{D}|\mathbf{B}^K=\mathbf{b}^K)\\
&\quad - H(\mathsf{S}_{n_c}\boldsymbol{X}_2^\mathsf{B},\mathsf{S}_{n_c}\boldsymbol{X}_2^\mathsf{C}|\mathbf{B}^K=\mathbf{b}^K)\\
&\quad + H(\boldsymbol{Y}_2^\mathsf{A}|\mathbf{B}^K=\mathbf{b}^K) + H(\boldsymbol{Y}_2^\mathsf{B},\boldsymbol{Y}_2^\mathsf{C}|\mathbf{B}^K=\mathbf{b}^K) + H(\boldsymbol{Y}_2^\mathsf{D}|\mathbf{B}^K=\mathbf{b}^K)\\
&\quad - H(\mathsf{S}_{n_c}\boldsymbol{X}_1^\mathsf{A},\mathsf{S}_{n_c}\boldsymbol{X}_1^\mathsf{B}|\mathbf{B}^K=\mathbf{b}^K)
\end{aligned} \tag{A.81}$$

where step (a) follows by the chain rule for entropy and because $(W_1,\mathbf{B}^K)$ determines $\boldsymbol{X}_1^K$, so we can subtract its contribution from the second and fourth entropy. Step (b) follows because conditioning reduces entropy. We next upper-bound (A.81) by combining the positive and negative entropies in areas B and C for user 1 and user 2; and areas A and B for user 2 and user 1:

$$\begin{aligned}
&I(W_1;\boldsymbol{Y}_1^K|\mathbf{B}^K=\mathbf{b}^K)+I(W_2;\boldsymbol{Y}_2^K|\mathbf{B}^K=\mathbf{b}^K)\\
&\overset{(a)}{\leq} H(\boldsymbol{Y}_1^{\mathsf{C}}|\mathbf{B}^K=\mathbf{b}^K)+H(\boldsymbol{Y}_1^{\mathsf{A}},\boldsymbol{Y}_1^{\mathsf{B}}|\mathsf{S}_{n_c}\boldsymbol{X}_1^{\mathsf{A}},\mathsf{S}_{n_c}\boldsymbol{X}_1^{\mathsf{B}},\mathbf{B}^K=\mathbf{b}^K)\\
&\quad+H(\boldsymbol{Y}_1^{\mathsf{D}}|\mathbf{B}^K=\mathbf{b}^K)+H(\boldsymbol{Y}_2^{\mathsf{A}}|\mathbf{B}^K=\mathbf{b}^K)\\
&\quad+H(\boldsymbol{Y}_2^{\mathsf{B}},\boldsymbol{Y}_2^{\mathsf{C}}|\mathsf{S}_{n_c}\boldsymbol{X}_2^{\mathsf{B}},\mathsf{S}_{n_c}\boldsymbol{X}_2^{\mathsf{C}},\mathbf{B}^K=\mathbf{b}^K)+H(\boldsymbol{Y}_2^{\mathsf{D}}|\mathbf{B}^K=\mathbf{b}^K)\\
&\leq H(\boldsymbol{Y}_1^{\mathsf{C}}|\mathbf{B}^K=\mathbf{b}^K)+H(\boldsymbol{Y}_1^{\mathsf{A}}|\mathsf{S}_{n_c}\boldsymbol{X}_1^{\mathsf{A}},\mathbf{B}^K=\mathbf{b}^K)\\
&\quad+H(\boldsymbol{Y}_1^{\mathsf{B}}|\mathsf{S}_{n_c}\boldsymbol{X}_1^{\mathsf{B}},\mathbf{B}^K=\mathbf{b}^K)+H(\boldsymbol{Y}_1^{\mathsf{D}}|\mathbf{B}^K=\mathbf{b}^K)\\
&\quad+H(\boldsymbol{Y}_2^{\mathsf{A}}|\mathbf{B}^K=\mathbf{b}^K)+H(\boldsymbol{Y}_2^{\mathsf{B}}|\mathsf{S}_{n_c}\boldsymbol{X}_2^{\mathsf{B}},\mathbf{B}^K=\mathbf{b}^K)\\
&\quad+H(\boldsymbol{Y}_2^{\mathsf{C}}|\mathsf{S}_{n_c}\boldsymbol{X}_2^{\mathsf{C}},\mathbf{B}^K=\mathbf{b}^K)+H(\boldsymbol{Y}_2^{\mathsf{D}}|\mathbf{B}^K=\mathbf{b}^K)
\end{aligned}\tag{A.82}$$

where step (a) follows because $H(F)-H(G)\leq H(F|G)$ for any random variables F and G. By maximizing the entropies in (A.82) over all input distributions, we obtain

$$\begin{aligned}
I(W_1;\boldsymbol{Y}_1^K|\mathbf{B}^K=\mathbf{b}^K)+&I(W_2;\boldsymbol{Y}_2^K\mathbf{B}^K=\mathbf{b}^K)\\
&\leq j_{\mathsf{A}}T[(n_d-n_c)^+ +\max(n_d,n_c)]\\
&+2j_{\mathsf{B}}T\max\{(n_d-n_c)^+,n_d\}\\
&+j_{\mathsf{C}}T[(n_d-n_c)^+ +\max(n_d,n_c)]+2j_{\mathsf{D}}T(n_d).
\end{aligned}\tag{A.83}$$

By dividing (A.83) by $N=KT$, and taking the limit as $K\to\infty$, we obtain

$$\begin{aligned}
&R_1+R_2\\
&\overset{(a)}{\leq}\lim_{K\to\infty}\frac{1}{KT}\sum_{\mathbf{b}^K}\mathcal{P}\{\mathbf{B}^K=\mathbf{b}^K\}\\
&\qquad\qquad\times\Big[I(W_1;\boldsymbol{Y}_1^K|\mathbf{B}^K=\mathbf{b}^K)+I(W_2;\boldsymbol{Y}_2^K|\mathbf{B}^K=\mathbf{b}^K)\Big]\\
&=\lim_{K\to\infty}\frac{1}{K}\Big[\mathsf{E}\big[j_{\mathsf{A}}[(n_c-n_d)^+ +\max(n_d,n_c)]+2j_{\mathsf{B}}[\max\{(n_d-n_c)^+,n_c\}]\big]\\
&\qquad+\mathsf{E}\big[j_{\mathsf{C}}[(n_d-n_c)^+ +\max(n_d,n_c)]+2j_{\mathsf{D}}n_d\big]\Big]
\end{aligned}\tag{A.84}$$

where (a) follows because $(\epsilon_{1K}+\epsilon_{2K})\to 0$ as $K\to\infty$. Next, we apply the dominated convergence theorem (DCT) [47, Sec. 1.34] to interchange limit and expectation. By the law of large numbers, we have that $\frac{j_\mathsf{A}}{K} \to p(1-p)$, $\frac{j_\mathsf{B}}{K} \to p^2$, $\frac{j_\mathsf{C}}{K} \to p(1-p)$, and $\frac{j_\mathsf{B}}{K} \to (1-p)^2$ almost surely as $K \to \infty$. By replacing these probabilities in (A.84), we thus obtain

$$\begin{aligned} R \leq\ & 2p(1-p)[(n_d - n_c)^+ + \max(n_d, n_c)] \\ & + 2p^2 \max\{(n_d - n_c)^+, n_c\} + 2(1-p)^2 n_d. \end{aligned} \tag{A.85}$$

This yields (4.51).

A.2.4 Proof of Theorem 17

In this section, we present the achievability schemes for global CSIRT and independent B_1^K and B_2^K. Let $\mathbf{b}^K$ denote the realizations of the interference states $\mathbf{B}^K$, and define $j_{\min} \triangleq \min(j_\mathsf{A}, j_\mathsf{B}, j_\mathsf{C})$. Consider the following achievable schemes.

A.2.4.1 Scheme 1 (MI, $0 \leq p \leq 1$)

Both transmitters employ uncoded transmission in the first $j_{\min}$ indices of regions A and C, respectively, and in the whole region D. Tx$_1$ copies the first $j_{\min}$ indices of region A in region B, while Tx$_2$ copies the first $j_{\min}$ indices of region C in B, aligned with those of user 1. The remaining indices are treated as a non-bursty IC attaining rate $r_{\text{ic}} = n_d - \frac{n_c}{2}$ [29].

To illustrate the decoding process, Figure A.5 shows the different normalized signals at the Rx$_1$ when $j_\mathsf{A} = j_\mathsf{B} = j_\mathsf{C} = j_\mathsf{D} = 1$. Tx$_1$ transmits the signals $\boxed{1}$, $\boxed{3}$, and $\boxed{4}$, in channel state A and B, C, and D, respectively. Similarly, Tx$_2$ transmits the signal $\boxed{2}$ in states B and C. Rx$_1$ has access to a clean copy of signal $\boxed{1}$ in region A, which can then be subtracted in state B to recover the interfering signal $\boxed{2}$. Since Tx$_2$ transmits the same signal in state C, the interference can then be canceled. Hence, signals $\boxed{3}$ and $\boxed{4}$ are recovered. For a given interference state and general A and B, C, and D, the rate attained by user i with this scheme is

$$R_i(\mathbf{b}^K) = n_d \tfrac{2j_{\min}}{K} + n_d \tfrac{j_\mathsf{D}}{K} + r_{\text{ic}} \tfrac{j_\mathsf{A} + j_\mathsf{B} + j_\mathsf{C} - 3j_{\min}}{K}. \tag{A.86}$$

Averaging (A.86) over $\mathbf{B}^K$, and letting $K \to \infty$, we obtain for the sum rate

$$\begin{aligned} R &= \lim_{K\to\infty} 2\mathsf{E}\left[n_d \tfrac{2j_{\min}}{K} + n_d \tfrac{j_\mathsf{D}}{K} + r_{\text{ic}} \tfrac{j_\mathsf{A} + j_\mathsf{B} + j_\mathsf{C} - 3j_{\min}}{K}\right] \\ &= 4n_d p_{\min} + 2n_d(1-p)^2 + \big(2n_d - n_c\big)\big(2p - p^2 - 3p_{\min}\big) \end{aligned} \tag{A.87}$$

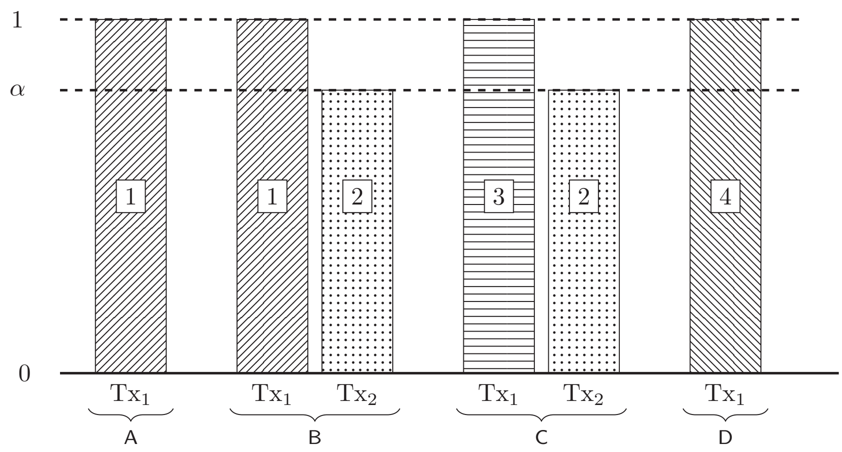


Figure A.5: Normalized by n_d signal levels at Rx_1 for MI and $j_{\mathsf{A}} = j_{\mathsf{B}} = j_{\mathsf{C}} = j_{\mathsf{D}}$.

where we changed the order of limit and expectation by appealing to the DCT, and used that, by the law of large numbers, $\frac{J_{\mathsf{A}}}{K} \to p(1-p)$, $\frac{J_{\mathsf{B}}}{K} \to p^2$, $\frac{J_{\mathsf{C}}}{K} \to p(1-p)$ and $\frac{J_{\mathsf{D}}}{K} \to (1-p)^2$ almost surely as $K \to \infty$.

A.2.4.2 Scheme 2 (SI, $0 \leq p \leq 1$)

Both transmitters employ uncoded transmission in the first $j_{\min}$ indices of states A and C. Tx_1 copies the lowest $2n_d - n_c$ sub-channels of the first $j_{\min}$ indices of region A into the highest $2n_d - n_c$ sub-channels and uses uncoded transmission in the lowest $n_c - n_d$ sub-channels of the corresponding sub-region in B. Tx_2 proceeds analogously but from region C to B. Both transmitters employ uncoded transmission in region D and treat the remaining indices as a non-bursty IC [29] with rate $\frac{n_c}{2}$.

