SELECTED TOPICS IN INFORMATION THEORETIC SECURITY AND COMPRESSION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Amine Laourine

May 2012

SELECTED TOPICS IN INFORMATION THEORETIC SECURITY AND COMPRESSION

Amine Laourine, Ph.D.

Cornell University 2012

The problems studied in this thesis fall within two different topics in network information theory. The first part of this dissertation will be about information theoretic security and the second about distributed source coding. In the following a brief description of these problems will be given.

In information theoretic security, we look at two different types of timing channels and we quantify the maximum rate at which a transmitter can communicate information reliably to a legitimate receiver while keeping an eavesdropper in the dark. The first timing channel that we study is the Poisson channel which is used to model certain direct detection optical communication systems. To transmit a message on this channel, the transmitter encodes information by modulating the intensity of an optical signal while the legitimate receiver and the eavesdropper use the arrival moments of the individual photons to decide which message was transmitted. The second timing channel studied is the exponential server queue. Here the transmitter encodes a message using a chosen sequence of packets inter-arrival times and both the legitimate receiver and the eavesdropper use the corresponding inter-departures from their respective exponential server queues to decode the transmitted message.

In distributed source coding, we consider a rate-distortion problem in which a decoder is interested in estimating two correlated Gaussian random variables with mean-square error distortion constraints on each of the reproductions. The

variables to be estimated are the roots of a given Gauss-Markov tree and each encoder observes one of the leaves of that tree. We show that a simple compression architecture that performs separate lossy quantization followed by Slepian-Wolf binning is sum-rate optimal for this problem.

BIOGRAPHICAL SKETCH

Amine Laourine was born in Jerba, Tunisia in 1981. He received a Master of Science degree in Electrical and Computer Engineering from Cornell University in 2011, a Master of Science degree in Telecommunications from the Institut National de la Recherche Scientifique (INRS) in 2007 and a diplôme d'ingénieur from the École Polytechnique de Tunisie (Tunisia Polytechnic School) in 2005. He received the best paper award at the wireless communication symposium in IEEE Globecom'07, the Irwin and Joan Jacobs's fellowship from Cornell University in 2007, the Rene-Fortier scholarship from Bell Canada and the Tunisian Government fellowship for academic excellence in 2006.

This document is dedicated to my family.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Professor Aaron Wagner. After I am gone, I will certainly not miss the cold winters of Ithaca but I will definitely miss his guidance. He has been of tremendous support during the past three years that I spent working with him. Without his help all the problems solved in this thesis would have almost surely remained unsolved. To put it in the language of information theory, my meetings with him were always entropy decreasing, i.e., after each single meeting I was a step closer to solving the problem than I was before.

I would like to express my gratitude to Professor Lang Tong. I had the privilege to work with him for a little more than a year and that period of my life at Cornell was very productive. With his guidance I learned a great deal about how to write better research papers and more importantly how to become a better researcher.

I would like to thank Professor Sidney Resnick who honored me by accepting to be part of my defense committee. I greatly enjoyed the class that he taught about Lévy processes and I also enjoyed reading two of his books: Adventures in Stochastic Processes and A Probability Path. He has the rare ability to present complex ideas in way that is easy for the students to comprehend.

I would like to acknowledge other educators that contributed to my knowledge through the valuable classes that they taught at Cornell: Professor Thomas Parks, Professor Victor Protsak, Professor Xin Guo, Doctor Maren Liese, Professor Philip Protter, Professor Kevin Tang, Professor Eugene Dynkin, Professor Salman Avestimehr and Professor Tibor Jánosi. My gratitude extends also to other educators with whom I had the opportunity to work before joining Cornell: Professor Sofiène Affes, Doctor Alex Stéphenne and Professor Mohamed-

Slim Alouini.

I am also thankful to Scott Coldren for being always helpful even when I bugged him with stupid administrative questions. I can not finish this section without thanking my friends who made my stay at Cornell more enjoyable, the list is long and I hope that I will not forget anyone (although this is bound to happen): Ebad Ahmed, Saifur Rahman, Yücel Altuğ, Benjamin Kelly, Hussam Abu-Libdeh, Ahmed Jaber, Mohamed Elhawary, Ahmad Helal, Anas Abognah ,Shiyao Chen, Nadine Hussami, Guilherme Pinto, Oliver Kosut, Jinsub Kim, Brandon Jones, Aaron Chuan Lei, Animashree Anandkumar, Stefan Geirhofer, Nithin Michael, Enrique Mallada ,Meng Wang, Alireza Vahid, and Ilan Shomorony.

Most importantly, I would like to thank my parents Moncef Laourine and Najet Ben Jemia, my siblings Foued Laourine and Fehmi Laourine and the rest of my family. This work is dedicated to them. Without their continuous support I would not have been able to finish this dissertation and write these words.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Information theory at its core revolves around two central questions: (1) What is the highest data rate at which reliable communication can take place over a given channel? and (2) What is the minimum rate required to describe a source within a given distortion constraint? The answer to both questions is known for most point-to-point communication systems. The picture however is far from being complete when we enter the realm of network information theory, i.e., when more than two participants (transmitter/receiver) are involved in the communication. The problems discussed in this thesis fall within this area.

This thesis has two distinct directions. In the first part we consider two problems in information theoretic security. For two different types of timing channels, we find the highest rate at which a transmitter can send information reliably to a legitimate receiver with zero information leakage to an eavesdropper. In the second part we solve a problem in distributed source coding. We find the minimum sum rate required to compress two remote correlated Gaussian sources — subject to mean-square error distortion constraints on each of the reconstructions — under a special covariance structure between the two sources and the observations of the encoders.

One can probably think of a convoluted (unnatural) way to link the two topics studied in this thesis. I will however refrain from such an attempt here. The reader will undeniably feel some discontinuity between the chapters, this discontinuity however brings with it some freedom as the chapters can be read in any particular order.

1.1 Information Theoretic Security

Like all design problems, constructing a communication system requires selecting from among several possible technologies by weighing their relative merits. The criteria of this comparison are well known: data rate; cost; complexity; bandwidth; reliability, both as a measure of fidelity and longevity; power; and latency. These parameters have received considerable attention within the context of classical communications engineering, and they are now reasonably well understood for most point-to-point communication scenarios.

More recently, *security* has been added to the list of criteria by which communication systems are evaluated. This is especially true in the context of wireless communication, which, when compared to wireline communication, is evidently more prone to casual eavesdropping. Security might appear to be a misplaced requirement of a communication system, because existing cryptographic protocols can always be added to any system that provides basic communication functionality. But these protocols are not provably secure, so there is no guarantee that they will not be broken in the future. At the same time, it is now recognized that communication systems themselves can provide intrinsic security even without the use of conventional cryptography. For instance, the transmitter can beamform to the intended receiver [1]-[2], channel randomness can be used for secret key generation [3], and coding can be employed to reduce the amount of information that is leaked to an eavesdropper [4]. We shall call the last of these techniques *wiretap coding*.

The idea behind wiretap coding dates back to the pioneering work of Wyner [4] and Csiszár and Körner [49], but it has received a surge of attention

of late with the rise of wireless communication [11]-[15]. The idea is that the intended receiver and any eavesdropper that is present will inevitably observe the transmitted signal through different channels. The transmitter can then encode the message so that it can be decoded by the intended receiver while the output of the eavesdropper's channel is virtually pure noise. Wiretap coding thereby provides provable security, but it typically requires reducing the data rate. It also requires the transmitter to know the law of both the legitimate receiver's and the eavesdropper's channel.

Most recent work in wiretap coding concentrated on finding the *secrecy capacity*—the maximum rate at which a transmitter can communicate reliably to a legitimate receiver with zero information leakage to an eavesdropper—of radio frequency channels. In this thesis we investigate the secrecy capacity of certain timing channels by looking at two different channels in this class, the first one is the Poisson channel and the second is the exponential server queue.

1.1.1 Secrecy Capacity of the Poisson channel

For radio-frequency (RF) channels with multipath fading, it can be nearly impossible to learn the eavesdropper's channel. Multipath fading makes the strength of the channel a sensitive function of the (unknown) position of the eavesdropper. In practice, it might be possible to guarantee that the eavesdropper is farther away from the transmitter than the legitimate receiver is, but this is clearly not sufficient to determine the eavesdropper's channel. Indeed, it is not even sufficient to guarantee that the eavesdropper's channel is weaker than the legitimate receiver's. Making matters worse, the eavesdropper can even control

her fading by moving around to find the strongest possible signal. Of course, not all RF channels are subject to multipath fading, but even the possibility that multipath might be present makes it difficult to provide guarantees on the relative strengths of the legitimate receiver's and the eavesdropper's channel.

Wireless communication systems operating in the optical and near-optical band are intrinsically more secure than those operating in the RF band. Infrared (IR) systems are very amenable to beamforming [5], and some ultraviolet (UV) systems experience high atmospheric attenuation, which enables secure short-range communication by guaranteeing that eavesdroppers beyond a certain distance are kept in the dark. Indeed, secure IR systems based on these advantages have been demonstrated [6]-[7] and UV systems are under development [8]-[9].

Despite this apparent advantage, relatively little is known about how to perform wiretap coding for near-optical systems. We examine the fundamental limits of coding for secure communication over optical channels by studying the secrecy capacity of the Poisson channel, a common model for direct detection optical communications systems. In such systems the transmitter sends information by modulating the intensity of an optical signal while the receiver observes the arrival moments of individual photons. The capacity of this channel has been determined under peak power constraint on the transmitted optical power by Kabanov [16] and under both average and peak power constraints by Davis [17]. Wyner [18] derived the reliability function of this channel for all rates below capacity and constructed exponentially optimal codes. Multiple-access Poisson channels were studied in [19]-[20] whereas broadcast Poisson channels were considered in [21] and [22]. The capacity of the Poisson channel has also been investigated in the presence of fading [24]. The second chapter

of this thesis will be dedicated to finding the secrecy capacity of the Poisson channel.

1.1.2 Secrecy Capacity of the Exponential Server Queue

Timing channels are by their own nature propitious to covert communication. That is because the information contained in the timing of events is usually not considered meaningful data and only the content of the event is of importance. Consider for instance a communication network, where information is transmitted using packets. Alice and Bob, as usual trying to evade Eve's prying eyes, can convey information to each other through the timing of the packets. Eve, who is just monitoring the content of the packets, will be fooled into thinking that Alice and Bob are talking about an innocuous subject like the weather whereas in fact they are exchanging love letters. In principle, if the packets experience deterministic delays in the network, the information capacity of this hidden channel can be arbitrarily large. Of course, packets in the network are usually subject to random delays, and so the capacity of this channel will be bounded. The question now is the following. What if Eve became aware of this little game, can Alice and Bob still be able to communicate secretly using this method? And if so, how much information can be transmitted securely between them? Clearly, if packet delays are deterministic, secure communication is going to be problematic. However, when randomness comes into play, the answer can be drastically different. Although generally considered a nuisance for reliable communication, randomness as research have shown [4] is actually a blessing in the context of secure communication.

In order to get a deeper understanding of this problem, we need to use a tractable model for abstracting the randomness affecting the transit times of the packets through the network. The first model that comes to mind is the exponential server queue or, using Kendall's notation, the $./M/1$ queue. The assumption that a packet will receive an exponentially distributed service time in a node of the network —independently from its service time in prior nodes— is attributed to Kleinrock. This assumption is in fact known as Kleinrock's independence assumption (or approximation) in queueing theory. Referring to this assumption, and in his own words, Kleinrock said [39]

> Now in networking, every pair of users is a different commodity. So from a theoretical point of view, you just can't solve these problems. So I asked, what can I do? And I made an assumption. I called it the independence assumption. And that cracked the problem wide open. The analysis became totally doable.

But even before this approximation, the exponential distribution has had a prominent role in the development of queueing networks. Indeed, most fundamental queueing theory results involve directly or indirectly $./M/1$ queues. Jackson's and Burke's theorems are just two concrete examples of such results.

The exponential distribution had also its share of contribution to the development of information theory. Arguably however the impact of this distribution on Shannon's theory is much less pronounced than its impact on Erlang's theory. Perhaps because it was eclipsed by the Gaussian distribution. Nevertheless, from an information theoretic perspective, the exponential distribution shares many of the interesting properties of its Gaussian counterpart [40]. In our biased opinion, probably the most prominent paper that exemplifies the

fundamental role that the $./M/1$ queue can play in information theory is [41]. In that paper, Anantharam and Verdú studied timing channels by looking at a single server queue where every packet has an independent service (transit) time. They showed that the information capacity of a queue with service rate μ is lower bounded by $\frac{\mu}{e}$ and that this lower bound is attained when the service time is exponentially distributed. That is the exponential server timing channel (ESTC) or the $\cdot/M/1$ queue has the lowest capacity. The $\cdot/M/1$ queue is also one of the rare non-trivial queues for which the capacity is known in closed-form. In this regard, the $\cdot/M/1$ queue plays for single server queues, the same role played by the additive white Gaussian noise (AWGN) channel for additive noise channels.

Due to the fundamental nature of the $./M/1$ queue, we will adopt this model to study the secrecy problem described in the beginning of this section. The third chapter of this thesis is dedicated to finding the secrecy capacity of the exponential server queue.

1.2 Distributed Source Coding: Compressing Neighbors in a Gauss-Markov Tree

One does not need to be an information theorist to appreciate the importance of knowing the compression limits of information sources and designing schemes to attain these limits. Algorithms for lossless and lossy data compression are pervasive today and everyone is using them without necessarily knowing how they function. The widespread use of these algorithms should not hide the fact that our understanding of the problems of compression is still very limited. An

area where progress is very much needed is distributed source coding. In distributed source coding, multiple correlated sources are observed by multiple encoders. These encoders communicate their messages over rate-constrained channels to one decoder (or potentially multiple decoders). The decoder then uses these messages to form an estimate of a subset (or all) of the sources. Applications for distributed source coding abound, for example to name just a few, sensor networks and video coding for multimedia applications. The objective of this part of this thesis is to gain a better understanding of this area by solving one particular distributed source coding problem that shall be described shortly.

In the following, we will restrict our attention to Gaussian sources. Since some practitioners might object to this simplification, we will give here two reasons that we believe are sufficient to alleviate their objections. The first one is that Gaussian sources are "friendly" objects in information theory. Indeed Gaussian random variables possess several fundamental properties that are very useful in proving theorems in information theory. The second reason— which will be probably more appealing to these practitioners— is that knowing the compression limits for Gaussian sources can usually serve as a benchmark for other sources. For instance, it is known that for a given variance, a Gaussian source has the largest rate-distortion function.

Now let us begin at the beginning. Consider a source of information that is generating a stream of independent and identically distributed zero mean Gaussian random variables with unit variance. Shannon showed in 1959 that the minimum number of bits per sample that an encoder needs to send to a

decoder to achieve a mean square error distortion d is[1]

$$R(d) = \frac{1}{2}\log^+\left(\frac{1}{d}\right) \text{ bits/sample.}$$

Now consider another problem where instead of having one source and one encoder, we have two correlated Gaussian sources with correlation coefficient ρ and two encoders each one of them observing one of the sources. The encoders send messages, without cooperating, to a central decoder whose purpose is to estimate the two sources within a maximum mean square error distortion d_1 on the reconstruction of the first source and d_2 on the second. If we denote by $R_i, i = 1, 2$ the number of bits per sample used by the ith encoder, then a natural question to ask is the following. What is the minimum sum rate (i.e., $\min R_1 + R_2$) required to achieve the two distortions d_1 and d_2? This question remained open until 2006, when Wagner et al. showed that this sum rate is given by

$$R(d_1, d_2) = \frac{1}{2}\log^+\left(\frac{1-\rho^2}{2d_1d_2}\left[1+\sqrt{1+\frac{4\rho^2 d_1 d_2}{(1-\rho^2)^2}}\right]\right) \text{ bits/sample.}$$

Our goal in this part of the thesis is to generalize this result in the following sense: while still assuming that there are two Gaussian sources that we would like to estimate, we allow more than two encoders and (more importantly) we assume that these encoders might not have direct access to the sources. By that we mean that each encoder now observes one of the sources through an additive white gaussian noise channel. Moreover the noise corrupting the observation of the ith encoder can be correlated with the noise corrupting the observation of the jth encoder. In chapter 4 of this thesis, we find the sum-rate for this problem under a specific covariance structure between the observations of the encoders.

[1]The logarithm is base two.

More specifically, we solve the problem when the two sources of interest are the roots of a Gauss-Markov tree and the encoders observe the leaves of that tree.

We believe that this problem is important on multiple levels. First because it is a natural generalization of several distributed source coding problems. Starting from the Gaussian CEO problem [62], the Gaussian two-encoder problem [54] and finally the Gaussian many-help-one problem with a tree structure [55]. All these problems can be seen as special cases of the one descried in chapter 4. The second reason is that studying remote or indirect multiterminal source coding problems usually allows us to find outer bounds on the rate region of other (direct) multiterminal source coding problems. For example, the Gaussian CEO problem was instrumental in the solution of the Gaussian two-encoder problem. We hope therefore that the sum-rate result provided here could be used by others to solve future distributed source coding problems. The third reason is that we think that the conditional independence structure — that is present in our problem — could potentially be a key in differentiating between "hard" problems and solvable ones. Indeed, most multiterminal source coding problems for which the rate region is known involve some sort of conditional independence (see for example [62], [63], [64]). This conditional independence is either present naturally in the source, e.g., [62] or is introduced artificially using source augmentation, e.g., [54]. Finally we think that the solution that we provide could be a stepping stone to solve the problem of reproducing three neighboring variables in the tree. By removing the distortion constraint on the middle variable of the triple, one could obtain a result for pairs of variables that are separated by at most one variable in the tree, which in particular would solve the Gaussian one-help-two problem under a tree constraint. The ultimate goal would be to handle distortion constraints on an arbitrary number of variables in the tree.

CHAPTER 2

SECRECY CAPACITY OF THE POISSON CHANNEL

2.1 Introduction

We study in this chapter the degraded Poisson wiretap channel. The legitimate receiver observes a doubly stochastic Poisson process with instantaneous rate $A_y X_t + \lambda_y$ where $\{X_t, 0 \leq t \leq T\}$ is the signal transmitted. The eavesdropper's observation is also a doubly stochastic Poisson process with instantaneous rate $A_z X_t + \lambda_z$. For degradedness we assume that[1] $A_y \geq A_z$ and $\lambda_y \leq \frac{A_y}{A_z}\lambda_z$. In Theorem 1 we provide a closed form expression of the secrecy capacity as a function of the parameters $(A_u, \lambda_u),\ u \in \{y, z\}$. This result is further extended by Theorem 5 which gives a full characterization of the rate equivocation region.

Our achievability proof uses stochastic encoding as well as the structured codes constructed by Wyner for the Poisson channel [18]. As for the converse, we will see that the infinite bandwidth nature of the Poisson channel makes it possible to prove the converse using only simple properties of the conditional expectation combined with basic information theoretical inequalities. This is to be contrasted with the converse of the (finite bandwidth) Gaussian channel which is proved using the entropy power inequality (EPI) [25] or the worst additive noise result developed in [26]. As an illustration for the basic ideas that underpin the converse for the Poisson channel, we will start by considering here the more familiar infinite bandwidth Gaussian channel and we will see also that for this channel the proof of the converse simplifies considerably.

[1]These conditions were shown to be sufficient for degradedness in [21]. The argument is reproduced in Lemma 1 below.

For this purpose, consider the continuous time Gaussian wiretap channel with bandwidth B (later we will let B tend to infinity) and with a power constraint P. This continuous time channel is equivalent to $2B$ uses per second of the discrete time Gaussian channel depicted in Figure 2.1. The input signal is power constrained, i.e., $\mathbb{E}[X^2] \leq P$, the legitimate receiver observes $Y = X + W_1$ and the eavesdropper receives $Z = Y + W_2 = X + W_1 + W_2$, where $W_i \sim \mathcal{N}(0, N_i B)$ and $W_1 \perp\!\!\!\perp W_2$.

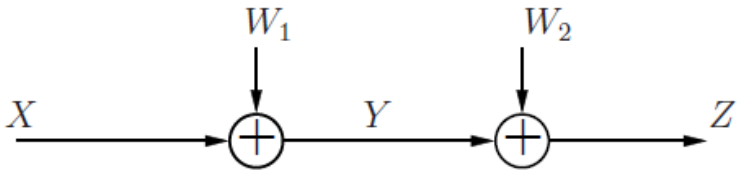

Figure 2.1: The discrete time Gaussian wiretap channel

Define $\tilde{N}$ by $\frac{1}{\tilde{N}} = \frac{1}{N_1} - \frac{1}{N_1+N_2}$ and observe that

$$\frac{\tilde{N}}{N_1} Y = X + \frac{N_1}{N_2} Z + \tilde{W}, \tag{2.1}$$

where $\tilde{W} = W_1 - \frac{N_1}{N_2} W_2$. It is easy to see that $\tilde{W} \sim \mathcal{N}(0, \tilde{N} B)$ and $\mathbb{E}[\tilde{W}(W_1 + W_2)] = 0$, it follows therefore that $\tilde{W} \perp\!\!\!\perp W_1 + W_2$ (since they are jointly Gaussian). For the discrete time Gaussian wiretap channel, it is known that the secrecy capacity is given by $\max_{p_X}(I(X;Y) - I(X;Z))$. For the continuous time channel counterpart with bandwidth B, the secrecy capacity becomes $C_s^B = 2B \max_{p_X}(I(X;Y) - I(X;Z))$.

In [25], using the celebrated EPI, a closed form expression for the secrecy capacity of the discrete time Gaussian wiretap channel was obtained. In just a few steps, we will see that the secrecy capacity of the infinite bandwidth ($B \to \infty$) Gaussian wiretap channel can be found much more simply. Starting with

(2.1) we obtain the following sequence of inequalities

$$\begin{aligned} I(X;Y) &= I(X; X + \frac{N_1}{N_2}Z + \tilde{W}) \\ &\overset{(a)}{\leq} I(X; X + \tilde{W}, Z) \\ &\overset{(b)}{=} h(X + \tilde{W}, Z) - h(X + \tilde{W}, Z|X) \\ &\overset{(c)}{\leq} h(X + \tilde{W}) + h(Z) - h(X + \tilde{W}, Z|X) \\ &\overset{(d)}{=} h(X + \tilde{W}) + h(Z) - h(X + \tilde{W}|X) - h(Z|X) \\ &\overset{(e)}{=} I(X; X + \tilde{W}) + I(X; Z), \end{aligned}$$

Inequality (a) follows from the data processing inequality, equalities in (b) and (e) are standard information theory identities, (c) follows from the independence bound on entropy and finally (d) holds because $X + \tilde{W}$ and $X + W_1 + W_2$ are conditionally independent given X (i.e., $(X+\tilde{W}) \perp\!\!\!\perp (X+W_1+W_2)|X$). Basically, the key identity needed to go from (a) to (e) is the following: if $Y_1 \perp\!\!\!\perp Y_2|X$, then we have $I(X; Y_1, Y_2) \leq I(X; Y_1) + I(X; Y_2)$. A proof of this simple inequality in a more general setting will be given later and will be used in part of the converse for the Poisson channel. Going back to the Gaussian problem, we see that

$$C_s^B = 2B \max_{P_X}(I(X;Y) - I(X;Z)) \leq 2B \max_{P_X} I(X; X + \tilde{W}) = B \ln(1 + \frac{P}{B\tilde{N}}). \tag{2.2}$$

For a fixed bandwidth B, this last inequality is not tight. But letting $B \to \infty$ we obtain

$$C_s^\infty \triangleq \lim_{B\to\infty} C_s^B \leq \frac{P}{\tilde{N}} = \frac{P}{N_1} - \frac{P}{N_1 + N_2}. \tag{2.3}$$

However, since $\max_{p_X}(I(X;Y) - I(X;Z)) \geq \max_{p_X} I(X;Y) - \max_{p_X} I(X;Z)$, we also have that

$$C_s^\infty \geq \lim_{B\to\infty} \left(B \ln(1 + \frac{P}{BN_1}) - B \ln(1 + \frac{P}{B(N_1 + N_2)}) \right) = \frac{P}{N_1} - \frac{P}{N_1 + N_2}. \tag{2.4}$$

It follows that $C_s^{\infty} = \frac{P}{N_1} - \frac{P}{N_1+N_2}$.

This remarkably simple approach will be useful for the Poisson channel. More specifically, when $\frac{\lambda_y}{A_y} = \frac{\lambda_z}{A_z}$, the eavesdropper's signal Z is a thinned version of the legitimate receiver's signal Y, i.e., [2] $Y = Z + \tilde{Z}$ where $Z \perp\!\!\!\perp \tilde{Z}|X$, the approach above gives that $I(X;Y) - I(X;Z) \leq I(X;\tilde{Z})$. Since $\tilde{Z}$ is itself a doubly stochastic Poisson process, the mutual information $I(X;\tilde{Z})$ can be maximized using the martingale techniques of Kabanov [16] and an (achievable) upperbound can be obtained on $I(X;Y) - I(X;Z)$. When $\frac{\lambda_y}{A_y} < \frac{\lambda_z}{A_z}$, a different bounding technique using only simple properties of the conditional expectation will be devised.

Although no "sophisticated" tools are required to prove the converse, we show in Section 2.7 that using some new results in information theory an alternative proof can be provided. This different proof hinges on the link that has been established between the mutual information (MI) and the minimum mean square estimation (MMSE) in Poisson channels [27]. It is worth noting at this point that the link between the MI and the MMSE in the Gaussian setting [28] has been also used recently for different Gaussian wiretap channels [29], [30].

One of the distinctive aspects of our work is that we do not resort to the Δ-discretization method introduced by Wyner [18]. This method was used to approximate the Poisson channel by a binary DMC thereby allowing the transposition of the widely known results for DMCs to the Poisson channel. This technique leads to extensive computations, especially when we are interested in the secrecy capacity as there are now two conflicting objectives involved, the maximization of the information rate at the legitimate receiver and the mini-

[2] The time dependence has been dropped to ease the notations. Refer to the converse part in this chapter for a mathematically precise statement.

mization of the information leakage at the eavesdropper. We circumvent the use of this method by using the techniques described above.

The rest of this chapter is organized as follows. Section 2.2 describes the setup of the problem and presents the main result of this chapter as well as some interpretations of the obtained result. The proof of the achievability of the secrecy capacity is given in Section 2.3 and the proof of the converse is presented in Section 2.4. In Section 2.5 we extend the main result by giving a complete characterization of the rate-equivocation region. Finally, in Section 2.6, some possible future directions are discussed.

2.2 Problem and Result Statement

The input process to the Poisson channel is a waveform denoted by $X_0^T \triangleq \{X_t, 0 \leq t \leq T\}$ satisfying $X_t \geq 0$ for all t. We further assume that the input process is peak power limited, i.e., $X_t \leq 1$ for all t. The received signal at the legitimate receiver Y_0^T is a doubly stochastic Poisson process with instantaneous rate $A_y X_t + \lambda_y$, i.e., given X_0^T the stochastic process Y_0^T has independent increments with $Y_0 = 0$ and for $0 \leq s \leq t \leq T$ we have

$$\Pr(Y_t - Y_s = k | X_0^T) = \frac{1}{k!}\Upsilon^k(s,t)e^{-\Upsilon(s,t)},\ k \in \mathbb{N},$$

where

$$\Upsilon(s,t) = \int_s^t (A_y X_\tau + \lambda_y) d\tau.$$

The parameter $A_y > 0$ accounts for possible signal attenuation at the receiver. The parameter $\lambda_y \geq 0$ is the dark current intensity which results from background noise and bears no information on the input process X_0^T. Similarly the

output process of the eavesdropper Z_0^T is a doubly stochastic Poisson process with instantaneous rate $A_z X_t + \lambda_z$.

In this chapter, the space of doubly stochastic Poisson processes on the interval $[0, T]$ will be denoted by $\mathcal{P}(T)$. Following the notation used in [27] the output process of the Poisson channel in the interval $[0, T]$ with instantaneous rate $\alpha X_t + \lambda$ will be denoted by $\mathcal{P}_0^T(\alpha X_0^T + \lambda)$. We use $\langle X_t \rangle_s$ to designate $E[X_t | \mathcal{P}_0^s(X_0^s)]$, as such $\langle X_t \rangle_t$ refers to the causal conditional mean estimate and $\langle X_t \rangle_T$ to the noncausal one.

All stochastic processes considered in this chapter are defined on a common measurable space $(\Omega, \mathcal{F})$. We use $\mathcal{F}_\xi^s$ to denote the internal history generated by the process ξ_0^s.

In this chapter we are interested in the degraded Poisson wiretap channel. Lapidoth et al. [21] gave conditions on the parameters (A_u, λ_u), $u \in \{y, z\}$ for stochastic degradedness. These conditions are presented in the following lemma. In order to prepare for the results to come we will also briefly go over the proof of this lemma.

Lemma 1 (Lapidoth, Telatar and Urbanke [21])**.** *The eavesdropper's channel is stochastically degraded with respect to the legitimate receiver's channel, if*

$$A_y \geq A_z, \tag{2.5}$$

and

$$\lambda_y \leq \frac{A_y}{A_z} \lambda_z. \tag{2.6}$$

Proof. Let $\widetilde{Y}_0^T$ (cf. Figure 2.2) be the process defined as follows[3]

$$\widetilde{Y}_t = Y_t + H_t,\ t \in [0, T], \tag{2.7}$$

where H_t is a homogeneous Poisson process with rate $\tilde{\lambda} = \frac{A_y}{A_z}\lambda_z - \lambda_y$ (note that $\tilde{\lambda} \geq 0$ by (2.6)) independent of (X_0^T, Y_0^T). It follows that $\widetilde{Y}_0^T$ is a doubly stochastic Poisson process with instantaneous rate $A_y X_t + \lambda_y + \tilde{\lambda} = A_y X_t + \frac{A_y}{A_z}\lambda_z$. The process Z_0^T is then obtained from $\widetilde{Y}_0^T$ by thinning with erasure probability $1 - \frac{A_z}{A_y}$ (note that because of (2.5) this quantity is ≥ 0). □

In the rest of this chapter we will assume that at least one of the inequalities (2.5) or (2.6) is strict. Note that this assumption can be made without losing generality for if there was an equality in (2.5) and (2.6) then the legitimate receiver's channel and the eavesdropper's channel will be identical and the secrecy capacity will be zero.

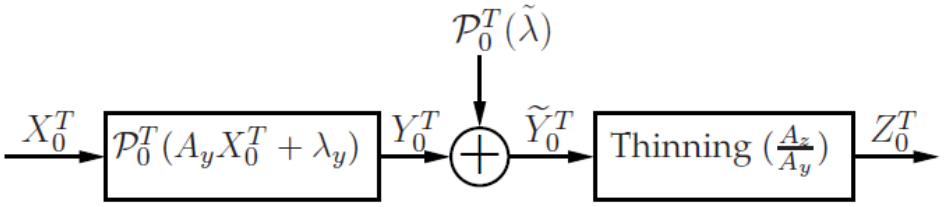

Figure 2.2: The degraded Poisson wiretap channel

We move now to the description of the information transmission aspect of the problem. The transmitter wishes to communicate a message U uniformly distributed on $\mathcal{U} = \{1, ..., M\}$. An (M, T) code (E_T, D_T) for the Poisson wiretap channel is a stochastic encoder E_T that maps a message U to a waveform X_0^T

[3] As depicted in Figure 2.2, the description provided here is that of a *physically* degraded Poisson channel and not of the more general *stochastically* degraded Poisson channel. This restriction is harmless however since the secrecy capacity depends only on the conditional marginals.

which satisfies the peak power constraint and a decoder $D_T : \mathcal{P}(T) \to \mathcal{U}$. The transmission rate of this code is

$$R = \frac{H(U)}{T} = \frac{1}{T} \ln M.$$

The average probability of error at the legitimate receiver is

$$P_e = \frac{1}{M} \sum_{m=1}^{M} \Pr\left(D_T(Y_0^T) \neq m | U = m\right). \tag{2.8}$$

The level of secrecy is measured by $\frac{1}{T} I(U; Z_0^T)$. This normalized mutual information quantifies the rate of the information leaked to the eavesdropper about the message U. As such our goal is to make this quantity as small as possible.

Definition A secrecy rate R_s is said to be *achievable*[4] for the Poisson wiretap channel if for all $\epsilon > 0$ and all sufficiently large T, there exists an (M, T) code such that

$$\begin{aligned} \frac{\ln M}{T} &\geq R_s - \epsilon \\ P_e &\leq \epsilon \\ \frac{1}{T} I(U; Z_0^T) &\leq \epsilon \end{aligned} \tag{2.9}$$

The supremum of achievable secrecy rates will be called the *secrecy capacity*. The main result of this chapter is the following.

Theorem 1. *The secrecy capacity of the degraded Poisson wiretap channel is given by*[5]

$$C_s = \alpha^*(A_y - A_z) + \ln\left(\frac{\lambda_y^{\lambda_y}}{\lambda_z^{\lambda_z}}\right) + \ln\left(\frac{(A_z\alpha^* + \lambda_z)^{\lambda_z}}{(A_y\alpha^* + \lambda_y)^{\lambda_y}}\right), \tag{2.10}$$

where α^* is the unique solution in $[0, 1]$ to the following equation

$$\frac{(A_y\alpha^* + \lambda_y)^{A_y}}{(A_z\alpha^* + \lambda_z)^{A_z}} = e^{A_z - A_y} \frac{(A_y + \lambda_y)^{A_y + \lambda_y}}{(A_z + \lambda_z)^{A_z + \lambda_z}} \frac{\lambda_z^{\lambda_z}}{\lambda_y^{\lambda_y}}. \tag{2.11}$$

[4] Equivalently, we say that R_s is achievable with *perfect secrecy*.

[5] If $\lambda = 0$, the convention is that $0^0 = 1$.

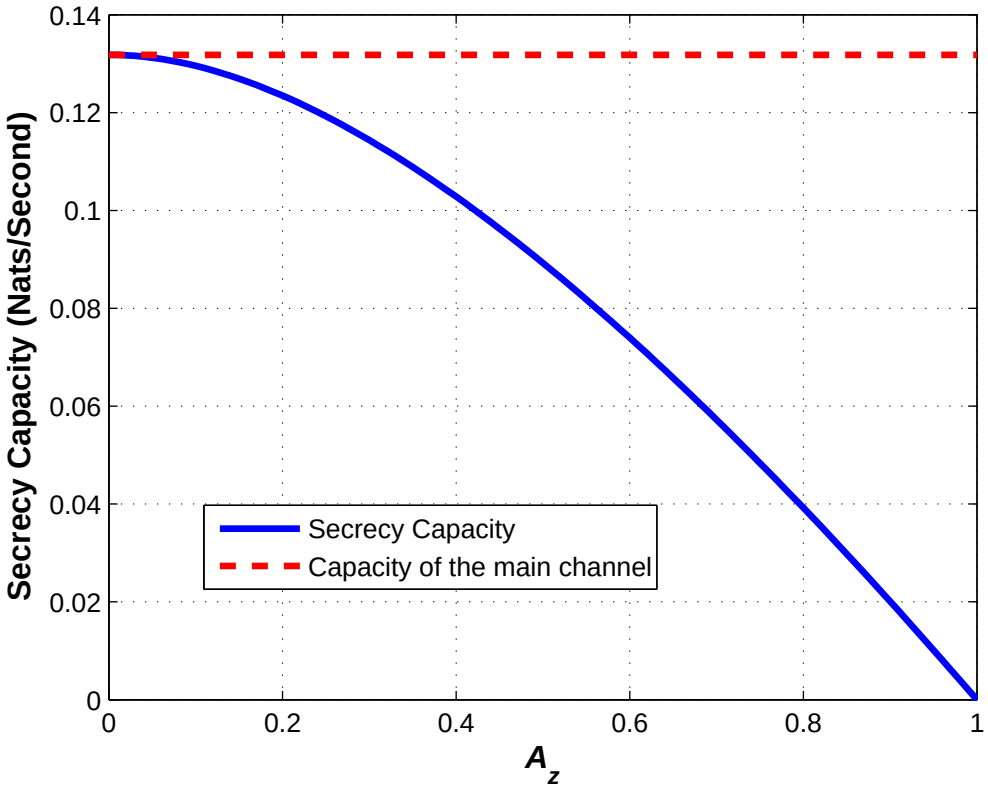

Figure 2.3: The secrecy capacity for $A_y = 1$, $\lambda_y = \lambda_z = 0.5$.

This result assumes that (A_z, λ_z) is known to the transmitter. Yet it follows that C_s is an achievable rate with perfect secrecy even if the eavesdropper observes $\mathcal{P}_0^T(A_z' X_0^T + \lambda_z')$, where A_z' and λ_z' are unknown but satisfy $A_z' \leq A_z$ and $\lambda_z' \geq \frac{A_z'}{A_z}\lambda_z$.[6] Thus, only one-sided estimates of A_z and λ_z are needed. In practice, an upper bound on A_z could be provided by guaranteeing that any potential eavesdropper is more than a certain distance away from the transmitter. A lower bound on the dark current λ_z could be provided using ambient noise measurements and the known physical limitations of existing receivers.

[6] If $V_0^T \sim \mathcal{P}_0^T(A_z' X_0^T + \lambda_z')$ where A_z' and λ_z' are unknown but satisfy $A_z' \leq A_z$ and $\lambda_z' \geq \frac{A_z'}{A_z}\lambda_z$ then, from Lemma 1, V_0^T can be viewed as a degraded version of Z_0^T. This means that $I(U;V_0^T) \leq I(U;Z_0^T)$ and a fortiori $\lim_{T\to\infty} \frac{1}{T} I(U;Z_0^T) = 0 \Rightarrow \lim_{T\to\infty} \frac{1}{T} I(U;V_0^T) = 0$.

Figure 2.3 depicts the secrecy capacity versus A_z for $A_y = 1$ and $\lambda_y = \lambda_z = 0.5$. As expected, the secrecy capacity is a deceasing function of A_z. When $A_z = 0$, the secrecy capacity is equal to the capacity of the main channel and it then decreases until it reaches zero when $A_z = A_y = 1$.

Worst case scenario: A particularly insightful case is when $\frac{\lambda_y}{A_y} = \frac{\lambda_z}{A_z} = \sigma$. This situation happens when the eavesdropper observes a thinned version of the signal of the legitimate receiver, i.e., $H_t \equiv 0$ in (2.7). In this case, after some algebraic manipulations, we obtain that

$$\alpha^* = \frac{(1+\sigma)^{1+\sigma}}{e\sigma^{\sigma}} - \sigma, \tag{2.12}$$

and the secrecy capacity reduces to

$$C_s = (\lambda_y - \lambda_z)\left(\frac{1}{e}\left(1+\frac{1}{\sigma}\right)^{1+\sigma} - (1+\sigma)\ln\left(1+\frac{1}{\sigma}\right)\right). \tag{2.13}$$

This is saying that the secrecy capacity is the difference between the capacity of the main channel (the channel between the transmitter and the legitimate receiver) and the capacity of the eavesdropper's channel. For instance, in the special case when there is no dark current $\lambda_y = \lambda_z = 0$, we find that $\alpha^* = \frac{1}{e}$ and the secrecy capacity reduces to

$$C_s = \frac{A_y - A_z}{e}. \tag{2.14}$$

For a degraded DMC, Wyner [4] showed that the secrecy capacity is equal to $\max_{p_X}(I(X;Y) - I(X;Z))$. Hence the following inequality is always satisfied

$$\text{Secrecy Capacity} \geq C_M - C_W,$$

where C_M is the capacity of the main channel and C_W is the capacity of the eavesdropper's channel. As shown in [31]-[32], there is equality in the inequality above if there is an input probability distribution p_X that simultaneously

maximizes $I(X;Y)$ and $I(X;Z)$. This is exactly what is happening here, when $\frac{\lambda_y}{A_y} = \frac{\lambda_z}{A_z}$ the mutual information $I(X_0^T;Y_0^T)$ and $I(X_0^T;Z_0^T)$ are both maximized by letting the input X_0^T cycle infinitely fast between its extreme values, i.e., the peak power 1 and 0 with $\Pr(X_t = 1) = 1 - \Pr(X_t = 0) = \alpha^* = \frac{(1+\sigma)^{1+\sigma}}{e\sigma^\sigma} - \sigma$. In order to achieve this limit the communication bandwidth must become very large. However, we are assuming here that this bandwidth is still negligible compared to the optical center frequency at which the communication system is operating. If this is not the case anymore, then the channel model that we are using is no longer appropriate.

