Communication and Third Parties: Costs, Cues, and Confidentiality

Krish Eswaran

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2009-142
http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-142.html

October 19, 2009

Communication and Third Parties: Costs, Cues, and Confidentiality

by

Krishnan Eswaran

B.S. (Cornell University) 2003
M.S. (University of California, Berkeley) 2005

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Philosophy

in

Engineering—Electrical Engineering and Computer Sciences

and the Designated Emphasis

in

Communication, Computation, and Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael Gastpar, Chair
Professor Kannan Ramchandran
Professor Sourav Chatterjee

Fall 2009

The dissertation of Krishnan Eswaran is approved:

Chair _____ Date _____

_____ Date _____

_____ Date _____

University of California, Berkeley

Communication and Third Parties: Costs, Cues, and Confidentiality

by

Krishnan Eswaran

Abstract

Communication and Third Parties: Costs, Cues, and Confidentiality

by

Krishnan Eswaran

Doctor of Philosophy in Engineering—Electrical Engineering and Computer Sciences

and the Designated Emphasis

in

Communication, Computation, and Statistics

University of California, Berkeley

Professor Michael Gastpar, Chair

This thesis tells the story of Alice and Bob, who wish to share information. However, Alice and Bob, motivated either by self-interest or simply good manners, also wish to respect constraints imposed by Candice, Damon, or Eve, who may be lurking in the background. This thesis takes an information-theoretic approach to determine the rates at which Alice and Bob can communicate while satisfying the constraints imposed by these third parties. The constraints in this thesis are inspired by wireless communication, in which multiple users share a common medium.

Signals over a common medium can interfere with one another, so a case study is introduced that models one aspect of this problem: Alice must send a message to Bob such that she does not interfere with communication from Candice to Damon. Furthermore, Candice and Damon are unaware of Alice and Bob, so there is no direct way for Alice to learn how much interference she causes. However, by relying on cues gained from eavesdropping on feedback signals from Damon to Candice, strategies are introduced that enable Alice to communicate with Bob and adaptively control

the average interference.

While the case study only accounts for the average interference, Damon may be more annoyed by interference arriving in bursts than interference spread out over time. To address this issue, a new model is introduced that forces Alice's transmissions to satisfy a cost constraint, where the cost at a given time can include memory about the past. A strategy is introduced that enables Alice to adjust her transmissions to communicate with Bob and satisfy the cost constraint. A converse is also proved that characterizes when this strategy is optimal.

Wireless environments allow characters like Eve, an eavesdropper, to intercept signals that are not intended for them. A model is considered in which Alice and Bob wish to keep the information they share secret from Eve. Specifically, Alice, Bob, and Eve share correlated observations, and there is a one-way noisy broadcast channel from Alice to Bob and Eve. Do Alice and Bob want to keep specific information secret, or is any information kept secret from Eve sufficient? It turns out that the rates Alice and Bob can achieve depend on the answer to this question, and a strategy is introduced that establishes a functional relationship between these two notions of secrecy. Furthermore, this strategy is shown to be optimal over a class of models that includes one applicable to wireless: namely, the case in which the channel noises are additive Gaussian and the correlated observations are jointly Gaussian.

Professor Michael Gastpar, Chair Date

Acknowledgements

This thesis owes a lot to the guidance of my two advisors: Michael Gastpar and Kannan Ramchandran. I met both of them during Berkeley's Visit Day, and my conversations with them during and after Visit Day were instrumental in my decision to come to Berkeley. Since then, both have provided complementary advising styles that have served me really well throughout the different stages of my graduate career. In particular, I want to thank Kannan for pushing me to carefully consider the motivations for and relevance of a research problem. Similarly, I want to thank Michael for his patience and for helping me keep up morale as I sometimes struggled to formulate and tackle research problems. Finally, I want to thank both of them for helping me communicate technical information in a clear and concise manner. I hope this thesis does justice to their efforts.

The comments and suggestions of Sourav Chatterjee, who was on both my quals committee and thesis committee, were very helpful in revising the presentation of the material found therein. Anant Sahai, who was also on my quals committee, gave his suggestions and input about several of the problems found in these chapters. I have also benefited from classes and meeting with other faculty in the department: David Tse, Venkat Anantharam, Martin Wainwright, Abhay Parekh, Howard Shelanski, and Christos Papadimitriou.