To illustrate the decoding process, Figure A.6 shows the different normalized signals at the Rx_1 when $j_{\mathsf{A}} = j_{\mathsf{B}} = j_{\mathsf{C}} = j_{\mathsf{D}} = 1$. Tx_1 transmits the signals $(\boxed{1}, \boxed{2})$, $(\boxed{1}, \boxed{3})$, $\boxed{5}$ and $\boxed{6}$ in channel state A and B, C, and D, respectively. Similarly, Tx_2 transmits the signal $(\boxed{4}, \boxed{7})$ and $(\boxed{4}, \boxed{8})$ in states B and C, respectively. Rx_1 has access to a clean copy of signals $\boxed{1}$ and $\boxed{2}$ in region A, signal $\boxed{1}$ can then be subtracted in state B to recover the interfering signals $\boxed{4}$ and $\boxed{7}$. In state B, Rx_1 has access to signal $\boxed{3}$. Since Tx_2 transmits signal $\boxed{4}$ in state C, the interference can then be canceled. Hence, signal $\boxed{5}$ can be recovered. Finally, signal $\boxed{6}$ is recovered without interference. For a given interference state, and

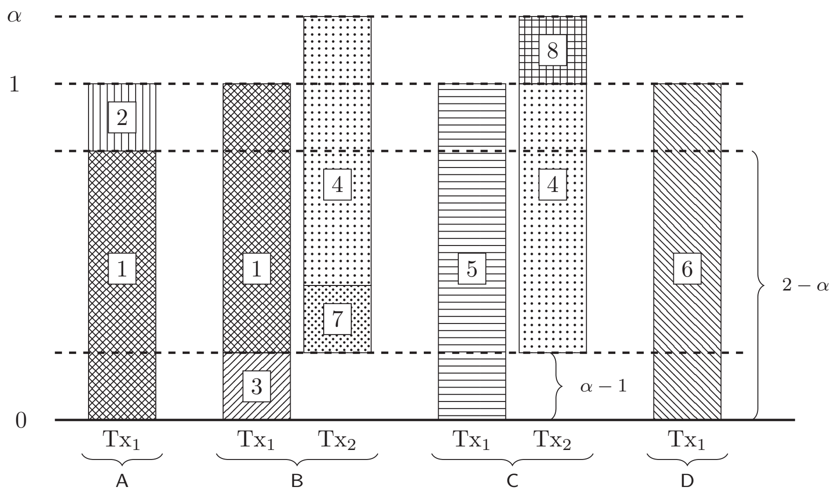


Figure A.6: Normalized by n_d signal levels at Rx_1 for SI.

general $j_{\mathsf{A}}, j_{\mathsf{B}}, j_{\mathsf{C}}, j_{\mathsf{D}}$, the rate attained by user i with this scheme is

$$R_i(\mathbf{b}^K) = (n_d + n_c)\tfrac{2j_{\min}}{K} + n_d\tfrac{j_{\mathsf{D}}}{K} + r_{\mathrm{ic}}\tfrac{j_{\mathsf{A}}+j_{\mathsf{B}}+j_{\mathsf{C}}-3j_{\min}}{K}. \tag{A.88}$$

Averaging (A.88) over $\mathbf{B}^K$, and letting $K \to \infty$, we obtain for the sum rate

$$\begin{aligned} R &= \lim_{K\to\infty} 2\mathsf{E}\left[(n_d + n_c)\tfrac{2j_{\min}}{K} + n_d\tfrac{j_{\mathsf{D}}}{K} + r_{\mathrm{ic}}\tfrac{j_{\mathsf{A}}+j_{\mathsf{B}}+j_{\mathsf{C}}-3j_{\min}}{K}\right] \\ &= 2(n_d + n_c)p_{\min} + 2n_d(1-p)^2 + n_c\big(2p - p^2 - 3p_{\min}\big). \end{aligned} \tag{A.89}$$

where we changed the order of limit and expectation by appealing to the DCT, and used that, by the law of large numbers, $\frac{J_{\mathsf{A}}}{K} \to p(1-p)$, $\frac{J_{\mathsf{B}}}{K} \to p^2$, $\frac{J_{\mathsf{C}}}{K} \to p(1-p)$ and $\frac{J_{\mathsf{D}}}{K} \to (1-p)^2$ almost surely as $K \to \infty$.

A.2.5 Proof of Theorem 18

The converse bound (4.57) for global CSIRT follows similar steps as in Appendix A.2.3.1 but considering $B_1^K = B_2^K = B^K$. We next present the converse bound (4.58) for global CSIRT when $B_1^K = B_2^K$. This bound follows by giving the extra information $(B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K)$ to Rx$_1$. By Fano's inequality, we have

$$
\begin{aligned}
N(R_1 - \epsilon_{1K}) &\leq I(W_1; \boldsymbol{Y}_1^K | B^K) \\
&\leq I(W_1; \boldsymbol{Y}_1^K, B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K) \\
&= I(W_1; B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K) + I(W_1; \boldsymbol{Y}_1^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) \\
&= H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K) + H(\boldsymbol{Y}_1^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) \\
&\quad - H(\boldsymbol{Y}_1^K | W_1, B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) \\
&= H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K) + H(\boldsymbol{Y}_1^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) \\
&\quad - H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K | B^K)
\end{aligned} \tag{A.90}
$$

where $\epsilon_{1K} \to 0$ as $K \to \infty$. Analogously, by giving the extra information $(B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K)$ to Rx$_2$, we obtain

$$
\begin{aligned}
N(R_2 - \epsilon_{2K}) \leq{}& H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K | B^K) + H(\boldsymbol{Y}_2^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K, B^K) \\
&- H(B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K | B^K)
\end{aligned} \tag{A.91}
$$

where $\epsilon_{2K} \to 0$ as $K \to \infty$. Thus, (A.90) and (A.91) yield

$$
\begin{aligned}
&N(R_1 + R_2) - N(\epsilon_{1K} + \epsilon_{2K}) \\
&\leq H(\boldsymbol{Y}_1^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) + H(\boldsymbol{Y}_2^K | B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K, B^K) \\
&= \sum_{k=1}^{K} \Big[H(\boldsymbol{Y}_{1,k} | \boldsymbol{Y}_1^{k-1}, B^K \mathsf{S}_{n_c} \boldsymbol{X}_1^K, B^K) \\
&\qquad + H(\boldsymbol{Y}_{2,k} | \boldsymbol{Y}_2^{k-1}, B^K \mathsf{S}_{n_c} \boldsymbol{X}_2^K, B^K) \Big] \\
&\leq \sum_{k=1}^{K} \left[H(\boldsymbol{Y}_{1,k} | B_k \mathsf{S}_{n_c} \boldsymbol{X}_{1,k}, B_k) + H(\boldsymbol{Y}_{2,k} | B_k \mathsf{S}_{n_c} \boldsymbol{X}_{2,k}, B_k) \right] \\
&\leq \sum_{k=1}^{K} \Big[(1-p) \left(H(\mathsf{S}_{n_d} \boldsymbol{X}_{1,k} | B_k = 0) + H(\mathsf{S}_{n_d} \boldsymbol{X}_{2,k} | B_k = 0) \right) \\
&\qquad + p(H(\boldsymbol{Y}_{1,k} | \mathsf{S}_{n_c} \boldsymbol{X}_{1,k}, B_k = 1) + H(\boldsymbol{Y}_{2,k} | \mathsf{S}_{n_c} \boldsymbol{X}_{2,k}, B_k = 1)) \Big]
\end{aligned} \tag{A.92}
$$

where we have used that conditioning reduces entropy. By maximizing the entropies in (A.92) over all input distributions, dividing by $N = KT$, and letting K tend to infinity, we obtain that

$$
R \leq 2(1-p)n_d + 2p \max\{(n_d - n_c)^+, n_c\}. \tag{A.93}
$$

This proves (4.58).

A.3 Achievability for Local CSIRT

In this appendix we present the achievability schemes for local CSIRT.

A.3.1 Very Weak Interference

The sum rate (4.27) coincides with that of local CSIR, which in this interference region is equal to the sum rate of global CSIRT. The achievability scheme presented in Section A.2.2.1 is thus optimal for local CSIRT and VWI.

A.3.2 Weak Interference

We follow a random-coding argument where the codebooks of Tx_1 and Tx_2 are drawn independent and identically distributed (IID) at random according to the distribution depicted in Figure A.7. Specifically, we divide the transmitted signal by Tx_1 into three regions. For each symbol (corresponding to a coherence block) we denote the bits in regions $\boxed{A}$, $\boxed{B}$ and $\boxed{C}$ by $\boldsymbol{X}_1^A$, $\boldsymbol{X}_1^B$ and $\boldsymbol{X}_1^C$, respectively. In each region the bits are IID, but they follow a different distribution.

- Regions $\boxed{A}$ and $\boxed{C}$: The bits $\boldsymbol{X}_1^A$ and $\boldsymbol{X}_1^C$ are IID with marginal probability mass function (pmf)

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \tag{A.94}$$

- Region $\boxed{B}$: The bits $\boldsymbol{X}_1^B$ are IID with marginal pmf

$$\begin{aligned} P_{X_1|B_1}(1|0) &= p_1 && \text{(A.95)}\\ P_{X_1|B_1}(1|1) &= p_2 && \text{(A.96)}\\ P_{X_1}(1) &= p_3 = (1-p)p_1 + pp_2. && \text{(A.97)} \end{aligned}$$

We further assume that $\boldsymbol{X}_1^A, \boldsymbol{X}_1^B$ and $\boldsymbol{X}_1^C$ are mutually independent. For Tx_2, the input distributions coincide with that of Tx_1 in the corresponding regions but with probabilities q_i instead of p_i, with $i = 1, 2$. Evaluating $I(\boldsymbol{X}_1; \boldsymbol{Y}_1|B_1)$ for

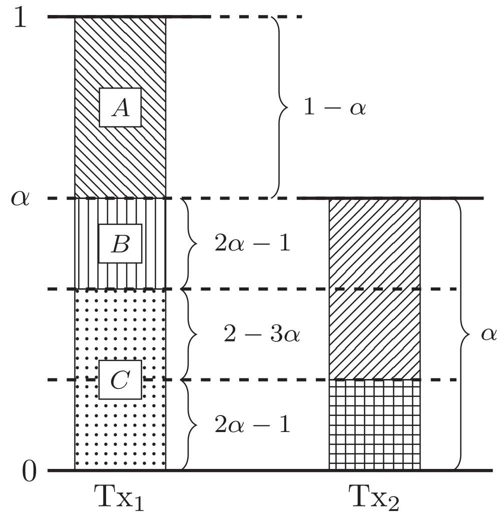


Figure A.7: Normalized signal levels at Rx_1 (WI).

these distributions, it follows that user 1 achieves the rate

$$\begin{aligned} R_1 =&(1-p)[(n_d-n_c)H_b(\tfrac{1}{2})+(2n_c-n_d)H_b(p_1)+(n_d-n_c)] \\ &+p(n_d-n_c)H_b(\tfrac{1}{2})+p(2n_c-n_d)\left[H_{\text{sum}}(p_2,\tfrac{1}{2})-H_b(q_3)\right] \\ &+p(2n_d-3n_c)(H_{\text{sum}}(\tfrac{1}{2},\tfrac{1}{2})-H_b(\tfrac{1}{2})) \\ &+p(2n_c-n_d)(H_{\text{sum}}(\tfrac{1}{2},q_3)-H_b(q_3)) \\ =&(n_d-n_c)+(1-p)[(n_d-n_c)+(2n_c-n_d)H_b(p_1)] \\ &+p(2n_c-n_d)(1-H_b(q_3)). \end{aligned} \tag{A.98}$$

Similarly, for user 2, we obtain (4.29).

A.3.3 Moderate Interference

We follow along similar lines to obtain the achievable rates for MI. However, in contrast to WI, for MI we need to consider different input distributions, depending on the value of α. In the proofs, we shall make use of the following auxiliary results, which can be proven by direct evaluation of the entropies considered.

Lemma 5 *Let X and $\tilde{X}$ be two binary random variables with joint pmf $P_{X\tilde{X}}(0,0)=P_{X\tilde{X}}(1,1)=\frac{\eta}{2}$, and $P_{X\tilde{X}}(0,1)=P_{X\tilde{X}}(1,0)=\frac{1-\eta}{2}$. Then,*

$$H(X|\tilde{X})=H(\tilde{X}|X)=H_b(\eta). \tag{A.99}$$

Lemma 6 *Let* $X, \tilde{X}$ *and* B *be binary random variables with joint pmf* $P_{X\tilde{X}B}(0,0,0) = P_{X\tilde{X}B}(1,1,0) = \frac{\eta_1}{2}(1-p)$, $P_{X\tilde{X}B}(0,1,0) = P_{X\tilde{X}B}(1,0,0) = \frac{1-\eta_1}{2}(1-p)$, $P_{X\tilde{X}B}(0,0,1) = P_{X\tilde{X}B}(1,1,1) = \frac{\eta_2}{2}p$, *and* $P_{X\tilde{X}B}(0,1,1) = P_{X\tilde{X}B}(1,0,1) = \frac{1-\eta_2}{2}p$. *Then,*

$$H(\tilde{X}|X,B) = (1-p)H_b(\eta_1) + pH_b(\eta_2) \tag{A.100}$$

and

$$H(\tilde{X}|X) = H_b\big((1-p)\eta_1 + p\eta_2\big). \tag{A.101}$$

Lemma 7 *Let* X_1 *and* $\tilde{X}_1$ *be two binary random variables with joint pmf* $P_{X_1\tilde{X}_1}(0,0) = P_{X_1\tilde{X}_1}(1,1) = \frac{\eta_1}{2}$ *and* $P_{X_1\tilde{X}_1}(0,1) = P_{X_1\tilde{X}_1}(1,0) = \frac{1-\eta_1}{2}$. *Similarly, let the pair of binary random variables* X_2 *and* $\tilde{X}_2$ *be independent of* X_1 *and* $\tilde{X}_1$ *have the same joint pmf but with parameter* η_2. *Further let* $Z \sim Ber(p_z)$. *Then,*

$$H(X_1|\tilde{X}_1 \oplus \tilde{X}_2, X_2) = H(\tilde{X}_1 \oplus \tilde{X}_2|X_1, X_2) = H_{sum}(\eta_1, \eta_2) \tag{A.102}$$

and

$$H(X_1 \oplus Z|\tilde{X}_1 \oplus \tilde{X}_2, X_2) = H_{sum}(p_z, \eta_1(1-\eta_2) + \eta_2(1-\eta_1)). \tag{A.103}$$

To derive the achievable rates for MI, we again follow a random-coding argument where the codebooks are drawn IID at random. We next describe the input distributions for different values of α:

A.3.3.1 MI, $\frac{2}{3} < \alpha \leq \frac{3}{4}$

Consider the regions shown in Figure A.8 for the received signal at Rx_1. For the transmitted signal X_1, we denote the bits in region $\boxed{j}$ by $\boldsymbol{X}_1^j$, $j = \{A, \ldots, F\}$. In each of these regions we consider the following input distributions:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: We group the bits $\boldsymbol{X}_1^A$ and $\boldsymbol{X}_1^{\tilde{A}}$ in pairs, and we let each of these pairs $(X_1, \tilde{X}_1)$ be IID and have the distribution from Lemma 6 with

$\eta_2 = 1$, i.e., their marginal pmf is

$$P_{X_1\tilde{X}_1|B_1}(0,0|0) = P_{X_1\tilde{X}_1|B_1}(1,1|0) = \frac{\eta_1}{2} \tag{A.104}$$
$$P_{X_1\tilde{X}_1|B_1}(0,1|0) = P_{X_1\tilde{X}_1|B_1}(1,0|0) = \frac{1-\eta_1}{2} \tag{A.105}$$
$$P_{X_1\tilde{X}_1|B_1}(0,0|1) = P_{X_1\tilde{X}_1|B_1}(1,1|1) = \frac{1}{2} \tag{A.106}$$
$$P_{X_1\tilde{X}_1|B_1}(0,1|1) = P_{X_1\tilde{X}_1|B_1}(1,0|1) = 0 \tag{A.107}$$
$$P_{\tilde{X}_1|X_1}(1|1) = \tilde{\eta} = p + \eta_1(1-p). \tag{A.108}$$

where $\frac{1}{2} \leq \eta_1 \leq 1$.