Before we proceed further with the presentation of the problem considered here, we give a lemma that will prove to be useful in the proofs of the achievability and the converse, a proof of this result can be found for instance in [16].

Lemma 2. *The mutual information between the input X_0^T and the output $\mathcal{P}_0^T(\alpha X_0^T + \lambda)$ can be upper bounded as follows*[7]

$$I(X_0^T; \mathcal{P}_0^T(\alpha X_0^T + \lambda)) \leq \int_0^T (\mathbb{E}[\vartheta(X_t)] - \vartheta(\mathbb{E}[X_t]))dt, \tag{2.15}$$

where $\vartheta(x) = (\alpha x + \lambda)\ln(\alpha x + \lambda)$.

2.3 Achievability of C_s

Our achievability proof relies on the structured codes that were designed for the Poisson channel by Wyner [18]. As pointed out by one of the reviewers, an alternative proof can be obtained by restricting the input process to the channel to be a piecewise constant binary waveform and using the lower bound on the

[7]Note that some authors use the function $\vartheta(x) = (\alpha x + \lambda)\ln(\alpha x + \lambda) - \lambda \ln \lambda$ instead but the constant term $\lambda \ln \lambda$ cancels out here.

secrecy capacity derived in [33],[34]. Before delving into the details of the proof, we will briefly describe the code construction and the properties inherited by this code.

Wyner codes $\mathcal{W}(T, M, k)$**:** Let T, M and k be given, and construct an $M \times \binom{M}{k}$ binary matrix Π as follows. The columns of Π are the $\binom{M}{k}$ binary M-vectors with exactly k ones and $M-k$ zeros. Now partition the interval $[0, T]$ into $\binom{M}{k}$ subintervals of equal length $\varpi_T \triangleq \frac{T}{\binom{M}{k}}$ and construct M waveforms $\{X_0^T(m)\}_{m=1}^M$ as follows

$$X_t(m) = \Pi(m, n), t \in ((n-1)\varpi_T, n\varpi_T],\ n = 1, ..., \binom{M}{k}. \tag{2.16}$$

For $\alpha = \frac{k}{M}$ fixed, these codes satisfy

$$\frac{1}{T}\mu\{t : X_t(m) = 1\} = \alpha, \quad \text{for all } m, \tag{2.17}$$

with μ being the Lebesgue measure. If moreover $M = e^{RT}$, for $T >> 1$, Wyner showed that for $m \neq m'$

$$\frac{1}{T}\mu\{t : X_t(m) = 1, X_t(m') = 0\} \approx \alpha(1-\alpha). \tag{2.18}$$

As such for T large enough the codewords $\{X_0^T(m)\}_{m=1}^M$ will behave as if they were chosen independently.

After this brief overview of Wyner codes we are in a position to state the achievability theorem and prove it.

Theorem 2. *Any secrecy rate $R_s < C_s$ is achievable.*

Proof. Let $\epsilon > 0$ be arbitrary and let $R_s = C_s - \epsilon$. Define

$$\begin{aligned} R_u = {} & \alpha^*(A_u + \lambda_u)\ln(A_u + \lambda_u) + (1-\alpha^*)\lambda_u \ln \lambda_u \\ & - (A_u\alpha^* + \lambda_u)\ln(A_u\alpha^* + \lambda_u),\ u \in \{y, z\}. \end{aligned} \tag{2.19}$$

After few algebraic manipulations, we can show that

$$C_s = R_y - R_z.$$

Given these parameters, the encoder-decoder pair considered here works as follows.

Encoding: Let $M = e^{R_s T}$ and let U be uniformly distributed on $\mathcal{U} = \{1, ..., M\}$. Define $M_y = e^{(R_y - \frac{3}{2}\epsilon)T}$ and following the steps described above construct a code[8] $\mathcal{C} = \mathcal{W}(T, M_y, \alpha^* M_y)$. Partition this code arbitrarily into M smaller subcodes, i.e., $\mathcal{C} = \cup_{i=1}^{M} \mathcal{C}_i$. The cardinality of each each subcode $\mathcal{C}_i$ will be equal to $M_z = \frac{M_y}{M} = e^{(R_z - \frac{\epsilon}{2})T}$.

The encoder works as follows, when the message $U = m$ is chosen, the codeword X_0^T is selected uniformly randomly from $\mathcal{C}_m$. Notice that every subcode $\mathcal{C}_m$ can be viewed as a code for the eavesdropper's channel with M_z codewords and a uniform prior distribution. The method used to build $\mathcal{C}$ mimics the construction used by Wyner in his original paper [4], where a code for the main channel is obtained by combining an appropriate number of subcodes that achieve the rate R_z with vanishing error probability over the eavesdropper's channel.

Decoding: The decoder considered here is the maximum likelihood decoder constructed by Wyner [18]. After observing Y_0^T, the decoder at the legitimate receiver computes the following metric

$$\Psi_m = \int_{S_m} dY_t, \tag{2.20}$$

where $S_m = \{t \in [0, T] : \ X_t(m) = 1\}$. Then $D_T(Y_0^T) = m$ if m maximizes Ψ_m, with ties resolved in favor of the smallest m.

[8]Note that even if α^* is not a rational number, it can be approximated arbitrary close by rationals.

Analysis of P_e: The fact that $P_e \to 0$, follows simply from the fact that Wyner codes with the peak power 1 and average power α^* are capacity achieving for the main channel.

Analysis of $\frac{1}{T}I(U; Z_0^T)$:

Notice first that for each m, the waveform $X_0^T(m)$ is piecewise constant. It follows that a sufficient statistic for making a decision is the number of arrivals during each subinterval $((n-1)\varpi_T, n\varpi_T]$, i.e., $Z_n = Z_{n\varpi_T} - Z_{(n-1)\varpi_T}$, $n = 1, ..., N_y$ with $N_y = \binom{M_y}{\alpha^* M_y}$. Consequently,

$$\begin{aligned} I(X_0^T; Z_0^T) &= I(\mathbf{X}; \mathbf{Z}) \\ I(U; Z_0^T) &= I(U; \mathbf{Z}) \end{aligned} \tag{2.21}$$

where $\mathbf{X} = [X_1, ..., X_{N_y}]$, $X_i = 0$ or 1 depending on the choice of the codeword and $\mathbf{Z} = [Z_1, ..., Z_{N_y}]$. The equalities above follows from the fact that $\mathbf{Z}$ is a sufficient statistic.

As a result of Lemma 2, we have

$$\frac{1}{T} I(X_0^T; Z_0^T) \leq \frac{1}{T} \int_0^T (\mathbb{E}[\phi_z(X_t)] - \phi_z(\mathbb{E}[X_t]))dt, \tag{2.22}$$

where $\phi_z(x) = (A_z x + \lambda_z) \ln(A_z x + \lambda_z)$. Because of the uniform choice in the encoding scheme and in view of (2.17) we must have that $\Pr[X_t = 1] = 1 - \Pr[X_t = 0] = \alpha^*$, hence we have

$$\begin{aligned} \mathbb{E}[\phi_z(X_t)] &= \alpha^* \phi_z(1) + (1 - \alpha^*)\phi_z(0) \\ &= \alpha^*(A_z + \lambda_z) \ln(A_z + \lambda_z) + (1 - \alpha^*)\lambda_z \ln \lambda_z. \end{aligned} \tag{2.23}$$

and

$$\phi_z(\mathbb{E}[X_t]) = \phi_z(\alpha^*) = (A_z \alpha^* + \lambda_z) \ln(A_z \alpha^* + \lambda_z). \tag{2.24}$$

Consequently, we deduce that

$$\frac{1}{T}I(\mathbf{X};\mathbf{Z}) = \frac{1}{T}I(X_0^T; Z_0^T) \leq R_z. \tag{2.25}$$

As stated before, every subcode $\mathcal{C}_m$ can be viewed as a code for the eavesdropper's channel. Define δ_m to be the probability of error for code $\mathcal{C}_m$ $(1 \leq m \leq M)$ with the (optimal) decoder described above. By the aforementioned code construction, the codewords of every subcode $\mathcal{C}_m$ $(1 \leq m \leq M)$ satisfy (2.17) and (4.6) (with α replaced by α^*). These two conditions dictate the pairwise error probability of the codewords in $\mathcal{C}_m$ [18]. Since the overall error probability of the code $\mathcal{C}_m$ is governed by the pairwise error probability [18], it follows that the error probability δ_m can be made arbitrarily small for every subcode $\mathcal{C}_m$. Now, from Fano's inequality we have

$$H(\mathbf{X}|\mathbf{Z}, U = m) \leq H(\delta_m) + \delta_m \ln M_z, \tag{2.26}$$

where $H(p) = -p\ln(p) - (1-p)\ln(1-p)$ is the binary entropy.

Since the codewords are uniformly distributed in each subcode, we deduce that $H(\mathbf{X}|U = m) = \ln M_z$. We conclude therefore that

$$\begin{aligned} I(\mathbf{X};\mathbf{Z}|U = m) &= H(\mathbf{X}|U = m) - H(\mathbf{X}|\mathbf{Z}, U = m) \\ &\geq \ln M_z - (H(\delta_m) + \delta_m \ln M_z). \end{aligned} \tag{2.27}$$

Let $\delta = \frac{1}{M}\sum_{m=1}^{M} \delta_m$, averaging over U and by using the concavity of $H(\cdot)$ we find that

$$I(\mathbf{X};\mathbf{Z}|U) \geq \ln M_z - (H(\delta) + \delta \ln M_z). \tag{2.28}$$

Notice also that $U \to \mathbf{X} \to \mathbf{Z}$ form a Markov chain, i.e.,

$$\begin{aligned} \frac{1}{T}I(U;\mathbf{Z}) &= \frac{1}{T}I(U,\mathbf{X};\mathbf{Z}) - \frac{1}{T}I(\mathbf{X};\mathbf{Z}|U) \\ &= \frac{1}{T}I(\mathbf{X};\mathbf{Z}) - \frac{1}{T}I(\mathbf{X};\mathbf{Z}|U) \end{aligned} \tag{2.29}$$

Combined with the last inequality this implies that

$$\frac{1}{T}I(U;\mathbf{Z}) \leq \frac{1}{T}I(\mathbf{X};\mathbf{Z}) - \frac{1}{T}\ln M_z + \frac{1}{T}(H(\delta) + \delta \ln M_z). \tag{2.30}$$

Inequalities (2.25) and (2.30) result in the following

$$\frac{1}{T}I(U;\mathbf{Z}) \leq R_z - \frac{1}{T}[\ln M_z - (H(\delta) + \delta \ln M_z)]. \tag{2.31}$$

As $M_z = e^{(R_z - \frac{\epsilon}{2})T}$, this gives

$$\frac{1}{T}I(U;Z_0^T) = \frac{1}{T}I(U;\mathbf{Z}) \leq \frac{\epsilon}{2} + \frac{1}{T}H(\delta) + \delta(R_z - \frac{\epsilon}{2}). \tag{2.32}$$

Since δ can be made arbitrarily small we can enforce that $\frac{1}{T}H(\delta) + \delta(R_z - \frac{\epsilon}{2}) \leq \frac{\epsilon}{2}$ by choosing T large enough. The previous inequality shows therefore that $\frac{1}{T}I(U;Z_0^T) \leq \epsilon$ and the desired secrecy condition is satisfied.

This shows that any secrecy rate $R_s < C_s$ can be achieved and completes the achievability proof. □

2.4 The Converse for the secrecy capacity

Before delving into the details of the converse we need the following technical lemma due to Wyner [35].

Lemma 3 (Wyner [35]). *If $\Gamma : \Omega \to F$ is a random variable such that F is a finite set and $\Lambda_0^T = \{\Lambda_t, 0 \leq t \leq T\}$ is a given stochastic process, then we have*

$$I(\Gamma;\Lambda_0^T) = H(\Gamma) - H(\Gamma|\Lambda_0^T), \tag{2.33}$$

where $H(\Gamma)$ is the usual entropy for discrete random variables and

$$H(\Gamma|\Lambda_0^T) = -\mathbb{E}\left[\sum_{\gamma \in F} \Pr[\Gamma = \gamma|\mathcal{F}_\Lambda^T] \ln \Pr[\Gamma = \gamma|\mathcal{F}_\Lambda^T]\right]. \tag{2.34}$$

This lemma is standard when all the random variables have discrete alphabets however this extension is needed here since we are dealing with continuous time stochastic processes.

The converse theorem will be proved through a sequence of Lemmas. The first one gives an inequality that must satisfied by every encoder-decoder pair (E_T, D_T).

Lemma 4. *For every (M, T) code with rate $R = \frac{\ln M}{T}$ we have*

$$R \leq \frac{1}{T(1-P_e)} \left(I(X_0^T; Y_0^T | Z_0^T) + I(U; Z_0^T) + H(P_e) \right). \tag{2.35}$$

Proof. Let $\hat{U} = D_T(Y_0^T)$ denote the output of the decoder at the legitimate receiver, so that $P_e = \Pr(U \neq \hat{U})$. We then have the following sequence of identities

$$\begin{aligned} RT = \ln M = H(U) &\overset{(a)}{=} H(U|Y_0^T) + I(U; Y_0^T) \\ &\overset{(b)}{\leq} H(U|\hat{U}) + I(U; Y_0^T) \\ &\overset{(c)}{\leq} H(P_e) + P_e \ln M + I(U; Y_0^T), \end{aligned} \tag{2.36}$$

the equality (a) follows from Wyner's lemma and the inequality (c) is an application of Fano's inequality. For the inequality (b), since $U \to Y_0^T \to \hat{U}$ is a Markov chain we deduce that[9] $I(U, Y_0^T) \geq I(U, \hat{U})$. Now by invoking Wyner's lemma again it follows that $H(U|Y_0^T) \leq H(U|\hat{U})$.

From Kolmogorov's formula (see Lemma 3.2 in [35]) we have[10]

$$I(U; Y_0^T, Z_0^T) = I(U; Y_0^T) + I(U; Z_0^T | Y_0^T), \tag{2.37}$$

[9]The data processing inequality extends to arbitrary random variables, see for instance Theorem 3.4 in [35].

[10]The definition of the conditional mutual information for arbitrary random variables can be found in [35].

since $U \to Y_0^T \to Z_0^T$ is a Markov chain we deduce that[11] $I(U; Z_0^T | Y_0^T) = 0$. By applying Kolmogorov's formula again we obtain

$$\begin{aligned} I(U; Y_0^T) = I(U; Y_0^T, Z_0^T) &= I(U; Z_0^T) + I(U; Y_0^T | Z_0^T) \\ &\leq I(U; Z_0^T) + I(X_0^T; Y_0^T | Z_0^T), \end{aligned} \tag{2.38}$$

where the last inequality follows from the fact that $U \to X_0^T \to Y_0^T \to Z_0^T$ form a Markov chain. Combining this last inequality with (c) and rearranging the terms yields the desired inequality. □

Lemma 5. *If* $\mathbb{E} \int_0^T |X_t \ln X_t| < \infty$, *then*

$$I(X_0^T; Y_0^T | Z_0^T) = I(X_0^T; Y_0^T) - I(X_0^T; Z_0^T) \tag{2.39}$$

Proof. Applying Kolmogorov's formula twice gives

$$\begin{aligned} I(X_0^T; Y_0^T, Z_0^T) &= I(X_0^T; Z_0^T) + I(X_0^T; Y_0^T | Z_0^T) \\ &= I(X_0^T; Y_0^T) + I(X_0^T; Z_0^T | Y_0^T). \end{aligned} \tag{2.40}$$

Since $X_0^T \to Y_0^T \to Z_0^T$ form a Markov chain, we have $I(X_0^T; Z_0^T | Y_0^T) = 0$. Consequently we deduce that

$$I(X_0^T; Y_0^T) = I(X_0^T; Z_0^T) + I(X_0^T; Y_0^T | Z_0^T). \tag{2.41}$$

The condition $\mathbb{E} \int_0^T |X_t \ln X_t| < \infty$ implies that $I(X_0^T; Z_0^T) < \infty$ and it follows that $I(X_0^T; Y_0^T) - I(X_0^T; Z_0^T) = I(X_0^T; Y_0^T | Z_0^T)$. □

The goal of the upcoming lemmas is to prove that $I(X_0^T; Y_0^T | Z_0^T) \leq TC_s$,

[11] Refer to Lemma 3.1. in [35].

where C_s is given by (2.10). We first decompose $I(X_0^T; Y_0^T | Z_0^T)$ as follows

$$\begin{aligned} I(X_0^T; Y_0^T | Z_0^T) &= I(X_0^T; Y_0^T) - I(X_0^T; Z_0^T) \\ &= I(X_0^T; Y_0^T) - I(X_0^T; \widetilde{Y}_0^T) \\ &+ I(X_0^T; \widetilde{Y}_0^T) - I(X_0^T; Z_0^T), \end{aligned} \tag{2.42}$$

where $\widetilde{Y}_0^T$ has been defined in (2.7). The next two lemmas will provide upper bounds on $I(X_0^T; Y_0^T) - I(X_0^T; \widetilde{Y}_0^T)$ and $I(X_0^T; \widetilde{Y}_0^T) - I(X_0^T; Z_0^T)$.

Lemma 6. *If* $\mathbb{E}\int_0^T |X_t \ln X_t| < \infty$, *then*

$$\begin{aligned} &I(X_0^T; Y_0^T) - I(X_0^T; \widetilde{Y}_0^T) \leq \\ &\int_0^T \left(\frac{A_y}{A_z} (\phi_z(\mathbb{E}[X_t]) - \mathbb{E}[\phi_z(X_t)]) - (\phi_y(\mathbb{E}[X_t]) - \mathbb{E}[\phi_y(X_t)]) \right) dt, \end{aligned} \tag{2.43}$$

where $\phi_y(x) = (A_y x + \lambda_y)\ln(A_y x + \lambda_y)$ *and* $\phi_z(x)$ *has been defined above analogously.*

Proof. Note first that [16], [48]

$$I(X_0^T; Y_0^T) = \int_0^T \left(\mathbb{E}[\phi_y(X_t)] - \mathbb{E}[\phi_y(\mathbb{E}[X_t | \mathcal{F}_Y^t])] \right) dt, \tag{2.44}$$

and

$$I(X_0^T; \widetilde{Y}_0^T) = \int_0^T \left(\mathbb{E}[\chi(X_t)] - \mathbb{E}[\chi(\mathbb{E}[X_t | \mathcal{F}_{\widetilde{Y}}^t])] \right) dt, \tag{2.45}$$

where $\chi(x) = (A_y x + \frac{A_y}{A_z}\lambda_z)\ln(A_y x + \frac{A_y}{A_z}\lambda_z)$. Consequently, using the fact that $\chi(x) = \frac{A_y}{A_z}\phi_z(x) + \ln(\frac{A_y}{A_z})(A_y x + \frac{A_y}{A_z}\lambda_z)$ and after simplifications, we deduce the following

$$\begin{aligned} I(X_0^T; Y_0^T) - I(X_0^T; \widetilde{Y}_0^T) = &\int_0^T \left(\mathbb{E}[\phi_y(X_t)] - \mathbb{E}[\phi_y(\mathbb{E}[X_t | \mathcal{F}_Y^t])] \right) dt \\ &- \frac{A_y}{A_z} \int_0^T \left(\mathbb{E}[\phi_z(X_t)] - \mathbb{E}[\phi_z(\mathbb{E}[X_t | \mathcal{F}_{\widetilde{Y}}^t])] \right) dt. \end{aligned} \tag{2.46}$$

Recall that $\widetilde{Y}_0^T = Y_0^T + H_0^T$, where H_0^T is a homogeneous Poisson process independent of (X_0^T, Y_0^T). Clearly, $\mathcal{F}_{\widetilde{Y}}^t \subset \mathcal{F}_Y^t \vee \mathcal{F}_H^t$, with $\mathcal{F}_Y^t \vee \mathcal{F}_H^t = \sigma(\mathcal{F}_Y^t \cup \mathcal{F}_H^t)$

being the smallest sigma-field containing $\mathcal{F}_Y^t \cup \mathcal{F}_H^t$. From the independence of (X_0^T, Y_0^T) from H_0^T, using the law of redundant conditioning (see, e.g. [48, pp. 281-282]), we deduce that

$$\mathbb{E}[X_t|\mathcal{F}_Y^t \vee \mathcal{F}_H^t] = \mathbb{E}[X_t|\mathcal{F}_Y^t] \quad \text{a.s.} \tag{2.47}$$

We can now establish the following sequence of identities

$$\begin{aligned}
\mathbb{E}[\phi_z(\mathbb{E}[X_t|\mathcal{F}_Y^t])] &\overset{(a)}{=} \mathbb{E}[\phi_z(\mathbb{E}[X_t|\mathcal{F}_Y^t \vee \mathcal{F}_H^t])] && (2.48)\\
&\overset{(b)}{=} \mathbb{E}[\mathbb{E}[\phi_z(\mathbb{E}[X_t|\mathcal{F}_Y^t \vee \mathcal{F}_H^t])|\mathcal{F}_{\widetilde{Y}}^t]] && (2.49)\\
&\overset{(c)}{\geq} \mathbb{E}[\phi_z(\mathbb{E}[\mathbb{E}[X_t|\mathcal{F}_Y^t \vee \mathcal{F}_H^t]|\mathcal{F}_{\widetilde{Y}}^t])] && (2.50)\\
&\overset{(d)}{=} \mathbb{E}[\phi_z(\mathbb{E}[X_t|\mathcal{F}_{\widetilde{Y}}^t])], && (2.51)
\end{aligned}$$

where (a) follows from (2.47), (b) follows from the smoothing property of the conditional expectation, (c) from Jensen's inequality applied to the convex function $\phi_z(\cdot)$ and (d) from the fact that $\mathcal{F}_{\widetilde{Y}}^t \subset \mathcal{F}_Y^t \vee \mathcal{F}_H^t$ and the smoothing property.

We deduce therefore that

$$\begin{aligned}
I(X_0^T; Y_0^T) - I(X_0^T; \widetilde{Y}_0^T) \leq &\int_0^T \left(\mathbb{E}[\phi_y(X_t)] - \mathbb{E}[\phi_y(\mathbb{E}[X_t|\mathcal{F}_Y^t])]\right) dt \\
&- \frac{A_y}{A_z} \int_0^T \left(\mathbb{E}[\phi_z(X_t)] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|\mathcal{F}_Y^t])]\right) dt. \qquad (2.52)
\end{aligned}$$

A simple derivation shows that the function $\pi(x) = \phi_y(x) - \frac{A_y}{A_z}\phi_z(x)$ is convex as

$$\pi''(x) = \frac{A_y(\lambda_z A_y - \lambda_y A_z)}{(A_y x + \lambda_y)(A_z x + \lambda_z)} \geq 0. \tag{2.53}$$

Now invoking again Jensen's inequality we obtain that

$$\mathbb{E}[\pi(\mathbb{E}[X_t|\mathcal{F}_Y^t])] \geq \pi(\mathbb{E}[\mathbb{E}[X_t|\mathcal{F}_Y^t]]) = \pi(\mathbb{E}[X_t]). \tag{2.54}$$

Using this last inequality and after rearranging the terms we obtain the desired

result, i.e.,

$$I(X_0^T;Y_0^T) - I(X_0^T;\widetilde{Y}_0^T) \leq \int_0^T (\mathbb{E}[\phi_y(X_t)] - \phi_y(\mathbb{E}[X_t]))\,dt \\ - \frac{A_y}{A_z}\int_0^T (\mathbb{E}[\phi_z(X_t)] - \phi_z(\mathbb{E}[X_t]))\,dt. \tag{2.55}$$

□

An alternative proof of this lemma using the link provided in [27] between the MMSE and the mutual information in Poisson channels is given in Section 2.7.1.

Lemma 7. *If* $\mathbb{E}\int_0^T |X_t \ln X_t| < \infty$, *then*

$$I(X_0^T;\widetilde{Y}_0^T) - I(X_0^T;Z_0^T) \\ \leq (\frac{A_y}{A_z} - 1)\int_0^T (\mathbb{E}[\phi_z(X_t)] - \phi_z(\mathbb{E}[X_t]))dt \tag{2.56}$$

Proof. Recall that Z_0^T was obtained from $\widetilde{Y}_0^T$ by thinning with erasure probability $1 - \frac{A_z}{A_y}$. Let the process $\tilde{Z}_0^T$ denote those points that were erased, hence we have that $\tilde{Z}_0^T$ is a doubly stochastic Poisson process with instantaneous rate $(A_y - A_z)X_t + (\frac{A_y}{A_z} - 1)\lambda_z$. Moreover Z_0^T and $\tilde{Z}_0^T$ are independent given X_0^T. We proceed with the proof of the lemma by showing that the following inequality holds

$$I(X_0^T;\widetilde{Y}_0^T) - I(X_0^T;Z_0^T) \leq I(X_0^T;\tilde{Z}_0^T). \tag{2.57}$$

Indeed, notice first that $X_0^T \to (Z_0^T, \tilde{Z}_0^T) \to \widetilde{Y}_0^T$ is a Markov chain, hence from the data processing inequality we deduce that

$$I(X_0^T;\widetilde{Y}_0^T) = I(X_0^T;Z_0^T + \tilde{Z}_0^T) \leq I(X_0^T;Z_0^T,\tilde{Z}_0^T). \tag{2.58}$$

Consider now two partitions of Ω, $\mathcal{Q}_Z = \{A_i\}_{i=1}^{N_1} \subseteq \mathcal{F}_Z^T$ and $\mathcal{Q}_{\tilde{Z}} = \{B_j\}_{j=1}^{N_2} \subseteq \mathcal{F}_{\tilde{Z}}^T$. Define two discrete random variables D and $\tilde{D}$ on Ω as follows $D(\omega) = i$ if

$\omega \in A_i$ and $\tilde{D}(\omega) = j$ if $\omega \in B_j$. The mutual information $I(X_0^T; Z_0^T, \tilde{Z}_0^T)$ can be computed as [35]

$$I(X_0^T; Z_0^T, \tilde{Z}_0^T) = \sup_{\mathcal{Q}_Z, \mathcal{Q}_{\tilde{Z}}} I(X_0^T; D, \tilde{D}), \tag{2.59}$$

where the supremum is taken over all such partitions of Ω. We proceed to prove (2.57) as follows

$$\begin{aligned} I(X_0^T; D, \tilde{D}) &\overset{(a)}{=} H(D, \tilde{D}) - H(D, \tilde{D}|X_0^T) \\ &\overset{(b)}{\leq} H(D) + H(\tilde{D}) - H(D, \tilde{D}|X_0^T) \\ &\overset{(c)}{=} H(D) + H(\tilde{D}) - H(D|X_0^T) - H(\tilde{D}|X_0^T) \\ &\overset{(d)}{=} I(D; X_0^T) + I(\tilde{D}; X_0^T), \end{aligned} \tag{2.60}$$

where (a) follows from Lemma 3 (Wyner's lemma) applied to the random variable $(D, \tilde{D})$, (d) is a also a direct instance of this lemma. The inequality (b) is the independence bound on the entropy (which holds here since the random variables D and $\tilde{D}$ are discrete). The equality (c) results from the fact that D and $\tilde{D}$ are conditionally independent given X_0^T, indeed $D \in \mathcal{F}_Z^T$ whereas $\tilde{D} \in \mathcal{F}_{\tilde{Z}}^T$ and $\mathcal{F}_Z^T$ and $\mathcal{F}_{\tilde{Z}}^T$ are conditionally independent given $\mathcal{F}_X^T$. Consequently we have

$$\begin{aligned} I(X_0^T; Z_0^T, \tilde{Z}_0^T) &= \sup_{\mathcal{Q}_Z, \mathcal{Q}_{\tilde{Z}}} I(X_0^T; D, \tilde{D}) \\ &\leq \sup_{\mathcal{Q}_Z, \mathcal{Q}_{\tilde{Z}}} (I(X_0^T; D) + I(X_0^T; \tilde{D})) \\ &= I(X_0^T; Z_0^T) + I(X_0^T; \tilde{Z}_0^T). \end{aligned} \tag{2.61}$$

Combining the last inequality with (2.58) we deduce that[12]

$$I(X_0^T; \widetilde{Y}_0^T) - I(X_0^T; Z_0^T) \leq I(X_0^T; \tilde{Z}_0^T). \tag{2.62}$$

Now using Lemma 2 we have

$$I(X_0^T; \tilde{Z}_0^T) \leq \int_0^T (\mathbb{E}[\varphi(X_t)] - \varphi(\mathbb{E}[X_t]))dt, \tag{2.63}$$

[12]Note that since $\mathbb{E}\int_0^T |X_t \ln X_t| < \infty$, then $I(X_0^T; Z_0^T) < \infty$ and hence the inequality is well defined.

where

$$\varphi(x) = ((A_y - A_z)x + (\frac{A_y}{A_z} - 1)\lambda_z)\ln((A_y - A_z)x + (\frac{A_y}{A_z} - 1)\lambda_z). \tag{2.64}$$

Notice now that

$$\varphi(x) = (\frac{A_y}{A_z} - 1)\phi_z(x) + (\frac{A_y}{A_z} - 1)\ln(\frac{A_y}{A_z} - 1)(A_z x + \lambda_z). \tag{2.65}$$

Plugging this identity in the inequality above, the linear term in x disappears and we are left with the inequality presented in the lemma. □

An alternative proof of this lemma using the link provided in [27] between the MMSE and the mutual information in Poisson channels is given in Section 2.7.2.

Theorem 3. *If* $\mathbb{E}\int_0^T |X_t \ln X_t| < \infty$, *then*

$$\frac{1}{T} I(X_0^T; Y_0^T | Z_0^T) \leq C_s \tag{2.66}$$

Proof. Combining (2.42) and the result of the two previous lemmas yields

$$I(X_0^T; Y_0^T) - I(X_0^T; Z_0^T) \leq \int_0^T (\mathbb{E}[K(X_t)] - K(\mathbb{E}[X_t]))dt, \tag{2.67}$$

where $K(x) = \phi_y(x) - \phi_z(x)$. A straightforward computation shows that

$$K''(x) = \frac{A_z A_y (A_y - A_z)x + \lambda_z A_y^2 - \lambda_y A_z^2}{(A_y x + \lambda_y)(A_z x + \lambda_z)}, \tag{2.68}$$

since $A_y \geq A_z$ and $\lambda_z A_y^2 \geq \lambda_y A_y A_z \geq \lambda_y A_z^2$ we deduce that $K''(x) \geq 0$. Moreover due to the assumption that at least one of the inequalities (2.5) or (2.6) is strict, we conclude that $K''(x) > 0$ (for $x > 0$) and $K(\cdot)$ is strictly convex.

Notice now that we have

$$\begin{aligned}
&\frac{1}{T}\int_0^T (\mathbb{E}[K(X_t)] - K(\mathbb{E}[X_t]))dt \\
&\overset{(a)}{\leq} \max_{0\leq\alpha\leq 1}\left(\max_{\rho:\int_0^1 x\rho(dx)=\alpha}\int_0^1 K(x)\rho(dx) - K(\alpha)\right) \\
&\overset{(b)}{=} \max_{0\leq\alpha\leq 1}\left(\alpha K(1) + (1-\alpha)K(0) - K(\alpha)\right),
\end{aligned} \tag{2.69}$$

where (a) follows from fixing $\mathbb{E}[X_t] = \alpha$ and maximizing over all distributions $\rho(x)$ on $[0,1]$ with mean α. Equality (b) follows from the convexity of $K(\cdot)$ (refer to [16] and [37]), i.e., the maximizing distribution ρ puts all the mass on the extremes $\{0,1\}$ and since the mean is α, the maximizing ρ assigns the mass α to 1 and $1-\alpha$ to 0.

The maximization of the last term shows that the optimal α^* is the unique solution to the equation

$$K'(\alpha^*) = K(1) - K(0),$$

which, after some algebraic manipulations, gives that α^* is the solution to (2.11). The existence of α^* follows simply from the mean value theorem, whereas the uniqueness is a consequence of the strict monotonicity of $K'(x)$.

Consequently, the following is true

$$\begin{aligned}
&\frac{1}{T}\int_0^T (\mathbb{E}[K(X_t)] - K(\mathbb{E}[X_t]))dt \\
&\leq \alpha^* K(1) + (1-\alpha^*)K(0) - K(\alpha^*) \\
&= \alpha^*(A_y - A_z) + \ln\left(\frac{\lambda_y^{\lambda_y}}{\lambda_z^{\lambda_z}}\right) + \ln\left(\frac{(A_z\alpha^* + \lambda_z)^{\lambda_z}}{(A_y\alpha^* + \lambda_y)^{\lambda_y}}\right).
\end{aligned} \tag{2.70}$$

This fact when combined with (2.67) gives the result announced in the theorem.

□

We are now in a position to prove the converse theorem.

Theorem 4 (Converse). *If R_s is an achievable secrecy rate then $R_s \leq C_s$.*

Proof. Since the secrecy rate R_s is achievable then for all $0 < \epsilon < \frac{1}{2}$ and sufficiently large T, there exists an (M, T) code such that $\frac{\ln M}{T} \geq R_s - \epsilon$, $P_e \leq \epsilon$ and $\frac{1}{T} I(U; Z_0^T) \leq \epsilon$. Hence we have

$$\begin{aligned} R_s &\leq \frac{\ln M}{T} + \epsilon \\ &\overset{(a)}{\leq} \frac{1}{T(1-P_e)} \left(I(X_0^T; Y_0^T | Z_0^T) + I(U; Z_0^T) + H(P_e) \right) + \epsilon \\ &\overset{(b)}{\leq} \frac{1}{1-P_e} \left(C_s + \frac{I(U; Z_0^T)}{T} + \frac{H(P_e)}{T} \right) + \epsilon \\ &\overset{(c)}{\leq} \frac{1}{1-\epsilon} \left(C_s + \epsilon + \frac{H(\epsilon)}{T} \right) + \epsilon, \end{aligned} \tag{2.71}$$

where inequality (a) follows from Lemma 4, inequality (b) from Theorem 3 and inequality (c) from the properties of the code. Now since ϵ is arbitrary, letting $\epsilon \to 0$ yields $R_s \leq C_s$. □

2.5 Rate-Equivocation region

In this section we turn our attention to the rate equivocation region of the degraded Poisson wiretap channel. The level of ignorance of the eavesdropper about the transmitted message U will be measured here by the normalized equivocation given by

$$\Delta_T = \frac{H(U|Z_0^T)}{H(U)}. \tag{2.72}$$

Definition A rate-equivocation pair (R, d) is said to be *achievable* for the Poisson wiretap channel if for all $\epsilon > 0$ and all sufficiently large T, there exists an (M, T)

code such that

$$\begin{aligned} \frac{\ln M}{T} &\geq R - \epsilon \\ P_e &\leq \epsilon \\ \frac{H(U|Z_0^T)}{H(U)} &\geq d - \epsilon \end{aligned} \tag{2.73}$$

The following theorem gives the rate equivocation region (that is the set of all achievable rate-equivocation pairs (R, d)) for the degraded Poisson Wiretap channel.

Theorem 5. *The rate-equivocation region is the set of all rate-equivocation pairs* (R, d) *for which there exits some* $\alpha \in [0, 1]$ *such that*

$$Rd \leq \alpha \ln\left(\frac{(A_y + \lambda_y)^{A_y+\lambda_y}}{(A_z + \lambda_z)^{A_z+\lambda_z}}\right) + (1-\alpha)\ln\left(\frac{\lambda_y^{\lambda_y}}{\lambda_z^{\lambda_z}}\right) - \ln\left(\frac{(A_y\alpha + \lambda_y)^{A_y\alpha+\lambda_y}}{(A_z\alpha + \lambda_z)^{A_z\alpha+\lambda_z}}\right) \tag{2.74}$$

$$R \leq \alpha \ln\left((A_y + \lambda_y)^{A_y+\lambda_y}\right) + (1-\alpha)\ln\left(\lambda_y^{\lambda_y}\right) - \ln\left((A_y\alpha + \lambda_y)^{A_y\alpha+\lambda_y}\right) \tag{2.75}$$

$$d \leq 1 \tag{2.76}$$

To ease the notations, using the functions $K(\cdot)$ and $\phi_y(\cdot)$, we can rewrite the two first inequalities as $Rd \leq \alpha K(1) + (1-\alpha)K(0) - K(\alpha)$ and $R \leq \alpha\phi_y(1) + (1-\alpha)\phi_y(0) - \phi_y(\alpha)$.

Proof. The main ingredients needed to prove this theorem has been already used to obtain the secrecy capacity. More specifically, for the achievability proof we will use stochastic encoding combined with Wyner codes for the Poisson channel, and for the converse we will use the key inequality (2.67) established by Lemma 6 and 7.

2.5.1 Direct result

Note first that for a fixed rate R, if the rate equivocation pair (R, d) is achievable then the pair $(R, \tilde{d})$ is achievable for all $0 \leq \tilde{d} \leq d$. Hence, in order to establish the direct result, it is enough to prove that any rate-equivocation pair (R, d) satisfying $Rd = \alpha K(1) + (1-\alpha)K(0) - K(\alpha)$, $R \leq \alpha\phi_y(1) + (1-\alpha)\phi_y(0) - \phi_y(\alpha)$ and $d \leq 1$ for some $\alpha \in [0, 1]$ is achievable.

Define

$$R_u = \alpha\phi_u(1) + (1-\alpha)\phi_u(0) - \phi_u(\alpha),\ u \in \{y, z\}. \tag{2.77}$$

Let $\epsilon > 0$ be arbitrary (small enough) and let $R = \frac{R_y - R_z - \epsilon R}{d}$ with $d \leq 1$ and $R \leq \alpha\phi_y(1) + (1-\alpha)\phi_y(0) - \phi_y(\alpha)$. The message U to be transmitted is selected uniformly randomly from $\mathcal{U} = \{1, ..., M\}$ with $M = e^{RT}$. Define $M_y = e^{(R_y - 3\epsilon\frac{R}{2})T}$ and, following the steps described for the achievability of the secrecy capacity, construct the Wyner code $\mathcal{C} = \mathcal{W}(T, M_y, \alpha M_y)$. Partition this code arbitrarily into M smaller subcodes, i.e., $\mathcal{C} = \cup_{i=1}^{M} \mathcal{C}_i$. The cardinality of each subcode $\mathcal{C}_i$ will be equal to $M_z = \frac{M_y}{M} = e^{(R_y - R - 3\epsilon\frac{R}{2})T}$. Notice that with this choice of parameters we have

$$\frac{1}{T}\ln M_z = R_y - R - 3\epsilon\frac{R}{2} \leq R_y - Rd - 3\epsilon\frac{R}{2} = R_z - \epsilon\frac{R}{2}. \tag{2.78}$$

The probability of error P_e of the legitimate receiver can be made less than ϵ because the Wyner code $\mathcal{C}$ can achieve the rate R_y.

The equivocation of the code $\mathcal{C}$ can be lower bounded using the same steps used to established the upper bound on $I(U; Z_0^T)$ for the secrecy capacity, as

follows

$$\Delta_T = \frac{H(U|Z_0^T)}{H(U)} = 1 - \frac{I(U;Z_0^T)}{RT} \tag{2.79}$$

$$\overset{(a)}{\geq} 1 - \frac{R_z}{R} + \frac{1}{RT}\ln M_z - \frac{1}{RT}\left(H(\delta) + \delta \ln M_z\right) \tag{2.80}$$

$$= 1 - \frac{R_z}{R} + \frac{R_y - R - 3\epsilon\frac{R}{2}}{R} - \frac{1}{RT}\left(H(\delta) + \delta \ln M_z\right) \tag{2.81}$$

$$\geq d - \frac{\epsilon}{2} - \frac{1}{RT}H(\delta) - \delta(\frac{R_z}{R} - \frac{\epsilon}{2}). \tag{2.82}$$

In the above, inequality (a) follows from (2.31) and $\delta = \frac{1}{M}\sum_{m=1}^{M}\delta_m$ where δ_m is the probability of error for the code $\mathcal{C}_m$ ($1 \leq m \leq M$) with the (optimal) decoder described previously.