The work in Chapter 2 was initially inspired by meetings of the Spectrum Sharing Group, which Kannan and Michael both encouraged me to attend during my first year of graduate school. The initial results became the basis for my qualifying exam and subsequent feedback from my committee, which included Anant Sahai and Sourav Chatterjee (Stat, UC Berkeley), helped improve the direction of the work. The work also benefited from discussions with Mubaraq Mishra and comments from reviewers

of a journal submission of the paper.

The motivations for Chapter 3 were in a series of discussions Michael and I had about when feedback provides a significant boost to the capacity of a channel. The initial work originated at the Information Theory School at Penn State in State College, PA. In fact, it was partially inspired by a brainteaser that Paul Cuff, then a graduate student at Stanford, asked. Ravi Tandon (University of Maryland – College Park), Pulkit Grover (Berkeley), and Amin Gohari (Berkeley) provided invaluable suggestions for the converse. A stimulating discussion with Venkat Anantharam is also gratefully acknowledged. Finally, I would like to thank Gireeja Ranade for carefully reading through several versions of this chapter, pinpointing errors in the proofs, and suggesting ways to improve the exposition of the work.

The work in Chapter 4 originated from a discussion with Kannan about secrecy. He mentioned some results about secrecy that he and Vinod Prabhakaran (UIUC) had presented at Allerton in 2007, which suggested an interesting separation architecture that combines the benefits of correlated source observations and a noisy broadcast channel for attaining secrecy. Kannan was curious to see whether such an architecture is always optimal or not, and subsequent work led to several fruitful collaborations with Vinod. Prakash Narayan (University of Maryland – College Park) provided several useful comments after seeing our presentation at ISIT 2008, which helped shape the direction of the research. I had similarly fruitful discussions with Sennur Ulukus (University of Maryland – College Park) and her students, as well as Aylin Yener (Penn State). The writing and exposition of the results presented in this chapter improved significantly after suggestions from Nebojsa Milosavljevic, who read several preprints from the research collaboration with Vinod.

While Berkeley can be bureaucratic at times, Ruth Gjerde and Mary Byrnes made this seem like a unsubstantiated rumor. I especially want to acknowledge Ruth, who provided invaluable help that went above and beyond the call of duty. For

instance, when I applied for a job at an unnamed company a few months ago, Ruth volunteered to proctor an exam required by that company after the designated person was unavailable.

Amy Ng, Kim Kail, Sim Kallan, and Annie Owen helped simplify life at the Eugene and Joan C. Wong Center for Communications Research, sometimes called Wireless Foundations or Wifo. I also want to thank Therese George and Loretta Lutcher for sharing their laughter, stories, and smiles.

It is difficult to quantify the type of support I have received from family over the past six years, but I think my life may have been starkly different up to this point without it. To my parents, P.K. and Gowri; Sandhya, Alex, and Maya; Lipica, Rajesh, Taarini, and Rudra; Aarthi and Aroon; Sai Ram; Nandini; Chinna Paati and Jagadish Thatha; Sumathy Periamma and Mani Periappa; Krishna Mama and Savithri Manni; Jayashree Aunty, Rao Uncle, Namrata, and Kirtana; Priya, Ganesh, Ananya, and Nitya; Nalini, Kumaran, Shravan, and Samyuktha; and so many others, thank you.

There are many people to thank for collaborations, interesting discussions, and stress relief. I've had several stimulating conversations with Bobak Nazer, Paolo Minero, and Anand Sarwate. In fact, a class project Anand and I collaborated on led to my first journal paper. Additionally, the figures contained in Chapter 1 all derive from figures created by Anand. The BASICS and SIPC groups provided me with a large academic family: Allie Fletcher, Abhik Majumdar, Dan Schonberg, Vinod Prabhakaran, Animesh Kumar, Wei Wang, Mark Johnson, Dan Hazen, June Wang, Ben Wild, Chuohao Zhao, Hao Zhang, Aline Roumy, Bobak Nazer, Anand Sarwate, Jérémie Ezri, Galen Reeves, Jiening Zhan, Birsen Sirkeci, Nebojsa Milosavljevic, and Naveen Goela.

I left my summer at the Broad Institute in Cambridge, MA with a different perspective about research, and there are several people to thank for making that ex-

perience memorable: Desmond Lun, Brian Weiner, James Galagan, Baris Nakigoblu, Mukul Agarwal, Lav Varshney, Kush Varshney, Megan Rokop, Sujay Sanghavi, and others.