- Regions $\boxed{B}$ and $\boxed{F}$: The bits $\boldsymbol{X}_1^B$ and $\boldsymbol{X}_1^F$ are IID with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) \quad = \quad \tfrac{1}{2}. \tag{A.109}$$

- Region $\boxed{C}$: The bits $\boldsymbol{X}_1^C$ are IID with marginal pmf

$$P_{X_1|B_1}(1|0) \quad = \quad p_1 \tag{A.110}$$
$$P_{X_1|B_1}(1|1) \quad = \quad p_2 \tag{A.111}$$
$$P_{X_1}(1) \quad = \quad p_3 = (1-p)p_1 + pp_2. \tag{A.112}$$

- Region $\boxed{D}$: The bits $\boldsymbol{X}_1^D$ are IID with marginal pmf

$$P_{X_1|B_1}(1|0) \quad = \quad \tilde{p}_1 \tag{A.113}$$
$$P_{X_1|B_1}(1|1) \quad = \quad \tilde{p}_2 \tag{A.114}$$
$$P_{X_1}(1) \quad = \quad \tilde{p}_3 = (1-p)\tilde{p}_1 + p\tilde{p}_2. \tag{A.115}$$

- Region $\boxed{E}$: The bits $\boldsymbol{X}_1^E$ are IID with marginal pmf

$$P_{X_1|B_1}(1|0) \quad = \quad \hat{p}_1 \tag{A.116}$$
$$P_{X_1|B_1}(1|1) \quad = \quad 0 \tag{A.117}$$
$$P_{X_1}(1) \quad = \quad \hat{p}_3 = (1-p)\hat{p}_1. \tag{A.118}$$

Furthermore, we assume that $\boldsymbol{X}_1^j$, $j = \{A, \ldots, F\}$ are independent. For user 2, the input distributions coincide with that of user 1 in the corresponding

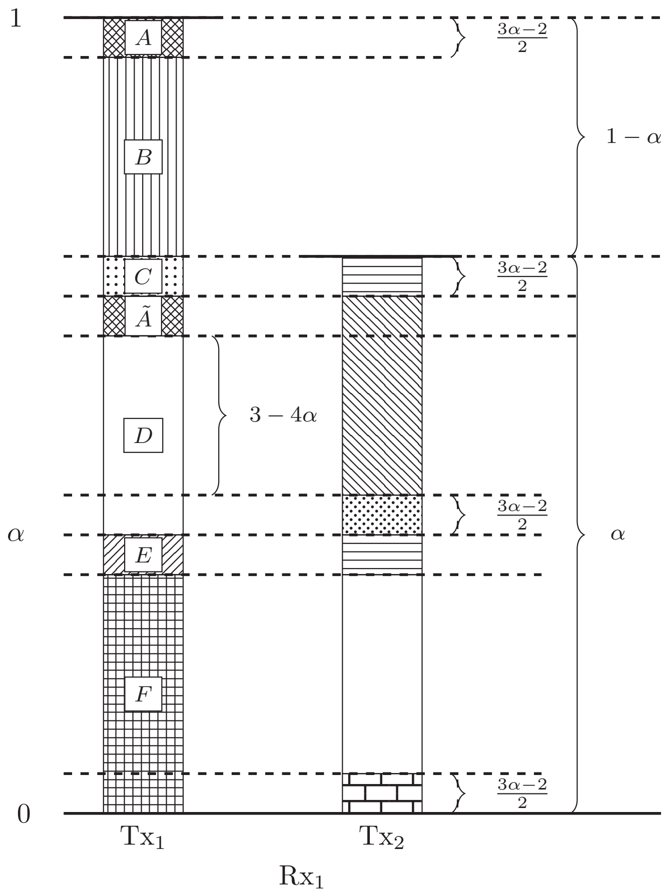


Figure A.8: Normalized signal levels at Rx_1 (MI) for $\frac{2}{3} < \alpha \leq \frac{3}{4}$.

regions, but with parameters q_i instead of p_i, $\tilde{q}_i$ instead of $\tilde{p}_i$, $\hat{q}_1$ instead of $\hat{p}_1$, and γ_i instead of η_i. From the random-coding argument, we know that the rate $R_1 = \frac{1}{N} I(\boldsymbol{X}_1^K; \boldsymbol{Y}_1^K | B_1)$ is achievable. Since the distributions considered are temporally IID, it suffices to evaluate $I(\boldsymbol{X}_1; \boldsymbol{Y}_1 | B_1)$ for one coherence block, obtaining

$$\begin{aligned}
TR_1 =& I(\boldsymbol{X}_1; \boldsymbol{Y}_1^A | B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^{\tilde{A}} | \boldsymbol{Y}_1^A, B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^E | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, B_1) \\
&+ I(\boldsymbol{X}_1; \boldsymbol{Y}_1^C | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^E, B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^B | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^E, \boldsymbol{Y}_1^C, B_1) \\
&+ I(\boldsymbol{X}_1; \boldsymbol{Y}_1^D | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^E, \boldsymbol{Y}_1^C, \boldsymbol{Y}_1^B, B_1) \\
&+ I(\boldsymbol{X}_1; \boldsymbol{Y}_1^F | \boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^E, \boldsymbol{Y}_1^C, \boldsymbol{Y}_1^B, \boldsymbol{Y}_1^D, B_1) \\
=& (1-p)[H(\boldsymbol{X}_1^A | B_1 = 0) + H(\boldsymbol{X}_1^{\tilde{A}} | \boldsymbol{X}_1^A, B_1 = 0) + H(\boldsymbol{X}_1^E | B_1 = 0) \\
&\quad + H(\boldsymbol{X}_1^C | B_1 = 0) + H(\boldsymbol{X}_1^B | B_1 = 0) + H(\boldsymbol{X}_1^D | B_1 = 0) \\
&\quad + H(\boldsymbol{X}_1^F | B_1 = 0)] \\
&+ p[H(\boldsymbol{X}_1^A | B_1 = 1) + H(\boldsymbol{X}_1^{\tilde{A}} \oplus \boldsymbol{X}_2^{\tilde{A}} | \boldsymbol{X}_1^A, B_1 = 1) - H(\boldsymbol{X}_2^{\tilde{A}}) \\
&\quad + H(\boldsymbol{X}_1^E \oplus \boldsymbol{X}_2^E | B_1 = 1) \\
&\quad - H(\boldsymbol{X}_2^E) + H(\boldsymbol{X}_1^C \oplus \boldsymbol{X}_2^C | \boldsymbol{X}_1^E \oplus \boldsymbol{X}_2^E, B_1 = 1) \\
&\quad - H(\boldsymbol{X}_2^C | \boldsymbol{X}_2^E) + H(\boldsymbol{X}_1^B | B_1 = 1) + H(\boldsymbol{X}_1^D \oplus \boldsymbol{X}_2^D | B_1 = 1) \\
&\quad - H(\boldsymbol{X}_2^D) + H(\boldsymbol{X}_1^F \oplus \boldsymbol{X}_2^F | B_1 = 1) - H(\boldsymbol{X}_2^F)].
\end{aligned} \tag{A.119}$$

By Lemma 6, we have that

$$H(\boldsymbol{X}_1^{\tilde{A}} | \boldsymbol{X}_1^A, B_1 = 0) = T \frac{3n_c - 2n_d}{2} H_b(\eta_1). \tag{A.120}$$

Furthermore, by Lemma 7, we have that

$$\begin{aligned}
H(\boldsymbol{X}_1^C \oplus \boldsymbol{X}_2^C | \boldsymbol{X}_1^E \oplus \boldsymbol{X}_2^E, B_1 = 1) &= H(\boldsymbol{X}_1^C \oplus \boldsymbol{X}_2^C | \boldsymbol{X}_2^E, B_1 = 1) \\
&= T \frac{3n_c - 2n_d}{2} H_{\text{sum}}(p_2, \tilde{\gamma})
\end{aligned} \tag{A.121}$$

because for the bits $\boldsymbol{X}_1^C$, $P_{X_1^C | B_1}(1|1) = 0$. Similarly, we have

$$H_{\text{sum}}(\boldsymbol{X}_1^{\tilde{A}} \oplus \boldsymbol{X}_2^{\tilde{A}} | \boldsymbol{X}_1^A, B_1 = 1) = T \frac{3n_c - 2n_d}{2} H_{\text{sum}}(1, \tfrac{1}{2}) = T \frac{3n_c - 2n_d}{2}. \tag{A.122}$$

The terms in the other regions follow analogously. Therefore, using (A.119) we obtain the rate

$$
\begin{aligned}
R_1 =& (n_d - n_c) + (1-p)\Big[\left(\tfrac{3n_c-2n_d}{2}\right)(H_b(\eta_1) + H_b(\hat{p}_1) + H_b(p_1)) \\
&\qquad\qquad + \left(\tfrac{4n_d-5n_c}{2}\right) H_b(\tilde{p}_1) + (n_d - n_c)\Big] \\
&+ p\Big[\left(\tfrac{3n_c-2n_d}{2}\right)(1 + H_{\mathrm{sum}}(p_2, \tilde{\gamma}) - H_b(\tilde{\gamma}) + H_{\mathrm{sum}}(\tilde{p}_2, q_3) - H_b(q_3) - H_b(\hat{q}_3)) \\
&\quad + \left(\tfrac{4n_d-5n_c}{2}\right)(1 - H_b(\tilde{q}_3))\Big].
\end{aligned}
\tag{A.123}
$$

Similarly, user 2 achieves the rate (4.31).

A.3.3.2 MI, $\frac{3}{4} \leq \alpha \leq \frac{4}{5}$

We use a similar transmission strategy as for the case where $\frac{2}{3} \leq \alpha \leq \frac{3}{4}$ (Section A.3.3.1), but where the regions have different sizes; see Figure A.9. Following the same steps as in Section A.3.3.1, we obtain the achievable rates (4.32) for R_1 and (4.33) for R_2.

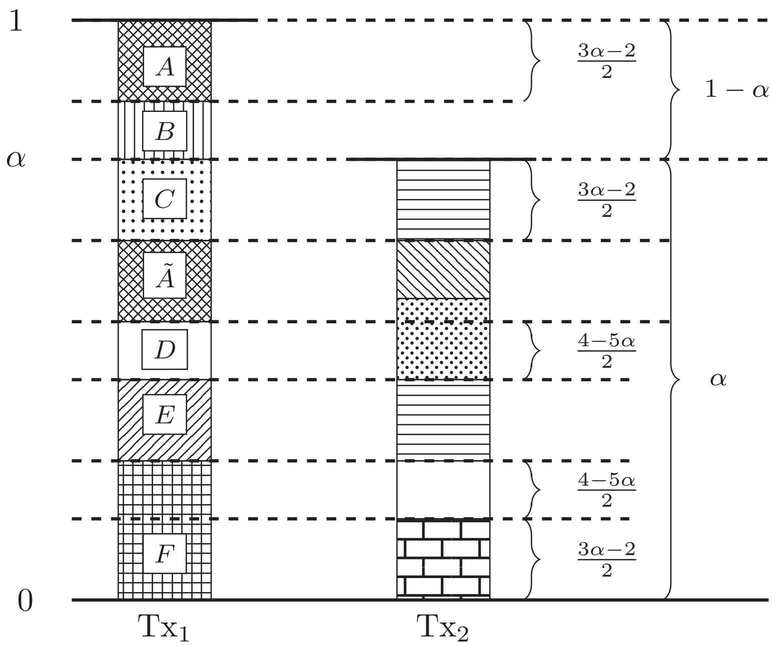


Figure A.9: Normalized signal levels at Rx_1 for $\frac{3}{4} \leq \alpha \leq \frac{4}{5}$ MI

A.3.3.3 MI, $\alpha = \frac{6}{7}$

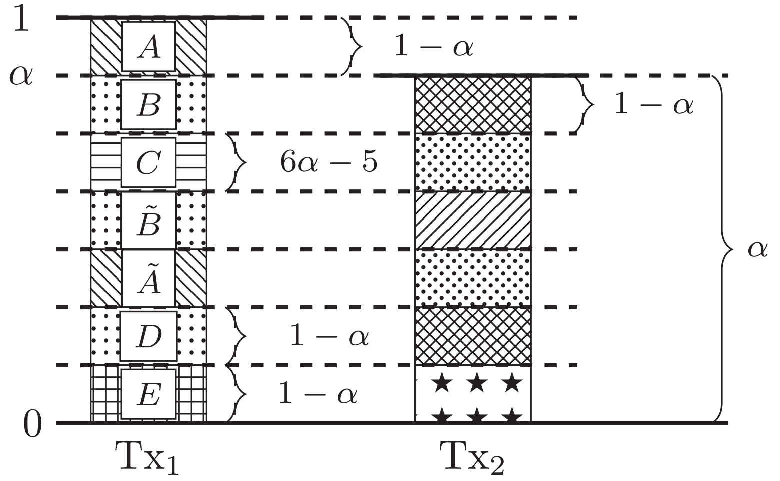


Figure A.10: Normalized signal levels at Rx_1 for $\alpha = \frac{6}{7}$ MI

In this subsection we consider the particular case $\alpha = \frac{6}{7}$. The proposed achievability scheme features two nested regions with a certain correlation. In particular, we consider the division of the bit-pipes for the transmitted signal Tx_1 in the subregions shown in Figure A.10. The input distributions considered in each of these regions are described next (for Tx_2, we shall consider the same input distributions parameterized by q_i, $\hat{q}_1$, γ_1 and γ', instead of p_i, $\hat{p}_1$, η_1 and η'):

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $\boldsymbol{X}_1^A$ and $\boldsymbol{X}_1^{\tilde{A}}$ are grouped in IID pairs with the marginal pmf given by (A.104)–(A.108).
- Regions $\boxed{B}$ and $\boxed{\tilde{B}}$: The bits $\boldsymbol{X}_1^B$ and $\boldsymbol{X}_1^{\tilde{B}}$ are grouped in IID pairs with marginal pmf

$$P_{X_1\tilde{X}_1|B_1}(0,0|0) = P_{X_1\tilde{X}_1|B_1}(1,1|0) = \frac{\eta'}{2} \tag{A.124}$$

$$P_{X_1\tilde{X}_1|B_1}(0,1|0) = P_{X_1\tilde{X}_1|B_1}(1,0|0) = \frac{1-\eta'}{2} \tag{A.125}$$

$$P_{X_1\tilde{X}_1|B_1}(0,0|1) = P_{X_1\tilde{X}_1|B_1}(1,1|1) = \frac{\eta'}{2} \tag{A.126}$$

$$P_{X_1\tilde{X}_1|B_1}(0,1|1) = P_{X_1\tilde{X}_1|B_1}(1,0|1) = \frac{1-\eta'}{2} \tag{A.127}$$

$$P_{\tilde{X}_1|X_1}(1|1) = \eta' \tag{A.128}$$

where $\frac{1}{2} \leq \eta' \leq 1$.