As was discussed before, the term $\frac{1}{RT}H(\delta) + \delta(\frac{R_z}{R} - \frac{\epsilon}{2})$ can be made less than $\frac{\epsilon}{2}$ for T large enough, which means that

$$\Delta_T = \frac{H(U|Z_0^T)}{H(U)} \geq d - \epsilon. \tag{2.83}$$

This establishes that the rate-equivocation pair (R, d) is achievable.

2.5.2 Converse

For every (M, T) code with rate $R_T = \frac{\ln M}{T}$ and equivocation $\Delta_T = \frac{H(U|Z_0^T)}{H(U)}$ we have

$$\begin{aligned} TR_T\Delta_T = H(U|Z_0^T) &= H(U) - I(U;Z_0^T) \\ &= H(U|Y_0^T) + I(U;Y_0^T) - I(U;Z_0^T) \\ &\leq H(U|\hat{U}) + I(U;Y_0^T|Z_0^T) \\ &\leq H(P_e) + P_e \ln M + I(X_0^T;Y_0^T|Z_0^T). \end{aligned} \tag{2.84}$$

From Lemma 6 and 7 (cf. (2.67)) we have that

$$I(X_0^T; Y_0^T | Z_0^T) \leq \int_0^T (\mathbb{E}[K(X_t)] - K(\mathbb{E}[X_t]))dt. \tag{2.85}$$

Consequently, we deduce that

$$R_T \Delta_T \leq \frac{H(P_e) + P_e \ln M}{T} + \frac{1}{T}\int_0^T (\mathbb{E}[K(X_t)] - K(\mathbb{E}[X_t]))dt \tag{2.86}$$

$$\leq \frac{H(P_e) + P_e \ln M}{T} + \alpha K(1) + (1-\alpha)K(0) - K(\alpha), \tag{2.87}$$

with $\alpha = \frac{1}{T}\int_0^T \mathbb{E}[X_t]dt$ and the last inequality follows from the convexity of the function $K(\cdot)$. Note that since $0 \leq X_t \leq 1$ it follows that $0 \leq \alpha \leq 1$.

Similarly, we have that

$$\begin{aligned} R_T = \frac{H(U)}{T} &= \frac{1}{T}H(U|Y_0^T) + \frac{1}{T}I(U; Y_0^T) \\ &\leq \frac{1}{T}H(U|\hat{U}) + \frac{1}{T}I(X_0^T; Y_0^T) \\ &\overset{(a)}{\leq} \frac{1}{T}(H(P_e) + P_e \ln M) + \frac{1}{T}\int_0^T (\mathbb{E}[\phi_y(X_t)] - \phi_y(\mathbb{E}[X_t]))dt \\ &\overset{(b)}{\leq} \frac{H(P_e) + P_e \ln M}{T} + \alpha\phi_y(1) + (1-\alpha)\phi_y(0) - \phi_y(\alpha), \end{aligned} \tag{2.88}$$

where (a) follows from Fano's inequality and Lemma 2 and (b) follows from the convexity of the function $\phi_y(\cdot)$.

Assume now that (R, d) is achievable, then for all $0 < \epsilon < \frac{1}{2}$ and all sufficiently large T, there exists an (M, T) code such that $R_T \geq R - \epsilon$, $P_e \leq \epsilon$ and $\Delta_T \geq d - \epsilon$. By definition $\Delta_T \leq 1$, and hence $d \leq 1 + \epsilon$ and in light of the previous inequalities we have

$$(R-\epsilon)(d-\epsilon) \leq \frac{H(\epsilon) + \epsilon \ln M}{T} + \alpha K(1) + (1-\alpha)K(0) - K(\alpha) \tag{2.89}$$

$$(R-\epsilon) \leq \frac{H(\epsilon) + \epsilon \ln M}{T} + \alpha\phi_y(1) + (1-\alpha)\phi_y(0) - \phi_y(\alpha). \tag{2.90}$$

Now since ϵ is arbitrary, letting $\epsilon \to 0$ yields the desired result. □

2.6 Conclusion and discussion

Motivated by the practical advantages of optical communication over RF for secure communication, we have derived the secrecy capacity and characterized the rate-equivocation region of the degraded Poisson wiretap channel.

Several interesting problems remain open and deserve further investigation. One is the non-degraded Poisson Wiretap channel. One can imagine a situation in which the eavesdropper is equipped with a powerful detector characterized by a negligible dark current (i.e., $\lambda_z = 0$). If the detector of the legitimate receiver has a higher received power from the transmitter but is more prone to dark current, then the channel will not be degraded. This is a practically-important situation but is not covered by our results. While the code construction used here would certainly give an achievable secrecy rate, determining the secrecy capacity is much more challenging in the non-degraded case. The difficulty is that, unless the legitimate user's channel is less noisy than the eavesdropper, finding the secrecy capacity requires optimizing over an auxiliary random variable [49], which complicates the analysis. One possible approach is to approximate the Poisson channel by a binary DMC [18]. In principle, one could then apply the classical result for the secrecy capacity of non-degraded DMCs [49] and take limits. This procedure seems to be quite onerous, however.

Another issue that we have not addressed is fading. As mentioned in the introduction, for wireless optical communications, atmospheric turbulence can induce random fluctuations of the intensity of the transmitted light beam [24]. This impairment has received considerable attention in the context of reliable communications, and it would be useful to determine its effect on secrecy.

MIMO Poisson channels have received some interest lately (see [38] and the references therein), and as has been done in the Gaussian setting, it would be interesting to see the impact of having multiple antennas on the secrecy capacity in the Poisson regime.

We believe that the results derived in this chapter and the tools used to derive them could be used to address these problems.

2.7 Alternative Proofs

2.7.1 An MMSE Proof for Lemma 6

We provide here an alternative proof for Lemma 6. This proof uses the link established in [27] between the MMSE and the mutual information in Poisson channels. Note first that since $\mathbb{E}\int_0^T |X_t \ln X_t| < \infty$, we have that $I(X_0^T; \mathcal{P}_0^T(A_y X_0^T + \lambda))$ is differentiable and Theorem 3 in [27] states that

$$\frac{d}{d\lambda} I(X_0^T; \mathcal{P}_0^T(A_y X_0^T + \lambda)) = \int_0^T \mathbb{E}\{\ln(A_y X_t + \lambda) - \ln\langle A_y X_t + \lambda\rangle_T\} dt. \tag{2.91}$$

Notice now that

$$I(X_0^T; \widetilde{Y}_0^T) - I(X_0^T; Y_0^T) = \int_{\lambda_y}^{\lambda_y + \tilde{\lambda}} \frac{d}{d\lambda} I(X_0^T; \mathcal{P}_0^T(A_y X_0^T + \lambda)) d\lambda. \tag{2.92}$$

Therefore

$$I(X_0^T;Y_0^T) - I(X_0^T;\widetilde{Y}_0^T) = \int_{\lambda_y}^{\lambda_y+\tilde{\lambda}} \left(\int_0^T (\mathbb{E}\{\ln\langle A_y X_t + \lambda\rangle_T\} - \mathbb{E}\{\ln(A_y X_t + \lambda)\}) dt \right) d\lambda. \tag{2.93}$$

Since the function $\ln(\cdot)$ is concave, using Jensen's inequality and the iterative conditioning property we have

$$\mathbb{E}\{\ln\langle A_y X_t + \lambda\rangle_T\} \leq \ln \mathbb{E}[\langle A_y X_t + \lambda\rangle_T] = \ln(A_y \mathbb{E}[X_t] + \lambda).$$

Making use of this inequality and the fact that $\lambda_y + \tilde{\lambda} = \frac{A_y}{A_z}\lambda_z$ we deduce that

$$I(X_0^T;Y_0^T) - I(X_0^T;\widetilde{Y}_0^T) \leq \int_0^T \left(\int_{\lambda_y}^{\frac{A_y}{A_z}\lambda_z} \ln(A_y \mathbb{E}[X_t] + \lambda) d\lambda - \mathbb{E}\{ \int_{\lambda_y}^{\frac{A_y}{A_z}\lambda_z} \ln(A_y X_t + \lambda) d\lambda \} \right) dt, \tag{2.94}$$

where we have also invoked Fubini's theorem to make the necessary exchanges between the integrals and the expectation operator. The desired inequality is then obtained after some algebraic manipulations using the elementary identity

$$\int \ln(A_y x + \lambda) d\lambda = (A_y x + \lambda)\ln(A_y x + \lambda) - \lambda. \tag{2.95}$$

2.7.2 An MMSE Proof for Lemma 7

Here we provide an alternative proof for Lemma 7. For ease of notations define $W_t = A_y X_t + \frac{A_y}{A_z}\lambda_z$. Using Theorem 4 in [27] we obtain that

$$\begin{aligned}
\frac{d}{d\alpha} I(W_0^T; \mathcal{P}_0^T(\alpha W_0^T)) &= \int_0^T \mathbb{E}[W_t \ln(\alpha W_t)] dt \\
&\quad - \int_0^T \mathbb{E}[\mathbb{E}[W_t|\mathcal{P}_0^T(\alpha W_0^T)] \ln(\mathbb{E}[\alpha W_t|\mathcal{P}_0^T(\alpha W_0^T)])] dt \\
&= \int_0^T \mathbb{E}[W_t \ln W_t] dt \\
&\quad - \int_0^T \mathbb{E}[\mathbb{E}[W_t|\mathcal{P}_0^T(\alpha W_0^T)] \ln(\mathbb{E}[W_t|\mathcal{P}_0^T(\alpha W_0^T)])] dt,
\end{aligned} \tag{2.96}$$

where the second equality is obtained after some simplifications using the identity $\mathbb{E}[\mathbb{E}[W_t|\mathcal{P}_0^T(\alpha W_0^T)]] = \mathbb{E}[W_t]$. Now by the convexity of the function $C(x) = x\ln(x)$, Jensen's inequality gives

$$\begin{aligned}\mathbb{E}[C(\mathbb{E}[W_t|\mathcal{P}_0^T(\alpha W_0^T)])] &\geq C(\mathbb{E}[\mathbb{E}[W_t|\mathcal{P}_0^T(\alpha W_0^T)]]) \\ &= C(\mathbb{E}[W_t]). \end{aligned} \tag{2.97}$$

It follows therefore that

$$\frac{d}{d\alpha} I(W_0^T; \mathcal{P}_0^T(\alpha W_0^T)) \leq \int_0^T \mathbb{E}[W_t \ln W_t] dt - \int_0^T \mathbb{E}[W_t] \ln \mathbb{E}[W_t] dt. \tag{2.98}$$

Clearly we have that $I(W_0^T; \mathcal{P}_0^T(\alpha W_0^T)) = I(X_0^T; \mathcal{P}_0^T(\alpha W_0^T))$. Also we have that $\widetilde{Y}_0^T = \mathcal{P}_0^T(W_0^T)$ and $Z_0^T = \mathcal{P}_0^T(\frac{A_z}{A_y} W_0^T)$. Consequently

$$I(X_0^T; \widetilde{Y}_0^T) - I(X_0^T; Z_0^T) = \int_{\frac{A_z}{A_y}}^1 \frac{d}{d\alpha} I(W_0^T; \mathcal{P}_0^T(\alpha W_0^T)) d\alpha. \tag{2.99}$$

Using the previous inequality, we conclude that

$$\begin{aligned} I(X_0^T; \widetilde{Y}_0^T) - I(X_0^T; Z_0^T) &\leq \int_{\frac{A_z}{A_y}}^1 \left(\int_0^T \mathbb{E}[W_t \ln W_t] dt - \int_0^T \mathbb{E}[W_t] \ln \mathbb{E}[W_t] dt \right) d\alpha \\ &= (1 - \frac{A_z}{A_y}) \left(\int_0^T \mathbb{E}[(A_y X_t + \frac{A_y}{A_z}\lambda_z) \ln(A_y X_t + \frac{A_y}{A_z}\lambda_z)] dt \right. \\ &\quad \left. - \int_0^T (A_y \mathbb{E}[X_t] + \frac{A_y}{A_z}\lambda_z) \ln(A_y \mathbb{E}[X_t] + \frac{A_y}{A_z}\lambda_z) dt \right). \end{aligned} \tag{2.100}$$

After some simplifications, the last inequality gives the desired result, i.e.,

$$I(X_0^T; \widetilde{Y}_0^T) - I(X_0^T; Z_0^T) \leq (\frac{A_y}{A_z} - 1) \int_0^T (\mathbb{E}[\phi_z(X_t)] - \phi_z(E[X_t])) \, dt. \tag{2.101}$$

CHAPTER 3

SECRECY CAPACITY OF THE EXPONENTIAL SERVER QUEUE

3.1 Introduction

The problem that we consider is depicted in Figure 3.1. The encoder (Bob), encodes a message U by using a given sequence of packet inter-arrival times represented here by the arrival process $X_0^T = (X_t, 0 \leq t \leq T)$. These packets are fed into two different $\cdot/M/1$ queues. The legitimate decoder (Alice) observes the departure process Y_0^T of the packets from the main queue which has a service rate μ_1. Similarly, the eavesdropper (Eve) observes the departure process Z_0^T from the second queue with a service rate μ_2. Our goal is to find the secrecy capacity for this model, that is the maximum rate at which Bob can send information to Alice while still keeping the information leakage to Eve arbitrarily close to zero. When $\mu_1 \leq \mu_2$, clearly the secrecy capacity is zero as the distortion affecting the packet arrival times is less severe for the eavesdropper. The interesting scenario is then when $\mu_2 < \mu_1$. The exact same problem represented in Figure 3.1 was studied in [42] and partial results are available. When $\mu_2 < \frac{\mu_1}{e}$, Dunn et. al. showed that the secrecy capacity is maximum and is equal to $\frac{\mu_1}{e}$, which is the capacity of the main queue. This can be achieved by choosing for input X_0^T a Poisson process with rate $\frac{\mu_1}{e}$. Doing so simultaneously achieves the capacity over the main queue and overloads the eavesdropper queue thereby making the output process Z_0^T behave asymptotically like a Poisson process with rate μ_2 that is independent of the input X_0^T. Dunn et. al. demonstrated also that when $\frac{\mu_1}{e} \leq \mu_2 < \mu_1$, the rate $\mu_2 \log\left(\frac{\mu_1}{\mu_2}\right)$ can be achieved in secrecy. The main result of this chapter is to show that this quantity is indeed the secrecy capacity for

$\frac{\mu_1}{e} \leq \mu_2 < \mu_1$.

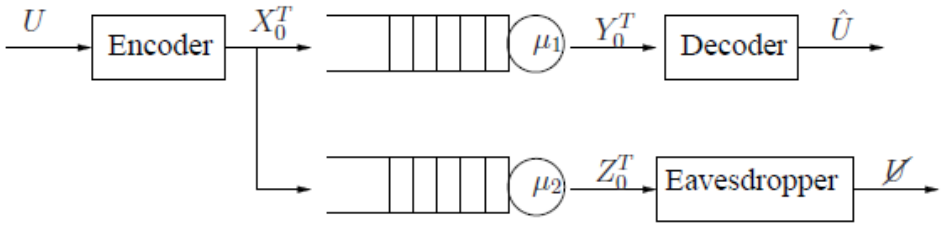

Figure 3.1: The wiretap exponential server timing channel.

To show the converse, we first prove that when $\mu_2 < \mu_1$, the eavesdropper's channel is stochastically degraded with respect to that of the legitimate receiver. That is we can assume that the Markov chain $X_0^T \to Y_0^T \to Z_0^T$ holds. We establish this by constructing an additional queue which takes as input the process Y_0^T and degrades it into the process Z_0^T. We show moreover that this result holds regardless of whether the queues start initially empty or in equilibrium. Using this and invoking the data processing inequality for relative entropy we obtain $D(\pi_1^Y||\pi_2^Y) \geq D(\pi_1^Z||\pi_2^Z)$ where $\pi_1^Y(dy)$ and $\pi_2^Y(dy)$ (respectively $\pi_1^Z(dz)$ and $\pi_2^Z(dz)$) are probability measures on the space of departure processes from the main queue (respectively the eavesdropper queue) induced respectively by $\nu_1(dx)$ and $\nu_2(dx)$ which are in turn probability measure on the space of arrival processes. If the queues are initially in equilibrium and we choose for $\nu_2(dx)$ the probability measure $P_\lambda(dx)$ corresponding to a Poisson process with rate $\lambda < \mu_2$, then Burke's theorem [45] implies that $\pi_2^Y = \pi_2^Z = P_\lambda$ which gives $D(\pi_1^Y||P_\lambda) \geq D(\pi_1^Z||P_\lambda)$. We show that this inequality combined with other intermediate results is enough to establish the converse.

We believe that the inequality

$$D(\pi_1^Y||P_\lambda) \geq D(\pi_1^Z||P_\lambda) \tag{3.1}$$

is in itself an interesting result. Prabhakar and Gallager [43, Theorem 1] showed that if $\mathbf{A} = \{A_n, n \in \mathbb{Z}\}$ is a sequence of i.i.d. inter-arrival times of packets with $E[A_1] = \frac{1}{\lambda}$ that are fed into a $\cdot/\mathrm{Geom}/1$ queue with mean service time $\frac{1}{\mu}$ (with $\lambda < \mu$), then the corresponding process of inter-departure times $\mathbf{D} = \{D_n, n \in \mathbb{Z}\}$ has an entropy rate that is no less than that of the process $\mathbf{A}$. Moreover they demonstrated [43, Corollary 2] that if $\mathbf{G} = \{G_n, n \in \mathbb{Z}\}$ is an i.i.d. sequence of geometric random variables with mean $\frac{1}{\lambda}$, then by letting $D_{ER}(\mathbf{A}||\mathbf{G})$ be the relative entropy rate between the processes $\mathbf{A}$ and $\mathbf{G}$, we have

$$D_{ER}(\mathbf{A}||\mathbf{G}) \geq D_{ER}(\mathbf{D}||\mathbf{G}). \tag{3.2}$$

They showed that Burke's theorem for the $\cdot/\mathrm{Geom}/1$ queue (which is the discrete time analog of the Poisson-in-Poisson-out property of the $\cdot/M/1$ queue) follows from this inequality. As mentioned by Anantharam and Verdú in [44], (3.2) can itself be established by using the discrete-time analog of Burke's theorem combined with the data processing inequality for relative entropy. The inequality that we establish here (3.1) can be seen as some sort of generalization of (3.2) for continuous time queues. However the novelty of (3.1) reside in the fact that it links relative entropies of the outputs of two different queues, whereas (3.2) is an inequality about the relative entropy between the input and output of one single queue.

The rest of this chapter is organized as follows. In Section 3.2, we go over some mathematical facts that will be used throughout the chapter. In Section 3.3, we present the main result of the chapter which is the secrecy capacity of the exponential server queue. In Section 3.4, we digress a little to talk about the achievability part of the main result. The converse part, which is our main contribution, is provided in Section 3.5.

3.2 Mathematical Prelude

In this section we will introduce some of the notation that will be used in this chapter. We will also recall some important facts and technical details that will serve as the main tools to prove the upcoming results. Most of the facts contained in this section can be found in [46]-[54]. They are reproduced here for the convenience of the reader.

3.2.1 Notation

For a given $T \in (0, \infty)$, we will use X_0^T to denote $(X_t : t \in [0, T])$. For $t \in [0, T]$, $\mathcal{F}_t^X = \sigma(X_s : s \in [0, t])$ is the internal history of the stochastic process X up to and including time t. For $t \in (0, T]$, we use $\mathcal{F}_{t-}^X = \sigma(X_s : s \in [0, t))$ to denote the strict past at time t of the stochastic process X. We will also use X_{t-} to denote the left limit of X at the time t, i.e., $X_{t-} = \lim_{s \uparrow t} X_s$. A stochastic process X will be left-continuous if $X_{t-} = X_t$ almost surely.

A process X is said to be adapted to the filtration $(\mathcal{F}_t : t \in [0, T])$ if X_t is $\mathcal{F}_t$-measurable for all $t \in [0, T]$. A process X is said to be predictable with respect to the filtration $(\mathcal{F}_t : t \in [0, T])$ if X can be written as the limit of left-continuous adapted processes. A more general definition for predictable processes can be found in [48, p. 8]. But for all practical purposes the definition given here is sufficient.

The input space considered here is the set of nondecreasing and right-continuous integer valued functions and will be denoted by $\mathcal{X}_T$. That is, $x \in \mathcal{X}_T$ if: (1) $x_t \in \mathbb{N}$ for all t, (2) $x_t = \lim_{s \downarrow t} x_s$ and (3) $x_{t_1} \leq x_{t_2}$ for $t_1 \leq t_2$. The space

$\mathcal{X}_T$ represents the set of arrival processes and x_t will track the number of packets arrived to the queue in the interval $[0, t]$. Note that this definition does not exclude having multiple arrivals at a given moment.

The output space is denoted by $\mathcal{D}_T$. This space is the set of all functions w_t on $[0, T]$ that are nondecreasing, right-continuous, have unit jumps and satisfy $w_0 = 0$. By unit jumps we mean that $\Delta w_t = w_t - w_{t-} \in \{0, 1\}$. The set $\mathcal{D}_T$ will represent the set of departure processes from the queue that can be observed by the decoder. Therefore w_t is the number of packets that have departed from the queue in the interval $[0, t]$.

3.2.2 Stochastic Intensity of Point Processes

The notion of stochastic intensity of point processes is crucial to the rest of this chapter. To define it, we start with a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $(\mathcal{F}_t : t \geq 0)$ on this space. The definition of stochastic intensity is the following.

Definition Let N be a point process (i.e., N is nondecreasing, right-continuous, have unit jumps and satisfy $N_0 = 0$) that is adapted to $\mathcal{F}_t$ and let α_t be a nonnegative process that is $\mathcal{F}_t$-predictable. The process N_t is said to have the $(P, \mathcal{F}_t)$-intensity α_t if the following is true[1]

- For all $t \geq 0$

$$\int_0^t \alpha_s ds < \infty \quad P - a.s. \tag{3.3}$$

[1]The general definition of a stochastic intensity of a point process requires only for α_t to be nonnegative and $\mathcal{F}_t$-progressive [48, p. 27]. However, this generality is not needed here as one can always find a predictable version of the intensity [48, p. 31].

- For all nonnegative $\mathcal{F}_t$-predictable processes C_t, we have

$$E\left[\int_0^\infty C_t dN_t\right] = E\left[\int_0^\infty C_t \alpha_t dt\right]. \tag{3.4}$$

If the last jump of N_t occurs before time T, then (3.4) is equivalent to $E\left[\int_0^T C_t dN_t\right] = E\left[\int_0^T C_t \alpha_t dt\right]$. Indeed in this case $dN_t = 0$ for $t > T$, which means that $E\left[\int_0^\infty C_t dN_t\right] = E\left[\int_0^T C_t dN_t\right]$. Moreover by choosing $C_t = 1_{\{t>T\}}$ in (3.4), we will have $0 = E\left[\int_T^\infty \alpha_t dt\right]$, which by the nonnegativity of the intensity implies that $\alpha_t = 0$ for $t > T$ almost surely. This in turn gives that $E\left[\int_0^\infty C_t \alpha_t dt\right] = E\left[\int_0^T C_t \alpha_t dt\right]$.

The following are two results pertaining to stochastic intensities that will be used later in the chapter. The first fact can be found in [48, p. 27] and the second one in [48, p. 28].

Fact 1: Let N_t be a point process with $\mathcal{F}_t$-intensity α_t. If $\mathcal{F}_t^N \subset \mathcal{G}_t \subset \mathcal{F}_t$ for $t \geq 0$ and α_t is $\mathcal{G}_t$-predictable, then α_t is also the $\mathcal{G}_t$-intensity of N_t.

Fact 2: Let N_t be a point process with $\mathcal{F}_t$-intensity α_t. If $\mathcal{G}_t$ is a filtration such that $\mathcal{G}_\infty$ is independent of $\mathcal{F}_t$ for all $t \geq 0$. Then α_t is also the $\mathcal{G}_t \vee \mathcal{F}_t$-intensity of N_t, where $\mathcal{G}_t \vee \mathcal{F}_t = \sigma(\mathcal{G}_t \cup \mathcal{F}_t)$.

Throughout this chapter, the notation P_λ will be reserved to the probability measure on $(D_T, \mathcal{F}_T^N)$ that makes N_t a point process with a constant intensity λ with respect to $\mathcal{F}_t^N$. That is on $(D_T, \mathcal{F}_T^N, P_\lambda)$, the point process N is Poisson with rate λ. The probability measure P_1 will be used as the reference measure in this chapter.

3.2.3 Channel Model

We describe in this section the transition probability from $\mathcal{X}_T$ to $\mathcal{D}_T$ that we use to capture the dynamics of the channel of the legitimate receiver and that of the eavesdropper. For the sake of brevity, some technical details will be omitted. For a more complete exposition, the interested reader is referred to [46]-[54]. Recall that the channels of the legitimate receiver and the eavesdropper are both exponential server timing channels (i.e., $\cdot/M/1$ queues) with respective service rates μ_1 and μ_2 and that the assumption here is that both of these queues are initially empty.

For a fixed arrival process $x \in \mathcal{X}_T$, the number of packets at time t in the main queue is given by

$$Q_t^{(1)} = x_t - Y_t \quad \text{for } t \in [0, T], \tag{3.5}$$

where Y_t is the number of departures seen by the legitimate party in the interval $[0, t]$. To model the dynamics of this queue, we would like to find a probability measure $P_x^{(1)}(dy)$ on $(\mathcal{D}_T, \mathcal{F}_T^Y)$ such that the point process Y_t admits the $(P_x^{(1)}, \mathcal{F}_t^Y)$-intensity

$$\alpha_t^{(1)} = \mu_1 I(Q_{t-}^{(1)} > 0) \quad \text{for } t \in (0, T]. \tag{3.6}$$

As described in [46] and [54], this can be accomplished by defining $P_x^{(1)}(dy)$ via the following Radon-Nikodym derivative

$$\frac{dP_x^{(1)}}{dP_1}(y) = L_x^{(1)}(y), \tag{3.7}$$

where

$$L_x^{(1)}(y) = \exp\left(\int_0^T [\log(\alpha_t^{(1)})dy_t + (1 - \alpha_t^{(1)})dt]\right). \tag{3.8}$$

To model the channel of the eavesdropper we follow a similar procedure. Namely, $Q_t^{(2)} = x_t - Z_t$ will be the size of the eavesdropper's queue at time t and Z_t will be the number of packets that have departed from this queue by time t. The channel transition probability will be $P_x^{(2)}(dz)$ and is defined by

$$\frac{dP_x^{(2)}}{dP_1}(z) = L_x^{(2)}(z) = \exp\left(\int_0^T [\log(\alpha_t^{(2)})dz_t + (1-\alpha_t^{(2)})dt]\right), \tag{3.9}$$

where

$$\alpha_t^{(2)} = \mu_2 I(Q_{t-}^{(2)} > 0) \quad \text{for } t \in (0,T]. \tag{3.10}$$

3.2.4 Mutual Information

A probability measure ν on the space $(\mathcal{X}_T, \mathcal{F}_T^X)$ combined with the channel transition probability $P_x^{(1)}(dy)$ given above will induce a probability measure $\pi^{(1)}(dy)$ on $(\mathcal{D}_T, \mathcal{F}_T^Y)$. It can be verified that $\pi^{(1)}(dy)$ has the following Radon-Nikodym derivative with respect to $P_1(dy)$

$$\frac{d\pi^{(1)}}{dP_1}(y) = \exp\left(\int_0^T [\log(\widehat{\alpha}_t^{(1)})dy_t + (1-\widehat{\alpha}_t^{(1)})dt]\right), \tag{3.11}$$

where

$$\widehat{\alpha}_t^{(1)} = E[\alpha_t^{(1)} | \mathcal{F}_{t-}^Y]. \tag{3.12}$$

Consequently, it can be shown that the normalized mutual information between an arrival process X_0^T and the corresponding departure process Y_0^T observed by the legitimate receiver can be written under the following form

$$\frac{1}{T} I(X_0^T; Y_0^T) = \frac{1}{T} E\left[\int_0^T (\Phi(\alpha_t^{(1)}) - \Phi(\widehat{\alpha}_t^{(1)}))dt\right], \tag{3.13}$$

where $\Phi(x) = x\ln(x)$ and $\Phi(0) = 0$. For the eavesdropper, similar objects can be defined analogously. Namely, $\pi^{(2)}(dz)$ will be the induced probability measure on $(\mathcal{D}_T, \mathcal{F}_T^Z)$ and $\widehat{\alpha}_t^{(2)} = E[\alpha_t^{(2)} | \mathcal{F}_{t-}^Z]$.

3.3 Main result

We proceed to the description of the main result of this chapter. We start with the information transmission aspect of the problem.

3.3.1 Encoding-Decoding

The transmitter wishes to communicate a message U uniformly distributed on $\mathcal{U} = \{1, ..., M\}$. An (n, M, T) code is constructed by mapping each message to an element $x \in \mathcal{X}_T$ having n arrivals over the interval $[0, T]$. The rate of this code will be $R = \frac{\ln M}{T}$. The encoding procedure could eventually be a stochastic one, the only restriction that we impose on the encoder is that the expected number of packets left in the main queue (averaged over the code-book and the queue distribution) at time T is no more than 1. This constraint will guarantee that the input rate and the output rate of the queue are asymptotically the same. The decoder after observing the departures during $[0, T]$ declares one of the M messages as transmitted. Note that since all the arrivals occur in the interval $[0, T]$, the departure process over $[0, T]$ is a sufficient statistic for deciding which message has been transmitted. The average probability of error of this code at the legitimate receiver will be denoted by P_{er} and the level of secrecy will be measured by $\frac{1}{T} I(U; Z_0^T)$.

Secrecy rates: A secrecy rate R_s is said to be *achievable* for this wiretap channel if for all $\epsilon > 0$ there exists a sequence of (n, M, T) codes with $T \to \infty$ such that

eventually we will have

$$\frac{1}{T}\ln M \geq R_s - \epsilon \tag{3.14}$$

$$P_{\text{er}} \leq \epsilon \tag{3.15}$$

$$\frac{1}{T}I(U; Z_0^T) \leq \epsilon \tag{3.16}$$

A secrecy rate R_s is said to be achievable at output rate λ if it is achievable using a sequence of (n, M, T_n) codes with

$$\lim_{n\to\infty} \frac{T_n}{n} = \frac{1}{\lambda}. \tag{3.17}$$

Secrecy Capacity: The supremum of achievable secrecy rates will be called the *secrecy capacity* and will be denoted by C_s. The supremum of achievable secrecy rates at output rate λ will be denoted by $C_s(\lambda)$.

From the above definitions, it can be verified that

$$C_s = \sup_{\lambda<\mu_1} C_s(\lambda). \tag{3.18}$$

3.3.2 Main result

The main result of this chapter is the characterization of the secrecy capacity C_s. Clearly, when $\mu_1 \leq \mu_2$, the secrecy capacity is zero. Moreover, when $\mu_2 < \frac{\mu_1}{e}$, it was shown in [42] that in this case the secrecy capacity is equal to the capacity of the main queue, i.e.,

$$C_s = \frac{\mu_1}{e}. \tag{3.19}$$

Throughout the rest of the chapter we will assume therefore that the condition $\frac{\mu_1}{e} \leq \mu_2 < \mu_1$ is satisfied. The main result of the chapter is the following theorem.

Theorem 6. *The secrecy capacity for* $\frac{\mu_1}{e} \leq \mu_2 < \mu_1$ *is given by*

$$C_s = \mu_2 \log \frac{\mu_1}{\mu_2}. \tag{3.20}$$

The main contribution of the present chapter is proving the converse part, i.e., $C_s \leq \mu_2 \log \frac{\mu_1}{\mu_2}$. The achievability part of the theorem was established in [42]. For the sake of completeness we will provide below a proof of the achievability of the rate $\mu_2 \log \left(\frac{\mu_1}{\mu_2}\right)$ using the point process channel model that we are using in this chapter.

3.4 Achievability of the rate $\mu_2 \log \left(\frac{\mu_1}{\mu_2}\right)$

To prove that the rate $\mu_2 \log \left(\frac{\mu_1}{\mu_2}\right)$ is achievable in secrecy, we need to show that the probability of error on the main channel and the information leakage to the eavesdropper can be both made arbitrary small. By choosing Poisson processes with intensity μ_2 as codewords, it can be shown [46] that the rate $\mu_2 \log \left(\frac{\mu_1}{\mu_2}\right)$ can be achieved reliably over the main channel. It remains to verify that the secrecy constraint is satisfied as well. The following lemma shows that in order to satisfy the secrecy constraint it suffices to overload the queue of the eavesdropper by choosing $\lambda \geq \mu_2$. As concluded in [42], what this shows is that stochastic encoding is not required to confuse the eavesdropper.

Lemma 8. *For* $\lambda \geq \mu_2$*, we have*

$$\lim_{T\to\infty} \frac{1}{T} I(U; Z_0^T) = 0. \tag{3.21}$$

Proof. Since the mutual information is always non negative, we clearly have $\lim_{T\to\infty} \frac{1}{T} I(U; Z_0^T) \geq 0$ and we only need to show the reverse inequality. Ap-

plying the data processing inequality on the Markov chain $U \leftrightarrow X_0^T \leftrightarrow Z_0^T$, we have

$$I(U; Z_0^T) \leq I(X_0^T; Z_0^T). \tag{3.22}$$

So to prove the lemma it suffices to show that

$$\lim_{T\to\infty} \frac{1}{T} I(X_0^T; Z_0^T) \leq 0. \tag{3.23}$$

As seen in Section 3.2, the mutual information between X_0^T and Z_0^T can be expressed via the following formula,

$$\frac{1}{T} I(X_0^T; Z_0^T) = \frac{1}{T} E\left[\int_0^T (\Phi(\alpha_t^{(2)}) - \Phi(\widehat{\alpha}_t^{(2)}))dt\right]. \tag{3.24}$$

Using Jensen's inequality combined with Fubini's theorem we can upper bound $\frac{1}{T} I(X_0^T; Z_0^T)$ as follows

$$\frac{1}{T} I(X_0^T; Z_0^T) \leq \frac{1}{T} \int_0^T E[\Phi(\alpha_t^{(2)})]dt - \Phi\left(\frac{1}{T}\int_0^T E[\widehat{\alpha}_t^{(2)}]dt\right). \tag{3.25}$$

Since $E[\widehat{\alpha}_t^{(2)}] = E[\alpha_t^{(2)}]$, the last inequality becomes

$$\frac{1}{T} I(X_0^T; Z_0^T) \leq \frac{1}{T} \int_0^T E[\Phi(\alpha_t^{(2)})]dt - \Phi\left(\frac{1}{T}\int_0^T E[\alpha_t^{(2)}]dt\right). \tag{3.26}$$

Note that

$$\begin{aligned}
\Phi(\alpha_t^{(2)}) &= \mu_2 I(Q_{t-}^{(2)} > 0) \ln(\mu_2 I(Q_{t-}^{(2)} > 0)) \\
&= \mu_2 \ln(\mu_2) I(Q_{t-}^{(2)} > 0) + \mu_2 I(Q_{t-}^{(2)} > 0) \ln(I(Q_{t-}^{(2)} > 0)) \\
&= \Phi(\mu_2) I(Q_{t-}^{(2)} > 0) + \mu_2 \Phi(I(Q_{t-}^{(2)} > 0)) \\
&= \Phi(\mu_2) I(Q_{t-}^{(2)} > 0),
\end{aligned} \tag{3.27}$$

where in the last equality we used the fact $\Phi(1) = \Phi(0) = 0$. It follows therefore that

$$\frac{1}{T} I(X_0^T; Z_0^T) \leq \frac{\Phi(\mu_2)}{T} \int_0^T P(Q_{t-}^{(2)} > 0)dt - \Phi\left(\frac{\mu_2}{T}\int_0^T P(Q_{t-}^{(2)} > 0)dt\right). \tag{3.28}$$

If the input rate is such that $\lambda \geq \mu_2$, the eavesdropper's queue will be unstable and for sufficiently large t we will have $P(Q_{t-}^{(2)} > 0) = 1$. This implies

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T P(Q_{t-}^{(2)} > 0) dt = 1. \tag{3.29}$$

The function $\Phi(\cdot)$ being continuous it follows that

$$\lim_{T\to\infty} \frac{1}{T} I(X_0^T; Z_0^T) \leq \Phi(\mu_2) - \Phi(\mu_2) = 0. \tag{3.30}$$

This completes the proof of the achievability. □

3.5 Converse

3.5.1 Preliminaries

We start this section with the following simple lemma.

Lemma 9. *When $\mu_2 \leq \lambda < \mu_1$, the secrecy capacity at output rate λ is upper-bounded as follows*

$$C_s(\lambda) \leq \mu_2 \log\left(\frac{\mu_1}{\mu_2}\right). \tag{3.31}$$

Proof. It is known that the capacity of the main channel at output rate $\lambda < \mu_1$ is [41],[46]

$$C(\lambda) = \lambda \log\left(\frac{\mu_1}{\lambda}\right). \tag{3.32}$$

As the secrecy capacity can never exceed the capacity of the main channel, a trivial upper bound on $C_s(\lambda)$ is then

$$C_s(\lambda) \leq C(\lambda) = \lambda \log\left(\frac{\mu_1}{\lambda}\right). \tag{3.33}$$

Now, note that

$$\frac{d}{d\lambda}C(\lambda) = \log\left(\frac{\mu_1}{e\lambda}\right). \tag{3.34}$$

Recalling the assumption $\frac{\mu_1}{e} \leq \mu_2$, we see that if $\mu_2 \leq \lambda$, we have

$$C_s(\lambda) \leq C(\lambda) \leq C(\mu_2) = \mu_2 \log\left(\frac{\mu_1}{\mu_2}\right) \quad \text{for} \quad \mu_2 \leq \lambda < \mu_1, \tag{3.35}$$

which is the result of the lemma. □

To establish the converse, we need to show that the inequality provided in Lemma 9 holds also for $\lambda < \mu_2$. We will in fact establish a tighter inequality given in the following theorem.

Theorem 7. *When $\lambda < \mu_2$, the secrecy capacity with output rate λ is upper-bounded by*

$$C_s(\lambda) \leq \lambda \log\left(\frac{\mu_1}{\mu_2}\right). \tag{3.36}$$

By combining the inequality in this theorem and the one given in the previous lemma we have

$$C_s = \sup_{\lambda<\mu_1} C_s(\lambda) \leq \mu_2 \log\left(\frac{\mu_1}{\mu_2}\right), \tag{3.37}$$

which establishes the converse and when combined with the achievability result proves Theorem 6. The rest of this chapter is dedicated to the proof of Theorem 7. Therefore, we are going to assume that $\lambda < \mu_2$. To prove Theorem 7, we will need several intermediate results. The first one, given in the lemma below, shows that the secrecy capacity when the queues are initially in equilibrium is an upper bound on the secrecy capacity when the queues are initially empty. Analyzing queuing systems is often easier when the queues are in equilibrium. The problem treated in this chapter is not an exception to this rule.

Lemma 10. *If a rate R_s is achievable with secrecy at output rate λ when the queues are initially empty then R_s is also achievable with secrecy when the queues start initially in equilibrium (with respect to a Poisson process with rate λ).*

Proof. Let $\mathcal{C}_n = (n, M, T_n)$ be a sequence of codes that achieve the secrecy rate R_s at output rate λ when the queues are initially empty. We would like to transform this sequence into another sequence of codes that can be used to communicate reliably and in secrecy when the queues start in equilibrium. In order to do that, we can imagine that at time zero, the encoder injects a synchronization packet (labeled packet 0) that bears a special mark. After observing the departure of packet 0 from the queue, the decoder will know that all the initial packets present in the queue have already departed. Consequently, all subsequent departing packets belong to the codeword sent by the encoder. Now let $(\Delta_n)_n$ be a sequence of positive numbers satisfying the following two conditions

$$\lim_{n\to\infty} \Delta_n = 0, \tag{3.38}$$

and

$$\lim_{n\to\infty} \Delta_n T_n = \infty. \tag{3.39}$$

We define a new sequence of codes that we shall call $\mathcal{C}_n^{\text{eq}}$ that is obtained from $\mathcal{C}_n$ by the following transformation. For every codeword in $\mathcal{C}_n$, we shift the arrival times of its n packets by $\Delta_n T_n$ and then add the special packet 0 that will arrive to the queue at time 0. We will show that the sequence of codes $\mathcal{C}_n^{\text{eq}}$ achieves the secrecy rate R_s when the queues start initially in equilibrium. From (3.38), we note that $\mathcal{C}_n$ and $\mathcal{C}_n^{\text{eq}}$ will have asymptotically the same rate. Indeed,

$$\frac{\text{Rate of } \mathcal{C}_n}{\text{Rate of } \mathcal{C}_n^{\text{eq}}} = 1 + \Delta_n \to 0 \text{ as } n \to \infty. \tag{3.40}$$

Let D_0 be the departure time of packet 0 from the main queue. The same decoder used for $\mathcal{C}_n$ is used again for $\mathcal{C}_n^{\text{eq}}$ with one modification; if $D_0 > \Delta_n T_n$, the entire

codeword is declared in error. The probability of error of this decoder can be bounded as follows

$$\text{Probability of Error} \leq P[D_0 > \Delta_n T_n] + P[\text{Decoding error} | D_0 \leq \Delta_n T_n]. \quad (3.41)$$

The term $P[D_0 > \Delta_n T_n]$ in (3.41) can be computed explicitly [41]

$$P[D_0 > \Delta_n T_n] = \exp(-(\mu_1 - \lambda)\Delta_n T_n). \quad (3.42)$$

Since $\lambda < \mu_2 < \mu_1$, this term converges to zero exponentially fast (cf. (3.39)). Given $D_0 \leq \Delta_n T_n$, since all the packets of the codeword were delayed by $\Delta_n T_n$, the first arriving packet finds the queue of the legitimate receiver empty. Hence, since $\mathcal{C}_n$ has a vanishing error probability, $P[\text{Decoding error} | D_0 \leq \Delta_n T_n]$ converges to 0. We conclude therefore that $\mathcal{C}_n^{\text{eq}}$ can be used to communicate reliably over the main queue.

To finish the proof of the lemma, it remains to verify that the sequence of codes $(\mathcal{C}_n^{\text{eq}})_n$ has a vanishing information leakage rate to the eavesdropper. This, however, simply follows from the fact that the sequence of codes $(\mathcal{C}_n)_n$ has a vanishing information leakage rate and that the initial number of packets in the eavesdropper's queue can only decrease the dependence between the transmitted message and the departure process from this queue. □

3.5.2 Degradation Lemma and consequences

Here we shall prove one of the key lemmas required to establish the converse. This lemma is stated below.

Lemma 11. *When $\mu_2 < \mu_1$ and the queues start at time zero in equilibrium with respect to a Poisson process with rate $\lambda \geq 0$, the eavesdropper's channel is stochastically*

degraded with respect to the legitimate receiver's channel.

Note that the case $\lambda = 0$ corresponds to the situation where the queues start empty at time zero. That is this lemma holds regardless of whether the queues start empty or in equilibrium. To prove this lemma, we will show that we can emulate an exponential server timing channel with parameter μ_2 (i.e., the departure process Z_0^T) by degrading the observation of the legitimate receiver (i.e., the departure process Y_0^T) with some operations that are independent of the transmitted codeword (i.e., the arrival process X_0^T). This result allow us to assume that the Markov chain

$$X_0^T \to Y_0^T \to Z_0^T$$

holds. Before proving this lemma, we will pause for a moment and go over two consequences of this Markov chain. The first byproduct is the following inequality.

Lemma 12. *For every (n, M, T)-code with rate $R = \frac{\ln M}{T}$, we have*

$$R \leq \frac{1}{T(1-P_{\text{er}})}(I(X_0^T; Y_0^T) - I(X_0^T; Z_0^T) + I(U; Z_0^T) + H(P_{\text{er}})), \tag{3.43}$$

where $H(P_{\text{er}})) = -P_{\text{er}} \log(P_{\text{er}}) - (1 - P_{\text{er}}) \log(1 - P_{\text{er}})$ is the binary entropy.

The proof of this lemma can be found in the previous chapter. The second consequence of the degradedness is the well know data processing inequality for relative entropy. To present this inequality in the context studied in this chapter, we need to introduce some further notation. Let $\nu_1(dx)$ and $\nu_2(dx)$ be two probability measures on the space of arrival processes $(\mathcal{X}_T, \mathcal{F}_T^X)$. Let $\pi_1^Y(dy)$ and $\pi_2^Y(dy)$ be the induced probability measures on the space of departure processes from the legitimate queue $(\mathcal{D}_T, \mathcal{F}_T^Y)$. These two measures are absolutely

continuous with respect to $P_\lambda(dy)$ and hence admit Radon-Nikodym derivatives $p_1^Y(y)$ and $p_2^Y(y)$ with respect to $P_\lambda(dy)$, i.e.,

$$p_i^Y(y) = \frac{\pi_i^Y(dy)}{P_\lambda(dy)} \quad i = 1, 2. \tag{3.44}$$

The relative entropy between π_1^Y and π_2^Y is then defined as follows

$$D(\pi_1^Y||\pi_2^Y) = \int p_1^Y(y) \log\left(\frac{p_1^Y(y)}{p_2^Y(y)}\right) P_\lambda(dy). \tag{3.45}$$

We define similarly $D(\pi_1^Z||\pi_2^Z)$, where $\pi_1^Z(dz)$ and $\pi_2^Z(dz)$ are probability measures on the space of departure processes from the eavesdropper queue $(\mathcal{D}_T, \mathcal{F}_T^Z)$, induced respectively by $\nu_1(dx)$ and $\nu_2(dx)$. An immediate consequence of the Markov chain $X_0^T \to Y_0^T \to Z_0^T$ is the data processing theorem for relative entropies stated as follows.

Lemma 13 (Data Processing Inequality)**.**

$$D(\pi_1^Y||\pi_2^Y) \geq D(\pi_1^Z||\pi_2^Z). \tag{3.46}$$

The proof of this inequality can be found for instance in [49]. We shall use this inequality combined with another well know theorem, but this time not in information theory but rather in queueing theory, Burke's theorem. More on this will come later. We present now the proof of the degradedness. Since this lemma is central to the results of this chapter and not all readers of this thesis are familiar with the stochastic intensity theory of point process, we present in Section 3.6 a different proof that we believe will be more appealing to those readers.

Proof of Lemma 11. Assume that at time zero the queues start in equilibrium with respect to a Poisson process with rate λ. In this case, the distribution of the

queue size at time zero $Q_0^{(1)}$ for the legitimate receiver and $Q_0^{(2)}$ for the eavesdropper is given by

$$P[Q_0^{(i)} = k] = (1-\rho_i)\rho_i^k, \quad k \geq 0, \quad i = 1, 2. \tag{3.47}$$

where $\rho_i = \frac{\lambda}{\mu_i}$. The case $\lambda = 0$ corresponds to the situation where the queues start empty since $P[Q_0^{(1)} = 0] = P[Q_0^{(2)} = 0] = 1$ when $\lambda = 0$. Now, let $\widetilde{Q}_0$ be a mixed geometric random variable (independent of all other random variables) defined as follows

$$P[\widetilde{Q}_0 = k] = \begin{cases} \beta, \text{ if } k = 0 \\ (1-\beta)(1-\rho_2)\rho_2^{k-1}, \text{ if } k \geq 1. \end{cases}$$

where

$$\beta = \frac{1-\rho_2}{1-\rho_1}. \tag{3.48}$$

The random variable $\widetilde{Q}_0$ has been chosen so that

$$Q_0^{(2)} \stackrel{d}{=} Q_0^{(1)} + \widetilde{Q}_0. \tag{3.49}$$

This can be easily verified by means of probability generating functions as follows

$$\begin{aligned} E\left[z^{Q_0^{(1)}+\widetilde{Q}_0}\right] = E\left[z^{Q_0^{(1)}}\right] E\left[z^{\widetilde{Q}_0}\right] &= \frac{1-\rho_1}{1-\rho_1 z}(\beta + (1-\beta)z\frac{1-\rho_2}{1-\rho_2 z}) \\ &= \frac{1-\rho_2}{1-\rho_1 z}(1 + \frac{(\rho_2-\rho_1)z}{1-\rho_2 z}) \\ &= \frac{1-\rho_2}{1-\rho_2 z} = E\left[z^{Q_0^{(2)}}\right]. \end{aligned} \tag{3.50}$$

As said above, the ESTC with parameter μ_2 will be constructed by degrading the departure process of an ESTC with parameter μ_1 (with $\mu_1 > \mu_2$). The degrading procedure that we use requires an extra ESTC with parameter μ_2, we will call this queue the degrading queue. At time zero, this queue will start with $\widetilde{Q}_0$ packets and we let $\widetilde{Q}_t$ be the number of packets in this queue at time t. The

input to this queue will be a thinned version of the departure process of the legitimate queue (i.e., Y_0^T). The thinning operation depends however on the state of the degrading queue $\widetilde{Q}_t$. This thinning operation works as follows; if a departure from the legitimate queue occurs at time t (i.e., $\Delta Y_t = Y_t - Y_{t-} = 1$) then

- If $\widetilde{Q}_{t-} > 0$, the packet enters the degrading queue.
- If $\widetilde{Q}_{t-} = 0$,
  - With probability $\frac{\mu_2}{\mu_1}$ the packet does not enter the degrading queue and leaves immediately the system.
  - With probability $1 - \frac{\mu_2}{\mu_1}$ the packet enters the degrading queue.

We define $\widetilde{Z}_t$ to be the output of this system. In other words, $\widetilde{Z}_t$ is the point process that counts the number of packets in the two streams: the steam of departures from the degrading queue (that we shall denote by V_t) plus the stream of packets that left the system immediately without entering the degrading queue (denoted by U_t). We define $Q_t = \widetilde{Q}_t + Q_t^{(1)}$ to be the total number of packets at time t in the system (i.e., in both queues). We will prove that this system behaves as an ESTC with service rate μ_2 by showing that the departure process $\widetilde{Z}_0^T$ from this system admits the $\mathcal{F}_t^Q$-intensity $\mu_2 I(Q_{t-} > 0)$. Since we also have that $Q_0 \stackrel{d}{=} Q_0^{(2)}$ and $\widetilde{Z}_0^T$ is constructed using only Y_0^T and the state of the degrading queue $\widetilde{Q}$ (which depends on X_0^T only through Y_0^T), this will be sufficient to prove the lemma.

In the following, we let d_n be the nth jump time of Y_t, i.e.,

$$Y_t = \sum_n I(d_n \leq t), \tag{3.51}$$

where $I(d_n \leq t)$ is equal to 1 if $d_n \leq t$ and is equal to 0 otherwise. On the same probability space we define a sequence of i.i.d. Bernoulli random variables $(B_n, n = 1, 2, ...)$ such that

$$P[B_n = 0] = 1 - P[B_n = 1] = \frac{\mu_2}{\mu_1}. \tag{3.52}$$

This sequence will be taken to be independent of all other random variables and we let $\mathcal{F}_\infty^B = \sigma(B_1, B_2, ...)$. On the same probability space we consider also an independent Poisson process N_t with $\mathcal{F}_t^N$-intensity μ_2 and we let k_n be the nth jump time of N_t, i.e.,

$$N_t = \sum_n I(k_n \leq t). \tag{3.53}$$

Now define

$$\mathcal{F}_t = \sigma(\widetilde{Q}_0) \vee \sigma(Q_0^{(1)}) \vee \mathcal{F}_t^X \vee \mathcal{F}_t^Y \vee \mathcal{F}_t^N \vee \mathcal{F}_\infty^B. \tag{3.54}$$

From Fact 2 in Section 3.2, we can see that the $\mathcal{F}_t$-intensity of N_t is also μ_2. This is because $\mathcal{F}_t^N$ is independent of $\sigma(\widetilde{Q}_0) \vee \sigma(Q_0^{(1)}) \vee \mathcal{F}_\infty^X \vee \mathcal{F}_\infty^Y \vee \mathcal{F}_\infty^B$. For similar reasons, the $\mathcal{F}_t$-intensity of Y_t is $\mu_1 I(Q_{t-}^{(1)} > 0)$.

The number of packets at time t in the degraded queue can be written as

$$\widetilde{Q}_t = \widetilde{Q}_0 + \widetilde{Y}_t - V_t \tag{3.55}$$

The process V_t represents the number of packets at time t that departed from the degrading queue. The process V_t can be constructed from the points of N_t as

$$V_t = \sum_n I(\widetilde{Q}_{k_n-} > 0) I(k_n \leq t). \tag{3.56}$$

The process $\widetilde{Y}_t$ represents the number of packets that entered the degrading queue at time t. Using the sequence of Benoulli random variables defined above, we can represent this process as follows

$$\widetilde{Y}_t = \sum_n (B_n I(\widetilde{Q}_{d_n-} = 0) + I(\widetilde{Q}_{d_n-} > 0)) I(d_n \leq t). \tag{3.57}$$

Now let U_t represent the number of packets at time t that left the system without receiving any service in the degrading queue, i.e.,

$$U_t = \sum_n (1 - B_n) I(\widetilde{Q}_{d_n-} = 0) I(d_n \leq t). \tag{3.58}$$

As one should expect, we must have

$$Y_t = \widetilde{Y}_t + U_t. \tag{3.59}$$

The output of the system $\widetilde{Z}_t$ is then obtained by combined the two streams, i.e.,

$$\widetilde{Z}_t = U_t + V_t. \tag{3.60}$$

Observe that U_t represents the common points of the process Y_t and $\widetilde{Z}_t$, i.e.,

$$U_t = \sum_{0<s\leq t} \Delta Y_s \Delta \widetilde{Z}_s. \tag{3.61}$$

We define now $\mathcal{G}_t = \mathcal{F}_t^{Q^{(1)}} \vee \mathcal{F}_t^{\widetilde{Q}}$ and we start by computing the $\mathcal{G}_t$-intensity of U_t. Let C_t be $\mathcal{G}_t$-predictable, from the definition of U_t we have

$$E\left[\int_0^\infty C_s dU_s\right] = \sum_n E[(1 - B_n) C_{d_n} I(\widetilde{Q}_{d_n-} = 0) I(d_n < \infty)]. \tag{3.62}$$

Observe that $C_{d_n} I(\widetilde{Q}_{d_n-} = 0) I(d_n < \infty)$ is $\mathcal{G}_{d_n-}$-measurable and that B_n is independent of $\mathcal{G}_{d_n-}$. Consequently, we have

$$\begin{aligned}
&E[(1 - B_n) C_{d_n} I(\widetilde{Q}_{d_n-} = 0) I(d_n < \infty)] = \\
&E[E[(1 - B_n) C_{d_n} I(\widetilde{Q}_{d_n-} = 0) I(d_n < \infty) | \mathcal{G}_{d_n-}]] = \\
&E[C_{d_n} I(\widetilde{Q}_{d_n-} = 0) I(d_n < \infty) E[(1 - B_n) | \mathcal{G}_{d_n-}]] = \\
&E[C_{d_n} I(\widetilde{Q}_{d_n-} = 0) I(d_n < \infty) E[(1 - B_n)]] = \\
&\frac{\mu_2}{\mu_1} E[C_{d_n} I(\widetilde{Q}_{d_n-} = 0) I(d_n < \infty)].
\end{aligned} \tag{3.63}$$

It follows that

$$E\left[\int_0^\infty C_s dU_s\right] = \frac{\mu_2}{\mu_1}\sum_n E[C_{d_n} I(\widetilde{Q}_{d_n-}=0) I(d_n<\infty)]$$
$$= \frac{\mu_2}{\mu_1} E\left[\int_0^\infty C_s I(\widetilde{Q}_{s-}=0) dY_s\right]. \quad (3.64)$$

The process $(C_t I(\widetilde{Q}_{t-}=0))_{t>0}$ is $\mathcal{G}_t$-predictable, since $\mathcal{G}_t \subset \mathcal{F}_t$, this process is also $\mathcal{F}_t$-predictable. Recall that the $\mathcal{F}_t$-intensity of Y_t is $\mu_1 I(Q^{(1)}_{t-}>0)$, therefore

$$E\left[\int_0^\infty C_s dU_s\right] = \frac{\mu_2}{\mu_1} E\left[\int_0^\infty C_s I(\widetilde{Q}_{s-}=0) dY_s\right]$$
$$= \mu_2 E\left[\int_0^\infty C_s I(\widetilde{Q}_{s-}=0) I(Q^{(1)}_{s-}>0) ds\right]. \quad (3.65)$$

The process $(\mu_2 I(\widetilde{Q}_{t-}=0) I(Q^{(1)}_{t-}>0))_{t>0}$ being $\mathcal{G}_t$-predictable we deduce that it is the $\mathcal{G}_t$-intensity of U_t.

We move now to the $\mathcal{G}_t$-intensity of V_t, from the definition of V_t we have

$$E\left[\int_0^\infty C_s dV_s\right] = E\left[\int_0^\infty C_s I(\widetilde{Q}_{s-}>0) dN_s\right]. \quad (3.66)$$

The process $(C_t I(\widetilde{Q}_{t-}>0))_{t>0}$ is $\mathcal{G}_t$-predictable, since $\mathcal{G}_t \subset \mathcal{F}_t$, this process is also $\mathcal{F}_t$-predictable. Recall that the $\mathcal{F}_t$-intensity of N_t is μ_2, we deduce therefore that

$$E\left[\int_0^\infty C_s dV_s\right] = \mu_2 E\left[\int_0^\infty C_s I(\widetilde{Q}_{s-}>0) ds\right]. \quad (3.67)$$

Since $(\mu_2 I(\widetilde{Q}_{t-}>0))_{t>0}$ is $\mathcal{G}_t$-predictable, it must be the $\mathcal{G}_t$-intensity of V_t.

Now recalling that $\widetilde{Z}_t = U_t + V_t$, we conclude that the $\mathcal{G}_t$-intensity of $\widetilde{Z}_t$ is $\mu_2[I(\widetilde{Q}_{t-}>0) + I(\widetilde{Q}_{t-}=0) I(Q^{(1)}_{t-}>0)]$. This intensity can be also written as $\mu_2 I(Q_{t-}>0)$. This can be verified easily through the enumeration of all the

possible cases or by the following

$$\begin{aligned} I(Q_{t-}>0) &= I(Q^{(1)}_{t-}+\widetilde{Q}_{t-}>0) \\ &= 1-I(Q^{(1)}_{t-}+\widetilde{Q}_{t-}=0) \\ &= 1-I(Q^{(1)}_{t-}=0)I(\widetilde{Q}_{t-}=0) \\ &= I(\widetilde{Q}_{t-}>0)+I(\widetilde{Q}_{t-}=0)-I(Q^{(1)}_{t-}=0)I(\widetilde{Q}_{t-}=0) \\ &= I(\widetilde{Q}_{t-}>0)+I(\widetilde{Q}_{t-}=0)(1-I(Q^{(1)}_{t-}=0)) \\ &= I(\widetilde{Q}_{t-}>0)+I(\widetilde{Q}_{t-}=0)I(Q^{(1)}_{t-}>0). \end{aligned} \tag{3.68}$$

Clearly $\mu_2 I(Q_{t-}>0)$ is $\mathcal{F}^Q_t$-predictable and $\mathcal{F}^Q_t \subset \mathcal{G}_t$; it follows from Fact 1 in Section 3.2 that $\mu_2 I(Q_{t-}>0)$ is the $\mathcal{F}^Q_t$-intensity of $\widetilde{Z}_t$. This concludes the proof of Lemma 11. □

3.5.3 Proof of Theorem 7

Before giving the proof of Theorem 7, we need some intermediate results.

Lemma 14. *Assume that the queues start in equilibrium (with respect to a Poisson process with arrival rate λ), then*

$$I(X_0^T;Y_0^T)-I(X_0^T;Z_0^T) \le \log(\frac{\mu_1}{e\lambda})E[Y_T]-\log(\frac{\mu_2}{e\lambda})E[Z_T]+H(Q_0^{(2)}), \tag{3.69}$$

where $H(Q_0^{(2)})$ is the entropy of $Q_0^{(2)}$, i.e.,

$$H(Q_0^{(2)}) = -\frac{1}{1-\rho_2}\left(\rho_2\log\rho_2+(1-\rho_2)\log(1-\rho_2)\right). \tag{3.70}$$

Proof. Using Kolmogorov's formula we have [35, Lemma 3.2]

$$\begin{aligned} I(X_0^T;Y_0^T)-I(X_0^T;Z_0^T) = I(X_0^T,Q_0^{(1)};Y_0^T)-I(Q_0^{(1)};Y_0^T|X_0^T) \\ -I(X_0^T,Q_0^{(2)};Z_0^T)+I(Q_0^{(2)};Z_0^T|X_0^T). \end{aligned} \tag{3.71}$$

Since $I(Q_0^{(1)}; Y_0^T | X_0^T) \geq 0$ and $I(Q_0^{(2)}; Z_0^T | X_0^T) \leq H(Q_0^{(2)})$, we have

$$I(X_0^T; Y_0^T) - I(X_0^T; Z_0^T) \leq I(X_0^T, Q_0^{(1)}; Y_0^T) - I(X_0^T, Q_0^{(2)}; Z_0^T) + H(Q_0^{(2)}). \quad (3.72)$$

We will show now that

$$I(X_0^T, Q_0^{(1)}; Y_0^T) - I(X_0^T, Q_0^{(2)}; Z_0^T) \leq \log(\frac{\mu_1}{e\lambda}) E[Y_T] - \log(\frac{\mu_2}{e\lambda}) E[Z_T]. \quad (3.73)$$

We remind the reader of the following general formula for computing mutual information. If you have a channel $A \xrightarrow{P_{B|A}} B$, then for any probability distribution W such that $P_B \ll W$ we have

$$I(A; B) = D(P_{B|A} || W | P_A) - D(P_B || W), \quad (3.74)$$

where

$$D(P_{B|A} || W | P_A) = E_A[D(P_{B|A=a} || W)], \quad (3.75)$$

and $E_A[\cdot]$ designates the expectation with respect to the distribution of A. We start working on the term $I(X_0^T, Q_0^{(1)}; Y_0^T)$. In order to simplify the notation, we will use X,Y and Q_0 instead of X_0^T, Y_0^T and $Q_0^{(1)}$. Applying the previous identity with $A = (X, Q_0)$, $B = Y$, $W = P_\lambda$ and $P_B = \pi^Y$, we obtain

$$I(X_0^T, Q_0^{(1)}; Y_0^T) = E_{X,Q_0} \left[D(P_{Y|X=x,Q_0=q_0} || P_\lambda) \right] - D(\pi^Y || P_\lambda). \quad (3.76)$$

By denoting by $p^Y_{x,q_0}(y)$ the Radon-Nikodym derivative of $P_{Y|X=x,Q_0=q_0}(dy)$ with respect to $P_\lambda(dy)$, we have

$$D(P_{Y|X=x,Q_0=q_0} || P_\lambda) = \int p^Y_{x,q_0}(y) \log(p^Y_{x,q_0}(y)) P_\lambda(dy) \quad (3.77)$$

$$= \int \log(p^Y_{x,q_0}(y)) P_{Y|X=x,Q_0=q_0}(dy). \quad (3.78)$$

Now, since $P_1(dy) \ll P_\lambda(dy)$ (in fact these two measures are equivalent), we have

$$\frac{dP_1}{dP_\lambda}(y) = \left(\frac{1}{\lambda}\right)^{y_T} e^{-(1-\lambda)T}. \quad (3.79)$$

Using the chain rule for Radon-Nikodym derivatives and the results in Section 3.2 we have

$$\begin{aligned}\log(p^Y_{x,q_0}(y)) &= \log\left(\frac{dP_{Y|X=x,Q_0=q_0}}{dP_1}(y)\right) + \log\left(\frac{dP_1}{dP_\lambda}(y)\right) \\ &= \int_0^T [\log(\alpha_t^{(1)})dy_t + (1-\alpha_t^{(1)})dt] + (\lambda-1)T - \log(\lambda)y_T \\ &= \int_0^T \log(\mu_1)dy_t + \int_0^T \log(I(q_{t-}^{(1)}>0))dy_t + T \\ &\quad - \int_0^T \mu_1 I(q_{t-}^{(1)}>0)dt + (\lambda-1)T - \log(\lambda)y_T, \end{aligned} \tag{3.80}$$

where $q_t^{(1)} = q_0 + x_t - y_t$. Note that

$$\int_0^T \log(I(q_{t-}^{(1)}>0))dy_t = 0. \tag{3.81}$$

This is because when $I(q_{t-}^{(1)}>0)=0$ no departures can happen at time t and so $dy_t = 0$. This leads to

$$\log(p^Y_{x,q_0}(y)) = \log\left(\frac{\mu_1}{\lambda}\right)y_T - \int_0^T \mu_1 I(q_{t-}^{(1)}>0)dt + \lambda T, \tag{3.82}$$

Consequently, we can write

$$E_{X,Q_0}\left[D(P_{Y|X=x,Q_0=q_0}||P_\lambda)\right] = \lambda T + \log\left(\frac{\mu_1}{\lambda}\right)E\left[Y_T\right] - E\left[\int_0^T \mu_1 I(Q_{t-}^{(1)}>0)dt\right], \tag{3.83}$$

where the expectation is over the joint distribution of (X, Q_0, Y). Recalling that $\mu_1 I(Q_{t-}^{(1)}>0)$ is the stochastic intensity of the process Y_t we obtain that

$$E\left[\int_0^T \mu_1 I(Q_{t-}^{(1)}>0)dt\right] = E\left[\int_0^T dY_t\right] = E[Y_T]. \tag{3.84}$$

That is

$$E_{X,Q_0}\left[D(P_{Y|X=x,Q_0=q_0}||P_\lambda)\right] = \lambda T + \log\left(\frac{\mu_1}{e\lambda}\right)E\left[Y_T\right]. \tag{3.85}$$

This leads to

$$I(X_0^T, Q_0^{(1)}; Y_0^T) = \lambda T + \log\left(\frac{\mu_1}{e\lambda}\right)E\left[Y_T\right] - D(\pi^Y||P_\lambda). \tag{3.86}$$

A similar analysis shows that

$$I(X_0^T, Q_0^{(2)}; Z_0^T) = \lambda T + \log\left(\frac{\mu_2}{e\lambda}\right) E\left[Z_T\right] - D(\pi^Z||P_\lambda). \tag{3.87}$$

We conclude that

$$\begin{aligned} I(X_0^T, Q_0^{(1)}; Y_0^T) - I(X_0^T, Q_0^{(2)}; Z_0^T) = \log(\frac{\mu_1}{e\lambda})E[Y_T] - \log(\frac{\mu_2}{e\lambda})E[Z_T] \\ - D(\pi^Y||P_\lambda) + D(\pi^Z||P_\lambda). \end{aligned} \tag{3.88}$$

The two queues under consideration are assumed to be in equilibrium with respect to a Poisson process with rate λ. It follows that P_λ is a fixed point for these queues [45] (see also [48, Theorem T1 pp. 123-124]). Using the data processing inequality given in Lemma 13 with $\pi_1^Y = \pi^Y$, $\pi_1^Z = \pi^Z$ and $\nu_2 = P_\lambda$, Burke's theorem guarantees that $\pi_2^Y = \pi_2^Z = P_\lambda$. Therefore

$$D(\pi^Y||P_\lambda) \geq D(\pi^Z||P_\lambda). \tag{3.89}$$

This shows that

$$I(X_0^T, Q_0^{(1)}; Y_0^T) - I(X_0^T, Q_0^{(2)}; Z_0^T) \leq \log(\frac{\mu_1}{e\lambda})E[Y_T] - \log(\frac{\mu_2}{e\lambda})E[Z_T], \tag{3.90}$$

which concludes the proof of the lemma. Note that equality in the previous inequality occurs when X_0^T is Poisson with rate λ. Poisson codewords are then optimal here. □

We have now all the ingredients necessary to prove Theorem 7.

Proof of Theorem 7. If R_s is achievable (when the queues are initially in equilibrium) at output rate λ then for all $0 < \epsilon < \frac{1}{2}$ and sufficiently large n, there exists an (n, M, T_n) code (with $\frac{T_n}{n} \to \frac{1}{\lambda}$) such that $\frac{\ln M}{T_n} \geq R_s - \epsilon$, $P_{\text{er}} \leq \epsilon$ and

$\frac{1}{T_n} I(U; Z_0^{T_n}) \leq \epsilon$. We have the following sequence of inequalities

$$\begin{aligned} R_s &\leq \frac{\ln M}{T_n} + \epsilon \\ &\leq \frac{1}{T_n(1-P_{\text{er}})}(I(X_0^{T_n}; Y_0^{T_n}) - I(X_0^{T_n}; Z_0^{T_n}) + I(U; Z_0^{T_n}) + H(P_{\text{er}})) + \epsilon \\ &\leq \frac{1}{1-\epsilon}\left(\log(\frac{\mu_1}{e\lambda})\frac{E[Y_{T_n}]}{T_n} - \log(\frac{\mu_2}{e\lambda})\frac{E[Z_{T_n}]}{T_n} + \frac{1}{T_n}H(Q_0^{(2)}) + \epsilon + \frac{1}{T_n}H(\epsilon)\right) + \epsilon. \end{aligned} \tag{3.91}$$

The second inequality comes from Lemma 12 and the third one from Lemma 14 and the code properties. Recall that both queues are stable since $\lambda < \mu_2 < \mu_1$. From the dynamics of the first queue, we have

$$\frac{E[Y_{T_n}]}{T_n} = \frac{E[Q_0^{(1)}]}{T_n} + \frac{E[X_{T_n}]}{T_n} - \frac{E[Q_{T_n}^{(1)}]}{T_n} \tag{3.92}$$

$$= \frac{1}{T_n}\frac{\rho_1}{1-\rho_1} + \frac{n}{T_n} - \frac{E[Q_{T_n}^{(1)}]}{T_n}. \tag{3.93}$$

As $\frac{E[Q_{T_n}^{(1)}]}{T_n} \to 0$, this shows that $\frac{E[Y_{T_n}]}{T_n} \to \lambda$. Similarly $\frac{E[Z_{T_n}]}{T_n} \to \lambda$. Therefore by letting $\epsilon \to 0$ and $n \to \infty$ in the above inequality yields the desired result, i.e.,

$$R_s \leq \lambda \log\left(\frac{\mu_1}{\mu_2}\right). \tag{3.94}$$

□

3.6 A different proof for the degradedness lemma

A model that is often used to describe the channel dynamics of the exponential server queue is the one that views the queue as an operator that maps a sequence of inter-arrivals into a corresponding sequence of inter-departures. This is the model that was used by Anantharam and Verdú in their seminal paper [41] and adopted afterwards by several other authors (see, e.g., [50]-[53]). Due to the

widespread use of this model and because of its relative simplicity compared to the point process approach used in this chapter, we deemed important to show in this appendix how to prove the degradedness lemma using this model. We believe that this can be potentially useful for researchers in this field. Moreover, this proof should extend easily to the $\cdot/\mathrm{Geom}/1$ queue which is the discrete time analog of the $\cdot/M/1$ queue studied here.

3.6.1 Degradedness when the queues are initially empty

We start by describing the channel dynamics under this model when the queues are initially empty. The sequence of inter-arrival times of the packets will be denoted by $A^n = (A_1, ..., A_n)$. The time of the k-th arrival will be therefore $a_k = \sum_{i=1}^{k} A_i$. The sequence of inter-departures from the legitimate receiver's queue (respectively the eavesdropper's queue) will be denoted by D^n (respectively E^n). Also, the time of the k-th departure from the legitimate receiver's queue (respectively the eavesdropper's queue) will be $d_k = \sum_{i=1}^{k} D_i$ (respectively $e_k = \sum_{i=1}^{k} E_i$).

We can express the kth inter-departure time D_k as follows

$$D_k = W_k + S_{1k}, \tag{3.95}$$

where $W_k = \left(\sum_{j=1}^{k} A_j - \sum_{j=1}^{k-1} D_j\right)^+$ is the kth idle time of the server of the main queue and $\{S_{1k}\}_{k\geq 1}$ is a sequence of i.i.d exponential random variables with parameter μ_1 representing the sequence of service times of the main server. Similarly

$$E_k = V_k + S_{2k}, \tag{3.96}$$

with $V_k = \left(\sum_{j=1}^{k} A_j - \sum_{j=1}^{k-1} E_j\right)^+$ and $\{S_{2k}\}_{k\geq 1}$ is a sequence of i.i.d exponential random variables with parameter μ_2.

The channel transition probability modeling the legitimate queue is

$$P_{D^n|A^n}(y^n|x^n) = \prod_{i=1}^{n} \exp_{\mu_1}(y_i - w_i) \tag{3.97}$$

where $w_i = \left(\sum_{j=1}^{i} x_j - \sum_{j=1}^{i-1} y_j\right)^+$ and $\exp_\mu(y-w)$ denotes

$$\exp_\mu(y-w) = \begin{cases} \mu \exp(-\mu(y-w)), & y \geq w, \\ 0, \quad \text{o.w.} \end{cases}$$

Similarly we have

$$P_{E^n|A^n}(z^n|x^n) = \prod_{i=1}^{n} \exp_{\mu_2}(z_i - v_i) \tag{3.98}$$

with $v_i = \left(\sum_{j=1}^{i} x_j - \sum_{j=1}^{i-1} z_j\right)^+$.

After this brief introduction to the channel model we can now proceed to the proof of the degradedness lemma.

Alternative proof for Lemma 11. We will provide a recursive construction of E^n using only D^n (and some other random variables that are independent of A^n) that will emulate the eavesdropper's channel dynamics (3.98). Let $S_2^n = (S_{21}, ...S_{2n})$ be a sequence of i.i.d. exponential random variables with parameter μ_2 (independent of all other random variables) and $M^n = (M_1, ..., M_n)$ be a sequence of i.i.d. mixed exponential random variables (independent of all other random variables) defined as follows

$$\begin{cases} P[M_1 = 0] = \frac{\mu_2}{\mu_1} \\ P[M_1 > x | M_1 > 0] = \exp(-\mu_2 x). \end{cases}$$

Using these random variables, the recursive construction of E^n is done as follows:

- Let $E_1 = D_1 + M_1$.
- For $k \geq 2$,
  - If $d_k > e_{k-1}$

$$E_k = d_k - e_{k-1} + M_k. \tag{3.99}$$

  - If $d_k \leq e_{k-1}$

$$E_k = S_{2k}. \tag{3.100}$$

This construction can be seen as if E^n were the inter-departures from a special queue whose inter-arrivals are D^n. Indeed, we can write

$$E_k = \left(\sum_{i=1}^{k} D_i - \sum_{i=1}^{k-1} E_i \right)^+ + T_k, \tag{3.101}$$

the first term in this equation represents the k-th idle time of the server of this special queue and T_k is the k-th service time. The caveat here is that the distribution of the service time is not independent of the idle time; indeed, as described above, if the idle time is zero then T_k will be exponentially distributed (with parameter μ_2) otherwise T_k will have a mixed exponential distribution. With this construction we shall prove that $P_{E^n|A^n}(z^n|x^n)$ is given by (3.98) which will prove the degradedness.

From the law of total probability, we have

$$\begin{aligned} P_{E^n|A^n}(z^n|x^n) &= \int_{\mathbb{R}_+^n} P_{E^n,D^n|A^n}(z^n, y^n|x^n)dy^n & (3.102)\\ &= \int_{\mathbb{R}_+^n} P_{E^n|D^n,A^n}(z^n|y^n, x^n)P_{D^n|A^n}(y^n|x^n)dy^n. & (3.103) \end{aligned}$$

From the recursive construction described above, we have $A^n \to D^n \to E^n$ form a Markov chain. It follows that

$$P_{E^n|A^n}(z^n|x^n) = \int_{\mathbb{R}_+^n} P_{E^n|D^n}(z^n|y^n)P_{D^n|A^n}(y^n|x^n)dy^n. \tag{3.104}$$

Also, we can see that

$$P_{E^n|D^n}(z^n|y^n) = \prod_{k=1}^{n} P_{E_k|D^n,E^{k-1}}(z_k|y^n, z^{k-1}) \tag{3.105}$$

$$= \prod_{k=1}^{n} P_{E_k|D^k,E^{k-1}}(z_k|y^k, z^{k-1}), \tag{3.106}$$

where in the last equality we have used the fact that E_k and $(D_{k+1}, ..., D_n)$ are independent given (D^k, E^{k-1}). Consequently,

$$P_{E^n|A^n}(z^n|x^n) = \int_{\mathbb{R}_+^n} \prod_{k=1}^{n} P_{E_k|D^k,E^{k-1}}(z_k|y^k, z^{k-1}) \exp_{\mu_1}(y_k - w_k) dy^n \tag{3.107}$$

$$= \int_{\mathbb{R}_+^{n-1}} \prod_{k=1}^{n-1} P_{E_k|D^k,E^{k-1}}(z_k|y^k, z^{k-1}) \times$$

$$\exp_{\mu_1}(y_k - w_k) I_n(z^n, y^{n-1}) dy^{n-1}, \tag{3.108}$$

where

$$I_n(z^n, y^{n-1}) = \int_{\mathbb{R}_+} P_{E_n|D^n,E^{n-1}}(z_n|y^n, z^{n-1}) \exp_{\mu_1}(y_n - w_n) dy_n \tag{3.109}$$

$$= \mu_1 \int_{w_n}^{\infty} P_{E_n|D^n,E^{n-1}}(z_n|y^n, z^{n-1}) e^{-\mu_1(y_n - w_n)} dy_n. \tag{3.110}$$

We will show that I_n depends on y^{n-1} only through x^{n-1}. Let $\alpha_{n-1} = \sum_{i=1}^{n-1} z_i - \sum_{i=1}^{n-1} y_i$ and recall that by our construction $\alpha_{n-1} \geq 0$. We proceed by noting that

- If $y_n > \alpha_{n-1}$

$$\Pr[E_n \leq z | D^n = y^n, E^{n-1} = z^{n-1}] = \Pr[M_n \leq z + \alpha_{n-1} - y_n]. \tag{3.111}$$

- And if $y_n \leq \alpha_{n-1}$

$$\Pr[E_n \leq z | D^n = y^n, E^{n-1} = z^{n-1}] = \Pr[S_{2n} \leq z]. \tag{3.112}$$

We can then write $P_{E_n|D^n,E^{n-1}}(z_n|y^n, z^{n-1})$ compactly as follows

$$P_{E_n|D^n,E^{n-1}}(z_n|y^n, z^{n-1}) = \mu_2 e^{-\mu_2 z_n}\mathbf{1}_{\{y_n\leq\alpha_{n-1}\}}+ \left[\frac{\mu_2}{\mu_1}\boldsymbol{\delta}(y_n-(z_n+\alpha_{n-1}))+(1-\frac{\mu_2}{\mu_1})\mu_2 e^{-\mu_2(z_n+\alpha_{n-1}-y_n)}\mathbf{1}_{\{y_n<z_n+\alpha_{n-1}\}}\right]\mathbf{1}_{\{\alpha_{n-1}<y_n\}} \tag{3.113}$$

where

$$\boldsymbol{\delta}(t) = \begin{cases} 1 & \text{if } t=0, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\mathbf{1}_C = \begin{cases} 1 & \text{if } C \text{ is true}, \\ 0 & \text{otherwise.} \end{cases}$$

To compute I_n, we need to distinguish between the two cases $\alpha_{n-1} \geq w_n$ and $\alpha_{n-1} \leq w_n$ with $w_n = \left(\sum_{j=1}^{n} x_j - \sum_{j=1}^{n-1} y_j\right)^+$.