- Region $\boxed{C}$: The bits $\boldsymbol{X}_1^C$ are IID with marginal pmf

$$\begin{aligned} P_{X_1|B_1}(1|0) &= p_1 && \text{(A.129)} \\ P_{X_1|B_1}(1|1) &= p_2 && \text{(A.130)} \\ P_{X_1}(1) &= p_3 = (1-p)p_1 + pp_2. && \text{(A.131)} \end{aligned}$$

- Region $\boxed{D}$: The bits $\boldsymbol{X}_1^D$ are IID with marginal pmf

$$\begin{aligned} P_{X_1|B_1}(1|0) &= \hat{p}_1 && \text{(A.132)} \\ P_{X_1|B_1}(1|1) &= 0 && \text{(A.133)} \\ P_{X_1}(1) &= \hat{p}_3 = (1-p)\hat{p}_1. && \text{(A.134)} \end{aligned}$$

- Region $\boxed{E}$: The bits $\boldsymbol{X}_1^E$ are IID with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \tag{A.135}$$

Furthermore, we assume that $\boldsymbol{X}_i^j$, i=1,2, $j = \{A, B, C, D, E\}$ are mutually independent. For the input distributions described above, we obtain for user 1 that

$$\begin{aligned} TR_1 =& I(\boldsymbol{X}_1; \boldsymbol{Y}_1^A|B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^{\tilde{A}}|\boldsymbol{Y}_1^A, B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^D|\boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, B_1) \\ &+ I(\boldsymbol{X}_1; \boldsymbol{Y}_1^B|\boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^D, B_1) + I(\boldsymbol{X}_1; \boldsymbol{Y}_1^{\tilde{B}}|\boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^D, \boldsymbol{Y}_1^B, B_1) \\ &+ I(\boldsymbol{X}_1; \boldsymbol{Y}_1^C|\boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^D, \boldsymbol{Y}_1^B, \boldsymbol{Y}_1^{\tilde{B}}, B_1) \\ &+ I(\boldsymbol{X}_1; \boldsymbol{Y}_1^E|\boldsymbol{Y}_1^A, \boldsymbol{Y}_1^{\tilde{A}}, \boldsymbol{Y}_1^D, \boldsymbol{Y}_1^B, \boldsymbol{Y}_1^{\tilde{B}}, \boldsymbol{Y}_1^C, B_1) \\ =& (1-p)[H(\boldsymbol{X}_1^A|B_1=0) + H(\boldsymbol{X}_1^{\tilde{A}}|\boldsymbol{X}_1^A, B_1=0) + H(\boldsymbol{X}_1^D|B_1=0) \\ &\quad + H(\boldsymbol{X}_1^B|B_1=0) + H(\boldsymbol{X}_1^{\tilde{B}}|\boldsymbol{X}_1^B, B_1=0) + H(\boldsymbol{X}_1^C|B_1=0) \\ &\quad + H(\boldsymbol{X}_1^E|B_1=0)] \qquad\qquad \text{(A.136)} \\ &+ p[H(\boldsymbol{X}_1^A|B_1=1) + H(\boldsymbol{X}_1^{\tilde{A}} \oplus \boldsymbol{X}_2^{\tilde{A}}|\boldsymbol{X}_1^A, B_1=1) \\ &\quad - H(\boldsymbol{X}_2^{\tilde{A}}) + H(\boldsymbol{X}_1^D \oplus \boldsymbol{X}_2^D|B_1=1) \\ &\quad - H(\boldsymbol{X}_2^D) + H(\boldsymbol{X}_1^B \oplus \boldsymbol{X}_2^B|\boldsymbol{X}_1^D \oplus \boldsymbol{X}_2^D, B_1=1) - H(\boldsymbol{X}_2^B|\boldsymbol{X}_2^D) \\ &\quad + H(\boldsymbol{X}_1^{\tilde{B}} \oplus \boldsymbol{X}_2^{\tilde{B}}|\boldsymbol{X}_1^B \oplus \boldsymbol{X}_2^B, \boldsymbol{X}_1^D \oplus \boldsymbol{X}_2^D, B_1=1) - H(\boldsymbol{X}_2^{\tilde{B}}) \\ &\quad + H(\boldsymbol{X}_1^C \oplus \boldsymbol{X}_2^C|\boldsymbol{X}_1^{\tilde{A}} \oplus \boldsymbol{X}_2^{\tilde{A}}, \boldsymbol{X}_1^A, B_1=1) - H(\boldsymbol{X}_2^C|\boldsymbol{X}_2^{\tilde{A}}) \\ &\quad + H(\boldsymbol{X}_1^E \oplus \boldsymbol{X}_2^E|B_1=1) - H(\boldsymbol{X}_2^E)]. \end{aligned}$$

We next evaluate the different terms in (A.136) by applying Lemmas 6 and 7 to obtain

$$H(\boldsymbol{X}_1^{\tilde{A}}|\boldsymbol{X}_1^{A}, B_1 = 0) = T(n_d - n_c)H_b(\eta_1) \tag{A.137}$$

$$H(\boldsymbol{X}_1^{\tilde{B}}|\boldsymbol{X}_1^{B}, B_1 = 0) = T(n_d - n_c)H_b(\eta') \tag{A.138}$$

$$\begin{aligned} H(\boldsymbol{X}_1^{\tilde{A}} \oplus \boldsymbol{X}_2^{\tilde{A}}|\boldsymbol{X}_1^{A}, B_1 = 1) &= T(n_d - n_c)H_{\text{sum}}(1, \tfrac{1}{2}) \\ &= T(n_d - n_c). \end{aligned} \tag{A.139}$$

Similarly, using Lemma 7, and since for $\boldsymbol{X}_1^D$ we have that $P_{X_1|B_1}(0|1) = 1$, we obtain

$$\begin{aligned} &H(\boldsymbol{X}_1^{\tilde{B}} \oplus \boldsymbol{X}_2^{\tilde{B}}|\boldsymbol{X}_1^{B} \oplus \boldsymbol{X}_2^{B}, \boldsymbol{X}_1^{D} \oplus \boldsymbol{X}_2^{D}, B_1 = 1) \\ &\quad = H(\boldsymbol{X}_1^{\tilde{B}} \oplus \boldsymbol{X}_2^{\tilde{B}}|\boldsymbol{X}_1^{B} \oplus \boldsymbol{X}_2^{B}, \boldsymbol{X}_2^{D}, B_1 = 1) \\ &\quad = T(n_d - n_c)H_{\text{sum}}(q_3, \eta'(1-\tilde{\gamma}) + (1-\eta')\tilde{\gamma}). \end{aligned} \tag{A.140}$$

Combining (A.137)–(A.140) with (A.136) yields

$$\begin{aligned} R_1 =& (n_d - n_c) \\ &+ (1-p)\big[(6n_c - 5n_d)H_b(p_1) \\ &\qquad + (n_d - n_c)\left(2 + H_b(\eta_1) + H_b(\eta') + H_b(\hat{p}_1)\right)\big] \\ &+ p\big[(n_d - n_c)\Big(2 - H_b(\tilde{\gamma}) - H_b(\hat{q}_3) \\ &\qquad + H_{\text{sum}}(\eta'(1-\tilde{\gamma}) + (1-\eta')\tilde{\gamma}, q_3) - H_b(q_3)\Big) \\ &\quad + (6n_c - 5n_d)\left(H_{\text{sum}}(p_2, \gamma') - H_b(\gamma')\right)\big]. \end{aligned} \tag{A.141}$$

Following along similar lines, it can be shown that user 2 achieves the rate

$$\begin{aligned} R_2 =& (n_d - n_c) \\ &+ (1-p)\big[(6n_c - 5n_d)H_b(q_1) \\ &\qquad + (n_d - n_c)\left(2 + H_b(\gamma_1) + H_b(\gamma') + H_b(\hat{q}_1)\right)\big] \\ &+ p\big[(n_d - n_c)\Big(2 - H_b(\tilde{\eta}) - H_b(\hat{p}_3) \\ &\qquad + H_{\text{sum}}(\gamma'(1-\tilde{\eta}) + (1-\gamma')\tilde{\eta}, p_3) - H_b(p_3)\Big) \\ &\quad + (6n_c - 5n_d)\left(H_{\text{sum}}(q_2, \eta') - H_b(\eta')\right)\big]. \end{aligned} \tag{A.142}$$

A.3.3.4 MI, $\frac{4}{5} < \alpha < \frac{6}{7}$

We consider the input distribution depicted in Figure A.11a with:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(\boldsymbol{X}_1^A, \boldsymbol{X}_1^{\tilde{A}})$ are IID, with marginal pmf given by (A.104)-(A.108).
- Regions $\boxed{B}$ and $\boxed{\tilde{B}}$: The bits $(\boldsymbol{X}_1^B, \boldsymbol{X}_1^{\tilde{B}})$ are IID, with marginal pmf given by (A.124)-(A.128).
- Region $\boxed{C}$: The bits $\boldsymbol{X}_1^C$ are IID with marginal pmf

$$\begin{aligned} P_{X_1|B_1}(1|0) &= p_1 && \text{(A.143)}\\ P_{X_1|B_1}(1|1) &= p_2 && \text{(A.144)}\\ P_{X_1}(1) &= p_3 = (1-p)p_1 + pp_2. && \text{(A.145)} \end{aligned}$$

- Region $\boxed{D}$: The bits $\boldsymbol{X}_1^D$ are IID with marginal pmf

$$\begin{aligned} P_{X_1|B_1}(1|0) &= \hat{p}_1 && \text{(A.146)}\\ P_{X_1|B_1}(1|1) &= 0 && \text{(A.147)}\\ P_{X_1}(1) &= \hat{p}_3 = (1-p)\hat{p}_1. && \text{(A.148)} \end{aligned}$$

- Region $\boxed{E}$: The bits $\boldsymbol{X}_1^E$ are IID with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2} \qquad \text{(A.149)}$$

Furthermore, we assume that $\boldsymbol{X}_1^j$, $j = \{A, B, C, D, E\}$ are independent. For Tx_2, the input distributions coincide with that of Tx_1 in the corresponding regions, but with parameters q_i instead of p_i, $\hat{q}_1$ instead of $\hat{p}_1$, γ_i instead of η_i and γ' instead of η'. Following similar steps as in the previous sections, we obtain (4.34) for R_1 and (4.35) for R_2.

A.3.3.5 MI, $\frac{6}{7} < \alpha < 1$

The transmission strategy is similar to the one for $\frac{4}{5} < \alpha < \frac{6}{7}$ (Section A.3.3.4), but with different sizes for the regions $\boxed{A}$ - $\boxed{E}$, see Figure A.11b. Following similar steps as in previous sections, we obtain (4.36) for R_1 and (4.37) for R_2.

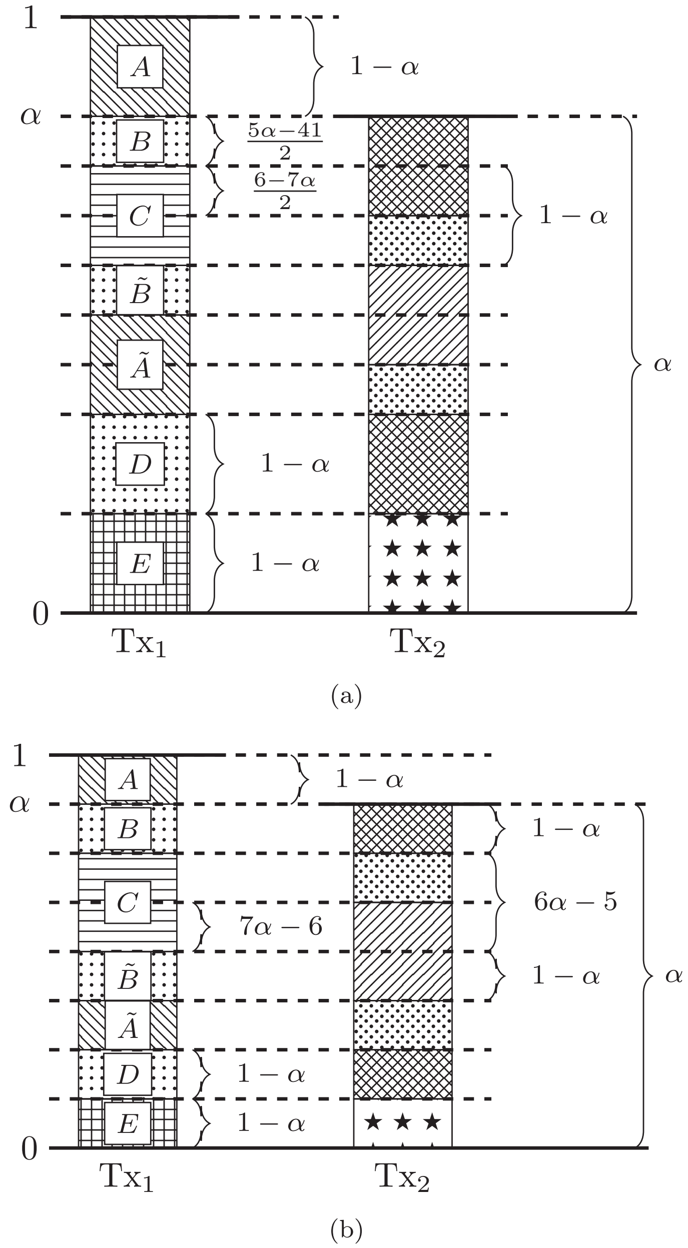


Figure A.11: Normalized signal levels at Rx_1. (**a**) (MI) for $\frac{4}{5} \leq \alpha \leq \frac{6}{7}$; (**b**) (MI) for $\frac{6}{7} \leq \alpha \leq 1$.

A.3.4 Strong Interference

To obtain the achievable rates for SI, we again need to consider different input distributions, depending on the value of α.