- If $\alpha_{n-1} \geq w_n$, I_n is computed as follows

$$\begin{aligned} I_n &= \mu_2 e^{-\mu_1(z_n+\alpha_{n-1}-w_n)} + \mu_1\mu_2(1-\frac{\mu_2}{\mu_1})\int_{\alpha_{n-1}}^{z_n+\alpha_{n-1}} e^{-\mu_1(y_n-w_n)}e^{-\mu_2(z_n+\alpha_{n-1}-y_n)}dy_n \\ &\quad + \mu_2 e^{-\mu_2 z_n}\int_{w_n}^{\alpha_{n-1}} \mu_1 e^{-\mu_1(y_n-w_n)}dy_n \qquad (3.114) \\ &= \mu_2 e^{-\mu_1(z_n+\alpha_{n-1}-w_n)} + \mu_2 e^{-\mu_1(\alpha_{n-1}-w_n)}e^{-\mu_2 z_n}(1-e^{(\mu_2-\mu_1)z_n}) \\ &\quad + \mu_2 e^{-\mu_2 z_n}(1-e^{-\mu_1(\alpha_{n-1}-w_n)}). \qquad (3.115) \end{aligned}$$

  After some cancelations, the last expression reduces to

$$I_n = \mu_2 \exp(-\mu_2 z_n). \tag{3.116}$$

- If $\alpha_{n-1} < w_n$, then I_n can be computed as follows[2]

$$I_n = \mu_2 e^{-\mu_1(z_n+\alpha_{n-1}-w_n)} + \mu_1\mu_2(1-\frac{\mu_2}{\mu_1})\int_{w_n}^{z_n+\alpha_{n-1}} e^{-\mu_1(y_n-w_n)}e^{-\mu_2(z_n+\alpha_{n-1}-y_n)}dy_n \tag{3.117}$$

$$= \mu_2 e^{-\mu_1(z_n+\alpha_{n-1}-w_n)} + \mu_2 e^{-\mu_2(z_n+\alpha_{n-1}-w_n)}(1-e^{(\mu_2-\mu_1)(z_n+\alpha_{n-1}-w_n)}) \tag{3.118}$$

$$= \mu_2 \exp(-\mu_2(z_n-(w_n-\alpha_{n-1}))). \tag{3.119}$$

Notice now that

$$\begin{aligned} w_n - \alpha_{n-1} &= \max(-\alpha_{n-1}, \sum_{i=1}^{n} x_i - \sum_{i=1}^{n-1} y_i - \alpha_{n-1}) \\ &= \max(-\alpha_{n-1}, \sum_{i=1}^{n} x_i - \sum_{i=1}^{n-1} z_i). \end{aligned} \tag{3.120}$$

Since $\alpha_{n-1} \geq 0$ it follows that $w_n \leq \alpha_{n-1}$ if and only if $\sum_{i=1}^{n} x_i \leq \sum_{i=1}^{n-1} z_i$ or equivalently

$$v_n \stackrel{\text{def}}{=} \left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n-1} z_i\right)^+ = 0 \tag{3.121}$$

Similarly, when $w_n > \alpha_{n-1}$, we have

$$w_n - \alpha_{n-1} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n-1} z_i = \left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n-1} z_i\right)^+ = v_n. \tag{3.122}$$

Therefore, we can rewrite I_n written compactly as

$$I_n = \exp_{\mu_2}(z_n - v_n). \tag{3.123}$$

This proves that I_n is only a function of x^n and z^{n-1} (which are fixed quantities here) and not a function of y^{n-1}. The conditional density $P_{E^n|A^n}(z^n|x^n)$ then simplifies to

$$P_{E^n|A^n}(z^n|x^n) = \exp_{\mu_2}(z_n - v_n)\int_{\mathbb{R}_+^{n-1}} \prod_{k=1}^{n-1} P_{E_k|D^k,E^{k-1}}(z_k|y^k, z^{k-1})\exp_{\mu_1}(y_k - w_k)dy^{n-1}. \tag{3.124}$$

[2]Note that the integral in the first line starts from w_n and not from α_{n-1} as in the previous case.

Using the same approach and proceeding inductively, we can deduce that

$$P_{E^n|A^n}(z^n|x^n) = I_1 \prod_{i=2}^{n} \exp_{\mu_2}(z_i - v_i), \tag{3.125}$$

where

$$I_1 = \int_{x_1}^{\infty} \mu_1 e^{-\mu_1(y_1 - x_1)} P_{E_1|D_1}(z_1|y_1) dy_1. \tag{3.126}$$

Since $E_1 = D_1 + M_1$, we have

$$P_{E_1|D_1}(z_1|y_1) = \frac{\mu_2}{\mu_1}\boldsymbol{\delta}(z_1 - y_1) + \mu_2(1 - \frac{\mu_2}{\mu_1})e^{-\mu_2(z_1 - y_1)}\mathbf{1}_{\{y_1 < z_1\}}. \tag{3.127}$$

After substituting this expression in the integral defining I_1 and carrying out the integration we obtain

$$I_1 = \exp_{\mu_2}(z_1 - v_1). \tag{3.128}$$

We conclude that

$$P_{E^n|A^n}(z^n|x^n) = \prod_{i=1}^{n} \exp_{\mu_2}(z_i - v_i), \tag{3.129}$$

which is the channel transition probability of the $./M/1$ queue with service rate μ_2. This proves that when $\mu_2 < \mu_1$, the eavesdropper's channel is stochastically degraded with respect to the legitimate receiver's channel. □

3.6.2 Degradedness when the queues are initially in equilibrium

We start this section by describing the communication protocol used when the queues start in equilibrium. The transmission of information begins at time 0 when the encoder injects a synchronization packet called packet zero. This special packet will find a random number $Q_0^{(1)}$ of packets in the main queue and $Q_0^{(2)}$ packets in the eavesdropper's queue. The departure time D_0 of packet zero

from the main queue is assumed to be observable by the legitimate decoder. Similarly the eavesdropper observes the corresponding departure time E_0 from his own queue. After packet 0, the encoder sends the codeword packets at inter-arrival times A^n and the legitimate receiver observes the inter-departures $D^n = (D_0, ..., D_n)$, this sequence is related to the inter-arrivals through the following channel transition probability [41]

$$P_{D^n|A^n}(y^n|x^n) = \exp_{\mu_1-\lambda}(y_0) \prod_{i=1}^{n} \exp_{\mu_1}(y_i - w_i). \tag{3.130}$$

with $w_i = \left(\sum_{j=1}^{i} x_j - \sum_{j=0}^{i-1} y_j\right)^+$. The eavesdropper similarly observes $E^n = (E_0, ..., E_n)$ and we have

$$P_{E^n|A^n}(z^n|x^n) = \exp_{\mu_2-\lambda}(z_0) \prod_{i=1}^{n} \exp_{\mu_2}(z_i - v_i). \tag{3.131}$$

with $v_i = \left(\sum_{j=1}^{i} x_j - \sum_{j=0}^{i-1} z_j\right)^+$. The proof of the degradedness here is very similar to the proof when the queues start initially empty. We will provide a recursive construction of $E^n = (E_0, ..., E_n)$ using only $D^n = (D_0, ..., D_n)$ and some other random variables that will emulate the eavesdropper's channel dynamics (3.131). From standard results in queuing theory, we know that the departure time D_0 of packet 0 is exponentially distributed with parameter $\mu_1 - \lambda$ [41]. Similarly E_0 is exponentially distributed with parameter $\mu_2 - \lambda$. Let $\widetilde{M}$ be a mixed exponential random variable (independent of all other random variables)

$$\begin{cases} P[\widetilde{M} = 0] = \frac{\mu_2-\lambda}{\mu_1-\lambda} \\ P[\widetilde{M} > x | \widetilde{M} > 0] = \exp(-(\mu_2 - \lambda)x). \end{cases}$$

We can construct E_0 using D_0 and $\widetilde{M}$ as follows

$$E_0 = D_0 + \widetilde{M}. \tag{3.132}$$

Indeed,

$$E[e^{iuE_0}] = E[e^{iuD_0}]E[e^{iu\widetilde{M}}] \tag{3.133}$$

$$= \frac{\mu_1 - \lambda}{\mu_1 - \lambda - iu}\left(\frac{\mu_2 - \lambda}{\mu_1 - \lambda} + \left(1 - \frac{\mu_2 - \lambda}{\mu_1 - \lambda}\right)\frac{\mu_2 - \lambda}{\mu_2 - \lambda - iu}\right) \tag{3.134}$$

$$= \frac{\mu_2 - \lambda}{\mu_1 - \lambda - iu}\left(1 + \frac{\mu_1 - \mu_2}{\mu_2 - \lambda - iu}\right) \tag{3.135}$$

$$= \frac{\mu_2 - \lambda}{\mu_2 - \lambda - iu}, \tag{3.136}$$

which is the characteristic function of an exponential random variable with parameter $\mu_2 - \lambda$. Now let $d_j = \sum_{i=0}^{j} D_i$ be the time of j-th departure from the legitimate receiver's queue (after the departure of packet 0) and analogously define $e_j = \sum_{i=0}^{j} E_i$ for the eavesdropper's queue. As we did in the previous section, we construct $(E_1, ..., E_n)$ recursively using (3.99) and (3.100). We shall prove that through this construction, $P_{E^n|A^n}$ is indeed given by (3.131).

Proceeding similarly to the proof in the case when the queues start empty, we have

$$P_{E^n|A^n}(z^n|x^n) = \int_{\mathbb{R}_+^{n+1}} P_{E^n,D^n|A^n}(z^n, y^n|x^n)dy^{n+1} \tag{3.137}$$

$$= \int_{\mathbb{R}_+^{n+1}} \exp_{\mu_1-\lambda}(y_0)\prod_{k=1}^{n}\exp_{\mu_1}(y_k - w_k)\times$$

$$\prod_{k=0}^{n} P_{E_k|D^k,E^{k-1}}(z_k|y^k, z^{k-1})dy^{n+1}. \tag{3.138}$$

This integral can be written as follows

$$P_{E^n|A^n}(z^n|x^n) = \int_{\mathbb{R}_+} \exp_{\mu_1-\lambda}(y_0)P_{E_0|D_0}(z_0|y_0)Q_n dy_0, \tag{3.139}$$

where

$$Q_n = \int_{\mathbb{R}_+^n}\prod_{k=1}^{n} P_{E_k|D^k,E^{k-1}}(z_k|y^k, z^{k-1})\exp_{\mu_1}(y_k - w_k)dy_1...dy_n. \tag{3.140}$$

We can invoke the result of the previous section and write that Q_n is simply given by

$$Q_n = \prod_{i=1}^{n} \exp_{\mu_2}(z_i - v_i). \tag{3.141}$$

Consequently,

$$P_{E^n|A^n}(z^n|x^n) = \left(\prod_{i=1}^{n} \exp_{\mu_2}(z_i - v_i) \right) I_0 \tag{3.142}$$

where

$$I_0 = \int_{\mathbb{R}_+} \exp_{\mu_1-\lambda}(y_0) P_{E_0|D_0}(z_0|y_0) dy_0. \tag{3.143}$$

If we can show that $I_0 = \exp_{\mu_2-\lambda}(z_0)$, the proof will be complete. Note that

$$P_{E_0|D_0}(z_0|y_0) = \frac{\mu_2 - \lambda}{\mu_1 - \lambda} \boldsymbol{\delta}(z_0 - y_0) + \frac{\mu_1 - \mu_2}{\mu_1 - \lambda} (\mu_2 - \lambda) e^{-(\mu_2-\lambda)(z_0-y_0)} \mathbf{1}\{y_0 < z_0\}. \tag{3.144}$$

Substituting this expression in I_0 we find

$$I_0 = (\mu_2 - \lambda) \left[e^{-(\mu_1-\lambda)z_0} + e^{-(\mu_2-\lambda)z_0} \int_0^{z_0} (\mu_1 - \mu_2) e^{-(\mu_1-\mu_2)y_0} dy_0 \right] \tag{3.145}$$

$$= (\mu_2 - \lambda) \left[e^{-(\mu_1-\lambda)z_0} + e^{-(\mu_2-\lambda)z_0} (1 - e^{-(\mu_1-\mu_2)z_0}) \right] \tag{3.146}$$

$$= (\mu_2 - \lambda) e^{-(\mu_2-\lambda)z_0}. \tag{3.147}$$

This shows that $P_{E^n|A^n}(z^n|x^n)$ is given by (3.131) and completes the proof.

CHAPTER 4

COMPRESSING NEIGHBORS IN A GAUSS-MARKOV TREE

4.1 Introduction

The topic of this chapter falls under the general umbrella of distributed compression of Gaussian sources. We focus here on a problem that can be classified as a special instance of the remote vector source coding problem. The problem considered is depicted in Figure 4.1. The main components are summarized in the bullets below.

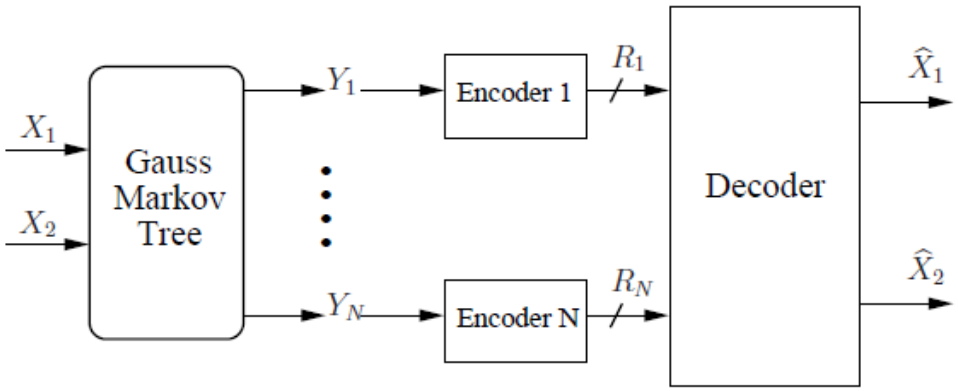

Figure 4.1: Source coding problem

- There are two sources of interest that are the roots of a given Gauss-Markov tree. These sources denoted here by $\{(X_1(t), X_2(t)), t = 1, 2, ...\}$ consist of a stream of i.i.d. jointly Gaussian random variables.
- A given number (say N) of encoders observe the leaves of the tree. Encoder i observes the stream $\{Y_i(t), t = 1, 2...\}$. The encoders are not allowed to cooperate and each one of them compresses its own observation

and sends a message to the decoder over a rate-constrained noiseless channel. The capacity of the channel linking the ith encoder to the decoder will be denoted by R_i.

- The decoder, after receiving the N messages, attempts to reconstruct the sources of interest (X_1, X_2) subject to separate distortion constraints on the time-average quadratic error of the estimate of X_1 and the estimate of X_2. We stress here that we are not trying to reconstruct the vector $\boldsymbol{X} = [X_1, X_2]^T$ subject to a matrix distortion constraint. Instead we constrain only the diagonal elements of the error covariance matrix of the reconstruction of $\boldsymbol{X} = [X_1, X_2]^T$.
- Problem Statement: In this chapter we characterize the minimum sum-rate (minimum of $\sum_{i=1}^{N} R_i$) required to achieve a target pair of distortions (d_1, d_2) on the error of reconstruction of X_1 and X_2. We show that a simple compression architecture that performs separate lossy quantization using Gaussian test channels followed by Slepian-Wolf binning is sum-rate optimal [60]. This scheme is sometimes referred to by some authors as the Berger-Tung (BT) scheme. In the rest of this chapter we will adopt the same terminology.

The approach used here is a more sophisticated version of the argument employed in [54] to obtain the minimum sum-rate for the quadratic Gaussian two-encoder source coding problem. The sum-rate of every code is lower bounded using two different methods. Whereas [54] uses the CEO bound [62], here we use the tree bound [55] which gives the sum-rate required to reconstruct an auxiliary Gaussian random variable X_0 that induces conditional independence between X_1 and X_2. For the other bound, [54] uses a simple cooperative bound, while here we use a recent bound derived by Oohama [56] for the problem of

compressing the vector $\boldsymbol{X} = [X_1, X_2]^T$ subject to a matrix distortion constraint on the error covariance matrix of the reconstruction. For some codes the first method yields a tighter bound whereas for others the second method is better. A composite lower bound is obtained by taking the maximum between these two bounds. We prove that this lower bound equals the minimum sum-rate achieved by the best Berger-Tung scheme when this one achieves both distortions constraints with equality. When only one of the distortion constraints is active for the best Berger-Tung scheme, we show that it is still sum-rate optimal by providing a separate lower bound on the sum-rate.

The determination of the rate region of correlated Gaussian hidden (or remote) sources has received a great deal of attention lately, in part because the remote source coding problem can be instrumental in determining the rate region for other multi-terminal source-coding problems [54], [56], [58]. Both Oohama [56] and Yang and Xiong [59], for instance, consider problems similar to the one treated here. The work closest to this one is Tavildar et al. [55], who show that the BT scheme with Gaussian auxiliary random variables achieves the entire rate region for the problem of reproducing a single variable that can be embedded in a Gauss-Markov tree with the observed variables. This chapter represents the first step in extending that result to the reproduction of a pair of sources that are neighbors in the tree. We are currently studying how to strengthen our result to fully subsume that of [55], by extending the sum-rate result provided here to a rate region result.

We shall see that the assumption that the two variables of interest are neighbors in the tree is crucial for our result. Thus a natural next step would be to study the problem of reproducing three neighboring variables in the tree. By re-

moving the distortion constraint on the middle variable of the triple, one could obtain a result for pairs of variables that are separated by at most one variable in the tree, which in particular would solve the Gaussian one-help-two problem under a tree constraint. The ultimate goal would be to handle distortion constraints on an arbitrary number of variables in the tree.

The rest of this chapter is organized as follows. The formulation of the problem and the main result are described in Section 4.2. Section 4.3 contains the direct part of the argument. The converse is given in Section 4.4. Specifically, in Section 4.4.1 we describe the converse when only one distortion constraint is active for the best BT scheme and in Section 4.4.2 we provide the converse when the best BT scheme achieves both distortion constraints with equality.

4.2 Problem Formulation and Main Result

To simplify the exposition, throughout the chapter we will assume that the Gauss-Markov tree, connecting the sources of interest and the observations of the encoders, is *binary*. This assumption can be made without any loss of generality. Indeed Tavildar et. al. [55] have shown that any Gauss-Markov tree can be transformed into a (potentially) larger binary Gauss-Markov tree. Binary trees have the desired property that they can be described with "minimal" notations which facilitates the presentation of the proof of our main result.

We will slightly change the notation used in the introduction to adopt a notation that is more suitable for the binary tree depicted in Figure 4.2. Since we will sometimes refer to [55], we purposely use the notation used by Tavildar et al. in that paper. The two sources of interest will be now denoted by

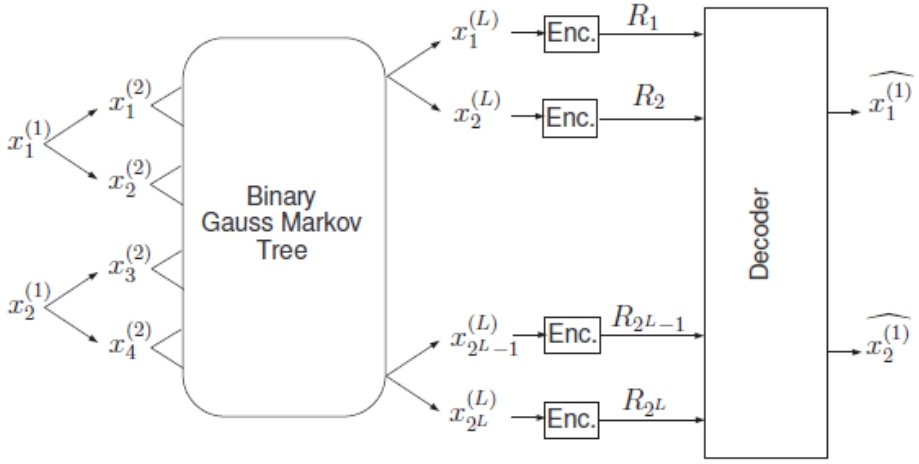

Figure 4.2: Source coding problem with a binary Gauss-Markov tree

$\left\{\boldsymbol{x}^{(1)}(t) = [x_1^{(1)}(t), x_2^{(1)}(t)]^T, t = 1, 2, ...\right\}$. This is a sequence of i.i.d. jointly Gaussian random variables with zero mean and covariance matrix

$$\boldsymbol{K}_{x^{(1)}} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \tag{4.1}$$

For a given tree depth $l \in \{1, ..., L-1\}$ and a given node index $i \in \{1, ..., 2^l\}$, the decedents of the ith node $x_i^{(l)}$ are constructed as follows

$$x_{2i-1}^{(l+1)}(t) = \alpha_{2i-1}^{(l+1)} x_i^{(l)}(t) + z_{2i-1}^{(l+1)}(t) \tag{4.2}$$

$$x_{2i}^{(l+1)}(t) = \alpha_{2i}^{(l+1)} x_i^{(l)}(t) + z_{2i}^{(l+1)}(t), \tag{4.3}$$

where $\{\alpha_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l\}$ is a given sequence of real numbers and $t = 1, 2, ...$ is the time index. The random variables $\{z_i^{(l)}(t), l = 2, ..., L, i = 1, ..., 2^l, t = 1, 2, ...\}$ are independent Gaussian random variables with zero mean and the variance of $z_i^{(l)}(t)$ will be denoted by $\sigma^2_{z_i^{(l)}}$. For the rest of the chapter we will assume that $\sigma^2_{z_i^{(l)}} > 0$ for all i and l. This assumption can be relaxed by invoking the continuity argument used in [55].

For $l = 1, ..., L$ and $i = 1, ..., 2^l$, a block of length n of $x_i^{(l)}$ will be denoted by

$$x_{i,n}^{(l)} = \left(x_i^{(l)}(1), ..., x_i^{(l)}(n) \right). \tag{4.4}$$

After observing $x_{i,n}^{(L)}$, encoder $i \in \{1, ..., 2^L\}$ sends a message to the decoder using a mapping

$$f_i^{(n)} : \mathbb{R}^n \mapsto \left\{ 1, ..., M_i^{(n)} \right\}.$$

The decoder combines all 2^L received messages to form the estimates $\widehat{x_{1,n}^{(1)}}$ and $\widehat{x_{2,n}^{(1)}}$ of $x_{1,n}^{(1)}$ and $x_{2,n}^{(1)}$ using the following mappings

$$\varphi_i^{(n)} : \left\{ 1, ..., M_1^{(n)} \right\} \times ... \times \left\{ 1, ..., M_{2^L}^{(n)} \right\} \mapsto \mathbb{R}^n \quad , i = 1, 2$$

Since the distortion metric that we will use here is the mean square error distortion, we may assume that $\varphi_1^{(n)}$ and $\varphi_2^{(n)}$ are minimum mean square error (MMSE) estimators.

Definition A rate distortion vector $(R_1, ..., R_{2^L}, d_1, d_2)$ is said to be strict-sense achievable if there exists a block length n, encoders $f_i^{(n)}, i = 1, .., 2^L$ and a decoder $(\varphi_1^{(n)}, \varphi_2^{(n)})$ such that

$$R_i \geq \frac{1}{n} \log M_i^{(n)} \quad \text{for} \quad i = 1, ..., 2^L. \tag{4.5}$$

$$d_i \geq \frac{1}{n} \sum_{t=1}^{n} E[(x_i^{(1)}(t) - \widehat{x_i^{(1)}}(t))^2] \quad \text{for} \quad i = 1, 2. \tag{4.6}$$

We denote the set of strict-sense achievable rate-distortion vectors by $\mathcal{RD}^s$. We define the set of achievable rate-distortion vectors to be the closure (denoted by $\overline{\mathcal{RD}^s}$) of $\mathcal{RD}^s$. For a given d_1 and d_2, we define the rate region of this problem to be

$$\mathcal{R}(d_1, d_2) = \left\{ (R_1, ..., R_{2^L}) : (R_1, ..., R_{2^L}, d_1, d_2) \in \overline{\mathcal{RD}^s} \right\}. \tag{4.7}$$

The sum rate of the problem is then

$$R_{\text{sum}}(d_1, d_2) = \inf \left\{ \sum_{i=1}^{2^L} R_i : (R_1, ..., R_{2^L}) \in \mathcal{R}(d_1, d_2) \right\}. \tag{4.8}$$

We will assume here that $\max(d_1, d_2) < \min(1, \rho^2 \min(d_1, d_2) + 1 - \rho^2)$.[1] Requiring that $\max(d_1, d_2) < 1$ should be clear. The condition $\max(d_1, d_2) < \rho^2 \min(d_1, d_2) + 1 - \rho^2$ however warrants further justifications. If this condition was not true, say for instance $d_2 \geq \rho^2 d_1 + 1 - \rho^2$, then the encoders can send information about the source $x_1^{(1)}$ and completely disregard $x_2^{(1)}$. The decoder then reconstructs $x_1^{(1)}$ with a distortion no greater than d_1. Now since we can write $x_2^{(1)} = \rho x_1^{(1)} + z$ for some $z \sim \mathcal{N}(0, 1 - \rho^2)$. The decoder will be able to reconstruct $x_2^{(1)}$ with a distortion less than $\rho^2 d_1 + 1 - \rho^2$ which is in turn less than d_2. Hence, if the condition $\max(d_1, d_2) < \rho^2 \min(d_1, d_2) + 1 - \rho^2$ is violated, the problem can be solved as if one of the sources $x_1^{(1)}$ or $x_2^{(1)}$ was not present. This problem reduces then to the Gaussian many-help-one distributed source coding problem addressed in [55].

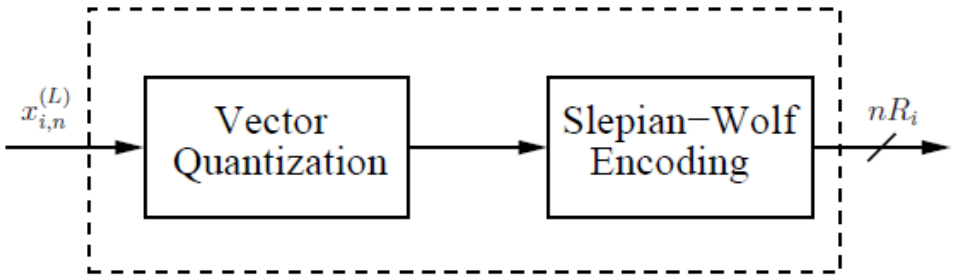

Figure 4.3: Separation scheme

There is a natural scheme that can employed to perform the compression. This method, often called the Berger-Tung scheme, separates the analog and digital aspects of the compression as in Figure 4.3. Specifically, each encoder first performs a vector quantization using a Gaussian test channel. This creates multiple correlated digital messages which can be encoding distributively us-

[1] In fact, not all values of d_1 and d_2 are attainable and so there is an extra condition that must be satisfied by d_1 and d_2. This condition states that there exists some covariance matrix $\boldsymbol{K}_{v^{(L)}}$ such that $d_1 > \boldsymbol{K}_{v^{(L)}}(1,1)$ and $d_2 > \boldsymbol{K}_{v^{(L)}}(2,2)$. This condition will be discussed later in Section 4.4.2.

ing Slepian-Wolf encoding (also known as binning). The minimum sum rate required by this scheme will be denoted by $R_{\mathcal{G}}(d_1, d_2)$. The main result of this chapter is the following theorem.

Theorem 8. *For the source coding problem depicted in Figure 4.1, the Berger-Tung scheme is sum-rate optimal, i.e.,*

$$R_{\text{sum}}(d_1, d_2) = R_{\mathcal{G}}(d_1, d_2). \tag{4.9}$$

The rest of this chapter is dedicated to proving this theorem. In the next section we start with the achievability part of the proof.

4.3 Direct part

The direct part of the main result, i.e., $R_{\mathcal{G}}(d_1, d_2) \geq R_{\text{sum}}(d_1, d_2)$ follows from standard results in information theory [60],[61]. The main goal of this section is to give an explicit characterization of $R_{\mathcal{G}}(d_1, d_2)$ that will be used in the converse. To describe a distributed Gaussian test channel, we associate to encoder $i \in \{1, ..., 2^L\}$ a random variable U_i such that $U_i = x_i^{(L)} + \Gamma_i$ where $\Gamma_i \sim \mathcal{N}(0, \sigma^2_{\Gamma_i})$ is independent of all other random variables. We introduce the following noise quantization rates

$$r_i^{(L)} = I(x_i^{(L)}; U_i | x_{\lceil \frac{i}{2} \rceil}^{(L-1)}) \quad i = 1, ..., 2^L, \tag{4.10}$$

and we construct recursively

$$r_i^{(l)} = f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \quad l = 2, ..., L-1 \quad i = 1, ..., 2^l \tag{4.11}$$

where

$$f_{x_i^{(l)}}(r_1, r_2) = \frac{1}{2} \log \left(1 + \sigma^2_{z_i^{(l)}} [s_{2i-1}^{(l+1)} (1 - 2^{-2r_1}) + s_{2i}^{(l+1)} (1 - 2^{-2r_2})] \right), \tag{4.12}$$

and

$$s_j^{(l+1)} = \left(\frac{\alpha_j^{(l+1)}}{\sigma_{z_j^{(l+1)}}} \right)^2 \quad j = 2i-1, 2i. \tag{4.13}$$

Define also

$$T_1(r_1^{(2)}, r_2^{(2)}) = \sum_{i=1}^{2} s_i^{(2)}(1 - 2^{-2r_i^{(2)}}) \tag{4.14}$$

$$T_2(r_3^{(2)}, r_4^{(2)}) = \sum_{i=3}^{4} s_i^{(2)}(1 - 2^{-2r_i^{(2)}}). \tag{4.15}$$

From (4.11), we can see that T_1 and T_2 are in fact functions of $r_1^{(L)}, ..., r_{2^{L-1}}^{(L)}$ and $r_{2^{L-1}+1}^{(L)}, ..., r_{2^L}^{(L)}$ respectively. Finally, define

$$C_1(r_1^{(L)}, ..., r_{2^{L-1}}^{(L)}) = \frac{(1-\rho^2)[1+(1-\rho^2)T_2(r_3^{(2)}, r_4^{(2)})]}{[1+(1-\rho^2)T_1(r_1^{(2)}, r_2^{(2)})][1+(1-\rho^2)T_2(r_3^{(2)}, r_4^{(2)})] - \rho^2} - d_1$$

$$C_2(r_{2^{L-1}+1}^{(L)}, ..., r_{2^L}^{(L)}) = \frac{(1-\rho^2)[1+(1-\rho^2)T_1(r_1^{(2)}, r_2^{(2)})]}{[1+(1-\rho^2)T_1(r_1^{(2)}, r_2^{(2)})][1+(1-\rho^2)T_2(r_3^{(2)}, r_4^{(2)})] - \rho^2} - d_2$$

$$\Psi(r_1^{(2)}, ..., r_4^{(2)}) = \frac{[1+(1-\rho^2)T_1(r_1^{(2)}, r_2^{(2)})][1+(1-\rho^2)T_2(r_3^{(2)}, r_4^{(2)})] - \rho^2}{1-\rho^2}.$$

The following lemma gives an expression for $R_{\mathcal{G}}(d_1, d_2)$ in terms of an optimization problem.

Lemma 15. *The sum rate achieved by the best Berger-Tung scheme is given by*

$$\begin{aligned} R_{\mathcal{G}}(d_1, d_2) = & \textit{minimize} \quad \frac{1}{2}\log \Psi(r_1^{(2)}, ..., r_4^{(2)}) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\ & \textit{subject to} \quad C_1(r_1^{(L)}, ..., r_{2^{L-1}}^{(L)}) \le 0, \\ & \qquad\qquad C_2(r_{2^{L-1}+1}^{(L)}, ..., r_{2^L}^{(L)}) \le 0 \\ & \qquad\qquad r_i^{(L)} \ge 0, i = 1, ..., 2^L. \end{aligned} \tag{4.16}$$

Proof. See Section 4.5. □

For the rest of this chapter we will use $\boldsymbol{o} = (o_1^{(L)}, ..., o_{2^L}^{(L)})$ to denote a solution to this non-convex optimization problem. It is not clear *a priori* what type of properties this optimizer must possess. It is easy to verify however that at least one of the distortion constraints must be saturated at $\boldsymbol{o}$. Two separate cases must be considered therefore. Case 1: the optimal Gaussian test channel achieves only one distortion constraint with equality (in this case we will assume without loss of generality that only the first constraint is saturated) and case 2: the optimal Gaussian test channel achieves both distortion constraints with equality. Throughout the rest of this chapter we will use the following variables

$$o_i^{(l)} = f_{x_i^{(l)}}(o_{2i-1}^{(l+1)}, o_{2i}^{(l+1)}), \quad l = 2, ..., L-1, i = 1, ..., 2^l, \tag{4.17}$$

and

$$T_1^* = T_1(o_1^{(2)}, o_2^{(2)}) \tag{4.18}$$

$$T_2^* = T_2(o_3^{(2)}, o_4^{(2)}) \tag{4.19}$$

$$C_1^* = C_1(o_1^{(L)}, ..., o_{2^{L-1}}^{(L)}) \tag{4.20}$$

$$C_2^* = C_2(o_{2^{L-1}+1}^{(L)}, ..., o_{2^L}^{(L)}) \tag{4.21}$$

We will see in the next section that we will need a separate conserve for each case. This will require a more elaborate characterization of $\boldsymbol{o}$. This will be established by showing that $\boldsymbol{o}$ satisfies a set of Karush-Kuhn-Tucker (KKT) conditions. For that we need the following technical lemma.

Lemma 16. *The optimal distributed Gaussian test channel $\boldsymbol{o} = (o_1^{(L)}, ..., o_{2^L}^{(L)})$ is regular.*

Proof. To prove this lemma, we show that $\boldsymbol{o}$ satisfies the linear independence constraint qualification [66, Theorem 12.1]. Refer to Section 4.6 for further details. □

Now that we know that $\boldsymbol{o}$ is regular, we move to the description of the KKT conditions satisfied by $\boldsymbol{o}$. The regularity of $\boldsymbol{o}$ implies that for some $(t_1, ..., t_{2^L}) \succeq 0$ and $\gamma_1', \gamma_2' \geq 0$, $\boldsymbol{o}$ must satisfy the following KKT conditions

$$0 = t_i o_i^{(L)} \quad i = 1, ..., 2^L, \tag{4.22a}$$

for $i = 1, ..., 2^{L-1}$:

$$t_i = \frac{C_1^* + d_1}{2\ln(2)} \frac{\partial T_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + 1 + \sum_{l=2}^{L-1} \frac{\partial r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + \gamma_1' \frac{\partial C_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + \gamma_2' \frac{\partial C_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, \tag{4.22b}$$

for $i = 2^{L-1} + 1, ..., 2^L$:

$$t_i = \frac{C_2^* + d_2}{2\ln(2)} \frac{\partial T_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + 1 + \sum_{l=2}^{L-1} \frac{\partial r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + \gamma_1' \frac{\partial C_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + \gamma_2' \frac{\partial C_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, \tag{4.22c}$$

$$0 = \gamma_i' C_i^*, \quad i = 1, 2. \tag{4.22d}$$

These KKT conditions will be examined more thoroughly in the next sections.

4.3.1 Case 1

We start with the case when the optimal Gaussian test channel $\boldsymbol{o}$ achieves only one distortion constraint with equality. In the following, we assume without loss of generality that at $\boldsymbol{o}$ only the first distortion is met with equality. In this case we have $C_1^* = 0$ and $C_2^* < 0$. From (4.22d) we can immediately see that $\gamma_2' = 0$. We would like now to simplify further (4.22b) and (4.22c). Since $C_1^* = 0$,

using (4.192), we have

$$\frac{\partial C_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -d_1^2 \frac{\partial T_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} \quad i = 1, ..., 2^{L-1}. \tag{4.23}$$

Let $\gamma_1 = 2\ln(2)\gamma_1'$, for $i = 1, ..., 2^{L-1}$ we have then

$$t_i = \frac{d_1}{2\ln(2)} \frac{\partial T_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + 1 + \sum_{l=2}^{L-1} \frac{\partial r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + \gamma_1' \frac{\partial C_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} \tag{4.24}$$

$$= 1 + \frac{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} \left(1 + \sum_{l=2}^{L-2} \frac{\partial r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}|_{\boldsymbol{r}=\boldsymbol{o}} + \frac{d_1 - \gamma_1 d_1^2}{2\ln(2)} \frac{\partial T_1}{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}|_{\boldsymbol{r}=\boldsymbol{o}} \right) \tag{4.25}$$

where in the second equality we have used

$$\frac{\partial T_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = \frac{\partial T_1}{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}|_{\boldsymbol{r}=\boldsymbol{o}} \frac{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, \tag{4.26}$$

and

$$\frac{\partial r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = \frac{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} \frac{\partial r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}|_{\boldsymbol{r}=\boldsymbol{o}}. \tag{4.27}$$

These two identities follow from the chain rule for partial derivative combined with (4.11). For a more elaborate explanation refer to Section 4.6. Now since

$$\frac{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = s_i^{(L)} \sigma^2_{z^{(L-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2o_i^{(L)}} 2^{-2o^{(L-1)}_{\lceil \frac{i}{2} \rceil}} \tag{4.28}$$

we can write

$$t_i = \Psi_i^{(L)} \quad i = 1, ..., 2^{L-1}, \tag{4.29}$$

with $\Psi_1^{(L)}, ..., \Psi_{2^{L-1}}^{(L)}$ defined recursively through

$$\Psi_k^{(l)} = 1 + s_k^{(l)} \sigma^2_{z^{(l-1)}_{\lceil \frac{k}{2} \rceil}} 2^{-2o_i^{(l)}} 2^{-2o^{(l-1)}_{\lceil \frac{k}{2} \rceil}} \Psi^{(l-1)}_{\lceil \frac{k}{2} \rceil}, l = 3, ..., L, \quad k = 1, ..., 2^{l-1} \tag{4.30}$$

with

$$\Psi_k^{(2)} = 1 - \frac{\nu_1}{2\ln(2)} \frac{\partial T_1}{\partial r_k^{(2)}}|_{\boldsymbol{r}=\boldsymbol{o}} = 1 - \nu_1 s_k^{(2)} 2^{-2o_k^{(2)}}, \quad k = 1, 2, \tag{4.31}$$

with $\nu_1 \stackrel{\text{def}}{=} \gamma_1 d_1^2 - d_1$.

Now for $i = 2^{L-1}+1, ..., 2^L$, using $C_1^* = 0$, we have

$$\frac{\partial C_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -\frac{\rho^2(1-\rho^2)^2}{((1+(1-\rho^2)T_1^*)(1+(1-\rho^2)T_2^*)-\rho^2)^2}\frac{\partial T_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} \tag{4.32}$$

$$= -\frac{\rho^2 d_1^2}{(1+(1-\rho^2)T_2^*)^2}\frac{\partial T_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}. \tag{4.33}$$

Also

$$C_2^* + d_2 = \frac{d_1(1+(1-\rho^2)T_1^*)}{1+(1-\rho^2)T_2^*} = \frac{1-\rho^2}{1+(1-\rho^2)T_2^*} + \frac{\rho^2 d_1}{(1+(1-\rho^2)T_2^*)^2} \tag{4.34}$$

Consequently, using transformations similar to those used above, (4.22c) becomes

$$t_i = 1 + \frac{\partial r_{\lceil \frac{i}{2} \rceil}^{(L-1)}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} \left(1 + \sum_{l=2}^{L-2} \frac{\partial r_{\lceil \frac{i}{2^{L-l}} \rceil}^{(l)}}{\partial r_{\lceil \frac{i}{2} \rceil}^{(L-1)}}|_{\boldsymbol{r}=\boldsymbol{o}} + \frac{(1-\rho^2)}{1+(1-\rho^2)T_2^*} \times \right.$$

$$\left. \left(1 - \frac{\rho^2}{1-\rho^2}\frac{\nu_1}{1+(1-\rho^2)T_2^*}\right)\frac{1}{2\ln(2)}\frac{\partial T_2}{\partial r_{\lceil \frac{i}{2} \rceil}^{(L-1)}}|_{\boldsymbol{r}=\boldsymbol{o}} \right). \tag{4.35}$$

We can therefore deduce that

$$t_i = \Psi_i^{(L)}, \quad i = 2^{L-1}+1, ..., 2^L \tag{4.36}$$

with $\Psi_{2^{L-1}+1}^{(L)}, ..., \Psi_{2^L}^{(L)}$ defined recursively as follows

$$\Psi_k^{(l)} = 1 + s_k^{(l)} \sigma^2_{z_{\lceil \frac{k}{2} \rceil}^{(l-1)}} 2^{-2o_k^{(l)}} 2^{-2o_{\lceil \frac{k}{2} \rceil}^{(l-1)}} \Psi_{\lceil \frac{k}{2} \rceil}^{(l-1)}, l = 3, ..., L, \quad k = 2^{l-1}+1, ..., 2^l \tag{4.37}$$

and

$$\Psi_k^{(2)} = 1 + \frac{(1-\rho^2)s_k^{(2)}2^{-2o_k^{(2)}}}{1+(1-\rho^2)T_2^*}\left(1 - \frac{\rho^2}{1-\rho^2}\frac{\nu_1}{1+(1-\rho^2)T_2^*}\right), \quad k = 3, 4. \tag{4.38}$$

4.3.2 Case 2

Here we assume that the optimal Gaussian test channel $\boldsymbol{o}$ achieves both distortion constraints with equality. In this situation we have $C_1^* = C_2^* = 0$ and (see

Section 4.6)

$$\frac{\partial C_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -d_1^2 \frac{\partial T_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, \quad i = 1, ..., 2^{L-1} \tag{4.39}$$

$$\frac{\partial C_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -d_1 d_2 (\theta^*)^2 \frac{\partial T_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, \quad i = 2^{L-1}+1, ..., 2^L. \tag{4.40}$$

and

$$\frac{\partial C_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -d_1 d_2 (\theta^*)^2 \frac{\partial T_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, \quad i = 1, ..., 2^{L-1} \tag{4.41}$$

$$\frac{\partial C_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -d_2^2 \frac{\partial T_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, \quad i = 2^{L-1}+1, ..., 2^L. \tag{4.42}$$

Replacing these identities in (4.22b) and (4.22c) we find

$$t_i = 1 - \frac{\nu_1}{2\ln(2)} \frac{\partial T_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + \sum_{l=2}^{L-1} \frac{\partial r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, \quad i = 1, ..., 2^{L-1} \tag{4.43}$$

$$t_i = 1 - \frac{\nu_2}{2\ln(2)} \frac{\partial T_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} + \sum_{l=2}^{L-1} \frac{\partial r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, \quad i = 2^{L-1}+1, ..., 2^L \tag{4.44}$$

where $\gamma_i = 2\ln(2)\gamma_i', i = 1, 2$ and

$$\nu_1 = \gamma_1 d_1^2 + \gamma_2 d_1 d_2 (\theta^*)^2 - d_1, \tag{4.45}$$

$$\nu_2 = \gamma_2 d_2^2 + \gamma_1 d_1 d_2 (\theta^*)^2 - d_2. \tag{4.46}$$

Proceeding as in case 1, we find that

$$t_i = \Psi_i^{(L)}, \quad i = 1, ..., 2^L \tag{4.47}$$

where $\Psi_i^{(L)}$ is defined recursively through (4.30) and (4.37) but with

$$\Psi_k^{(2)} = 1 - \nu_1 s_k^{(2)} 2^{-2o_k^{(2)}}, \quad k = 1, 2 \tag{4.48}$$

$$\Psi_k^{(2)} = 1 - \nu_2 s_k^{(2)} 2^{-2o_k^{(2)}}, \quad k = 3, 4. \tag{4.49}$$

4.4 Converse

4.4.1 Case 1 Converse

Theorem 9. *When the optimal Gaussian test channel achieves only the first distortion with equality we have*

$$R_{\text{sum}}(d_1, d_2) \geq R_{\mathcal{G}}(d_1, d_2). \tag{4.50}$$

Proof. The converse in case 1 is based on the outer bound derived in [55]. Let $\boldsymbol{C} = (C_1, ..., C_{2^L})$ designate all the messages sent to the decoder. Then, using the conditional independence relations imposed by the tree structure and the chain rule for mutual information we obtain the following sequence of identities

$$\sum_{i=1}^{2^L} R_i \geq \frac{1}{n} H(\boldsymbol{C}) \tag{4.51}$$

$$= \frac{1}{n} I(\boldsymbol{C}; \{x_{i,n}^{(l)}, l = 1, ..., L, i = 1, ..., 2^l\}) \tag{4.52}$$

$$= \frac{1}{n} I(\boldsymbol{C}; x_{1,n}^{(1)}) + \frac{1}{n} I(C_{2^{L-1}+1}, ..., C_{2^L}; x_{2,n}^{(1)} | x_{1,n}^{(1)})$$

$$+ \sum_{l=2}^{L} \sum_{i=1}^{2^l} \frac{1}{n} I(\boldsymbol{C}_{\mathcal{O}(x_i^{(l)})}; x_{i,n}^{(l)} | x_{\lceil \frac{i}{2} \rceil, n}^{(l-1)}) \tag{4.53}$$

$$\geq \frac{1}{2} \log \frac{1}{d_1} + r + \sum_{l=2}^{L} \sum_{i=1}^{2^l} r_i^{(l)}, \tag{4.54}$$

where in the last step we have used the classical inequality[2]

$$\frac{1}{n} I(\boldsymbol{C}; x_{1,n}^{(1)}) \geq \frac{1}{2} \log \frac{1}{d_1}$$

and defined

$$r_i^{(l)} \stackrel{\text{def}}{=} \frac{1}{n} I(\boldsymbol{C}_{\mathcal{O}(x_i^{(l)})}; x_{i,n}^{(l)} | x_{\lceil \frac{i}{2} \rceil, n}^{(l-1)})$$

[2]See page 575 in [55].

and

$$r \stackrel{\text{def}}{=} \frac{1}{n} I(C_{2^{L-1}+1}, ..., C_{2^L}; x_{2,n}^{(1)} | x_{1,n}^{(1)}).$$

Since $x_{2,n}^{(1)}, x_{1,n}^{(2)}, x_{2,n}^{(2)}$ are conditionally independent given $x_{1,n}^{(1)}$, using [62, Lemmas 2 and 3] we can obtain the following inequality

$$\frac{1}{d_1} \leq 2^{\frac{2}{n} I(\boldsymbol{C}; x_{1,n}^{(1)})} \tag{4.55}$$

$$\leq 1 + \sum_{i=1}^{2} s_i^{(2)} \left(1 - 2^{-2r_i^{(2)}}\right) + \frac{\rho^2}{1-\rho^2}(1 - 2^{-2r}). \tag{4.56}$$

Note that we can write $x_2^{(1)}$ as $x_2^{(1)} = \rho x_1^{(1)} + z$ for some $z \sim \mathcal{N}(0, 1-\rho^2)$ that is independent of $x_1^{(1)}$. Since for $i = 3, 4$ we have $x_i^{(2)} = \alpha_i^{(2)} x_2^{(1)} + z_i^{(2)}$ with $z_i^{(2)}$ independent of $x_1^{(1)}$ we can use Lemma 4 in [55] to guarantee that $r, r_3^{(2)}, r_4^{(2)}$ satisfy the following inequality

$$r \leq \frac{1}{2} \log \left(1 + (1-\rho^2) \sum_{i=3}^{4} s_i^{(2)} (1 - 2^{-2r_i^{(2)}})\right). \tag{4.57}$$

Also from Lemma 4 in [55], for all $l = 2, ..., L-1$ and $i = 1, ..., 2^l$, we have

$$r_i^{(l)} \leq f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \tag{4.58}$$

where $f_{x_i^{(l)}}$ is the function defined in (4.12). Consequently, we arrive to the following lower bound on the sum rate

$$R_{\text{sum}}(d_1, d_2) \geq \mathcal{L}_{\text{sum}}, \tag{4.59}$$

where $\mathcal{L}_{\text{sum}}$ is defined as follows

$$
\begin{aligned}
\mathcal{L}_{\text{sum}} = \text{minimize} \quad & \frac{1}{2}\log\frac{1}{d_1} + r + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\
\text{subject to} \quad & r \le \frac{1}{2}\log\left(1 + (1-\rho^2)\sum_{i=3}^{4} s_i^{(2)}(1-2^{-2r_i^{(2)}})\right), \\
& \frac{1}{d_1} \le 1 + \sum_{i=1}^{2} s_i^{(2)}\left(1-2^{-2r_i^{(2)}}\right) + \frac{\rho^2}{1-\rho^2}(1-2^{-2r}), \\
& r_i^{(l)} \le f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \quad l = 2, ..., L-1 \quad i = 1, ..., 2^l, \\
& r, r_i^{(l)} \ge 0 \quad l = 2, ..., L \quad i = 1, ..., 2^l.
\end{aligned}
$$

The optimization problem above is convex and the KKT conditions here are both necessary and sufficient. These conditions read as follows, $(r, r_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l)$ is optimal if and only if there exists $(\nu, \eta, \kappa, \kappa_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l) \succeq$

0 and $(\tau_i^{(l)}, l = 2, ..., L-1, i = 1, ..., 2^l) \succeq 0$ such that[3]

$$0 = 1 - \kappa + \nu - \eta \frac{\rho^2}{1-\rho^2} 2^{-2r}, \tag{4.60a}$$

$$0 = 1 - \kappa_i^{(2)} + \tau_i^{(2)} - \eta s_i^{(2)} 2^{-2r_i^{(2)}}, \quad i = 1, 2 \tag{4.60b}$$

$$0 = 1 - \kappa_i^{(2)} + \tau_i^{(2)} - \nu \frac{(1-\rho^2) s_i^{(2)} 2^{-2r_i^{(2)}}}{1 + (1-\rho^2) \sum_{i=3}^{4} s_i^{(2)} (1 - 2^{-2r_i^{(2)}})}, \quad i = 3, 4 \tag{4.60c}$$

$$0 = 1 - \kappa_i^{(l)} + \tau_i^{(l)} - \tau_{\lceil \frac{i}{2} \rceil}^{(l-1)} s_i^{(l)} \sigma^2_{z^{(l-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2r_i^{(l)}} 2^{-2f_{x^{(l-1)}_{\lceil \frac{i}{2} \rceil}}}, \quad l = 3, ..., L-1, i = 1, ..., 2^l \tag{4.60d}$$

$$0 = 1 - \kappa_i^{(L)} - \tau_{\lceil \frac{i}{2} \rceil}^{(L-1)} s_i^{(L)} \sigma^2_{z^{(L-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2r_i^{(L)}} 2^{-2f_{x^{(L-1)}_{\lceil \frac{i}{2} \rceil}}}, \quad i = 1, ..., 2^L \tag{4.60e}$$

$$0 = \kappa r, \quad 0 = \kappa_i^{(l)} r_i^{(l)}, \quad l = 2, ..., L, i = 1, ..., 2^l \tag{4.60f}$$

$$0 = \nu \left(r - \frac{1}{2} \log \left(1 + (1-\rho^2) \sum_{i=3}^{4} s_i^{(2)} (1 - 2^{-2r_i^{(2)}}) \right) \right) \tag{4.60g}$$

$$0 = \eta \left(\frac{1}{d_1} - 1 - \sum_{i=1}^{2} s_i^{(2)} \left(1 - 2^{-2r_i^{(2)}} \right) - \frac{\rho^2}{1-\rho^2} (1 - 2^{-2r}) \right) \tag{4.60h}$$

$$0 = \tau_i^{(l)} \left(r_i^{(l)} - f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}) \right), \quad l = 2, ..., L-1 \quad i = 1, ..., 2^l \tag{4.60i}$$

An examination of this system shows that the Lagrange multiplier η must be positive. Indeed, if $\eta = 0$, then from (4.60a) and (4.60b) we will have $\kappa, \kappa_1^{(2)}, \kappa_2^{(2)} > 0$ which in turn means that $r = r_1^{(2)} = r_2^{(2)} = 0$. This is not possible since $d_1 < 1$. We will now identify a solution to this system using the optimal Gaussian test channel $\boldsymbol{o} = (o_1^{(L)}, ..., o_{2^L}^{(L)})$ (which satisfies the system of equations in (4.22)). Define $o^* = \frac{1}{2} \log (1 + (1-\rho^2) T_2^*)$. Using the characterization of $\boldsymbol{o}$ given by (4.22), we can verify that by setting $r = o^*$ and $r_i^{(l)} = o_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l$ the KKT conditions in (4.60) are satisfied with

[3]The primal feasibility conditions have been omitted. This will be usually the norm throughout the chapter.

an appropriate choice of KKT multiplies. For this choice, first set

$$\kappa = 0, \tag{4.61}$$

$$\eta = \nu_1, \tag{4.62}$$

$$\nu = \frac{\rho^2}{1-\rho^2}\nu_1 2^{-2o^*} - 1. \tag{4.63}$$

Then $1 + \nu = \eta\frac{\rho^2}{1-\rho^2}2^{-2o^*}$ and therefore (4.60a) is satisfied. However, since ν must be nonnegative, in order for this coupling to be valid we must have $\left(1 - \frac{\rho^2}{1-\rho^2}\nu_1 2^{-2o^*}\right) \leq 0$.[4] In fact we will establish that

$$1 - \frac{\rho^2}{1-\rho^2}\nu_1 2^{-2o^*} < 0.$$

To show this inequality we give a proof by contradiction. Assume that $1 - \frac{\rho^2}{1-\rho^2}\nu_1 2^{-2o^*} \geq 0$, then we must have $\Psi_3^{(2)}, \Psi_4^{(2)} > 0$ which in turn gives that $t_i = \Psi_i^{(L)} > 0$ for $i = 2^{L-1} + 1, ..., 2^L$. By (4.22a) it follows then that $o_i^{(L)} = 0$ for $i = 2^{L-1} + 1, ..., 2^L$. Recalling that $o_i^{(l)} = f_{x_i^{(l)}}(o_{2i-1}^{(l+1)}, o_{2i}^{(l+1)})$, we see that $o_i^{(l)} = 0$ for $l = 2, ..., L-1$ and $i = 2^{l-1} + 1, ..., 2^l$. In particular $o_3^{(2)} = o_4^{(2)} = 0$ and $T_2^* = 0$. The condition $C_1^* = 0$ implies then

$$d_1(1 + (1-\rho^2)T_1^*) = \frac{\rho^2 d_1}{1 + (1-\rho^2)T_2^*} + 1 - \rho^2 = \rho^2 d_1 + 1 - \rho^2. \tag{4.64}$$

However the second distortion constraint is inactive at $\boldsymbol{o}$, i.e., $C_2^* < 0$ or equivalently

$$d_1(1 + (1-\rho^2)T_1^*) < d_2(1 + (1-\rho^2)T_2^*) \tag{4.65}$$

$$= d_2. \tag{4.66}$$

We conclude that

$$\rho^2 d_1 + (1-\rho^2) < d_2. \tag{4.67}$$

[4]Note that this will in turn prove that $\nu_1 \geq 0$ and so the choice $\eta = \nu_1$ will be also valid.

This contradicts however our initial assumption that $d_2 \leq \rho^2 d_1 + (1-\rho^2)$. Therefore $1 - \frac{\rho^2}{1-\rho^2}\nu_1 2^{-2o^*} < 0$ and the coupling above is legitimate.

We continue now the task of showing that o^* and $o_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l$ satisfy the KKT conditions in (4.60). It should be clear to the reader that (4.60g)-(4.60i) are satisfied. For $i = 1, ..., 4$, set

$$\tau_i^{(2)} = \max(0, -\Psi_i^{(2)}), \tag{4.68}$$

$$\kappa_i^{(2)} = \max(0, \Psi_i^{(2)}). \tag{4.69}$$

Then clearly (4.60b) and (4.60c) are satisfied with this choice. Before tackling (4.60d), we consider (4.60e) first. Let

$$\tau_i^{(L-1)} = \max(0, -\Psi_i^{(L-1)}), \quad i = 1, ..., 2^{L-1} \tag{4.70}$$

$$\kappa_i^{(L)} = \min(1, t_i) \quad i = 1, ..., 2^L. \tag{4.71}$$

As $t_i = \Psi_i^{(L)}$, we can verify using (4.30) and (4.37) that (4.60e) holds with this choice. Now for (4.60d), we set

$$\tau_i^{(l)} = \max(0, -\Psi_i^{(l)}), \quad l = 3, ..., L-2, i = 1, ..., 2^l \tag{4.72}$$

$$\kappa_i^{(l)} = \max(1, 1 - \Psi_i^{(l)}) - \max(0, 1 - \Psi_i^{(l)}) \quad l = 3, ..., L-1 \quad i = 1, ..., 2^l. \tag{4.73}$$

Then

$$\begin{aligned} 1 + \tau_i^{(l)} - \tau_{\lceil \frac{i}{2} \rceil}^{(l-1)} s_i^{(l)} \sigma^2_{z^{(l-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2o_i^{(l)}} 2^{-2o^{(l-1)}_{\lceil \frac{i}{2} \rceil}} &= \max(1, 1 - \Psi_i^{(l)}) \\ &\quad - \max(0, -s_i^{(l)} \sigma^2_{z^{(l-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2o_i^{(l)}} 2^{-2o^{(l-1)}_{\lceil \frac{i}{2} \rceil}} \Psi^{(l)}_{\lceil \frac{i}{2} \rceil}) \end{aligned} \tag{4.74}$$

$$= \max(1, 1 - \Psi_i^{(l)}) - \max(0, 1 - \Psi_i^{(l)}) \tag{4.75}$$

$$= \kappa_i^{(l)}, \tag{4.76}$$

which shows that (4.60d) is verified with this choice. It remains to check the complementary slackness conditions $\kappa_i^{(l)} o_i^{(l)} = 0$ for $l = 2, ..., L, i = 1, ..., 2^l$. For $l = 2$, we will make the verification for $i = 1$ only, the rest of the cases are similar. If $\Psi_1^{(2)} \leq 0$, then $\kappa_1^{(2)} = 0$ and a fortiori $\kappa_1^{(2)} o_1^{(2)} = 0$. Assume that $\Psi_1^{(2)} > 0$, then from the definition of $\Psi^{(L)}$ we have $\Psi_j^{(L)} > 0$ for $j = 1, ..., 2^{L-2}$ or equivalently $t_j > 0$ for $j = 1, ..., 2^{L-2}$ which implies that $o_j^{(L)} = 0$ for $j = 1, ..., 2^{L-2}$. Recalling that $o_i^{(l)} = f_{x_i^{(l)}}(o_{2i-1}^{(l+1)}, o_{2i}^{(l+1)})$, we see that $o_i^{(l)} = 0$ for $l = 2, ..., L-1$ and $i = 1, ..., 2^{l-2}$. In particular $o_1^{(2)} = 0$ and hence $\kappa_1^{(2)} o_1^{(2)} = 0$.

Let now $l \in \{3, ..., L-1\}$, if $\Psi_i^{(l)} \leq 0$, then $\kappa_i^{(l)} = 0$ and $\kappa_i^{(l)} o_i^{(l)} = 0$. If $\Psi_i^{(l)} > 0$, then an argument similar to the one above shows that $o_i^{(l)} = 0$ and hence $\kappa_i^{(l)} o_i^{(l)} = 0$. Finally, for $l = L$, we have

$$\kappa_i^{(L)} o_i^{(L)} = \min(o_i^{(L)}, o_i^{(L)} t_i) = \min(o_i^{(L)}, 0) = 0. \tag{4.77}$$

This establishes the optimality of $(o^*, o_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l)$, i.e.,

$$\begin{aligned}
\mathcal{L}_{\text{sum}} &= \frac{1}{2} \log \frac{1}{d_1} + o^* + \sum_{l=2}^{L} \sum_{i=1}^{2^l} o_i^{(l)} \\
&= \frac{1}{2} \log \frac{1}{d_1} + \frac{1}{2} \log \left(1 + (1-\rho^2) T_2^*\right) + \sum_{l=2}^{L} \sum_{i=1}^{2^l} o_i^{(l)} \\
&\overset{(a)}{=} \frac{1}{2} \log \left(\frac{[1 + (1-\rho^2) T_1^*][1 + (1-\rho^2) T_2^*] - \rho^2}{1-\rho^2} \right) + \sum_{l=2}^{L} \sum_{i=1}^{2^l} o_i^{(l)} \\
&= R_{\mathcal{G}}(d_1, d_2),
\end{aligned}$$

where (a) follows from the fact that $C_1^* = 0$. This shows that $R_{\text{sum}}(d_1, d_2) \geq R_{\mathcal{G}}(d_1, d_2)$ and completes the proof of the converse in this case. □

4.4.2 Case 2 Converse

The following technical lemma will be needed to prove the converse.

Lemma 17. *Let*

$$f(x) = \rho^2[\nu_1 + (1-x)(1+\lambda_1\nu_1)][1+(1-x)\lambda_1], \tag{4.78}$$

$$g(x) = [x + (x-\rho^2)\lambda_2][\nu_2 x + (x-\rho^2)(1+\lambda_2\nu_2)]. \tag{4.79}$$

Then, if the optimal Gaussian test channel ***o*** *achieves both distortions with equality, the equation* $f(x) = g(x)$ *has a unique root in the interval* $[\rho^2, 1]$*. Moreover this root is given by*

$$x^* = \rho\frac{(1+\lambda_1)\sqrt{\nu_1+d_1} + \rho\lambda_2\sqrt{\nu_2+d_2}}{(1+\lambda_2)\sqrt{\nu_2+d_2} + \rho\lambda_1\sqrt{\nu_1+d_1}}, \tag{4.80}$$

Proof. Refer to Section 4.7. □

Theorem 10. *When the optimal Gaussian test channel achieves both distortions with equality we have*

$$R_{\text{sum}}(d_1, d_2) \geq R_{\mathcal{G}}(d_1, d_2). \tag{4.81}$$

Proof. The converse argument will depend on the value of x^*. We will distinguish two different cases: $x^* \in \{\rho^2, 1\}$ and $\rho^2 < x^* < 1$.

The case $\mathbf{x^* \in \{\rho^2, 1\}}$

We will consider only the case $x^* = 1$, the case $x^* = \rho^2$ can be solved along the same lines. From the proof of Theorem 9, we know that

$$R_{\text{sum}}(d_1, d_2) \geq \mathcal{L}_{sum}. \tag{4.82}$$

When $x^* = 1$, we will show that $R_{\mathcal{G}}(d_1, d_2) = \mathcal{L}_{sum}$. To establish this a similar approach to the one used in the previous section will be employed. Define first

$$o^* = \frac{1}{2}\log\left(1 + (1-\rho^2)T_2^*\right) = \frac{1}{2}\log(1 + (1-\rho^2)\lambda_2). \tag{4.83}$$

Using the characterization of $\boldsymbol{o}$, we can verify that by setting $r = o^*$ and $r_i^{(l)} = o_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l$ the KKT conditions in (4.60) are satisfied with an appropriate choice of parameters. Set

$$\kappa = 0 \tag{4.84}$$

$$\eta = \nu_1 \tag{4.85}$$

$$\nu = \frac{\nu_2(1 + (1-\rho^2)\lambda_2)}{1-\rho^2}. \tag{4.86}$$

We start by verifying (4.60a). Since, $f(1) = g(1)$, we have

$$\rho^2\nu_1 = (1 + (1-\rho^2)\lambda_2)(1 - \rho^2 + \nu_2(1 + (1-\rho^2)\lambda_2)). \tag{4.87}$$

This means that

$$1 + \nu = 1 + \frac{\nu_2(1 + (1-\rho^2)\lambda_2)}{1-\rho^2}, \tag{4.88}$$

$$= \frac{\rho^2}{1-\rho^2}\frac{\nu_1}{1 + (1-\rho^2)\lambda_2}, \tag{4.89}$$

$$= \eta\frac{\rho^2}{1-\rho^2}2^{-2o^*}, \tag{4.90}$$

which shows that (4.60a) holds with this choice. However, since η and ν must be nonnegative, for the above choice to be valid we need to have $\nu_1, \nu_2 \geq 0$. This is not hard to verify and in fact $\nu_1, \nu_2 > 0$. Indeed, in Section 4.6, we have shown that there must exist an $i^* \in \{1, ..., 2^{L-1}\}$ and a $j^* \in \{2^{L-1} + 1, ..., 2^L\}$ such that $o_{i^*}^{(L)}, o_{j^*}^{(L)} > 0$. In other words there must exist an $i^* \in \{1, ..., 2^{L-1}\}$ and a $j^* \in \{2^{L-1} + 1, ..., 2^L\}$ such that $t_{i^*} = t_{j^*} = 0$. This is only possible when $\nu_1, \nu_2 > 0$. Because if for instance $\nu_1 \leq 0$, then from the definition of $\Psi^{(L)}$ we must have $\Psi_i^{(L)} > 0$ for $i = 1, ..., 2^{L-1}$, i.e., $t_i > 0$ for $i = 1, ..., 2^{L-1}$.

For the rest of the KKT parameters, the selection is exactly similar to the one described in case 1, i.e.,

$$\tau_i^{(l)} = \max(0, -\Psi_i^{(l)}), \quad l = 2, ..., L-1, i = 1, ..., 2^l \tag{4.91}$$

$$\kappa_i^{(2)} = \max(0, \Psi_i^{(2)}), \quad i = 1, ..., 4 \tag{4.92}$$

$$\kappa_i^{(L)} = \min(1, t_i) \quad i = 1, ..., 2^L \tag{4.93}$$

$$\kappa_i^{(l)} = \max(1, 1 - \Psi_i^{(l)}) - \max(0, 1 - \Psi_i^{(l)}), \quad l = 3, ..., L-1, i = 1, ..., 2^l. \tag{4.94}$$

The interested reader can check that with this choice the vector $(o^*, o_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l)$ satisfies the system of equations in (4.60). This establishes the optimality of $(o^*, o_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l)$, i.e.,

$$\begin{aligned}
\mathcal{L}_{\text{sum}} &= \frac{1}{2}\log\frac{1}{d_1} + o^* + \sum_{l=2}^{L}\sum_{i=1}^{2^l} o_i^{(l)} \\
&\overset{(a)}{=} \frac{1}{2}\log\frac{1-\rho^2}{d_1 d_2(1-(\theta^*)^2)} + \sum_{l=2}^{L}\sum_{i=1}^{2^l} o_i^{(l)} \\
&= R_{\mathcal{G}}(d_1, d_2),
\end{aligned}$$

where in (a) we have used the fact that (see Appendices B and C) $\frac{1+(1-\rho^2)\lambda_2}{d_1} = \frac{1-\rho^2}{d_1 d_2(1-(\theta^*)^2)}$. This shows that $R_{\text{sum}}(d_1, d_2) \geq R_{\mathcal{G}}(d_1, d_2)$ and completes the proof of the converse in this case.

The case $\rho^2 < \mathbf{x}^* < 1$

We will provide a lower bound on the sum rate $R_{\text{sum}}(d_1, d_2)$ and then we will show that this lower bound equals $R_{\mathcal{G}}(d_1, d_2)$ when $\boldsymbol{o}$ achieves both distortion constraints. The lower bound that we use to prove the converse is a combination of two other lower bounds derived in [55] and [56].

Assume $(R_1, ..., R_{2^L}, d_1, d_2)$ is strict sense achievable, then there exist encoders $f_i^{(n)}, i = 1, .., 2^L$ and a decoder $(\varphi_1^{(n)}, \varphi_2^{(n)})$ satisfying (4.5) and (4.6). This code will have some error covariance matrix $\widehat{\boldsymbol{D}}$

$$\widehat{\boldsymbol{D}} = \frac{1}{n}\sum_{t=1}^{n}\widehat{\boldsymbol{D}}(t), \tag{4.95}$$

where

$$\widehat{\boldsymbol{D}}(t) = E\left[\left(\boldsymbol{x}^{(1)}(t) - \widehat{\boldsymbol{x}^{(1)}}(t)\right)\left(\boldsymbol{x}^{(1)}(t) - \widehat{\boldsymbol{x}^{(1)}}(t)\right)^T\right]. \tag{4.96}$$

Tree Bound Define $(\alpha_1^{(1)}, \alpha_2^{(1)})$ as follows

$$\alpha_1^{(1)} = \sqrt{x^*} = \sqrt{\rho\frac{(1+\lambda_1)\sqrt{\nu_1+d_1} + \rho\lambda_2\sqrt{\nu_2+d_2}}{(1+\lambda_2)\sqrt{\nu_2+d_2} + \rho\lambda_1\sqrt{\nu_1+d_1}}}, \tag{4.97}$$

$$\alpha_2^{(1)} = \frac{\rho}{\alpha_1^{(1)}}. \tag{4.98}$$

With this choice we have $\alpha_1^{(1)}, \alpha_2^{(1)} \in (\rho, 1)$. Consider now three independent zero mean Gaussian random variables $x^{(0)}$, $z_1^{(1)}$ and $z_2^{(1)}$ with respective variance $\sigma^2_{x^{(0)}} = 1$, $\sigma^2_{z_1^{(1)}} = 1 - (\alpha_1^{(1)})^2$ and $\sigma^2_{z_2^{(1)}} = 1 - (\alpha_2^{(1)})^2$. It is easy to check that the covariance matrix of $(\alpha_1^{(1)}x^{(0)} + z_1^{(1)}, \alpha_2^{(1)}x^0 + z_2^{(1)})$ is $\boldsymbol{K}_{x^{(1)}}$. Therefore we can couple these variables to $\left(x_1^{(1)}, x_2^{(1)}\right)$ using

$$x_1^{(1)} = \alpha_1^{(1)}x^{(0)} + z_1^{(1)} \tag{4.99}$$

$$x_2^{(1)} = \alpha_2^{(1)}x^{(0)} + z_2^{(1)} \tag{4.100}$$

This auxiliary random variable $x^{(0)}$ can be written in terms of $(x_1^{(1)}, x_2^{(1)})$ as follows

$$x^{(0)} = \mu_1 x_1^{(1)} + \mu_2 x_2^{(1)} + z^{(0)}, \tag{4.101}$$

where $z^{(0)}$ is a Gaussian random variable independent of $x_1^{(1)}$ and $x_2^{(1)}$ with variance $\sigma^2_{z^{(0)}}$

$$\sigma^2_{z^{(0)}} = \left(1 + \left(\frac{\alpha_1^{(1)}}{\sigma_{z_1^{(1)}}}\right)^2 + \left(\frac{\alpha_2^{(1)}}{\sigma_{z_2^{(1)}}}\right)^2\right)^{-1}, \tag{4.102}$$

and

$$\mu_i = \frac{\sigma^2_{z^{(0)}}}{\sigma^2_{z_i^{(1)}}} \alpha_i^{(1)} \quad i = 1, 2. \tag{4.103}$$

It is easy to verify that $z^{(0)}$ is also independent of $x_1^{(L)}, ..., x_{2^L}^{(L)}$ and a fortiori independent of the messages $C_1, ..., C_{2^L}$ sent by the encoders. Using $\widehat{\boldsymbol{x}^{(1)}}(t)$, we can then obtain the following estimate of $x^{(0)}(t)$

$$\widehat{x}^{(0)}(t) = \boldsymbol{\mu}^T \widehat{\boldsymbol{x}^{(1)}}(t), \tag{4.104}$$

where $\boldsymbol{\mu} = [\mu_1, \mu_2]^T$. Now notice that

$$\frac{1}{n} \sum_{t=1}^{n} E\left[\left(\boldsymbol{\mu}^T \boldsymbol{x}^{(1)}(t) - \boldsymbol{\mu}^T \widehat{\boldsymbol{x}^{(1)}}(t) \right)^2 \right] = \boldsymbol{\mu}^T \widehat{\boldsymbol{D}} \boldsymbol{\mu}. \tag{4.105}$$

Consequently, from the independence of $z^{(0)}$ and $(x_1^{(L)}, \ldots, x_{2^L}^{(L)})$, we deduce

$$\frac{1}{n} \sum_{t=1}^{n} E\left[\left(x^{(0)}(t) - \widehat{x^{(0)}}(t) \right)^2 \right] = \boldsymbol{\mu}^T \widehat{\boldsymbol{D}} \boldsymbol{\mu} + \sigma^2_{z^{(0)}} \stackrel{\text{def}}{=} d. \tag{4.106}$$

Let $x_n^{(0)} = (x^{(0)}(1), ..., x^{(0)}(n))$, then we can obtain the following sequence of identities

$$\begin{aligned}
\sum_{i=1}^{2^L} R_i &\geq \frac{1}{n} H(\boldsymbol{C}) && (4.107)\\
&= \frac{1}{n} I(\boldsymbol{C}; \{x_{i,n}^{(l)}, l = 1, ..., L, i = 1, ..., 2^l\}, x_n^{(0)}) && (4.108)\\
&= \frac{1}{n} I(\boldsymbol{C}; x_n^{(0)}) + \sum_{l=1}^{L} \sum_{i=1}^{2^l} \frac{1}{n} I(\boldsymbol{C}_{\mathcal{O}(x_i^{(l)})}; x_{i,n}^{(l)} | x_{\lceil \frac{i}{2} \rceil, n}^{(l-1)}) && (4.109)\\
&\geq \frac{1}{2} \log \frac{1}{d} + \sum_{l=1}^{L} \sum_{i=1}^{2^l} r_i^{(l)}, && (4.110)
\end{aligned}$$

where we have defined

$$\begin{aligned}
r_1^{(1)} &\stackrel{\text{def}}{=} \frac{1}{n} I(C_1, ..., C_{2^{L-1}}; x_{1,n}^{(1)} | x_n^{(0)}) && (4.111)\\
r_2^{(1)} &\stackrel{\text{def}}{=} \frac{1}{n} I(C_{2^{L-1}+1}, ..., C_{2^L}; x_{2,n}^{(1)} | x_n^{(0)}) && (4.112)\\
r_i^{(l)} &\stackrel{\text{def}}{=} \frac{1}{n} I(\boldsymbol{C}_{\mathcal{O}(x_i^{(l)})}; x_{i,n}^{(l)} | x_{\lceil \frac{i}{2} \rceil, n}^{(l-1)}) \quad \text{for } l = 2, ..., L \text{ and } i = 1, ..., 2^l. && (4.113)
\end{aligned}$$

From Lemma 4 in [55], for all $l = 1, ..., L-1$ and $i = 1, ..., 2^l$, we have

$$r_i^{(l)} \leq f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \tag{4.114}$$

where

$$f_{x_i^{(l)}}(r_1, r_2) = \frac{1}{2}\log\left(1 + \sigma^2_{z_i^{(l)}}[s_{2i-1}^{(l+1)}(1-2^{-2r_1}) + s_{2i}^{(l+1)}(1-2^{-2r_2})]\right). \tag{4.115}$$

Similarly using [62, Lemmas 2 and 3] we have

$$\frac{1}{d} \leq 1 + \sum_{i=1}^{2} s_i^{(1)}\left(1 - 2^{-2r_i^{(1)}}\right), \tag{4.116}$$

where $s_i^{(1)} = \left(\frac{\alpha_i^{(1)}}{\sigma_{z_i^{(1)}}}\right)^2$. Consequently, we can lower bound $\sum_{i=1}^{2^L} R_i$ by

$$\sum_{i=1}^{2^L} R_i \geq R_{\text{tree}}(d), \tag{4.117}$$

where $R_{\text{tree}}(d)$ is given by

$$\begin{aligned} R_{\text{tree}}(d) = \text{minimize} \quad & \frac{1}{2}\log\frac{1}{d} + \sum_{l=1}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\ \text{subject to} \quad & \frac{1}{d} \leq 1 + \sum_{i=1}^{2} s_i^{(1)}\left(1 - 2^{-2r_i^{(1)}}\right) \\ & r_i^{(l)} \leq f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \quad l = 1, ..., L-1 \quad i = 1, ..., 2^l, \\ & r_i^{(l)} \geq 0 \quad l = 1, ..., L \quad i = 1, ..., 2^l. \end{aligned}$$

Matrix-distortion bound We start by verifying that $\widehat{\boldsymbol{D}}$ is positive definite. Note first that $h(\boldsymbol{x}_n^{(1)}|\boldsymbol{C}) > -\infty$, this follows from

$$n\sum_{i=1}^{2^L} R_i \geq H(\boldsymbol{C}) \geq h(\boldsymbol{x}_n^{(1)}) - h(\boldsymbol{x}_n^{(1)}|\boldsymbol{C}). \tag{4.118}$$

By a classical argument we have

$$\frac{1}{n}h(\boldsymbol{x}_n^{(1)}|\boldsymbol{C}) \leq \frac{1}{n}h(\boldsymbol{x}_n^{(1)} - \widehat{\boldsymbol{x}_n^{(1)}}) \leq \frac{1}{2}\log\left((2\pi e)^2|\widehat{\boldsymbol{D}}|\right). \tag{4.119}$$

From this we can conclude that $\widehat{\boldsymbol{D}}$ is positive definite. Again by invoking the conditional independence relations inherited from the Gauss-Markov tree, we obtain the following sequence of identities

$$\sum_{i=1}^{2^L} R_i \geq \frac{1}{n} H(\boldsymbol{C}) \tag{4.120}$$

$$= \frac{1}{n} I(\boldsymbol{C}; \{x_{i,n}^{(l)}, l = 1, ..., L, i = 1, ..., 2^l\}) \tag{4.121}$$

$$= \frac{1}{n} I(\boldsymbol{C}; \boldsymbol{x}_n^{(1)}) + \sum_{l=2}^{L} \sum_{i=1}^{2^l} \frac{1}{n} I(\boldsymbol{C}_{\mathcal{O}(x_i^{(l)})}; x_{i,n}^{(l)} | x_{\lceil \frac{i}{2} \rceil, n}^{(l-1)}) \tag{4.122}$$

$$\geq \frac{1}{2} \log \left(\frac{|\boldsymbol{K}_{x^{(1)}}|}{|\widehat{\boldsymbol{D}}|} \right) + \sum_{l=2}^{L} \sum_{i=1}^{2^l} r_i^{(l)}, \tag{4.123}$$

As before, from Lemma 4 in [55], for all $l = 2, ..., L-1$ and $i = 1, ..., 2^l$, we have

$$r_i^{(l)} \leq f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}). \tag{4.124}$$

Now, note that

$$\boldsymbol{x}^{(2)} = \boldsymbol{\alpha}^{(2)} \boldsymbol{x}^{(1)} + \boldsymbol{z}^{(2)}, \tag{4.125}$$

where

$$\boldsymbol{\alpha}^{(2)} = \begin{bmatrix} \alpha_1^{(2)} & \alpha_2^{(2)} & 0 & 0 \\ 0 & 0 & \alpha_3^{(2)} & \alpha_4^{(2)} \end{bmatrix}^T \tag{4.126}$$

Using Lemma 7 in [56] we obtain[5]

$$h(x_{1,n}^{(1)} | x_{2,n}^{(1)}, \boldsymbol{C}) \leq \frac{n}{2} \log \left(\frac{2\pi e}{\widehat{\boldsymbol{D}}^{-1}(1,1)} \right). \tag{4.127}$$