A.3.4.1 SI, $1 \leq \alpha \leq \frac{6}{5}$

We consider the input distribution depicted in Figure A.12a with:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(\boldsymbol{X}_1^A, \boldsymbol{X}_1^{\tilde{A}})$ are IID, with marginal pmf given by (A.104)–(A.108).
- Regions $\boxed{B}$ and $\boxed{\tilde{B}}$: The bits $(\boldsymbol{X}_1^B, \boldsymbol{X}_1^{\tilde{B}})$ are IID, with marginal pmf given by (A.124)–(A.128).
- Region $\boxed{C}$: The bits $\boldsymbol{X}_1^C$ are IID with marginal pmf

$$P_{X_1|B_1}(1|0) = p_1 \tag{A.150}$$

$$P_{X_1|B_1}(1|1) = p_2 \tag{A.151}$$

$$P_{X_1}(1) = p_3 = (1-p)p_1 + pp_2. \tag{A.152}$$

Furthermore, we assume that $\boldsymbol{X}_1^j$, $j = \{A, B, C\}$ are independent. For Tx_2, the input distributions coincide with that of Tx_1 in the corresponding regions, but with parameters q_i instead of p_i, γ_1 instead of η_1 and γ' instead of η'. Following similar steps as in previous sections, we obtain the achievable rate pair (4.38) and (4.39).

A.3.4.2 SI, $\frac{6}{5} \leq \alpha \leq \frac{4}{3}$

We consider the input distribution depicted in Figure A.12b with the following distributions:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(\boldsymbol{X}_1^A, \boldsymbol{X}_1^{\tilde{A}})$ are IID, with marginal pmf given by (A.104)–(A.108).
- Regions $\boxed{B}$ and $\boxed{D}$: The bits $\boldsymbol{X}_1^B$ and $\boldsymbol{X}_1^D$ are independent and temporally IID with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \tag{A.153}$$

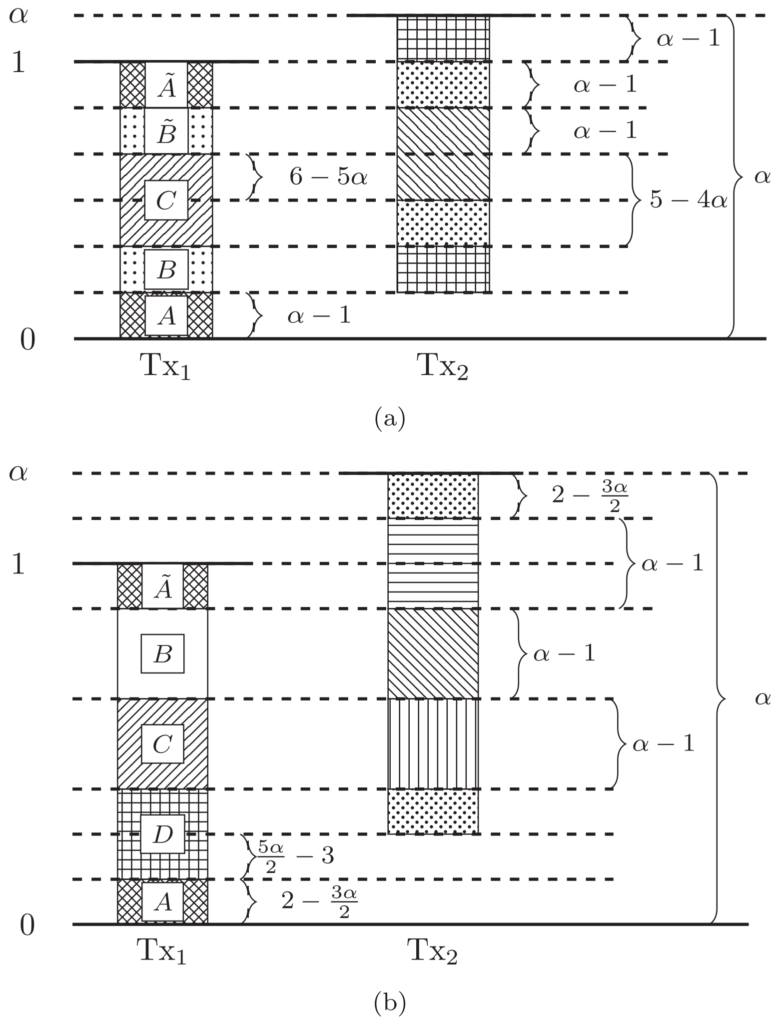


Figure A.12: Normalized signal levels at Rx_1. (**a**) (SI) for $1 \leq \alpha \leq \frac{6}{5}$.; (**b**) (SI) for $\frac{6}{5} \leq \alpha \leq \frac{4}{3}$.

- Region $\boxed{C}$: The bits $\boldsymbol{X}_1^C$ are IID with marginal pmf

$$P_{X_1|B_1}(1|0) = p_1 \quad (A.154)$$
$$P_{X_1|B_1}(1|1) = p_2 \quad (A.155)$$
$$P_{X_1}(1) = p_3 = (1-p)p_1 + pp_2. \quad (A.156)$$

Furthermore, we assume that $\boldsymbol{X}_1^j$, $j = \{A, B, C, D\}$ are independent. For Tx_2, the input distributions coincide with that of Tx_1 in the corresponding regions, but with parameters q_i instead of p_i, $\hat{q}_1$ instead of $\hat{p}_1$ and γ_1 instead of η_1. Following similar steps as in previous sections, we obtain the achievable rate pair (4.40) and (4.41).

A.3.4.3 SI, $\frac{4}{3} \le \alpha \le \frac{3}{2}$

We consider the input distribution depicted in Figure A.13a with the following distributions:

- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(\boldsymbol{X}_1^A, \boldsymbol{X}_1^{\tilde{A}})$ are IID, with marginal pmf given by (A.104)–(A.108).

- Regions $\boxed{B}$, $\boxed{C}$, $\boxed{E}$ and $\boxed{F}$: The bits $\boldsymbol{X}_1^B$, $\boldsymbol{X}_1^C$, $\boldsymbol{X}_1^E$ and $\boldsymbol{X}_1^F$ are independent and temporally IID with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \quad (A.157)$$

- Region $\boxed{D}$: The bits $\boldsymbol{X}_1^D$ are IID with marginal pmf

$$P_{X_1|B_1}(1|0) = \hat{p}_1 \quad (A.158)$$
$$P_{X_1|B_1}(1|1) = \hat{p}_1 \quad (A.159)$$
$$P_{X_1}(1) = \hat{p}_3 = \hat{p}_1. \quad (A.160)$$

Furthermore, we assume that $\boldsymbol{X}_1^j$, $j = \{A, B, C, D, E, F\}$ are independent. For Tx_2, the input distributions coincide with that of Tx_1 in the corresponding regions, but with parameters q_i instead of p_i and γ_1 instead of η_1. Following similar steps as in previous sections, we obtain an achievable rate pair for $\frac{4}{3} < \alpha \le \frac{3}{2}$ which is given by (4.42) and (4.43).

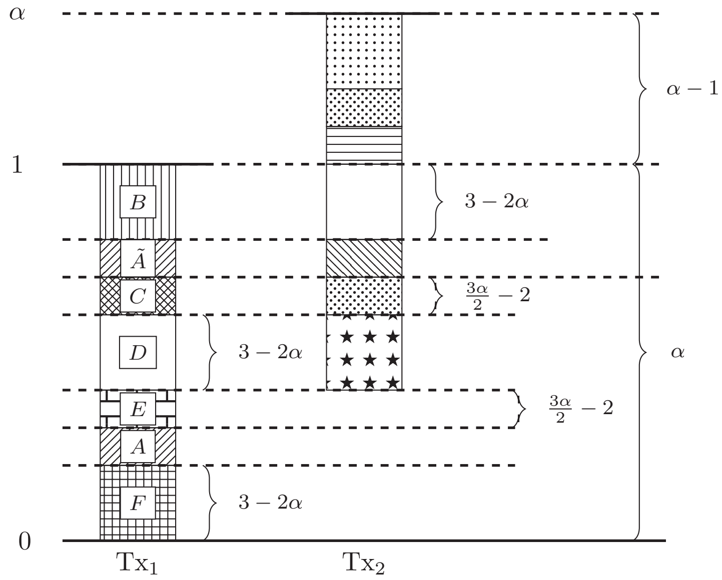
A.3.4.4 SI, $\frac{3}{2} \leq \alpha \leq 2$

We consider the input distribution depicted in Figure A.13b with the following distributions:

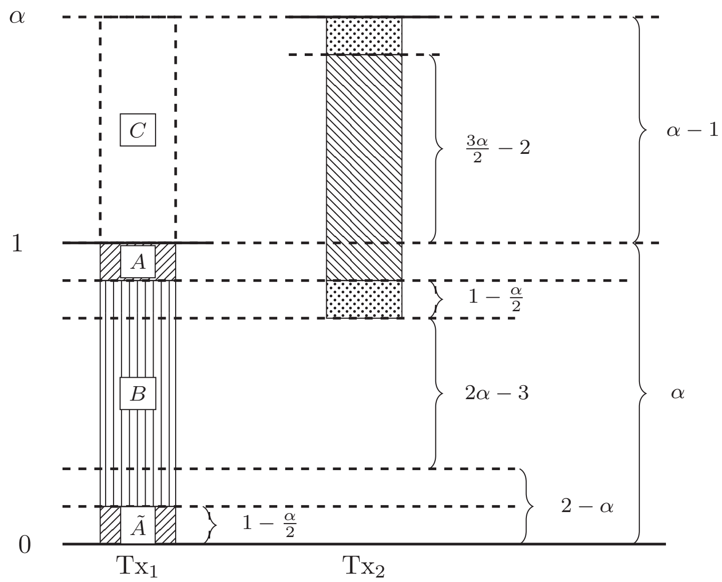
- Regions $\boxed{A}$ and $\boxed{\tilde{A}}$: The bits $(\boldsymbol{X}_1^A, \boldsymbol{X}_1^{\tilde{A}})$ are IID, with marginal pmf given by (A.104)–(A.108).
- Region $\boxed{B}$: The bits are IID with marginal pmf

$$P_{X_1|B_1}(1|0) = P_{X_1|B_1}(1|1) = \tfrac{1}{2}. \tag{A.161}$$

Furthermore, we assume that $\boldsymbol{X}_1^j$, $j = \{A, B\}$ are independent. For Tx_2, the input distributions coincide with that of Tx_1 in the corresponding regions, but with parameters q_i instead of p_i, $\hat{q}_1$ instead of $\hat{p}_1$ and γ_1 instead of η_1. Proceeding as in the previous sections we obtain the achievable rate pair (4.44) and (4.45).



(a)



(b)

Figure A.13: Normalized signal levels at Rx_1. (**a**) (SI) for $\frac{4}{3} \leq \alpha \leq \frac{3}{2}$; (**b**) (SI) for $\frac{3}{2} \leq \alpha \leq 2$.

B

Proof of Lemma 3 in Appendix A.1

B.1 Proof of Lemma 3

In this appendix, we prove the Lemma 3. To this end, we first introduce definitions and properties that will be used in the proof of the lemma.

Definition 17 (Sup-entropy rate) *The sup-entropy rate $\overline{H}(Y)$ is defined as the limsup in probability of $\frac{1}{N}\log\frac{1}{P_{Y^N}(Y^N)}$. Analogously, the conditional sup-entropy rate $\overline{H}(Y|X)$ is the limsup in probability (according to $\{P_{X^N Y^N}\}$) of $\frac{1}{N}\log\frac{1}{P_{Y^N|X^N}(Y^N|X^N)}$.*

Lemma 8 (Sup-entropy rate properties) *Suppose (X,Y) takes values in $(\mathcal{X},\mathcal{Y})$. The sup-entropy rate has the following properties:*

$$\overline{H}(Y|X) < \overline{H}(Y) \tag{B.1}$$

$$0 \leq \overline{H}(Y) < \log|\mathcal{Y}| \tag{B.2}$$

where $|\mathcal{Y}|$ *denotes the cardinality of* Y.

Proof: Property (B.1) follows directly from properties (c) and (d) of [62, Th. 8]. Property (B.2) is equal to property (e) in [62, Th. 8]. ■

We recall the information densities $\mathrm{i}_1(\mathbf{x}_1^N, \mathbf{y}_1^N, \mathbf{b})$ and $\mathrm{i}_2(\mathbf{x}_2^N, \mathbf{y}_2^N, \mathbf{b})$ defined in (A.1) and (A.2), respectively. By decomposing the logarithms and applying the Bayes rule to both probability terms, we obtain

$$\begin{aligned} \mathrm{i}_i(\mathbf{x}_i^N, \mathbf{y}_i^N, \mathbf{b}) &= \log P_{\mathbf{Y}_i^N|\mathbf{X}_i^N,\mathbf{B}}(\mathbf{y}_i^N|\mathbf{x}_i^N, \mathbf{b}) - \log P_{\mathbf{Y}_i^N|\mathbf{B}}(\mathbf{y}_i^N|\mathbf{b}) \\ &= \log P_{\mathbf{X}_i^N|\mathbf{Y}_i^N,\mathbf{B}}(\mathbf{x}_i^N|\mathbf{y}_i^N, \mathbf{b}) - \log P_{\mathbf{X}_i^N|\mathbf{B}}(\mathbf{x}_i^N|\mathbf{b}). \end{aligned} \tag{B.3}$$

To shorten notation, we shall omit the arguments and write $\mathrm{i}_i \triangleq \mathrm{i}_i(\mathbf{x}_i^N, \mathbf{y}_i^N, \mathbf{b})$, $i = 1, 2$ wherever the arguments are clear from the context.

Recall the error events $\mathcal{E}_i(\Gamma_i) \triangleq \left\{\frac{1}{n}\mathrm{i}_i \leq \Gamma_i\right\}$, $i = 1, 2$, and $\mathcal{E}_{12}(\Gamma) \triangleq \left\{\frac{1}{n}\mathrm{i}_1 + \frac{1}{n}\mathrm{i}_2 \leq \Gamma\right\}$, with $\Gamma = \Gamma_1 + \Gamma_2$, as defined in (A.3) and (A.4), respectively. We first note that

$$\mathcal{E}_1 \cap \mathcal{E}_2 \quad \subseteq \quad \mathcal{E}_{12} \tag{B.4}$$

$$\mathcal{E}_1 \cap \mathcal{E}_2 \quad = \quad \mathcal{E}_1 \setminus \{\mathcal{E}_1 \cap \mathcal{E}_2^c\} \supseteq \mathcal{E}_1 \setminus \{\mathcal{E}_2^c\} \tag{B.5}$$

where (B.4) follows because the conditions $\frac{1}{N}\mathrm{i}_1 \leq \Gamma_1$ and $\frac{1}{N}\mathrm{i}_2 \leq \Gamma_2$ imply that $\frac{1}{N}(\mathrm{i}_1 + \mathrm{i}_2) \leq \Gamma_1 + \Gamma_2$. Then, (B.5) follows by applying basic set operations. Using (B.4) and (B.5), and computing the probability of the corresponding events, we obtain

$$\Pr\{\mathcal{E}_{12}\} \geq \Pr\{\mathcal{E}_1\} - \Pr\{\mathcal{E}_2^c\}. \tag{B.6}$$

For clarity of exposition, we define

$$\epsilon_{\mathbf{b}} \triangleq \Pr\left\{\frac{1}{N}(\mathrm{i}_1 + \mathrm{i}_2) \leq \Gamma \;\middle|\; \mathbf{B} = \mathbf{b}\right\} \tag{B.7}$$

and analyze the necessary conditions on Γ such that $\epsilon_{\mathbf{b}} \to 0$ as $N \to \infty$. We next consider separately the four possible realizations of $\mathbf{B} = \mathbf{b}$.