Moreover, given $\boldsymbol{x}_n^{(1)} = (x_{1,n}^{(1)}, x_{2,n}^{(1)})$ the quantities $(x_{1,n}^{(2)}, \boldsymbol{C}_{\mathcal{O}(x_1^{(2)})}), ..., (x_{4,n}^{(2)}, \boldsymbol{C}_{\mathcal{O}(x_4^{(2)})})$ are independent. Using this condition it is not hard to see that the result of lemma 8 in [56] remains valid, that is

$$\frac{n}{2} \log \left(\frac{2\pi e}{\boldsymbol{K}(1,1)} \right) \leq h(x_{1,n}^{(1)} | x_{2,n}^{(1)}, \boldsymbol{C}), \tag{4.128}$$

[5]Note there is a typo in Lemma 7 in [56, page 19]. This typo is corrected in [56, page 26].

where

$$\boldsymbol{K} = \boldsymbol{K}_{x^{(1)}}^{-1} + \operatorname{diag}\left(\sum_{i=1}^{2} s_i^{(2)}(1 - 2^{-2r_i^{(2)}}), \sum_{i=3}^{4} s_i^{(2)}(1 - 2^{-2r_i^{(2)}})\right). \tag{4.129}$$

Combining these inequalities leads to

$$\widehat{\boldsymbol{D}}^{-1}(1,1) \leq \boldsymbol{K}_{x^{(1)}}^{-1}(1,1) + \sum_{i=1}^{2} s_i^{(2)}(1 - 2^{-2r_i^{(2)}}) \tag{4.130}$$

The vectors $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ being jointly Gaussian, we can find a Gaussian vector $\boldsymbol{v}^{(2)}$ that is independent of $\boldsymbol{x}^{(2)}$ such that

$$\boldsymbol{x}^{(1)} = \boldsymbol{K}_{x^{(1)}x^{(2)}} \boldsymbol{K}_{x^{(2)}}^{-1} \boldsymbol{x}^{(2)} + \boldsymbol{v}^{(2)}. \tag{4.131}$$

The inverse of the covariance matrix of $\boldsymbol{v}^{(2)}$ can be computed as follows

$$\boldsymbol{K}_{v^{(2)}}^{-1} = \boldsymbol{K}_{x^{(1)}|x^{(2)}}^{-1} = \boldsymbol{K}_{x^{(1)}}^{-1} + (\boldsymbol{\alpha}^{(2)})^T \boldsymbol{K}_{z^{(2)}}^{-1} \boldsymbol{\alpha}^{(2)} \tag{4.132}$$

$$= \boldsymbol{K}_{x^{(1)}}^{-1} + \operatorname{diag}(a, b), \tag{4.133}$$

where $a \stackrel{\text{def}}{=} \sum_{i=1}^{2} s_i^{(2)}$ and $b \stackrel{\text{def}}{=} \sum_{i=3}^{4} s_i^{(2)}$. We deduce that

$$\boldsymbol{K}_{v^{(2)}}^{-1}(1,1) = \boldsymbol{K}_{x^{(1)}}^{-1}(1,1) + a. \tag{4.134}$$

Using this identity we can rewrite (4.130) as

$$\sum_{i=1}^{2} s_i^{(2)} 2^{-2r_i^{(2)}} \leq \boldsymbol{K}_{v^{(2)}}^{-1}(1,1) - \widehat{\boldsymbol{D}}^{-1}(1,1). \tag{4.135}$$

A similar approach shows that

$$\sum_{i=3}^{4} s_i^{(2)} 2^{-2r_i^{(2)}} \leq \boldsymbol{K}_{v^{(2)}}^{-1}(2,2) - \widehat{\boldsymbol{D}}^{-1}(2,2). \tag{4.136}$$

Consequently, we can lower bound $\sum_{i=1}^{2^L} R_i$ by

$$\sum_{i=1}^{2^L} R_i \geq R_{\text{vec}}(\widehat{\boldsymbol{D}}), \tag{4.137}$$

where $R_{\text{vec}}(\widehat{\boldsymbol{D}})$ is given by

$$
\begin{aligned}
R_{\text{vec}}(\widehat{\boldsymbol{D}}) = \text{minimize} \quad & \frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{x^{(1)}}|}{|\widehat{\boldsymbol{D}}|}\right) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\
\text{subject to} \quad & \sum_{i=1}^{2} s_i^{(2)} 2^{-2r_i^{(2)}} \leq \boldsymbol{K}^{-1}_{v^{(2)}}(1,1) - \widehat{\boldsymbol{D}}^{-1}(1,1), \\
& \sum_{i=3}^{4} s_i^{(2)} 2^{-2r_i^{(2)}} \leq \boldsymbol{K}^{-1}_{v^{(2)}}(2,2) - \widehat{\boldsymbol{D}}^{-1}(2,2), \\
& r_i^{(l)} \leq f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \quad l = 2, ..., L-1 \quad i = 1, ..., 2^l, \\
& r_i^{(l)} \geq 0 \quad l = 2, ..., L \quad i = 1, ..., 2^l.
\end{aligned}
$$

Combination of the two bounds In this section we will couple the two bounds found in parts (a) and (b) to provide a single lower bound on the sum rate. Since $\widehat{\boldsymbol{D}}$ is positive definite, we can write it as

$$
\widehat{\boldsymbol{D}} = \begin{bmatrix} \hat{d}_1 & \phi\sqrt{\hat{d}_1\hat{d}_2} \\ \phi\sqrt{\hat{d}_1\hat{d}_2} & \hat{d}_2 \end{bmatrix}, \tag{4.138}
$$

where $\hat{d}_1 \leq d_1$ and $\hat{d}_2 \leq d_2$ and $\phi \in (-1, 1)$. Now let $\varphi = \frac{\phi\sqrt{\hat{d}_1\hat{d}_2}}{\sqrt{d_1 d_2}}$ and

$$
\boldsymbol{D}_\varphi = \begin{bmatrix} d_1 & \varphi\sqrt{d_1 d_2} \\ \varphi\sqrt{d_1 d_2} & d_2 \end{bmatrix}. \tag{4.139}
$$

Then $\boldsymbol{D}_\varphi - \widehat{\boldsymbol{D}} = \text{diag}(d_1 - \hat{d}_1, d_2 - \hat{d}_2)$ and it follows that

$$
\widehat{\boldsymbol{D}} \preceq \boldsymbol{D}_\varphi. \tag{4.140}
$$

As a consequence of this inequality, the following identities hold

$$
|\widehat{\boldsymbol{D}}| \leq |\boldsymbol{D}_\varphi|, \tag{4.141}
$$

$$
\widehat{\boldsymbol{D}}^{-1}(1,1) \geq \boldsymbol{D}_\varphi^{-1}(1,1), \tag{4.142}
$$

$$
\widehat{\boldsymbol{D}}^{-1}(2,2) \geq \boldsymbol{D}_\varphi^{-1}(2,2), \tag{4.143}
$$

$$
d \stackrel{\text{def}}{=} \boldsymbol{\mu}^T \widehat{\boldsymbol{D}} \boldsymbol{\mu} + \sigma^2_{z^{(0)}} \leq d(\varphi) \stackrel{\text{def}}{=} \boldsymbol{\mu}^T \boldsymbol{D}_\varphi \boldsymbol{\mu} + \sigma^2_{z^{(0)}}. \tag{4.144}
$$

Since the functions $R_{\text{vec}}(\cdot)$ and $R_{\text{tree}}(\cdot)$ are decreasing we deduce that

$$R_{\text{vec}}(\widehat{\boldsymbol{D}}) \geq R_{\text{vec}}(\boldsymbol{D}_{\varphi}) \tag{4.145}$$

$$R_{\text{tree}}(d) \geq R_{\text{tree}}(d(\varphi)). \tag{4.146}$$

Combining the previous two bounds we obtain that for some $\varphi \in (-1, 1)$

$$\sum_{i=1}^{2^L} R_i \geq \max\left(R_{\text{vec}}(\boldsymbol{D}_{\varphi}), R_{\text{tree}}(d(\varphi))\right). \tag{4.147}$$

Since $(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(L)})$ are jointly Gaussian, we can write

$$\boldsymbol{x}^{(1)} = \boldsymbol{A}\boldsymbol{x}^{(L)} + \boldsymbol{v}^{(L)}, \tag{4.148}$$

with $\boldsymbol{A} = \boldsymbol{K}_{x^{(1)}x^{(L)}}\boldsymbol{K}^{-1}_{x^{(L)}}$ and $\boldsymbol{v}^{(L)} \perp\!\!\!\perp \boldsymbol{x}^{(L)}$, we deduce

$$\widehat{\boldsymbol{D}} = \boldsymbol{K}_{v^{(L)}} + \boldsymbol{A}\widehat{\boldsymbol{D}}_{x^{(L)}}\boldsymbol{A}^T, \tag{4.149}$$

with

$$\widehat{\boldsymbol{D}}_{x^{(L)}} = \frac{1}{n}\sum_{t=1}^{n} E\left[\left(\boldsymbol{x}^{(L)}(t) - \widehat{\boldsymbol{x}^{(L)}}(t)\right)\left(\boldsymbol{x}^{(L)}(t) - \widehat{\boldsymbol{x}^{(L)}}(t)\right)^T\right]. \tag{4.150}$$

It can also be verified that $\widehat{\boldsymbol{D}}_{x^{(L)}}$ is nonsingular and hence positive definite. Since $\boldsymbol{A}$ is full row-rank ($\text{rank}(\boldsymbol{A}) = 2$), we also have that $\boldsymbol{A}\widehat{\boldsymbol{D}}_{x^{(L)}}\boldsymbol{A}^T$ is positive definite[6]. We deduce that[7] $\widehat{\boldsymbol{D}} \succ \boldsymbol{K}_{v^{(L)}}$ and as a consequence $\boldsymbol{D}_{\varphi} \succ \boldsymbol{K}_{v^{(L)}}$. Hence if we let $\mathcal{S} = \{\theta \in (-1, 1) : \boldsymbol{D}_{\theta} \succ \boldsymbol{K}_{v^{(L)}}\}$, we have

$$\sum_{i=1}^{2^L} R_i \geq \inf_{\theta \in \mathcal{S}} \max\left(R_{\text{vec}}(\boldsymbol{D}_{\theta}), R_{\text{tree}}(d(\theta))\right). \tag{4.151}$$

We will now prove that

$$\inf_{\theta \in \mathcal{S}} \max\left(R_{\text{vec}}(\boldsymbol{D}_{\theta}), R_{\text{tree}}(d(\theta))\right) = R_{\text{vec}}(\boldsymbol{D}_{\theta^*}) = R_{\text{tree}}(d(\theta^*)), \tag{4.152}$$

[6]This is easy to check since if $\boldsymbol{u}$ is a nonzero vector such that $\boldsymbol{A}\widehat{\boldsymbol{D}}_{x^{(L)}}\boldsymbol{A}^T\boldsymbol{u} = \boldsymbol{0}$ then $\boldsymbol{u}^T\boldsymbol{A}\widehat{\boldsymbol{D}}_{x^{(L)}}\boldsymbol{A}^T\boldsymbol{u} = 0$. Since $\widehat{\boldsymbol{D}}_{x^{(L)}}$ is positive definite it follows $\boldsymbol{A}^T\boldsymbol{u} = \boldsymbol{0}$ from which we can conclude that $\boldsymbol{u} = \boldsymbol{0}$ since $\boldsymbol{A}$ has a full row rank.

[7]Note that this imposes conditions on the distortions d_1 and d_2. In particular we must have $d_1 > \boldsymbol{K}_{v^{(L)}}(1,1)$ and $d_2 > \boldsymbol{K}_{v^{(L)}}(2,2)$.

where θ^* was defined in (4.204) and is given by

$$\theta^* = \frac{\sqrt{(1-\rho^2)^2 + 4\rho^2 d_1 d_2} - (1-\rho^2)}{2\rho\sqrt{d_1 d_2}}. \tag{4.153}$$

We start first by verifying that $\theta^* \in \mathcal{S}$. To see this, recall that $\boldsymbol{U} = \boldsymbol{x}^{(L)} + \boldsymbol{\Gamma}$. This implies that

$$\boldsymbol{K}_{x^{(1)}|U} \succ \boldsymbol{K}_{x^{(1)}|x^{(L)}} = \boldsymbol{K}_{v^{(L)}}. \tag{4.154}$$

But when the optimal Gaussian channels achieves both distortion constraints with equality we have (see Section 4.6) $\boldsymbol{K}_{x^{(1)}|U} = \boldsymbol{D}_{\theta^*}$ establishing therefore $\boldsymbol{D}_{\theta^*} \succ \boldsymbol{K}_{v^{(L)}}$, i.e., $\theta^* \in \mathcal{S}$.

It is easy to check that $d(\theta)$ is an increasing function of θ. As a consequence, the function $R_{\text{tree}}(d(\theta))$ is decreasing. Hence if $\theta \leq \theta^*$

$$\max\left(R_{\text{vec}}(\boldsymbol{D}_\theta), R_{\text{tree}}(d(\theta))\right) \geq R_{\text{tree}}(d(\theta)) \geq R_{\text{tree}}(d(\theta^*)). \tag{4.155}$$

In Section 4.8, we show that

$$R_{\mathcal{G}}(d_1, d_2) = R_{\text{tree}}(d(\theta^*)). \tag{4.156}$$

It follows that

$$\max\left(R_{\text{vec}}(\boldsymbol{D}_\theta), R_{\text{tree}}(d(\theta))\right) \geq R_{\mathcal{G}}(d_1, d_2) \quad \forall \theta \leq \theta^*. \tag{4.157}$$

We will now show that for $\theta \geq \theta^*$, $R_{\text{vec}}(\boldsymbol{D}_\theta) \geq R_{\text{vec}}(\boldsymbol{D}_{\theta^*})$. Define

$$\begin{aligned} M(\theta) = \Bigg\{ (r_1^{(2)}, ..., r_4^{(2)}) : \sum_{i=1}^{2} s_i^{(2)} 2^{-2r_i^{(2)}} \leq \frac{\boldsymbol{K}_{v^{(2)}}(2,2)}{|\boldsymbol{K}_{v^{(2)}}|} - \frac{d_2}{|\boldsymbol{D}_\theta|}, \\ \sum_{i=3}^{4} s_i^{(2)} 2^{-2r_i^{(2)}} \leq \frac{\boldsymbol{K}_{v^{(2)}}(1,1)}{|\boldsymbol{K}_{v^{(2)}}|} - \frac{d_1}{|\boldsymbol{D}_\theta|} \Bigg\}. \end{aligned} \tag{4.158}$$

Using $M(\theta)$, we can rewrite $R_{\text{vec}}(\boldsymbol{D}_\theta)$ as

$$\begin{aligned} R_{\text{vec}}(\boldsymbol{D}_\theta) = \text{minimize} \quad & \frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{x^{(1)}}|}{|\boldsymbol{D}_\theta|}\right) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\ \text{subject to} \quad & (r_1^{(2)}, ..., r_4^{(2)}) \in M(\theta) \\ & r_i^{(l)} \leq f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \quad l = 2, ..., L-1 \quad i = 1, ..., 2^l, \\ & r_i^{(l)} \geq 0 \quad l = 2, ..., L \quad i = 1, ..., 2^l. \end{aligned}$$

Since the functions $\frac{\boldsymbol{K}_{v^{(2)}}(2,2)}{|\boldsymbol{K}_{v^{(2)}}|} - \frac{d_2}{|\boldsymbol{D}_\theta|}$ and $\frac{\boldsymbol{K}_{v^{(2)}}(1,1)}{|\boldsymbol{K}_{v^{(2)}}|} - \frac{d_1}{|\boldsymbol{D}_\theta|}$ are decreasing we have $M(\theta) \subset M(\theta^*)$ for $\theta \geq \theta^*$. Moreover since $\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{x^{(1)}}|}{|\boldsymbol{D}_\theta|}\right) \geq \frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{x^{(1)}}|}{|\boldsymbol{D}_{\theta^*}|}\right)$, we conclude that $R_{\text{vec}}(\boldsymbol{D}_\theta) \geq R_{\text{vec}}(\boldsymbol{D}_{\theta^*})$. It follows that for $\theta \geq \theta^*$

$$\max\left(R_{\text{vec}}(\boldsymbol{D}_\theta), R_{\text{tree}}(d(\theta))\right) \geq R_{\text{vec}}(\boldsymbol{D}_\theta) \geq R_{\text{vec}}(\boldsymbol{D}_{\theta^*}). \tag{4.159}$$

We have proved in Section 4.9 that $R_{\text{vec}}(\boldsymbol{D}_{\theta^*}) = R_{\mathcal{G}}(d_1, d_2)$. Consequently,

$$\max\left(R_{\text{vec}}(\boldsymbol{D}_\theta), R_{\text{tree}}(d(\theta))\right) \geq R_{\mathcal{G}}(d_1, d_2) \quad \forall \theta \geq \theta^*. \tag{4.160}$$

We conclude then that

$$\sum_{i=1}^{2^L} R_i \geq \inf_{\theta \in \mathcal{S}} \max\left(R_{\text{vec}}(\boldsymbol{D}_\theta), R_{\text{tree}}(d(\theta))\right) = R_{\mathcal{G}}(d_1, d_2). \tag{4.161}$$

This being true for every strict sense achievable $(R_1, ..., R_{2^L}, d_1, d_2)$, we deduce

$$R_{\text{sum}}(d_1, d_2) \geq R_{\mathcal{G}}(d_1, d_2). \tag{4.162}$$

This completes the proof of the converse. □

4.5 Proof of Lemma 15

For $l \in \{1, ..., L\}$, define $\boldsymbol{x}^{(l)} = [x_1^{(l)}, ..., x_{2^l}^{(l)}]^T$. Relying on the tree structure, we can use the chain rule for mutual information to compute the total rate of the

Gaussian distributed test channel $\boldsymbol{U} = [U_1, ..., U_{2^L}]^T$ as follows

$$I(\boldsymbol{x}^{(L)}; \boldsymbol{U}) = I(\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(L)}; \boldsymbol{U}) \tag{4.163}$$

$$= I(\boldsymbol{x}^{(1)}; \boldsymbol{U}) + \sum_{l=2}^{L} I(\boldsymbol{x}^{(l)}; \boldsymbol{U}|\boldsymbol{x}^{(l-1)}) \tag{4.164}$$

$$= \frac{1}{2} \log \left(\frac{|\boldsymbol{K}_{x^{(1)}}|}{|\boldsymbol{D}|} \right) + \sum_{l=2}^{L} \sum_{i=1}^{2^l} I(x_i^{(l)}; \boldsymbol{U}_{\mathcal{O}(x_i^{(l)})}|x_{\lceil \frac{i}{2} \rceil}^{(l-1)}), \tag{4.165}$$

where

$$\boldsymbol{D} = E\left[(\boldsymbol{x}^{(1)} - E[\boldsymbol{x}^{(1)}|\boldsymbol{U}])(\boldsymbol{x}^{(1)} - E[\boldsymbol{x}^{(1)}|\boldsymbol{U}])^T\right]. \tag{4.166}$$

is the error covariance matrix achieved by $\boldsymbol{U}$ and $\mathcal{O}(x_i^{(l)})$ is the set of indices of the variable nodes observed by the encoders (i.e., at level L) that are decedents of $x_i^{(l)}$. For instance $\mathcal{O}(x_1^{(1)}) = \{1, ..., 2^{L-1}\}$ and $\mathcal{O}(x_1^{(L-2)}) = \{1, ..., 4\}$, that is $\boldsymbol{U}_{\mathcal{O}(x_1^{(1)})} = [U_1, ..., U_{2^{L-1}}]^T$ and $\boldsymbol{U}_{\mathcal{O}(x_1^{(L-2)})} = [U_1, ..., U_4]^T$. Now define

$$r_i^{(l)} \stackrel{\text{def}}{=} I(x_i^{(l)}; \boldsymbol{U}_{\mathcal{O}(x_i^{(l)})}|x_{\lceil \frac{i}{2} \rceil}^{(l-1)}). \tag{4.167}$$

For $l = 2, ..., L-1$ and $i = 1, ..., 2^l$, we will establish the following recurrence formula

$$r_i^{(l)} = \frac{1}{2} \log \left(1 + \sigma^2_{z_i^{(l)}}[s_{2i-1}^{(l+1)}(1 - 2^{-2r_{2i-1}^{(l+1)}}) + s_{2i}^{(l+1)}(1 - 2^{-2r_{2i}^{(l+1)}})]\right). \tag{4.168}$$

To show this, we note that using the tree structure, we can write the following identity

$$\boldsymbol{U}_{\mathcal{O}(x_j^{(l+1)})} = \boldsymbol{a}_j^{(l+1)} x_j^{(l+1)} + \boldsymbol{N}_j^{(l+1)} \quad j = 2i-1, 2i, \tag{4.169}$$

where $\boldsymbol{a}_{2i-1}^{(l+1)}, \boldsymbol{a}_{2i}^{(l+1)}$ are two given column vectors and $\boldsymbol{N}_{2i-1}^{(l+1)}, \boldsymbol{N}_{2i}^{(l+1)}$ are two independent Gaussian vectors. Since $x_j^{(l+1)} = \alpha_j^{(l+1)} x_i^{(l)} + z_j^{(l+1)}$ for $j = 2i-1, 2i$, we deduce

$$\boldsymbol{U}_{\mathcal{O}(x_j^{(l+1)})} = \alpha_j^{(l+1)} \boldsymbol{a}_j^{(l+1)} x_i^{(l)} + \boldsymbol{w}_j^{(l+1)} \quad j = 2i-1, 2i, \tag{4.170}$$

where $\boldsymbol{w}_j^{(l+1)} = \boldsymbol{a}_j^{(l+1)} z_j^{(l+1)} + \boldsymbol{N}_j^{(l+1)}$. Using the relation $x_i^{(l)} = \alpha_i^{(l)} x_{\lceil \frac{i}{2} \rceil}^{(l-1)} + z_i^{(l)}$, we can write

$$x_{\lceil \frac{i}{2} \rceil}^{(l-1)} = \beta x_i^{(l)} + v, \tag{4.171}$$

where $\beta = E\left[x_i^{(l)} x_{\lceil \frac{i}{2} \rceil}^{(l-1)}\right] / \sigma^2_{x_i^{(l)}}$ and v is a Gaussian random variable independent of $x_i^{(l)}$. Note that v is also going to be independent of $(\boldsymbol{w}_{2i-1}^{(l+1)}, \boldsymbol{w}_{2i}^{(l+1)})$. Hence, since $\boldsymbol{U}_{\mathcal{O}(x_i^{(l)})} = [\boldsymbol{U}^T_{\mathcal{O}(x_{2i-1}^{(l+1)})}, \boldsymbol{U}^T_{\mathcal{O}(x_{2i}^{(l+1)})}]^T$, using [65, pp. 77] we have

$$\frac{1}{\sigma^2_{x_i^{(l)} | \boldsymbol{U}_{\mathcal{O}(x_i^{(l)})}, x_{\lceil \frac{i}{2} \rceil}^{(l-1)}}} = \frac{1}{\sigma^2_{x_i^{(l)}}} + \frac{\beta^2}{\sigma_v^2} + \sum_{j=2i-1}^{2i} \left(\alpha_j^{(l+1)}\right)^2 (\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{w_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)} \tag{4.172}$$

$$= \frac{1}{\sigma^2_{z_i^{(l)}}} + \sum_{j=2i-1}^{2i} \left(\alpha_j^{(l+1)}\right)^2 (\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{w_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)}, \tag{4.173}$$

where in the last equality we have used the fact that

$$\frac{1}{\sigma^2_{z_i^{(l)}}} = \frac{1}{\sigma^2_{x_i^{(l)} | x_{\lceil \frac{i}{2} \rceil}^{(l-1)}}} = \frac{1}{\sigma^2_{x_i^{(l)}}} + \frac{\beta^2}{\sigma_v^2}. \tag{4.174}$$

By the independence of $z_j^{(l+1)}$ and $\boldsymbol{N}_j^{(l+1)}$ and the matrix inversion lemma we have

$$\boldsymbol{K}^{-1}_{w_j^{(l+1)}} = \boldsymbol{K}^{-1}_{N_j^{(l+1)}} - \boldsymbol{K}^{-1}_{N_j^{(l+1)}} \frac{\sigma^2_{z_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)} (\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{N_j^{(l+1)}}}{1 + \sigma^2_{z_j^{(l+1)}} (\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{N_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)}}, \quad j = 2i-1, 2i. \tag{4.175}$$

This leads to

$$(\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{w_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)} = \frac{(\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{N_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)}}{1 + \sigma^2_{z_j^{(l+1)}} (\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{N_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)}}, \quad j = 2i-1, 2i. \tag{4.176}$$

We deduce that

$$\frac{1}{\sigma^2_{x_i^{(l)} | \boldsymbol{U}_{\mathcal{O}(x_i^{(l)})}, x_{\lceil \frac{i}{2} \rceil}^{(l-1)}}} = \frac{1}{\sigma^2_{z_i^{(l)}}} + \sum_{j=2i-1}^{2i} \frac{(\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{N_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)}}{1 + \sigma^2_{z_j^{(l+1)}} (\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{N_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)}}. \tag{4.177}$$

A similar approach gives

$$\frac{1}{\sigma^2_{x_j^{(l+1)}|\boldsymbol{U}_{\mathcal{O}(x_j^{(l+1)}),x_i^{(l)}}}} = \frac{1}{\sigma^2_{z_j^{(l+1)}}} + (\boldsymbol{a}_j^{(l+1)})^T \boldsymbol{K}^{-1}_{N_j^{(l+1)}} \boldsymbol{a}_j^{(l+1)}, \quad j = 2i-1, 2i. \tag{4.178}$$

Combining these identities with the following two facts

$$r_i^{(l)} = \frac{1}{2}\log\left(\frac{\sigma^2_{z_i^{(l)}}}{\sigma^2_{x_i^{(l)}|\boldsymbol{U}_{\mathcal{O}(x_i^{(l)}),x^{(l-1)}_{\lceil \frac{i}{2} \rceil}}}}\right) \tag{4.179}$$

$$r_j^{(l+1)} = \frac{1}{2}\log\left(\frac{\sigma^2_{z_j^{(l+1)}}}{\sigma^2_{x_j^{(l+1)}|\boldsymbol{U}_{\mathcal{O}(x_j^{(l+1)}),x_i^{(l)}}}}\right) \quad j = 2i-1, 2i \tag{4.180}$$

establishes (4.168). An expression for $R_{\mathcal{G}}(d_1, d_2)$ can be then obtained through the following optimization problem

$$\begin{aligned} R_{\mathcal{G}}(d_1, d_2) = \text{minimize} \quad & \frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{x^{(1)}}|}{|\boldsymbol{D}|}\right) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\ \text{subject to} \quad & \boldsymbol{D}(1,1) \leq d_1, \\ & \boldsymbol{D}(2,2) \leq d_2 \\ & r_i^{(L)} \geq 0, i = 1, ..., 2^L. \end{aligned}$$

We will try to make this optimization problem more explicit by expressing $\boldsymbol{D}$ in terms of the the noise quantization rates $r_1^{(2)}, ..., r_4^{(2)}$. We start by noting that

$$\boldsymbol{U}_{\mathcal{O}(x_1^{(1)})} = \begin{bmatrix} \alpha_1^{(2)}\boldsymbol{a}_1^{(2)} \\ \alpha_2^{(2)}\boldsymbol{a}_2^{(2)} \end{bmatrix} x_1^{(1)} + \begin{bmatrix} \boldsymbol{w}_1^{(2)} \\ \boldsymbol{w}_2^{(2)} \end{bmatrix} \tag{4.181}$$

and

$$\boldsymbol{U}_{\mathcal{O}(x_2^{(1)})} = \begin{bmatrix} \alpha_3^{(2)}\boldsymbol{a}_3^{(2)} \\ \alpha_4^{(2)}\boldsymbol{a}_4^{(2)} \end{bmatrix} x_2^{(1)} + \begin{bmatrix} \boldsymbol{w}_3^{(2)} \\ \boldsymbol{w}_4^{(2)} \end{bmatrix}. \tag{4.182}$$

We can write these 2 equations compactly as

$$\boldsymbol{U} = \boldsymbol{\alpha}\boldsymbol{x}^{(1)} + \boldsymbol{w}^{(2)}. \tag{4.183}$$

where

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1^{(2)}\boldsymbol{a}_1^{(2)} & \mathbf{0} \\ \alpha_2^{(2)}\boldsymbol{a}_2^{(2)} & \mathbf{0} \\ \mathbf{0} & \alpha_3^{(2)}\boldsymbol{a}_3^{(2)} \\ \mathbf{0} & \alpha_4^{(2)}\boldsymbol{a}_4^{(2)} \end{bmatrix}, \tag{4.184}$$

and $\boldsymbol{w}^{(2)} = [(\boldsymbol{w}_1^{(2)})^T, ..., (\boldsymbol{w}_4^{(2)})^T]^T$. Since $\boldsymbol{x}^{(1)} \perp\!\!\!\perp \boldsymbol{w}^{(2)}$, we have (see for instance [65, pp. 77])

$$\boldsymbol{D}^{-1} = \boldsymbol{K}_{x^{(1)}}^{-1} + \boldsymbol{\alpha}^T \boldsymbol{K}_{w^{(2)}}^{-1} \boldsymbol{\alpha}. \tag{4.185}$$

By the independence of the vectors $\boldsymbol{w}_1^{(2)}, ..., \boldsymbol{w}_4^{(2)}$, the covariance matrix $\boldsymbol{K}_{w^{(2)}}$ is going to be block diagonal. In this case, it can be verified that $\boldsymbol{\alpha}^T \boldsymbol{K}_{\boldsymbol{w}^{(2)}}^{-1} \boldsymbol{\alpha}$ is also a diagonal matrix, i.e., $\boldsymbol{\alpha}^T \boldsymbol{K}_{w^{(2)}}^{-1} \boldsymbol{\alpha} = \mathrm{diag}(T_1, T_2)$ with $T_1, T_2 \geq 0$ given by

$$T_1 = \sum_{i=1}^{2} \left(\alpha_i^{(2)}\right)^2 (\boldsymbol{a}_i^{(2)})^T \boldsymbol{K}_{w_i^{(2)}}^{-1} \boldsymbol{a}_i^{(2)} \tag{4.186}$$

$$T_2 = \sum_{i=3}^{4} \left(\alpha_i^{(2)}\right)^2 (\boldsymbol{a}_i^{(2)})^T \boldsymbol{K}_{w_i^{(2)}}^{-1} \boldsymbol{a}_i^{(2)}, \tag{4.187}$$

which can be written in terms of $r_1^{(2)}, ..., r_4^{(2)}$ (following the approach described above) as

$$T_1(r_1^{(2)}, r_2^{(2)}) = \sum_{i=1}^{2} s_i^{(2)}(1 - 2^{-2r_i^{(2)}}) \tag{4.188}$$

$$T_2(r_3^{(2)}, r_4^{(2)}) = \sum_{i=3}^{4} s_i^{(2)}(1 - 2^{-2r_i^{(2)}}). \tag{4.189}$$

Now let

$$\Psi(r_1^{(2)}, ..., r_4^{(2)}) = \frac{[1 + (1-\rho^2)T_1(r_1^{(2)}, r_2^{(2)})][1 + (1-\rho^2)T_2(r_3^{(2)}, r_4^{(2)})] - \rho^2}{1 - \rho^2}. \tag{4.190}$$

Then after some algebraic manipulations, by using (4.185) we can rewrite

$R_{\mathcal{G}}(d_1, d_2)$ in terms of T_1 and T_2 as follows

$$
\begin{aligned}
R_{\mathcal{G}}(d_1, d_2) = \text{minimize} \quad & \frac{1}{2}\log\Psi(r_1^{(2)}, ..., r_4^{(2)}) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\
\text{subject to} \quad & \frac{(1-\rho^2)[1+(1-\rho^2)T_2(r_3^{(2)}, r_4^{(2)})]}{[1+(1-\rho^2)T_1(r_1^{(2)}, r_2^{(2)})][1+(1-\rho^2)T_2(r_3^{(2)}, r_4^{(2)})]-\rho^2} \leq d_1, \\
& \frac{(1-\rho^2)[1+(1-\rho^2)T_1(r_1^{(2)}, r_2^{(2)})]}{[1+(1-\rho^2)T_1(r_1^{(2)}, r_2^{(2)})][1+(1-\rho^2)T_2(r_3^{(2)}, r_4^{(2)})]-\rho^2} \leq d_2 \\
& r_i^{(L)} \geq 0, i = 1, ..., 2^L.
\end{aligned}
\tag{4.191}
$$

which is the result claimed in Lemma 15.

4.6 Proof of Lemma 16: The regularity of the optimal Gaussian test channel

As mentioned before there are two case that must be distinguished. Case 1: the optimal Gaussian test channel achieves only one distortion constraint with equality and case 2: the optimal Gaussian test channel achieves both distortion constraints with equality. In order to show that $\boldsymbol{o}$ is regular we will show that $\boldsymbol{o}$ satisfies the linear independence constraint qualification [66, Theorem 12.1]. For that we need to check that the gradients of the active inequality constraints and the gradients of the equality constraints are linearly independent at $\boldsymbol{o}$. In the following we let $\mathcal{A} = \{i : o_i^{(L)} = 0\}$ and $\boldsymbol{e}_i$ be the ith vector of the standard basis of $\mathbb{R}^{2^L}$. To simplify the notations we will usually drop the argument of $C_1(r_1^{(L)}, ..., r_{2^{L-1}}^{(L)})$ and write simply C_1, this applies also for C_2, T_1 and T_2.

4.6.1 Case 1:

Recall that we are assuming here (without loss of generality) that $C_1^* = 0$ and $C_2^* < 0$. We need to show that $(\nabla_{\boldsymbol{r}} C_1|_{\boldsymbol{r}=\boldsymbol{o}}, \{\boldsymbol{e}_i, i \in \mathcal{A}\})$ are linearly independent. We start by computing $\nabla_{\boldsymbol{r}} C_1|_{\boldsymbol{r}=\boldsymbol{o}}$. For $i = 1, ..., 2^{L-1}$, we have

$$\frac{\partial C_1}{\partial r_i^{(L)}} = -\left(\frac{(1-\rho^2)[1+(1-\rho^2)T_2]}{[1+(1-\rho^2)T_1][1+(1-\rho^2)T_2]-\rho^2}\right)^2 \frac{\partial T_1}{\partial r_i^{(L)}}, \tag{4.192}$$

with

$$\frac{\partial T_1}{\partial r_i^{(L)}} = \frac{\partial T_1(r_1^{(2)}, r_2^{(2)})}{\partial r^{(2)}_{\lceil \frac{i}{2^{L-2}} \rceil}} \prod_{l=2}^{L-1} \frac{\partial r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}{\partial r^{(l+1)}_{\lceil \frac{i}{2^{L-l-1}} \rceil}} \tag{4.193}$$

$$= 2\ln(2) s^{(2)}_{\lceil \frac{i}{2^{L-2}} \rceil} 2^{-2r^{(2)}_{\lceil \frac{i}{2^{L-2}} \rceil}} \prod_{l=2}^{L-1} \sigma^2_{z^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}} s^{(l+1)}_{\lceil \frac{i}{2^{L-l-1}} \rceil} 2^{-2r^{(l+1)}_{\lceil \frac{i}{2^{L-l-1}} \rceil}} 2^{-2r^{(l)}_{\lceil \frac{i}{2^{L-l}} \rceil}}. \tag{4.194}$$

This last identity follows from the chain rule for partial derivative combined with (4.11). From (4.11) and (4.14), we see that $T_1 = F(r_1^{(L-1)}, ..., r_{2^{L-2}}^{(L-1)})$ for some function F, hence from the chain rule we have

$$\frac{\partial T_1}{\partial r_i^{(L)}} = \sum_{j=1}^{2^{L-2}} \frac{\partial T_1}{\partial r_j^{(L-1)}} \frac{\partial r_j^{(L-1)}}{\partial r_i^{(L)}}. \tag{4.195}$$

Using (4.11) again we can see that $\frac{\partial r_j^{(L-1)}}{\partial r_i^{(L)}} = 0$ for all $j \neq \lceil \frac{i}{2} \rceil$ which establishes

$$\frac{\partial T_1}{\partial r_i^{(L)}} = \frac{\partial T_1}{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}} \frac{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}{\partial r_i^{(L)}} \tag{4.196}$$

$$= s_i^{(L)} \sigma^2_{z^{(L-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2r_i^{(L)}} 2^{-2r^{(L-1)}_{\lceil \frac{i}{2} \rceil}} \frac{\partial T_1}{\partial r^{(L-1)}_{\lceil \frac{i}{2} \rceil}}. \tag{4.197}$$

Proceeding forward by induction we arrive at the claimed equality. For $i = 2^{L-1} + 1, ..., 2^L$, we have similarly

$$\frac{\partial C_1}{\partial r_i^{(L)}} = -\left(\frac{\rho(1-\rho^2)}{[1+(1-\rho^2)T_1][1+(1-\rho^2)T_2]-\rho^2}\right)^2 \frac{\partial T_2}{\partial r_i^{(L)}} \tag{4.198}$$

with $\frac{\partial T_2}{\partial r_i^{(L)}}$ given by a similar expression as the one above. We can see therefore that all the components of $\nabla_{\boldsymbol{r}} C_1|_{\boldsymbol{r}=\boldsymbol{o}}$ are nonzero. Since $\mathcal{A} \neq \{1, ..., 2^L\}$, we must have that $(\nabla_{\boldsymbol{r}} C_1|_{\boldsymbol{r}=\boldsymbol{o}}, \{\boldsymbol{e}_i, i \in \mathcal{A}\})$ are linearly independent and hence $\boldsymbol{o}$ is regular.

4.6.2 Case 2:

Here we have $C_1^* = 0$ and $C_2^* = 0$. To establish the regularity of $\boldsymbol{o}$ we must now show that $(\nabla_{\boldsymbol{r}} C_1|_{\boldsymbol{r}=\boldsymbol{o}}, \nabla_{\boldsymbol{r}} C_2|_{\boldsymbol{r}=\boldsymbol{o}}, \{\boldsymbol{e}_i, i \in \mathcal{A}\})$ are linearly independent. We have computed $\nabla_{\boldsymbol{r}} C_1$ above, we will also need to compute $\nabla_{\boldsymbol{r}} C_2$. For $i = 1, ..., 2^{L-1}$, we have

$$\frac{\partial C_2}{\partial r_i^{(L)}} = -\left(\frac{\rho(1-\rho^2)}{[1+(1-\rho^2)T_1][1+(1-\rho^2)T_2]-\rho^2}\right)^2 \frac{\partial T_1}{\partial r_i^{(L)}}. \tag{4.199}$$

And for $i = 2^{L-1}+1, ..., 2^L$,

$$\frac{\partial C_2}{\partial r_i^{(L)}} = -\left(\frac{(1-\rho^2)[1+(1-\rho^2)T_1]}{[1+(1-\rho^2)T_1][1+(1-\rho^2)T_2]-\rho^2}\right)^2 \frac{\partial T_2}{\partial r_i^{(L)}}. \tag{4.200}$$

To simplify the upcoming derivations, let $q_i \stackrel{\text{def}}{=} \frac{\partial T_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, i = 1, ..., 2^{L-1}$ and $q_i \stackrel{\text{def}}{=} \frac{\partial T_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}}, i = 2^{L-1}+1, ..., 2^L$. And note that $q_i \neq 0$ for all i. Using the fact that the two distortion constraints are active at $\boldsymbol{o}$, i.e., $C_1^* = 0$ and $C_2^* = 0$, we can find an expression for T_1^* and T_2^* in terms of ρ, d_1 and d_2. Let $\Upsilon_j = 1 + (1-\rho^2)T_j^*$, $j = 1, 2$. Since $C_1^* = C_2^* = 0$, we can see immediately that

$$d_1 \Upsilon_1 = d_2 \Upsilon_2. \tag{4.201}$$

and

$$d_2 \Upsilon_2^2 - (1-\rho^2)\Upsilon_2 - \rho^2 d_1 = 0. \tag{4.202}$$

By solving these 2 equations we find that

$$T_j^* = \lambda_j \stackrel{\text{def}}{=} \frac{1}{d_j(1-(\theta^*)^2)} - \frac{1}{1-\rho^2}, j = 1, 2, \tag{4.203}$$

where

$$\theta^* = \frac{\sqrt{(1-\rho^2)^2 + 4\rho^2 d_1 d_2} - (1-\rho^2)}{2\rho\sqrt{d_1 d_2}}. \tag{4.204}$$

We can now find simplified expressions for $\nabla_{\boldsymbol{r}} C_1|_{\boldsymbol{r}=\boldsymbol{o}}$ and $\nabla_{\boldsymbol{r}} C_2|_{\boldsymbol{r}=\boldsymbol{o}}$. Using the fact that $C_1^* = 0$, we have

$$\frac{\partial C_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -d_1^2 q_i, \quad i = 1, ..., 2^{L-1}. \tag{4.205}$$

and using $C_2^* = 0$ we find that

$$\frac{\partial C_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -d_2^2 q_i, \quad i = 2^{L-1}+1, ...2^L. \tag{4.206}$$

Now recall that in Section 4.5 we have proved

$$\boldsymbol{D}^{-1} = \boldsymbol{K}_{x^{(1)}}^{-1} + \text{diag}(T_1^*, T_2^*). \tag{4.207}$$

This means that

$$\boldsymbol{D}^{-1} = \begin{bmatrix} \frac{1}{d_1(1-(\theta^*)^2)} & -\frac{\rho}{1-\rho^2} \\ -\frac{\rho}{1-\rho^2} & \frac{1}{d_2(1-(\theta^*)^2)} \end{bmatrix}. \tag{4.208}$$

Using the identity

$$1 - (\theta^*)^2 = \frac{(1-\rho^2)\theta^*}{\rho\sqrt{d_1 d_2}}. \tag{4.209}$$

We deduce that $\boldsymbol{D} = \boldsymbol{D}_{\theta^*}$, where

$$\boldsymbol{D}_\theta = \begin{bmatrix} d_1 & \theta\sqrt{d_1 d_2} \\ \theta\sqrt{d_1 d_2} & d_2 \end{bmatrix}. \tag{4.210}$$

and as such

$$\left(\frac{\rho(1-\rho^2)}{[1+(1-\rho^2)T_1^*][1+(1-\rho^2)T_2^*]-\rho^2}\right)^2 = \left(\rho\frac{|\boldsymbol{D}_{\theta^*}|}{|\boldsymbol{K}_{x^{(1)}}|}\right)^2 \tag{4.211}$$

$$= \left(\rho\frac{d_1 d_2(1-(\theta^*)^2)}{1-\rho^2}\right)^2 \tag{4.212}$$

$$= (\theta^*)^2 d_1 d_2. \tag{4.213}$$

We have then

$$\frac{\partial C_2}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -(\theta^*)^2 d_1 d_2 q_i \quad i = 1, ..., 2^{L-1}, \tag{4.214}$$

$$\frac{\partial C_1}{\partial r_i^{(L)}}|_{\boldsymbol{r}=\boldsymbol{o}} = -(\theta^*)^2 d_1 d_2 q_i \quad i = 2^{L-1}+1, ..., 2^L. \tag{4.215}$$

Consequently,

$$\nabla_{\boldsymbol{r}} C_1|_{\boldsymbol{r}=\boldsymbol{o}} = -[d_1^2 q_1, ..., d_1^2 q_{2^{L-1}}, (\theta^*)^2 d_1 d_2 q_{2^{L-1}+1}, ..., (\theta^*)^2 d_1 d_2 q_{2^L}]^T, \tag{4.216}$$

and

$$\nabla_{\boldsymbol{r}} C_2|_{\boldsymbol{r}=\boldsymbol{o}} = -[(\theta^*)^2 d_1 d_2 q_1, ..., (\theta^*)^2 d_1 d_2 q_{2^{L-1}}, d_2^2 q_{2^{L-1}+1}, ..., d_2^2 q_{2^L}]^T. \tag{4.217}$$

Using our initial assumption that $d_2 < \rho^2 d_1 + 1 - \rho^2$, we will argue that there exists $i^* \in \{1, ..., 2^{L-1}\}$ and $j^* \in \{2^{L-1}+1, ..., 2^L\}$ such that $o_{i^*}^{(L)}, o_{j^*}^{(L)} > 0$. Indeed, if for instance $o_j^{(L)} = 0$ for all $2^{L-1}+1 \leq j \leq 2^L$ then it follows that $o_3^{(2)} = o_4^{(2)} = 0$ that is $T_2^* = T_2(o_3^{(2)}, o_4^{(2)}) = 0$ and $\Upsilon_2 = 1$. According to (4.202) this means that $d_2 = \rho^2 d_1 + 1 - \rho^2$ which is impossible in view of the assumption. Since i^* and j^* do not belong to $\mathcal{A}$, if $(\nabla_{\boldsymbol{r}} C_1|_{\boldsymbol{r}=\boldsymbol{o}}, \nabla_{\boldsymbol{r}} C_2|_{\boldsymbol{r}=\boldsymbol{o}}, \{\boldsymbol{e}_i, i \in \mathcal{A}\})$ were to be linearly dependent there must exist $(c_1, c_2) \neq (0, 0)$ such that

$$c_1 d_1^2 + c_2 (\theta^*)^2 d_1 d_2 = 0, \tag{4.218}$$

$$c_1 (\theta^*)^2 d_1 d_2 + c_2 d_2^2 = 0. \tag{4.219}$$

This is however not possible since $\theta^* < 1$, i.e., the matrix $\begin{bmatrix} d_1 & (\theta^*)^2 d_2 \\ (\theta^*)^2 d_1 & d_2 \end{bmatrix}$ is invertible.

4.7 Proof of Lemma 17

We start by remarking that $f(x) = \rho^2[\nu_1 + (1-x)(1+\lambda_1\nu_1)][1+(1-x)\lambda_1]$ is decreasing on $[\rho^2, 1]$ whereas $g(x) = [x + (x-\rho^2)\lambda_2][\nu_2 x + (x-\rho^2)(1+\lambda_2\nu_2)]$ is increasing. To show that the equation $f(x) = g(x)$ has only one solution in the interval $[\rho^2, 1]$ when the optimal Gaussian test channel achieves both distortions with equality, we will show that

$$f(1) \leq g(1), \tag{4.220}$$

$$g(\rho^2) \leq f(\rho^2), \tag{4.221}$$

and at most one of the inequalities above can be an equality. To establish this recall first that

$$\nu_1 = \gamma_1 d_1^2 + \gamma_2 (\theta^*)^2 d_1 d_2 - d_1 \tag{4.222}$$

$$\nu_2 = \gamma_2 d_2^2 + \gamma_1 (\theta^*)^2 d_1 d_2 - d_2, \tag{4.223}$$

with $\gamma_1, \gamma_2 \geq 0$. After some transformations we can show that this system of equations is equivalent to

$$d_2(\nu_1 + d_1) = (\theta^*)^2 d_1(\nu_2 + d_2) + \gamma_1 d_1^2 d_2(1 - (\theta^*)^4), \tag{4.224}$$

$$d_1(\nu_2 + d_2) = (\theta^*)^2 d_2(\nu_1 + d_1) + \gamma_2 d_2^2 d_1(1 - (\theta^*)^4). \tag{4.225}$$

Since $\gamma_1, \gamma_2 \geq 0$ and $\theta^* < 1$ we must have

$$d_2(d_1 + \nu_1) \geq (\theta^*)^2 d_1(\nu_2 + d_2), \tag{4.226a}$$

$$d_1(d_2 + \nu_2) \geq (\theta^*)^2 d_2(\nu_1 + d_1). \tag{4.226b}$$

Clearly γ_1 and γ_2 can not be both equal to zero, because in this case $t_i > 0$ for $i \in \{1, ..., 2^L\}$ and hence $o_i^{(L)} = 0$ for all i (refer to (4.22)). This means that at most one of the inequalities above can be an equality.

Using the following identities

$$1 + (1 - \rho^2)\lambda_1 = \frac{\rho}{\theta^*}\sqrt{\frac{d_2}{d_1}}, \tag{4.227}$$

$$1 + (1 - \rho^2)\lambda_2 = \frac{\rho}{\theta^*}\sqrt{\frac{d_1}{d_2}}, \tag{4.228}$$

$$d_1 d_2(1 - (\theta^*)^2) = (1 - \rho^2)\frac{\theta^*\sqrt{d_1 d_2}}{\rho}. \tag{4.229}$$

The two inequalities in (4.226) can be written as follows

$$(1 + (1 - \rho^2)\lambda_1)((1 + (1 - \rho^2)\lambda_1)\nu_1 + (1 - \rho^2)) \geq \rho^2 \nu_2, \tag{4.230}$$

$$(1 + (1 - \rho^2)\lambda_2)((1 + (1 - \rho^2)\lambda_2)\nu_2 + (1 - \rho^2)) \geq \rho^2 \nu_1, \tag{4.231}$$

or equivalently

$$f(\rho^2) \geq g(\rho^2) \tag{4.232}$$

$$g(1) \geq f(1). \tag{4.233}$$

Moreover, as we said above, at most one of the inequalities above can be an equality. This proves that the equation $f(x) = g(x)$ has only one root in the interval $[\rho^2, 1]$. A tedious computation[8] shows that this root is given by

$$x^* = \rho \frac{(1+\lambda_1)\sqrt{\nu_1 + d_1} + \rho\lambda_2\sqrt{\nu_2 + d_2}}{(1+\lambda_2)\sqrt{\nu_2 + d_2} + \rho\lambda_1\sqrt{\nu_1 + d_1}}. \tag{4.234}$$

4.8 $R_{\text{tree}}(d(\theta^*)) = R_{\mathcal{G}}(d_1, d_2)$

In this section we will prove that $R_{\text{tree}}(d(\theta^*)) = R_{\mathcal{G}}(d_1, d_2)$. Recall that $\alpha_1^{(1)}$ is selected to be equal to $\sqrt{x^*}$ where x^* is the solution in the interval $(\rho^2, 1)$ to the quadratic equation

$$\rho^2[\nu_1 + (1-x)(1+\lambda_1\nu_1)][1 + (1-x)\lambda_1] = [x + (x-\rho^2)\lambda_2][\nu_2 x + (x-\rho^2)(1+\lambda_2\nu_2)] \tag{4.235}$$

Recall also that

$$\begin{aligned} R_{\text{tree}}(d(\theta^*)) = \text{minimize} \quad & \frac{1}{2}\log\frac{1}{d(\theta^*)} + \sum_{l=1}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\ \text{subject to} \quad & \frac{1}{d(\theta^*)} \leq 1 + \frac{x^*}{1-x^*}\left(1 - 2^{-2r_1^{(1)}}\right) + \frac{\rho^2}{x^* - \rho^2}\left(1 - 2^{-2r_2^{(1)}}\right) \\ & r_i^{(l)} \leq f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \quad l = 1, ..., L-1 \quad i = 1, ..., 2^l, \\ & r_i^{(l)} \geq 0 \quad l = 1, ..., L \quad i = 1, ..., 2^l. \end{aligned}$$

[8] This can be proved by showing that x^* is also a root to the equation $\rho^2(\nu_1 + d_1)(1 + (1-x)\lambda_1)^2 = (\nu_2 + d_2)(x + (x-\rho^2)\lambda_2)^2$.

The optimization problem defining $R_{\text{tree}}(d(\theta^*))$ is convex, the KKT conditions can be then used as a certificate of optimality. After some simplifications, the KKT conditions for this optimization problem state that the tuple $(r_i^{(l)}, l = 1, ...L, i = 1, ..., 2^l)$ is optimal if and only if we can find $\zeta \geq 0$, $(\zeta_i^{(l)}, l = 1, ...L-1, i = 1, ..., 2^l) \succeq 0$ and $(\eta_i^{(l)}, l = 1, ...L, i = 1, ..., 2^l) \succeq 0$ such that

$$0 = 1 + \zeta_1^{(1)} - \eta_1^{(1)} - \frac{\zeta x^* 2^{-2r_1^{(1)}}}{1 - x^*} \tag{4.236a}$$

$$0 = 1 + \zeta_2^{(1)} - \eta_2^{(1)} - \frac{\zeta \rho^2 2^{-2r_2^{(1)}}}{x^* - \rho^2} \tag{4.236b}$$

$$0 = 1 + \zeta_i^{(l)} - \eta_i^{(l)} - \zeta_{\lceil \frac{i}{2} \rceil}^{(l-1)} s_i^{(l)} \sigma^2_{z^{(l-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2r_i^{(l)}} 2^{-2f_{x^{(l-1)}_{\lceil \frac{i}{2} \rceil}}}, l = 2, ..., L-1, i = 1, ..., 2^l \tag{4.236c}$$

$$0 = 1 - \eta_i^{(L)} - \zeta_{\lceil \frac{i}{2} \rceil}^{(L-1)} s_i^{(L)} \sigma^2_{z^{(L-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2r_i^{(L)}} 2^{-2f_{x^{(L-1)}_{\lceil \frac{i}{2} \rceil}}}, i = 1, ..., 2^L \tag{4.236d}$$

$$0 = \eta_i^{(l)} r_i^{(l)}, \quad l = 1, ...L, i = 1, ..., 2^l. \tag{4.236e}$$

We will now identify the solution to this system using the optimal Gaussian channel $\boldsymbol{o}$. Define first $o_1^{(1)}$ and $o_2^{(1)}$ as follows

$$o_1^{(1)} \stackrel{\text{def}}{=} \frac{1}{2} \log \left(1 + (1 - x^*) \sum_{i=1}^{2} s_i^{(2)} (1 - 2^{-2o_i^{(2)}}) \right) = \frac{1}{2} \log(1 + (1 - x^*)\lambda_1) \tag{4.237}$$

$$o_2^{(1)} \stackrel{\text{def}}{=} \frac{1}{2} \log \left(1 + (1 - \frac{\rho^2}{x^*}) \sum_{i=3}^{4} s_i^{(2)} (1 - 2^{-o_i^{(2)}}) \right) = \frac{1}{2} \log(1 + (1 - \frac{\rho^2}{x^*})\lambda_2). \tag{4.238}$$

Using the characterization of $\boldsymbol{o}$ derived in section 4.3.2, we can verify that $(o_i^{(l)}, l = 1, ..., L, i = 1, ..., 2^l)$ satisfy the KKT conditions above with an appropriate choice of parameters. But before describing this choice we note here that $(o_i^{(l)}, l = 1, ..., L, i = 1, ..., 2^l)$ is feasible for the optimization problem defining $R_{\text{tree}}(d(\theta^*))$. Indeed, $(o_i^{(l)}, l = 1, ..., L, i = 1, ..., 2^l)$ satisfies all the constraints defining the feasible set of $R_{\text{tree}}(d(\theta^*))$ and in particular

$$1 + \frac{x^*}{1 - x^*}(1 - 2^{-2o_1^{(1)}}) + \frac{\rho^2}{x^* - \rho^2}(1 - 2^{-2o_2^{(1)}}) = \frac{1}{d(\theta^*)}. \tag{4.239}$$

Now for the choice of the KKT parameters, we select $\eta_1^{(1)} = \eta_2^{(1)} = 0$,

$$\zeta_1^{(1)} = \frac{(1 + (1 - x^*)\lambda_1)\nu_1}{1 - x^*}, \tag{4.240}$$

$$\zeta_2^{(1)} = \frac{(1 + (1 - \frac{\rho^2}{x^*})\lambda_2)\nu_2}{1 - \frac{\rho^2}{x^*}}. \tag{4.241}$$

and

$$\zeta = \frac{1}{x^*}(1 - x^*)(1 + \zeta_1^{(1)})(1 + (1 - x^*)\lambda_1). \tag{4.242}$$

With this choice we can see that (4.236a) and (4.236b) are verified by $o_1^{(1)}$ and $o_2^{(1)}$. Indeed, for (4.236a) just notice that

$$\frac{\zeta x^* 2^{-2o_1^{(1)}}}{1 - x^*} = 1 + \zeta_1^{(1)}. \tag{4.243}$$

For verifying (4.236b), we use the fact that x^* is a solution of the quadratic equation (4.235). This allow us to show that

$$\zeta = \frac{1}{\rho^2}(x^* - \rho^2)(1 + \zeta_2^{(1)})(1 + (1 - \frac{\rho^2}{x^*})\lambda_2). \tag{4.244}$$

From this, we see that

$$\frac{\zeta \rho^2 2^{-2o_2^{(1)}}}{x^* - \rho^2} = 1 + \zeta_2^{(1)}, \tag{4.245}$$

which is what we need for (4.236b). The rest of the KKT parameters are selected as follows

$$\zeta_i^{(l)} = \max(0, -\Psi_i^{(l)}), \quad l = 2, ..., L-1, i = 1, ..., 2^l \tag{4.246}$$

$$\eta_i^{(2)} = \max(0, \Psi_i^{(2)}), \quad i = 1, ..., 4 \tag{4.247}$$

$$\eta_i^{(l)} = \max(1, 1 - \Psi_i^{(l)}) - \max(0, 1 - \Psi_i^{(l)}), \quad l = 3, ..., L-2, i = 1, ..., 2^l \tag{4.248}$$

$$\eta_i^{(L)} = \min(1, t_i) \quad i = 1, ..., 2^L \tag{4.249}$$

The interested reader can easily verify that with this choice $(o_i^{(l)}, l = 1, ..., L, i = 1, ..., 2^l)$ satisfy the KKT conditions above, we omit this verification here since

it is very similar to the one performed in IV.A. This shows the optimality of $(o_i^{(l)}, l = 1, ..., L, i = 1, ..., 2^l)$, i.e.,

$$R_{\text{tree}}(d(\theta^*)) = \frac{1}{2}\log\left(\frac{1}{d(\theta^*)}\right) + o_1^{(1)} + o_2^{(1)} + \sum_{l=2}^{L}\sum_{i=1}^{2^l} o_i^{(l)}. \tag{4.250}$$

From the chain rule of mutual information and standard calculations we have

$$\begin{aligned}\frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{x^{(1)}}|}{|\boldsymbol{D}_{\theta^*}|}\right) = I(\boldsymbol{x}^{(1)}, \boldsymbol{U}) &= I(x^{(0)}, \boldsymbol{U}) + I(x_1^{(1)}; \boldsymbol{U}|x^{(0)}) + I(x_2^{(1)}; \boldsymbol{U}|x^{(0)}) \\ &= \frac{1}{2}\log\left(\frac{1}{d(\theta^*)}\right) + o_1^{(1)} + o_2^{(1)}. \end{aligned} \tag{4.251}$$

We conclude therefore that

$$R_{\mathcal{G}}(d_1, d_2) = R_{\text{tree}}(d(\theta^*)). \tag{4.252}$$

4.9 $R_{\text{vec}}(\boldsymbol{D}_{\theta^*}) = R_{\mathcal{G}}(d_1, d_2)$

In this section we will prove that $R_{\text{vec}}(\boldsymbol{D}_{\theta^*}) = R_{\mathcal{G}}(d_1, d_2)$. To establish this equality one can follow a similar approach to the one descried in Section 4.8. However we will provide here a somewhat different method that exploits the similarities between the optimization problems defining $R_{\text{vec}}(\boldsymbol{D}_{\theta^*})$ and $R_{\mathcal{G}}(d_1, d_2)$. When the optimal Gaussian test channel achieves both distortion constraints with equality, the optimization problem defining $R_{\mathcal{G}}(d_1, d_2)$ is equivalent to

$$\begin{aligned} R_{\mathcal{G}}(d_1, d_2) = \text{minimize} \quad & \frac{1}{2}\log\Psi(r_1^{(2)}, ..., r_4^{(2)}) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\ \text{subject to} \quad & C_1(r_1^{(L)}, ..., r_{2^{L-1}}^{(L)}) = 0 \\ & C_2(r_{2^{L-1}+1}^{(L)}, ..., r_{2^L}^{(L)}) = 0 \\ & r_i^{(L)} \geq 0, i = 1, ..., 2^L. \end{aligned} \tag{4.253}$$

As described in Section 4.6, the conditions $C_1(r_1^{(L)},...,r_{2^{L-1}}^{(L)}) = 0$ and $C_2(r_{2^{L-1}+1}^{(L)},...,r_{2^L}^{(L)}) = 0$ are equivalent to $T_1(r_1^{(2)}, r_2^{(2)}) = \lambda_1$ and $T_2(r_3^{(2)}, r_4^{(2)}) = \lambda_2$. Which is similar to

$$\sum_{i=1}^{2} s_i^{(2)} 2^{-2r_i^{(2)}} = a - \lambda_1, \tag{4.254}$$

$$\sum_{i=3}^{4} s_i^{(2)} 2^{-2r_i^{(2)}} = b - \lambda_2. \tag{4.255}$$

Moreover,

$$\frac{1}{2}\log\left(\frac{[1+(1-\rho^2)T_1(r_1^{(2)}, r_2^{(2)})][1+(1-\rho^2)T_2(r_3^{(2)}, r_4^{(2)})] - \rho^2}{1-\rho^2}\right) = \frac{1}{2}\log\left(\frac{1-\rho^2}{d_1 d_2(1-(\theta^*)^2)}\right) \tag{4.256}$$

We can write therefore that[9]

$$\begin{aligned} R_{\mathcal{G}}(d_1,d_2) = \text{minimize} \quad & \frac{1}{2}\log\left(\frac{1-\rho^2}{d_1 d_2(1-(\theta^*)^2)}\right) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\ \text{subject to} \quad & \sum_{i=1}^{2} s_i^{(2)} 2^{-2r_i^{(2)}} = a - \lambda_1, \\ & \sum_{i=3}^{4} s_i^{(2)} 2^{-2r_i^{(2)}} = b - \lambda_2 \\ & r_i^{(L)} \geq 0, i = 1,...,2^L. \end{aligned} \tag{4.257}$$

[9]It can be noted here that this optimization problem can be decoupled into two subproblems. In the first one we would minimize $\sum_{l=2}^{L}\sum_{i=1}^{2^{l-1}} r_i^{(l)}$ subject to $\sum_{i=1}^{2} s_i^{(2)} 2^{-2r_i^{(2)}} = a - \lambda_1$ and in the second, we minimize $\sum_{l=2}^{L}\sum_{i=2^{l-1}+1}^{2^l} r_i^{(l)}$ subject to $\sum_{i=3}^{4} s_i^{(2)} 2^{-2r_i^{(2)}} = b - \lambda_2$.

Now recall that

$$\begin{aligned} R_{\text{vec}}(\boldsymbol{D}_{\theta^*}) = \text{minimize} \quad & \frac{1}{2}\log\left(\frac{|\boldsymbol{K}_{x^{(1)}}|}{|\boldsymbol{D}_{\theta^*}|}\right) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\ \text{subject to} \quad & \sum_{i=1}^{2} s_i^{(2)} 2^{-2r_i^{(2)}} \leq \frac{\boldsymbol{K}_{v^{(2)}}(2,2)}{|\boldsymbol{K}_{v^{(2)}}|} - \frac{d_2}{|\boldsymbol{D}_{\theta^*}|}, \\ & \sum_{i=3}^{4} s_i^{(2)} 2^{-2r_i^{(2)}} \leq \frac{\boldsymbol{K}_{v^{(2)}}(1,1)}{|\boldsymbol{K}_{v^{(2)}}|} - \frac{d_1}{|\boldsymbol{D}_{\theta^*}|} \\ & r_i^{(l)} \leq f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \quad l = 2, ..., L-1 \quad i = 1, ..., 2^l, \\ & r_i^{(l)} \geq 0 \quad l = 2, ..., L \quad i = 1, ..., 2^l. \end{aligned}$$

Recall also that

$$\boldsymbol{K}_{v^{(2)}}^{-1} = \boldsymbol{K}_{x^{(1)}|x^{(2)}}^{-1} = \boldsymbol{K}_{x^{(1)}}^{-1} + \text{diag}(a, b). \tag{4.258}$$

This shows

$$\frac{\boldsymbol{K}_{v^{(2)}}(2,2)}{|\boldsymbol{K}_{v^{(2)}}|} - \frac{d_2}{|\boldsymbol{D}_{\theta^*}|} = \frac{1}{1-\rho^2} + a - \frac{1}{d_1(1-(\theta^*)^2)} = a - \lambda_1 \tag{4.259}$$

$$\frac{\boldsymbol{K}_{v^{(2)}}(1,1)}{|\boldsymbol{K}_{v^{(2)}}|} - \frac{d_1}{|\boldsymbol{D}_{\theta^*}|} = \frac{1}{1-\rho^2} + b - \frac{1}{d_2(1-(\theta^*)^2)} = b - \lambda_2. \tag{4.260}$$

That is

$$\begin{aligned} R_{\text{vec}}(\boldsymbol{D}_{\theta^*}) = \text{minimize} \quad & \frac{1}{2}\log\left(\frac{1-\rho^2}{d_1 d_2(1-(\theta^*)^2)}\right) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} r_i^{(l)} \\ \text{subject to} \quad & \sum_{i=1}^{2} s_i^{(2)} 2^{-2r_i^{(2)}} \leq a - \lambda_1, \\ & \sum_{i=2}^{4} s_i^{(2)} 2^{-2r_i^{(2)}} \leq b - \lambda_2, \\ & r_i^{(l)} \leq f_{x_i^{(l)}}(r_{2i-1}^{(l+1)}, r_{2i}^{(l+1)}), \quad l = 2, ..., L-1 \quad i = 1, ..., 2^l, \\ & r_i^{(l)} \geq 0 \quad l = 2, ..., L \quad i = 1, ..., 2^l. \end{aligned} \tag{4.261}$$

We start by showing that $R_{\mathcal{G}}(d_1, d_2) \geq R_{\text{vec}}(\boldsymbol{D}_{\theta^*})$. By comparing (4.257) and (4.261), we can clearly see that $(o_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l)$ is in the feasible set

of the optimization problem defining $R_{\text{vec}}(\boldsymbol{D}_{\theta^*})$. Hence

$$R_{\mathcal{G}}(d_1, d_2) = \frac{1}{2}\log\left(\frac{1-\rho^2}{d_1 d_2(1-(\theta^*)^2)}\right) + \sum_{l=2}^{L}\sum_{i=1}^{2^l} o_i^{(l)} \tag{4.262}$$

$$\geq R_{\text{vec}}(\boldsymbol{D}_{\theta^*}). \tag{4.263}$$

We would like now to show the reverse inequality. Let $(b_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l)$ be the optimal solution to the convex optimization problem defining $R_{\text{vec}}(\boldsymbol{D}_{\theta^*})$ in (4.261). To prove that $R_{\mathcal{G}}(d_1, d_2) \leq R_{\text{vec}}(\boldsymbol{D}_{\theta^*})$, we will show that $(b_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l)$ lies in the feasible set of the optimization problem defining $R_{\mathcal{G}}(d_1, d_2)$. The optimization problem in (4.261) is convex, we can therefore use the KKT conditions to identify $(b_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l)$. The KKT conditions for this problem state that there exists $(\varrho_1, \varrho_2, \varsigma_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l) \succeq 0$ and $(\upsilon_i^{(l)}, l = 2, ..., L-1, i = 1, ..., 2^l) \succeq 0$ such that

$$0 = 1 - \varsigma_i^{(2)} + \upsilon_i^{(2)} - \varrho_1 s_i^{(2)} 2^{-2b_i^{(2)}}, \quad i = 1, 2 \tag{4.264a}$$

$$0 = 1 - \varsigma_i^{(2)} + \upsilon_i^{(2)} - \varrho_2 s_i^{(2)} 2^{-2b_i^{(2)}}, \quad i = 3, 4 \tag{4.264b}$$

$$0 = 1 - \varsigma_i^{(l)} + \upsilon_i^{(l)} - \upsilon_{\lceil \frac{i}{2} \rceil}^{(l-1)} s_i^{(l)} \sigma^2_{z^{(l-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2b_i^{(l)}} 2^{-2f_{x^{(l-1)}_{\lceil \frac{i}{2} \rceil}}}, \quad l = 3, ..., L-1, i = 1, ..., 2^l, . \tag{4.264c}$$

$$0 = \varrho_1 (a - \lambda_1 - \sum_{i=1}^{2} s_i^{(2)} 2^{-2b_i^{(2)}}) \tag{4.264d}$$

$$0 = \varrho_2 (b - \lambda_2 - \sum_{i=3}^{4} s_i^{(2)} 2^{-2b_i^{(2)}}) \tag{4.264e}$$

$$0 = 1 - \varsigma_i^{(L)} - \upsilon_{\lceil \frac{i}{2} \rceil}^{(L-1)} s_i^{(L)} \sigma^2_{z^{(L-1)}_{\lceil \frac{i}{2} \rceil}} 2^{-2b_i^{(L)}} 2^{-2f_{x^{(L-1)}_{\lceil \frac{i}{2} \rceil}}}, \quad i = 1, ..., 2^L \tag{4.264f}$$

$$0 = \upsilon_i^{(l)} \left(b_i^{(l)} - f_{x_i^{(l)}}(b_{2i-1}^{(l+1)}, b_{2i}^{(l+1)}) \right), l = 2, ..., L-1, i = 1, ..., 2^l \tag{4.264g}$$

$$0 = \varsigma_i^{(l)} b_i^{(l)}, \quad l = 2, ...L, i = 1, ..., 2^l. \tag{4.264h}$$

Assume $\varrho_1 = 0$, then from (4.264a) we must have $\varsigma_i^{(2)} > 0$ for $i = 1, 2$. From

complementary slackness this implies that $b_1^{(2)} = b_2^{(2)} = 0$. This is not possible[10] since we must have $a - \lambda_1 \geq \sum_{i=1}^{2} s_i^{(2)} 2^{-2b_i^{(2)}}$. This proves that $\varrho_1 > 0$. A similar argument shows that $\varrho_2 > 0$. Since $\varrho_1, \varrho_2 > 0$, we conclude from (4.264d) and (4.264e) that $a - \lambda_1 = \sum_{i=1}^{2} s_i^{(2)} 2^{-2b_i^{(2)}}$ and $b - \lambda_2 = \sum_{i=3}^{4} s_i^{(2)} 2^{-2b_i^{(2)}}$.

We will verify now that

$$f_{x_i^{(l)}}(b_{2i-1}^{(l+1)}, b_{2i}^{(l+1)}) = b_i^{(l)}, l = 2, ..., L-1, i = 1, ..., 2^l. \tag{4.265}$$

First, if $\upsilon_i^{(l)} > 0$, then (4.265) will simply follow from (4.264g). Now assume $\upsilon_i^{(l)} = 0$, then from (4.264c) we deduce that $\varsigma_{2i-1}^{(l+1)}, \varsigma_{2i}^{(l+1)} > 0$ and from complementary slackness $b_{2i-1}^{(l+1)} = b_{2i}^{(l+1)} = 0$. Since $0 \leq b_i^{(l)} \leq f_{x_i^{(l)}}(b_{2i-1}^{(l+1)}, b_{2i}^{(l+1)})$, we will have $b_i^{(l)} = 0$ and hence (4.265) is trivially satisfied. This proves that $(b_i^{(l)}, l = 2, ..., L, i = 1, ..., 2^l)$ is in the feasible set of the optimization problem in (4.257) and therefore

$$R_{\mathcal{G}}(d_1, d_2) \leq \frac{1}{2} \log \left(\frac{1 - \rho^2}{d_1 d_2 (1 - (\theta^*)^2)} \right) + \sum_{l=2}^{L} \sum_{i=1}^{2^l} b_i^{(l)}, \tag{4.266}$$

$$= R_{\text{vec}}(\boldsymbol{D}_{\theta^*}). \tag{4.267}$$

This combined with the previous inequality shows that $R_{\mathcal{G}}(d_1, d_2) = R_{\text{vec}}(\boldsymbol{D}_{\theta^*})$.

[10]Recall that $\lambda_1, \lambda_2 > 0$ which is a consequence of the assumption $\max(d_1, d_2) < \rho^2 \min(d_1, d_2) + 1 - \rho^2$.

BIBLIOGRAPHY

[1] R. Negi and S. Goel, "Secret communication using artificial noise," in *Proc. Vehic. Tech. Conf.*, Dallas, TX, pp. 1906-1910, September 2005.

[2] S. Goel and R. Negi, "Secret communication in presence of colluding eavesdroppers," in *Proc. IEEE Military Commun. Conf.*, Atlantic City, NJ, pp. 1501-1506, 2005.

[3] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. B. Mandayam, "Information- theoretically secret key generation for fading wireless channels," *IEEE Trans. Inf. Forensics and Security*, vol. 5, no. 2, pp. 240-254, June 2010.

[4] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355-1387, October 1975.

[5] J. R. Barry, "Wireless Infrared Communications", *Kluwer Academic Publishers*, Boston, MA, 1994.

[6] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, pp. 265-298, Feburary 1997.

[7] R. J. Green, "Secure communications: the Infrared alternative," in *ICTON Mediterranean Winter Conference*, 2007.

[8] Z. Xu and B. M. Sadler, "Ultraviolet communications: Potential and state-of-the-art," *IEEE Communications Magazine*, pp. 67-73, May 2008.

[9] S. Adee, "Ultraviolet radios beam to life," *IEEE Spectrum*, May 2009.

[10] I. Csisár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339-348, May 1978.

[11] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas: The MIMOME channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2547-2553, June 2009.

[12] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," in *Proc. IEEE Int. Symp. Information Theory*, Toronto, ON, Canada, pp. 524-528, July 2008.

[13] T. Liu and S. Shamai (Shitz), "A note on the secrecy capacity of the multi-antenna wiretap channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2547-2553, June 2009.

[14] P. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4687-4698, October 2008.

[15] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2453-2469, June 2008.

[16] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and its Appl.*, vol. 23, pp. 143-147, 1978.

[17] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inf. Theory*, vol. 26, pp. 710-715, November 1980.

[18] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel: Part I," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1449-1461, November 1988.

[19] A. Lapidoth and S. Shamai (Shitz), "The Poisson multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, pp. 488-502, March 1998.

[20] S. I. Bross, M. V. Burnashev and S. Shamai (Shitz), "Error exponents for the two-user Poisson multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1999-2016, July 2001.

[21] A. Lapidoth, E. Telatar and R. Urbanke, "On wide-band broadcast channels", *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3250-3258, December 2003.

[22] A. Sokolovsky and S.I. Bross, "Attainable error exponents for the Poisson broadcast channel with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 364-374, January 2005.

[23] S. M. Haas and J. H. Shapiro, "Capacity of wireless optical communications," *IEEE Journal. Selec. Areas Comm.*, vol. 21, no. 8, pp. 1346-1357, October 2003.

[24] K. Chakraborty and P. Narayan, "The Poisson fading channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2349-2364, July 2007.

[25] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451-456, July 1978.

[26] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072-3081, November 2001.

[27] D. Guo, S. Shamai (Shitz) and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels", *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1837-1849, May 2008.

[28] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261-1283, April 2005.

[29] E. Ekrem, and S. Ulukus "Secrecy Capacity Region of the Gaussian Multi-Receiver Wiretap Channel," in *Proc. of IEEE International Symposium on Information Theory*, Seoul, Korea, June 2009.

[30] R. Bustin, R. Liu, H. V. Poor and S. Shamai (Shitz), "An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel," *EURASIP Journal on Wireless Communications and Networking*, 2009.

[31] S. K. Leung-Yan-Cheong, "On a special class of wiretap channels," *IEEE Trans. Inf. Theory*, vol. 23, no. 5, pp. 625-627, September 1977.

[32] M. Van Dijk, "On a special class of broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 712-714, Mars 1997.

[33] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562-1575, April 2006.

[34] M. Bloch and J. N. Laneman, "On the secrecy capacity of arbitrary wiretap channels," in *Proc. of 46th Allerton Conference on Communication, Control, and Computing* Monticello, IL, pp. 818-825, September 2008.

[35] A.D. Wyner, "A definition of conditional mutual information for arbitrary ensembles", *Inform. Contr.*, vol. 38, pp. 51-59, 1978.

[36] P. Brémaud, Point Processes and queues: Martingale Dynamics, Springer-Verlag, New York, 1981.

[37] M. Krein and A. Nudelman, "The Markov moment problem and extremal problems," *Translations of Mathematical Monographs Providence*, RI: Amer. Math. Soc., vol. 5, 1977.

[38] K. Chakraborty, S. Dey, M. Franceschetti, "Outage capacity of MIMO Poisson fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4887-4907, November 2008.

[39] http://www.computer.org/portal/web/computingnow/internet40/login

[40] S. Verdú, "The Exponential Distribution in Information Theory," *Problems of Information Transmission*, vol. 32, no. 1, pp. 86-95, January-March 1996.

[41] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 1290-1301, January 1996.

[42] B. P. Dunn, M. Bloch, and J. N. Laneman, "Secure bits through queues," in *Proc. IEEE Information Theory Workshop*, Volos, Greece, June 2009.

[43] B. Prabhakar, R. Gallager, "Entropy and the timing capacity of discrete queues," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 357-370, February 2003.

[44] V. Anantharam and S. Verdú, "Reflections on the 1998 information theory society paper award: bits through queues," *IEEE Information Theory Society Newsletter*, vol. 49, no. 4, Dec. 1999.

[45] P. Burke, "The output of a queueing system," *Operations Reserach*, vol. 4, pp. 699704, 1956.

[46] R. Sundaresan and S. Verdú, "Capacity of queues via point-process channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2697-2709, June 2006.

[47] A. B. Wagner and V. Anantharam, "Zero-rate reliability of the exponential-server timing channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 447-465, Feb. 2005.

[48] P. Brémaud, "Point processes and queues: martingale dynamics," New York: Springer-Verlag, 1981.

[49] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observation," *Studia scientiarum mathematicarum Hungarica*, vol. 2, pp. 229-318, 1967.

[50] A. S. Bedekar and M. Azizoğlu, "The information-theoretic capacity of discrete-time queues," *IEEE Trans. Inform. Theory*, vol. 44, pp. 446461, 1998.

[51] R. Sundaresan and Sergio Verdú, "Sequential decoding for the exponential server timing channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 705-709, March 2000.

[52] E. Arikan, "On the reliability exponent of the exponential timing channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1681-1689, June 2002.

[53] G. Nakibly and S. I. Bross, "On the reliability exponents of two discrete-time timing channel models," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp.4320-4335, Spetember 2006.

[54] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate Region of the Quadratic Gaussian Two-Encoder Source-Coding Problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938-1961, May 2008.

[55] S. Tavildar, P. Viswanath, and A. B. Wagner, "The Gaussian Many-Help-One Distributed Source Coding Problem," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 564-581, January 2010.

[56] Y. Oohama, "Distributed Source Coding of Correlated Gaussian Sources," *IEEE Trans. Inf. Theory*, submitted, available at http://arxiv.org/abs/1007.4418.

[57] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471-480, July 1973.

[58] J. Wang, J. Chen and X. Wu, "On the minimum sum rate of Gaussian multiterminal source coding: New proofs and results," *IEEE Trans. Inf. Theory*, vol. 56, No.8, pp. 3946-3960, Aug. 2010.

[59] Y. Yang and Z. Xiong, "The generalized quadratic Gaussian CEO problem: New cases with tight rate region and applications," in *Proc. ISIT*, pp. 21-25, June 2010.

[60] T. Berger, "Multiterminal source coding," in The Information Theory Approach to Communications, ser. CISM Courses and Lectures, G. Longo, Ed. Springer-Verlag, 1978, vol. 229, pp. 171-231.

[61] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Electrical Engineering, Cornell University, Ithaca, NY, May 1978.

[62] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577-2593, July 2005.

[63] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2762-2768, November 2004.

[64] A. B. Wagner and V. Anantharam, "Improved outer bound for multiterminal source coding," *IEEE Transactions on Information Theory*, vol. 54, pp. 1919-1937, May 2008.

[65] A. H. Sayed, "Fundamentals of Adaptive Filtering," Wiley, New York, USA, 2003.

[66] J. Nocedal and S. J. Wright, "Numerical Optimization," Second Edition, Springer Verlag, New York, 2006.