B.1.1 Case $\mathbf{B} = [0, 0]$

When $\mathbf{B} = [0, 0]$, the channel corresponds to two parallel channels with no interference links. Then, the underlying distribution of the probability (B.7) is

$$\begin{aligned} &P_{\mathbf{X}_1^N,\mathbf{X}_2^N,\mathbf{Y}_1^N,\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{x}_1^N, \mathbf{x}_2^N, \mathbf{y}_1^N, \mathbf{y}_2^N|\mathbf{b}) \\ &= P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b}) P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{x}_2^N|\mathbf{b}) \mathbb{1}\{\mathbf{y}_1^N = \mathsf{S}_{n_d}\mathbf{x}_1^N\} \mathbb{1}\{\mathbf{y}_2^N = \mathsf{S}_{n_d}\mathbf{x}_2^N\} \end{aligned} \tag{B.8}$$

as the outputs $\mathbf{y}_1^N$ and $\mathbf{y}_2^N$ must coincide with the corresponding inputs according to the deterministic model. To prove the constraint (A.6), we use (B.3) in (B.7) to obtain

$$\epsilon_{00} = \Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B}) \le \Gamma \,\middle|\, \mathbf{B} = [0, 0] \right\} \tag{B.9}$$

where we used that, according to (B.8), $\log P_{\mathbf{X}_i^N|\mathbf{Y}_i^N,\mathbf{B}}(\mathbf{X}_i^N|\mathbf{Y}_i^N, \mathbf{B}) = 0$ with probability (w.p.) 1, for $i = 1, 2$.

We consider now the conditional sup-entropy rates $\overline{H}(\mathbf{X}_i^N|\mathbf{B})$, $i = 1, 2$. According to (B.2) in Lemma 8, we have that $\overline{H}(\mathbf{X}_i^N|\mathbf{B}) < n_d$, $i = 1, 2$. With these considerations, if we set $\Gamma = 2n_d + 2\delta$ for some arbitrary $\delta > 0$ in (B.9), we obtain

$$\begin{aligned} \epsilon_{00} \ge &\Pr\Bigg\{ -\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B} \\ &\qquad\qquad \le 2n_d + 2\delta \,\Big|\, \mathbf{B} = [0, 0] \Bigg\} \\ \ge &\Pr\Bigg\{ -\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B}) \\ &\qquad\qquad < \overline{H}(\mathbf{X}_1^N|\mathbf{B}) + \overline{H}(\mathbf{X}_2^N|\mathbf{B}) + 2\delta \,\Big|\, \mathbf{B} = [0, 0] \Bigg\} \\ \ge &\Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) < \overline{H}(\mathbf{X}_1^N|\mathbf{B}) + \delta \,\middle|\, \mathbf{B} = [0, 0] \right\} \\ &- \Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B}) \ge \overline{H}(\mathbf{X}_2^N|\mathbf{B}) + \delta \,\middle|\, \mathbf{B} = [0, 0] \right\} \end{aligned} \tag{B.10}$$

where the last step follows from (B.6).

Recalling the definitions of the conditional sup-entropy rates $\overline{H}(\mathbf{X}_i^N|\mathbf{B})$ we have that, for any $\delta > 0$,

$$\lim_{N\to\infty} \Pr\left\{ -\frac{1}{N}\log P_{\mathbf{X}_i^N|\mathbf{B}}(\mathbf{X}_i^N|\mathbf{B}) \ge \overline{H}(\mathbf{X}_i^N|\mathbf{B}) + \delta \,\middle|\, \mathbf{B} = [0, 0] \right\} = 0, \quad i = 1, 2. \tag{B.11}$$

This implies that the first probability on the right-hand side (RHS) of (B.10) tends to 1 as $N \to \infty$, and the second probability on the RHS of (B.10) tends to

0 as $N \to \infty$. We conclude that for any $\Gamma > 2n_d$ the lower bound in (B.10) tends to 1 as $N \to \infty$. Thus, $\epsilon_{00} \to 0$ as $N \to \infty$ only if $\Gamma \leq 2n_d$.

B.1.2 Case $\mathbf{B} = [0, 1]$

When $\mathbf{B} = [0, 1]$, the channel corresponds to a two-user IC where only one of the transmitters interferes its non-intended receiver. In this case, the underlying distribution in (B.7) is given by

$$\begin{aligned}
&P_{\mathbf{X}_1^N,\mathbf{X}_2^N,\mathbf{Y}_1^N,\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{x}_1^N,\mathbf{x}_2^N,\mathbf{y}_1^N,\mathbf{y}_2^N|\mathbf{b}) \\
&= P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b})P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{x}_2^N|\mathbf{b})\mathbb{1}\{\mathbf{y}_1^N = \mathsf{S}_{n_d}\mathbf{x}_1^N\}\mathbb{1}\{\mathbf{y}_2^N = \mathsf{S}_{n_d}\mathbf{x}_2^N \oplus \mathsf{S}_{n_c}\mathbf{x}_1^N\}
\end{aligned} \tag{B.12}$$

We next prove the constraints (A.6) and (A.7) in Lemma 3.

B.1.2.1 Proof of Constraint (A.6)

We lower-bound the probability ϵ_{01} by that of 2 parallel channels and follow the steps in Appendix B.1.1. Indeed, by using (B.3) in (B.7) and lower-bounding $\log P_{\mathbf{X}_i^N|\mathbf{Y}_i^N,\mathbf{B}}(\mathbf{X}_i^N|\mathbf{Y}_i^N,\mathbf{B}) \leq 0$, $i = 1, 2$, we obtain that

$$\epsilon_{01} \geq \Pr\left\{-\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B}) \leq \Gamma \,\middle|\, \mathbf{B} = [0, 1]\right\}. \tag{B.13}$$

The RHS of (B.13) coincides with (B.9) conditioned in $\mathbf{B} = [0, 1]$. The proof then follows the one in Appendix B.1.1, with the probabilities and sup-entropy rates conditioned on $\mathbf{B} = [0, 1]$ instead of $\mathbf{B} = [0, 0]$.

B.1.2.2 Proof of Constraint (A.7)

According to (B.12), the following identities hold w.p. 1:

(i1) $\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N = \mathsf{S}_{n_c}\mathbf{X}_1^N$

(i2) $P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{X}_2^N, [0, 1]) = P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N|\mathbf{B} = [0, 1])$

(i3) $P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{X}_1^N, [0, 1]) = 1$

Using (B.3) in (B.7) and the identities (i1)–(i3), we obtain

$$\begin{aligned}\epsilon_{01} = \Pr\bigg\{ & \frac{1}{N}\log P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{Y}_1^N|\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{B}) \\ & + \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) \le \Gamma \bigg| \mathbf{B} = [0,1]\bigg\} \\ = \Pr\bigg\{ & -\frac{1}{N}\log P_{\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B}) \\ & + \frac{1}{N}\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) \le \Gamma \bigg| \mathbf{B} = [0,1]\bigg\}\end{aligned} \tag{B.14}$$

We next define $\tilde{\mathsf{L}}_d \triangleq \mathsf{L}_d\mathsf{S}_{n_d}$ and apply the chain rule of probability to obtain

$$\begin{aligned}&\log P_{\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_d}\mathbf{x}_1^N|\mathbf{b}) \\ &= \log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_1^N|\mathbf{b}) + \log P_{\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{x}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b}).\end{aligned} \tag{B.15}$$

Using (B.15) in (B.14) and canceling the term $\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B})$, we obtain

$$\begin{aligned}\epsilon_{01} =\Pr\Bigg\{ & -\frac{1}{N}\log P_{\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}) \\ & - \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) \le \Gamma \;\Big|\; \mathbf{B} = [0,1]\Bigg\}.\end{aligned} \tag{B.16}$$

Consider the sup-entropy rates $\overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}\big)$ and $\overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big)$. By (B.1) and (B.2) in Lemma 8, we have that

$$\overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}\big) \le \overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathbf{B}) < (n_d-n_c)^+ \tag{B.17}$$

$$\overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big) < \max(n_d,n_c). \tag{B.18}$$

Let $\Gamma = (n_d-n_c)^+ + \max(n_d,n_c) + 2\delta$ for some arbitrary $\delta > 0$. It follows that $\Gamma \ge \overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}) + \overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big) + 2\delta$, so (B.16) can be lower-bounded as

$$
\begin{aligned}
&\epsilon_{01}\\
&\geq \Pr\Bigg\{-\frac{1}{N}\log P_{\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B})-\frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B})\\
&\qquad\qquad < \overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}\big)+\overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big)+2\delta\;\Big|\;\mathbf{B}=[0,1]\Bigg\}\\
&\geq \Pr\Bigg\{-\frac{1}{N}\log P_{\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B})\\
&\qquad\qquad < \overline{H}\big(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}\big)+\delta\;\Big|\;\mathbf{B}=[0,1]\Bigg\}\\
&\quad -\Pr\left\{-\frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B})\geq\overline{H}\big(\mathbf{Y}_2^N|\mathbf{B}\big)+\delta\;\Big|\;\mathbf{B}=[0,1]\right\}
\end{aligned}\tag{B.19}
$$

where the second step follows from (B.6). By the definition of the conditional sup-entropy rate, it follows that the first probability on the RHS of (B.19) tends to 1 as $N\to\infty$, and the second probability on the RHS of (B.19) tends to 0 as $N\to\infty$. This implies that $\epsilon_{01}\to 0$ as $N\to\infty$ only if $\Gamma\leq(n_d-n_c)^+ + \max(n_d,n_c)$ and proves conditions (A.6) and (A.7) in Lemma 3.

Remark 15 *Given the symmetry of the problem, the constraints* (A.6) *and* (A.7) *for* $\mathbf{B}=[1,0]$ *are proven by swapping the roles of users 1 and 2, and following the same steps as for* $\mathbf{B}=[0,1]$.

B.1.3 Case $\mathbf{B}=[1,1]$

This scenario corresponds to a non-bursty IC. The underlying distribution in (B.7) is given by

$$
\begin{aligned}
&P_{\mathbf{X}_1^N,\mathbf{X}_2^N,\mathbf{Y}_1^N,\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{x}_1^N,\mathbf{x}_2^N,\mathbf{y}_1^N,\mathbf{y}_2^N|\mathbf{b})\\
&= P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b})P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{x}_2^N|\mathbf{b})\mathbb{1}\{\mathbf{y}_1^N=\mathsf{S}_{n_d}\mathbf{x}_1^N\oplus\mathsf{S}_{n_c}\mathbf{x}_2^N\}\mathbb{1}\{\mathbf{y}_2^N=\mathsf{S}_{n_d}\mathbf{x}_2^N\oplus\mathsf{S}_{n_c}\mathbf{x}_1^N\}
\end{aligned}\tag{B.20}
$$

where the last step follows from the deterministic model since, for given $\mathbf{x}_1^N$ and $\mathbf{x}_2^N$, the outputs $\mathbf{y}_1^N$ and $\mathbf{y}_2^N$ are given by the equations appearing in the corresponding indicator functions. We next obtain the constraints (A.6)–(A.8) in Lemma 3.

B.1.3.1 Proof of Constraint (A.6)

To prove this constraint, we lower-bound the probability ϵ_{11} by that of 2 parallel channels. Indeed, using (B.3) in (B.7), we obtain that

$$\begin{aligned}\epsilon_{11} = \Pr\Bigg\{\frac{1}{N}\Bigg(&\log P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}) - \log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) \\ &+ \log P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}) - \log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B})\Bigg) \le \Gamma \Bigg| \,\mathbf{B}=[1,1]\Bigg\} \\ \ge \Pr\Bigg\{&-\frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{X}_1^N|\mathbf{B}) - \frac{1}{N}\log P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{X}_2^N|\mathbf{B}) \le \Gamma \Bigg| \,\mathbf{B}=[1,1]\Bigg\}\end{aligned} \tag{B.21}$$

where the inequality follows because $\log P_{\mathbf{X}_i^N|\mathbf{Y}_i^N,\mathbf{B}}(\mathbf{X}_i^N|\mathbf{Y}_i^N,\mathbf{B}) \le 0,\ i=1,2$. As this expression coincides with (B.9) conditioned on $\mathbf{B}=[1,1]$, the proof then follows the one in Appendix B.1.1, with the probabilities and sup-entropy rates conditioned on $\mathbf{B}=[1,1]$ instead of $\mathbf{B}=[0,0]$.

B.1.3.2 Proof of Constraint (A.7)

We next lower-bound the probability ϵ_{11} by that of an interference channel, in which only one of the transmitters interferes its non-intended receiver. Using the information densities i_1 and i_2 in (B.3), we have that

$$\begin{aligned}&\mathrm{i}_1 + \mathrm{i}_2 \\ &= \log P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathbf{y}_1^N|\mathbf{x}_1^N,\mathbf{b}) - \log P_{\mathbf{Y}_1^N|\mathbf{B}}(\mathbf{y}_1^N|\mathbf{b}) + \log P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_2^N|\mathbf{x}_2^N,\mathbf{b}) \\ &\quad - \log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{y}_2^N|\mathbf{b}) \\ &= \log P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_1^N|\mathbf{x}_1^N,\mathbf{x}_2^N,\mathbf{b}) - \log P_{\mathbf{Y}_1^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_1^N|\mathbf{x}_2^N,\mathbf{b}) \\ &\quad + \log P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_2^N|\mathbf{x}_2^N,\mathbf{b}) - \log P_{\mathbf{Y}_2^N,\mathbf{B}}(\mathbf{y}_2^N|\mathbf{b}) \\ &\quad - \log \frac{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,\mathbf{b})}\end{aligned} \tag{B.22}$$

where the second step follows from adding and subtracting

$$\frac{1}{N}\log\frac{P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_1^N|\mathbf{x}_1^N,\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{Y}_1^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_1^N|\mathbf{x}_2^N,\mathbf{b})}$$

and simplifying the resulting terms via the Bayes rule and using that $P_{\mathbf{X}_1^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{x}_2^N,\mathbf{b}) = P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b})$ since $\mathbf{X}_1^N$ and $\mathbf{X}_2^N$ are independent conditioned on $\mathbf{B}$.

According to the underlying distribution (B.20), the following identities hold w.p. 1:

(i1) $\mathbf{Y}_1^N \oplus \mathsf{S}_{n_c}\mathbf{X}_2^N = \mathsf{S}_{n_d}\mathbf{X}_1^N$

(i2) $\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N = \mathsf{S}_{n_c}\mathbf{X}_1^N$

(i3) $P_{\mathbf{Y}_1^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{X}_2^N, \mathbf{B} = [1,1]) = P_{\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B}}(\mathbf{Y}_1^N \oplus \mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B} = [1,1])$

(i4) $P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{X}_2^N, \mathbf{B} = [1,1]) = P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N|\mathbf{B} = [1,1])$

(i5) $P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathbf{X}_2^N,\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{X}_1^N, \mathbf{X}_2^N, \mathbf{B} = [1,1]) = 1$

(i6) $P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{X}_2^N,\mathbf{B}}(\mathbf{X}_1^N|\mathbf{Y}_1^N, \mathbf{X}_2^N, \mathbf{B} = [1,1]) = 1$.

Using (B.22) and the identities (i1)–(i6), we obtain for (B.7)

$$\begin{aligned}\epsilon_{11} = \Pr\Big\{&-\frac{1}{N}\log P_{\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B}) + \frac{1}{N}\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}) \\ &- \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B}) + \frac{1}{N}\log P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{X}_1^N|\mathbf{Y}_1^N, \mathbf{B}) \le \Gamma \;\Big|\; \mathbf{B} = [1,1]\Big\}.\end{aligned} \tag{B.23}$$

Using (B.15) in (B.23), canceling the term $\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B})$, and using that $\log P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{X}_1^N|\mathbf{Y}_1^N, \mathbf{B}) \le 0$, we obtain the lower bound

$$\begin{aligned}\epsilon_{11} \ge \Pr\Big\{\Big(&-\frac{1}{N}\log P_{\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\tilde{\mathsf{L}}_{(n_d-n_c)^+}\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N, \mathbf{B}) \\ &- \frac{1}{N}\log P_{\mathbf{Y}_2^N|\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{B})\Big) \le \Gamma \Big| \;\mathbf{B} = [1,1]\Big\}.\end{aligned} \tag{B.24}$$

The RHS of (B.24) coincides with (B.16) conditioned on $\mathbf{B} = [1,1]$. The proof then follows the one in Appendix B.1.2.2, with the probabilities and sup-entropy rates conditioned on $\mathbf{B} = [1,1]$ instead of $\mathbf{B} = [0,1]$.

B.1.3.3 Proof of Constraint (A.8)

We begin this proof by using (B.3) to write

$$
\begin{aligned}
&\mathrm{i}_1 + \mathrm{i}_2 \\
&= \log \frac{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b})} + \log \frac{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{x}_2^N|\mathbf{b})} \\
&\overset{(a)}{=} \log \frac{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b})} + \log \frac{P_{\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|\mathbf{B}}(\mathbf{x}_1^N|\mathbf{b})} \\
&\quad + \log \frac{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b})} + \log \frac{P_{\mathbf{X}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|\mathbf{B}}(\mathbf{x}_2^N|\mathbf{b})} \\
&\quad - \log \frac{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,\mathbf{b})} \\
&\quad - \log \frac{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,\mathbf{b})} \qquad \text{(B.25)} \\
&\overset{(b)}{=} \log P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{y}_1^N|\mathbf{x}_1^N,\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b}) - \log P_{\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{y}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b}) \\
&\quad + \log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_1^N|\mathbf{x}_1^N,\mathbf{b}) - \log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_1^N|\mathbf{b}) \\
&\quad + \log P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_2^N|\mathbf{x}_2^N,\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b}) - \log P_{\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b}) \\
&\quad + \log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_2^N|\mathbf{x}_2^N,\mathbf{b}) - \log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_2^N|\mathbf{b}) \\
&\quad - \log \frac{P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_1^N|\mathbf{x}_1^N,\mathbf{y}_1^N,\mathbf{b})}{P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_1^N|\mathbf{y}_1^N,\mathbf{b})} \\
&\quad - \log \frac{P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_2^N|\mathbf{x}_2^N,\mathbf{y}_2^N,\mathbf{b})}{P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{x}_2^N|\mathbf{y}_2^N,\mathbf{b})}
\end{aligned}
$$

where (a) follows by adding and subtracting

$$
\begin{aligned}
&\frac{1}{N}\log \frac{P_{\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathbf{y}_1^N,\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b})}{P_{\mathbf{X}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{x}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b})} \quad \text{and} \\
&\frac{1}{N}\log \frac{P_{\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathbf{y}_2^N,\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b})}{P_{\mathbf{X}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{x}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b})}
\end{aligned}
$$

and by rearranging terms. Step (b) follows by applying the Bayes rule and by decomposing the logarithm terms.

We analyze the second and the seventh terms in (B.25). To this end, we define $n_- \triangleq \min\{(n_d - n_c)^+, n_c\}$ and $n_+ \triangleq \max\{(n_d - n_c)^+, n_c\}$ and apply the chain rule of probability to obtain

$$\begin{aligned}
&P_{\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{y}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b}) \\
&\quad = P_{\mathsf{S}_{n_-}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathsf{S}_{n_-}\mathbf{y}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathbf{b}) \\
&\qquad \times P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{y}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathsf{S}_{n_-}\mathbf{y}_1^N,\mathbf{b}) \\
&\quad = P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{y}_1^N|\mathsf{S}_{n_c}\mathbf{x}_1^N,\mathsf{S}_{n_-}\mathbf{y}_1^N,\mathbf{b})
\end{aligned} \tag{B.26}$$

and

$$\begin{aligned}
&P_{\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{y}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b}) \\
&\quad = P_{\mathsf{S}_{n_-}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathsf{S}_{n_-}\mathbf{y}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathbf{b}) \\
&\qquad \times P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{y}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathsf{S}_{n_-}\mathbf{y}_2^N,\mathbf{b}) \\
&\quad = P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{y}_2^N|\mathsf{S}_{n_c}\mathbf{x}_2^N,\mathsf{S}_{n_-}\mathbf{y}_2^N,\mathbf{b}).
\end{aligned} \tag{B.27}$$

The probabilities (B.26) and (B.27) were simplified by recalling the underlying distribution (B.20). Indeed, we have w.p. 1 that $P_{\mathsf{S}_{n_-}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathsf{S}_{n_-}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}) = 1$ and $P_{\mathsf{S}_{n_-}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathsf{S}_{n_-}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}) = 1$, since $\mathsf{S}_{n_-}\mathbf{Y}_i^N$, $i = 1,2$ is not affected by interference, so it is determined by $\mathsf{S}_{n_c}\mathbf{X}_i^N$, $i = 1,2$. Similarly, we have that $P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B}) = 1$ and $P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B}) = 1$.

We next note that, for the underlying distribution in (B.20), the following identities hold w.p. 1:

(i1) $\mathbf{Y}_1^N \oplus \mathsf{S}_{n_d}\mathbf{X}_1^N = \mathsf{S}_{n_c}\mathbf{X}_2^N$

(i2) $\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N = \mathsf{S}_{n_c}\mathbf{X}_1^N$

(i3) $P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{X}_1^N,\mathbf{B} = [1,1]) = 1$

(i4) $P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{X}_2^N,\mathbf{B} = [1,1]) = 1$

(i5) $P_{\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B}}(\mathbf{Y}_1^N|\mathbf{X}_1^N,\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathbf{B} = [1,1])$
$= P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}}(\mathbf{Y}_1^N \oplus \mathsf{S}_{n_d}\mathbf{X}_1^N|\mathbf{B} = [1,1])$

(i6) $P_{\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B}}(\mathbf{Y}_2^N|\mathbf{X}_2^N,\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathbf{B} = [1,1])$
$= P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathbf{Y}_2^N \oplus \mathsf{S}_{n_d}\mathbf{X}_2^N|\mathbf{B} = [1,1])$

We combine these identities with (B.7), (B.25)–(B.27) to obtain

$$
\begin{aligned}
\epsilon_{11} &= \Pr\left\{\frac{1}{N}(\mathrm{i}_1+\mathrm{i}_2)\leq\Gamma\,\Big|\,\mathbf{B}=[1,1]\right\}\\
&= \Pr\Bigg\{\frac{1}{N}\Bigg(\log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B})\\
&\qquad -\log P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B})\\
&\qquad -\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B})+\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B})\\
&\qquad +\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B})\\
&\qquad -\log P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B})\\
&\qquad -\log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B})\\
&\qquad +\log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B})\Bigg)\leq\Gamma\,\Big|\,\mathbf{B}=[1,1]\Bigg\}\\
&\geq \Pr\Bigg\{-\frac{1}{N}\Bigg(\log P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B})\\
&\qquad -\log P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B})\Bigg)\\
&\qquad\qquad \leq\Gamma\,\Big|\,\mathbf{B}=[1,1]\Bigg\}
\end{aligned}
\tag{B.28}
$$

where in the last step we canceled the terms $\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{B})$ and $\log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{B})$ and we used that, w.p. 1, $\log P_{\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_1^N|\mathbf{Y}_1^N,\mathbf{B})\leq 0$ and $\log P_{\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{S}_{n_c}\mathbf{X}_2^N|\mathbf{Y}_2^N,\mathbf{B})\leq 0$.

By (B.1) and (B.2) in Lemma 8, the conditional sup-entropy rates satisfy

$$
\overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_i^N|\mathsf{S}_{n_c}\mathbf{X}_i^N,\mathsf{S}_{n_-}\mathbf{Y}_i^N,\mathbf{B})\leq\overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_i^N|\mathbf{B})<\max\{(n_d-n_c)^+,n_c\}
\tag{B.29}
$$

where $i=1,2$.

Then, setting $\Gamma=2\max\{(n_d-n_c)^+,n_c\}+2\delta$ for some arbitrary $\delta>0$, we obtain from (B.6) that (B.28) can be lower-bounded by

$$\begin{aligned}\epsilon_{11} \geq \Pr\Bigg\{ & -\frac{1}{N}\log P_{\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}) \\ & < \overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B}) + \delta \;\Big|\; \mathbf{B}=[1,1]\Bigg\} \\ - \Pr\Bigg\{ & \frac{1}{N}\log P_{\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_-}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}) \\ & \geq \overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B}) + \delta \;\Big|\; \mathbf{B}=[1,1]\Bigg\}.\end{aligned} \tag{B.30}$$

By the definition of $\overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_1^N|\mathsf{S}_{n_c}\mathbf{X}_1^N,\mathsf{S}_{n_-}\mathbf{Y}_1^N,\mathbf{B})$, the fist probability on the RHS of (B.30) tends to 1 as $N \to \infty$. Similarly, by the definition of $\overline{H}(\mathsf{L}_{n_+}\mathbf{Y}_2^N|\mathsf{S}_{n_c}\mathbf{X}_2^N,\mathsf{S}_{n_-}\mathbf{Y}_2^N,\mathbf{B})$, the second probability on the RHS of (B.30) tends to 0 as $N \to \infty$. This demonstrates that if $\Gamma > 2\max\{(n_d - n_c)^+, n_c\}$, then the lower bound in (B.28) tends to 1 as $N \to \infty$. Thus, $\epsilon_{11} \to 0$ as $N \to \infty$ only if

$$\Gamma \leq 2\max\{(n_d - n_c)^+, n_c\}. \tag{B.31}$$

C

Appendix to Chapter 5

In this appendix, we demonstrate that ordering the interferers according to α_ℓ gives rise to the largest ρ satisfying (5.2),

$$\frac{\alpha_{\ell+1}}{\alpha_\ell} \geq \rho. \tag{C.1}$$

To this end, we assume that $1 \triangleq \alpha_0 \geq \alpha_1 \geq \alpha_2 \geq \ldots$ and show that the ratios

$$\frac{\alpha_{\ell_1}}{\alpha_{m_1}}, \frac{\alpha_{\ell_2}}{\alpha_{m_2}}, \frac{\alpha_{\ell_3}}{\alpha_{m_3}}, \ldots \tag{C.2}$$

are largest if $\ell_i = i + 1$ and $m_i = i$.

To prove this result, we note that, if $m_i \leq m_j$ and $\ell_i \geq \ell_j$, then

$$\min\left\{\frac{\alpha_{\ell_i}}{\alpha_{m_i}}, \frac{\alpha_{\ell_j}}{\alpha_{m_j}}\right\} \leq \min\left\{\frac{\alpha_{\ell_j}}{\alpha_{m_i}}, \frac{\alpha_{\ell_i}}{\alpha_{m_j}}\right\}. \tag{C.3}$$

Indeed, since the coefficients $\{\alpha_\ell\}$ are ordered, it follows that $\alpha_{m_i} \leq \alpha_{m_j}$ and

$\alpha_{\ell_i} \geq \alpha_{\ell_j}$. Consequently,

$$\frac{\alpha_{\ell_j}}{\alpha_{m_j}} \leq \min\left\{\frac{\alpha_{\ell_j}}{\alpha_{m_i}}, \frac{\alpha_{\ell_i}}{\alpha_{m_j}}\right\} \tag{C.4}$$

from which (C.3) follows.

Using (C.3), we can then show that the minimum of the ratios

$$\frac{\alpha_1}{\alpha_0}, \frac{\alpha_2}{\alpha_1}, \frac{\alpha_3}{\alpha_2}, \ldots \tag{C.5}$$

is not smaller than the minimum of the ratios

$$\frac{\alpha_{\ell_1}}{\alpha_{m_1}}, \frac{\alpha_{\ell_2}}{\alpha_{m_2}}, \frac{\alpha_{\ell_3}}{\alpha_{m_3}}, \ldots \tag{C.6}$$

for any $\{\ell_i\}$ and $\{m_i\}$. Indeed, for any sequences $\{\ell_i\}$ and $\{m_i\}$ we can find a pair of indices (i, j) such that $m_i \leq m_j$ and $\ell_i \geq \ell_j$ when $\{\ell_i\}$ and $\{m_i\}$ are given by

$$\begin{aligned} \ell_i &= i + 1 \\ m_i &= i \end{aligned}$$

or any permutation thereof. However, if $m_i \leq m_j$ and $\ell_i \geq \ell_j$, then the minimum of the ratios is not reduced by swapping $\ell_i \leftrightarrow \ell_j$. Since we can repeat this process until both sequences of indices are ordered, it follows that the ratios (C.5) have the largest minimum. This proves the claim.

References

[1] R. Ahlswede. The capacity region of a channel with two senders and two receivers. *The Annals of Probability*, pages 805–814, 1974.

[2] V. S. Annapureddy and V. V. Veeravalli. Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region. *IEEE Trans. Inf. Theory*, 55(7):3032–3050, July 2009.

[3] Robert B. Ash. *Information Theory*. Dover Books on Mathematics. Mineola, NY: Courier Dover Publications, 1990.

[4] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse. A deterministic model for wreless relay networks an its capacity. In *IEEE Inf. Theory Workshop on Inf. Theory for Wireless Networks*, pages 1–6, July 2007.

[5] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse. Wireless network information flow: A deterministic approach. *IEEE Trans. Inf. Theory*, 57(4):1872–1905, April 2011.

[6] P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Transactions on Information Theory*, 19(2):197–207, March 1973.

[7] E. Biglieri, J. Proakis, and S. Shamai (Shitz). Fading channels: information-theoretic and communications aspects. *IEEE Trans. Inf. Theory*, 44(6):2619–2692, October 1998.

[8] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *The Annals of Mathematical Statistics*, 30(4):1229–1241, 1959.

[9] G. Bresler and D. N. C. Tse. The two-user Gaussian interference channel: A deterministic view. *Trans. Emerging Tel. Tech.*, 19(4):333–354, 2008.

[10] V.R. Cadambe and S.A. Jafar. Interference alignment and degrees of freedom of the k-user interference channel. *IEEE Trans. Inf. Theory*, 54(8):3425–3441, August 2008.

[11] G. Caire and S. Shamai (Shitz). On the capacity of some channels with channel state information. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 42–, August 1998.

[12] A. Carleial. A case where interference does not reduce capacity. *IEEE Trans. Inf. Theory*, 21(5):569–570, 1975.

[13] T. Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18(1):2–14, January 1972.

[14] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006.

[15] I. Csiszár and J. Körner. *"Information Theory: Coding Theorems for Discrete Memoryless Systems"*. Cambridge University Press, 2011.

[16] S. N. Diggavi and D. N. C. Tse. On opportunistic codes and broadcast codes with degraded message sets. In *Proc. IEEE Inf. Theory Workshop (ITW)*, pages 227–231, March 2006.

[17] R. L. Dobrushin. General formulation of Shannon's main theorem in information theory. *Amer. Math. Soc. Trans*, 33(2):323–438, 1963.

[18] M. Effros, A. Goldsmith, and Y. Liang. Generalizing capacity: New definitions and capacity theorems for composite channels. *IEEE Trans. Inf. Theory*, 56(7):3069–3087, July 2010.

[19] A. El Gamal and M. Costa. The capacity region of a class of deterministic interference channels. *IEEE Trans. Inf. Theory*, 28(2):343–346, 1982.

[20] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2011.

[21] R. H. Etkin, D. N. C. Tse, and H. Wang. Gaussian interference channel capacity to within one bit. *IEEE Trans. Inf. Theory*, 54(12):5534–5562, Dec 2008.

[22] A. Ganti, A. Lapidoth, and I. E. Telatar. Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit. *IEEE Trans. Inf. Theory*, 46(7):2315–2328, November 2000.

[23] S. Gelfand and M. Pinsker. Coding for channel with random parameters. *Probl. Contr. and Inf. Theory*, 9(1):19–31, 1980.

[24] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. Inf. Theory*, 43(6):1986–1992, November 1997.

[25] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products.* Elsevier/Academic Press, Amsterdam, seventh edition, 2007.

[26] R. M. Gray. *Entropy and Information Theory.* Springer-Verlag, Berlin, Heidelberg, 1990.

[27] Te Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Trans. Inf. Theory*, 27(1):49–60, 1981.

[28] C. Heegard and A. El Gamal. On the capacity of computer memory with defects. *IEEE Trans. Inf. Theory*, 29(5):731–739, 1983.

[29] S. A Jafar and S. Vishwanath. Generalized degrees of freedom of the symmetric Gaussian K-user interference channel. *IEEE Trans. Inf. Theory*, 56(7):3297–3303, 2010.

[30] D. T. H. Kao and A. Sabharwal. Two-user interference channels with local views: On capacity regions of TDM-dominating policies. *IEEE Trans. Inf. Theory*, 59(11):7014–7040, November 2013.

[31] F. Kelly. Charging and rate control for elastic traffic. *Eur. Trans. Telecomm.*, 8(1):33–37, 1997.

[32] N. Khude, V. Prabhakaran, and P. Viswanath. Harnessing bursty interference. In *Proc. IEEE Inf. Theory Workshop (ITW)*, pages 13–16, June 2009.

[33] N. Khude, V. Prabhakaran, and P. Viswanath. Opportunistic interference management. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 2076–2080, June 2009.

[34] Y. Kim. A coding theorem for a class of stationary channels with feedback. *IEEE Trans. Inf. Theory*, 54(4):1488–1499, April 2008.

[35] T. Koch and A. Lapidoth. On multipath fading channels at high SNR. *IEEE Trans. Inf. Theory*, 56(12):5945–5957, December 2010.

[36] J. Körner and K. Marton. General broadcast channels with degraded message sets. *IEEE Trans. Inf. Theory*, 23(1):60–64, January 1977.

[37] G. Kramer. Outer bounds on the capacity of Gaussian interference channels. *IEEE Trans. Inf. Theory*, 50(3):581–586, 2004.

[38] A. Kuznetsov and B. Tsybakov. Coding in a memory with defective cells. *Problemy peredachi informatsii*, 10(2):52–60, 1974.

[39] A. Lapidoth. On the asymptotic capacity of stationary Gaussian fading channels. *IEEE Trans. Inf. Theory*, 51(2):437–446, February 2005.

[40] A. Lapidoth and S. Moser. Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels. *IEEE Trans. Inf. Theory*, 49(10):2426–2467, October 2003.

[41] A. Lapidoth and P. Narayan. Reliable communication under channel uncertainty. *IEEE Trans. Inf. Theory*, 44(6):2148–2177, October 1998.

[42] A. Lapidoth and I. E. Telatar. The compound channel capacity of a class of finite-state channels. *IEEE Trans. Inf. Theory*, 44(3):973–983, May 1998.

[43] A. Lozano, R.W. Heath, and J.G. Andrews. Fundamental limits of cooperation. *IEEE Trans. Inf. Theory*, 59(9):5213–5226, September 2013.

[44] S. Mishra, I.-H. Wang, and S. N. Diggavi. Opportunistic interference management for multicarrier systems. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 389–393, July 2013.

[45] S. Mishra, I.-H Wang, and S. N. Diggavi. Harnessing bursty interference in multicarrier systems with output feedback. *IEEE Trans. Inf. Theory*, 63(7):4430–4452, 2017.

[46] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner. Information theoretic considerations for cellular mobile radio. *IEEE Trans. Veh. Tech.*, 43(2):359–378, May 1994.

[47] W. Rudin. *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., New York, NY, USA, 1987.

[48] M Salehi. Capacity and coding for memories with real-time noisy defect information at encoder and decoder. *IEEE Proc. I-Comm., Speech and Vision*, 139(2):113–117, 1992.

[49] H. Sato. Two-user communication channels. *IEEE Trans. Inf. Theory*, 23(3):295–304, May 1977.

[50] H. Sato. The capacity of the Gaussian interference channel under strong interference. *IEEE Trans. Inf. Theory*, 27(6):786–788, 1981.

[51] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[52] C. E. Shannon. Channels with side information at the transmitter. *IBM journal of Research and Development*, 2(4):289–293, 1958.

[53] S. Shamai (Shitz). A broadcast strategy for the Gaussian slowly fading channel. In *IEEE Int. Symp. Inf. Theory (ISIT)*, pages 150–, June 1997.

[54] C. Suh and D. N. C. Tse. Feedback capacity of the Gaussian interference channel to within 2 bits. *IEEE Trans. Inf. Theory*, 57(5):2667–2685, May 2011.

[55] D. N. C. Tse and P. Viswanath. *Fundamentals of Wireless Communication.* Cambridge University Press, 2005.

[56] A. Vahid and R. Calderbank. When does spatial correlation add value to delayed channel state information? In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 2624–2628, July 2016.

[57] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr. Interference channel with binary fading: Effect of delayed network state information. In *Proc. 49th Annual Allerton Conf. Comm., Control, and Comp. (Allerton)*, pages 894–901, Sept 2011.

[58] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr. Binary fading interference channel with no CSIT. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 666–670, June 2014.

[59] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr. Capacity results for binary fading interference channels with delayed CSIT. *IEEE Trans. Inf. Theory*, 60(10):6093–6130, October 2014.

[60] A. Vahid, M. A. Maddah-Ali, A. S. Avestimehr, and Y. Zhu. Binary fading interference channel with no CSIT. *IEEE Trans. Inf. Theory*, 63(6):3565–3578, June 2017.

[61] S. Vembu, S. Verdú, and Y. Steinberg. The source-channel separation theorem revisited. *IEEE Trans. Inf. Theory*, 41(1):44–54, Jan 1995.

[62] S. Verdú and T. S. Han. A general formula for channel capacity. *IEEE Trans. Inf. Theory*, 40(4):1147–1157, 1994.

[63] Grace Villacrés, Tobias Koch, Aydin Sezgin, and Gonzalo Vazquez-Vilar. Robust signaling for bursty interference. *Entropy*, 20(11):870, 2018.

[64] G. Villacrés and T. Koch. Wireless networks of bounded capacity. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 2584–2588, July 2016.

[65] I. H. Wang, C. Suh, S. Diggavi, and P. Viswanath. Bursty interference channel with feedback. In *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pages 21–25, July 2013.

[66] I. H. Wang, C. Suh, S. Diggavi, and P. Viswanath. Bursty interference channel with feedback. https://sites.google.com/site/ihsiangw/isit13preprintburstyic, January 2013.

[67] Jacob Wolfowitz. *Coding Theorems of Information Theory*, volume 31. Springer Science & Business Media, 2012.

[68] S.-Y. Yeh and I.H. Wang. Degrees of Freedom of the Bursty MIMO X Channel without Feedback. `arXiv:1611.07630 [cs.IT]`, November 2016.

[69] X. Yi and H. Sun. Opportunistic treating interference as noise. `arXiv:1808.08926 [cs.IT]`, August 2018.

Errata

Chapter 3: Interference Channel

- Equations (3.15)–(3.17) should read: (missing $\frac{1}{2}$ factor)

$$\begin{aligned} R_1 &\leq \tfrac{1}{2}\log(1+\mathrm{SNR}_1) & (3.15)\\ R_2 &\leq \tfrac{1}{2}\log(1+\mathrm{SNR}_2) & (3.16)\\ R_1+R_2 &\leq \tfrac{1}{2}\min\{\log(1+\mathrm{SNR}_1+\mathrm{INR}_1),\log(1+\mathrm{SNR}_2+\mathrm{INR}_2)\}. & (3.17)\end{aligned}$$

- Equations (3.18)–(3.19) should read: (missing $\frac{1}{2}$ factor)

$$\begin{aligned} R_1 &\leq \tfrac{1}{2}\log(1+\mathrm{SNR}_1) & (3.18)\\ R_2 &\leq \tfrac{1}{2}\log(1+\mathrm{SNR}_2). & (3.19)\end{aligned}$$

- Equations (3.21)–(3.22) should read: (missing $\frac{1}{2}$ factor)

$$\begin{aligned} R_1 &< \tfrac{1}{2}\log\left(1+\frac{\mathrm{SNR}_1}{1+\mathrm{INR}_1}\right) & (3.21)\\ R_2 &< \tfrac{1}{2}\log\left(1+\frac{\mathrm{SNR}_2}{1+\mathrm{INR}_2}\right). & (3.22)\end{aligned}$$

- Equations (3.23)–(3.25) should read: (missing $\frac{1}{2}$ factor)

$$\begin{aligned} R_1 &<\tfrac{1}{2}\log(1+\mathrm{SNR}_1) & (3.23)\\ R_2 &<\tfrac{1}{2}\log(1+\mathrm{SNR}_2) & (3.24)\\ R_1+R_2 &<\tfrac{1}{2}\min\{\log(1+\mathrm{SNR}_1+\mathrm{INR}_1),\log(1+\mathrm{SNR}_2+\mathrm{INR}_2)\}. & (3.25)\end{aligned}$$

- Equations (3.26)–(3.27) should read: (missing $\frac{1}{2}$ factor)

$$\begin{aligned} R_1 &< \tfrac{1}{2}\tau\log\left(1+\frac{\mathrm{SNR}_1}{\tau}\right) & (3.26)\\ R_2 &< \tfrac{1}{2}(1-\tau)\log\left(1+\frac{\mathrm{SNR}_2}{(1-\tau)}\right). & (3.27)\end{aligned}$$

- Equations (3.29) should read: (missing $\frac{1}{2}$ factor)

$$\mathcal{D}(\alpha) = \lim_{\substack{\mathrm{SNR}\to\infty:\\ \mathrm{INR}=\mathrm{SNR}^{\alpha}}} \frac{C_{\mathrm{sym}}}{\frac{1}{2}\log(1+\mathrm{SNR})} \tag{3.29}$$

- After equation (3.31) p. 51, should read:
  where S *is a* $q\times q$ *down-shift matrix as defined in Section 1.3 and with the correspondence* $n=\frac{1}{2}\log\mathrm{SNR}$.

- Equations (3.34)–(3.35) should read: (missing $\frac{1}{2}$ factor)

$$\begin{aligned} n_d &= \lfloor\log_2|h_d|^2\rfloor = \lfloor\tfrac{1}{2}\log_2\mathrm{SNR}\rfloor & (3.34)\\ n_c &= \lfloor\log_2|h_c|^2\rfloor = \lfloor\tfrac{1}{2}\log_2\mathrm{INR}\rfloor. & (3.35)\end{aligned}$$

- Theorem 8 should be read: (for real valued two-user Gaussian IC)

  Theorem 8 [9, Th. 2] *The capacity of the real-valued two-user Gaussian IC with signal and interference to noise ratios* SNR *and* INR *is within 18.6 bits per user of the capacity of the linear deterministic IC with gains* $n_d = \lfloor \frac{1}{2} \log_2 \mathrm{SNR} \rfloor$, $n_c = \lfloor \frac{1}{2} \log_2 \mathrm{INR} \rfloor$.