Graduate school would have been much less enjoyable without friends, several of whom I met during my first couple years of graduate school: Bobak Nazer, Anand Sarwate, Rahul Tandra, June Wang, Drew Carlson, Niels Hoven, John Secord, Steven Stafford, James Fung, Jingyi Shao, Renaldi Winoto, and Alex Dimakis. Salman Avestimehr, Lenny Grokop, Salman Avestimehr, and Hari Palaiyanur. I learned a lot from senior graduate students and post docs, as well. In particular, Vinod Prabhakaran, Prasad Santhanam, Lara Dolecek, Cheng Chang, Aaron Wagner, and Parv Venkitasubramaniam all served as positive role models. Not all of my friends made it into Cory, and I wanted to single out the following friends outside the department: Neelam Ihsanullah, Julia Nefsky, Alex Rennet, Justin Bledin, Michelle Tremaine, Emily Chu, Omar Nayeem, Ritu Mahajan, Neel Shah, Igor Piryazev, Saurabh Bhargava, Rozy Brar, and Gia Skoumbis.

Finally, I want to thank some friends for making my final year of graduate school special. I want to thank Gireeja Ranade for one of my most memorable experiences from last year: co-teaching a Geometry class at San Quentin. Gireeja also provided endless encouragement and enthusiasm as I worked to finish. Justin Bledin was an excellent roommate who would often have dinner waiting during busier nights of my last semester. Last but certainly not least, Pulkit Grover and Kris Woyach offered the perfect balance of talking about research and taking a break when it became necessary.

To anyone I have forgotten, thank you so much for all your help, and please forgive me for the omission.

To Maya Nandini Bean, for bringing happiness to my life and those of many I love.

Contents

Chapter 1

Introduction

The Concert

The music starts before Alice and Bob find their spot: row 4, seats A and B. The concert has already begun, and Alice frets she may have to wait until it's over to tell Bob the Big News. She briefly considers shouting over the crescendo of music, but she hesitates, looking around the audience. They're enjoying themselves, and Alice does not want to attract any attention.

Should I just wait? Alice thinks to herself. Or maybe I could just whisper to Bob.

"Bob," Alice starts, "there's something–"

Candice, seated to Bob's right, interrupts Alice with a harsh whisper, "Shh!"

Alice stops talking, but it doesn't end there. Candice switches seats with Damon, and grumbling ensues in the rows behind them. As Alice is about to start talking again, she notices that Eve is eyeing her. Will she overhear us? worries Alice. Alice wishes she were better equipped to handle these problems.

1.1 Problem Overview

A classical communication problem is to send the maximum number of bits reliably from Alice to Bob. However, at the concert, third parties like Candice, Damon, and Eve add a dimension of complexity to the vanilla problem of reliably sending bits. This thesis develops models and strategies that enable Alice and Bob to use the resources available to them to address these additional complexities. For instance, if Alice and Bob have a pen and paper at the concert, they might prefer this channel of communication over others that cause a greater disturbance.

The problems faced by Alice and Bob are not uncommon in communication networks. In a cellular network or the Internet, two users want to communicate over a shared medium. The strategy they use to communicate must therefore take into account both the performance of the two users and their behavior within the system. For instance, strategies should mitigate system congestion and provide security for the users. However, there are a couple key differences between these two systems. First, in a cellular network, communication takes place over a wireless medium, but the Internet is largely a wired medium. Second, in a cellular network, the entire system is under the control of a single provider, and the strategies users employ can be tailored tightly to follow these standards.

Since there is no centralized control of the Internet, devices may employ different strategies or have different capabilities. Despite such variation, most devices follow guidelines that ensure the Internet remains stable and functional. At the concert, these guidelines can be thought of as the manners and etiquette that Alice and Bob choose to respect in the presence of the rest of the audience.

The impetus for the topics studied in this thesis, however, do not come from the wired Internet (nor our desire to help out confused concertgoers), but rather from the wireless setting. On November 4, 2008, the Federal Communications Commission

(FCC), which regulates the spectrum in the United States of America, announced that it would modify its rules to allow adaptable wireless devices to find and reuse unoccupied broadcast television bands[1]. Before this rule change, only the broadcaster with the license for a particular band could use it. The problem was that not all license holders chose to exercise this right, leading to underutilized spectrum. Some consider the rule change a first step to allow unlicensed, adaptable wireless devices to coexist with licensed, legacy communication systems.

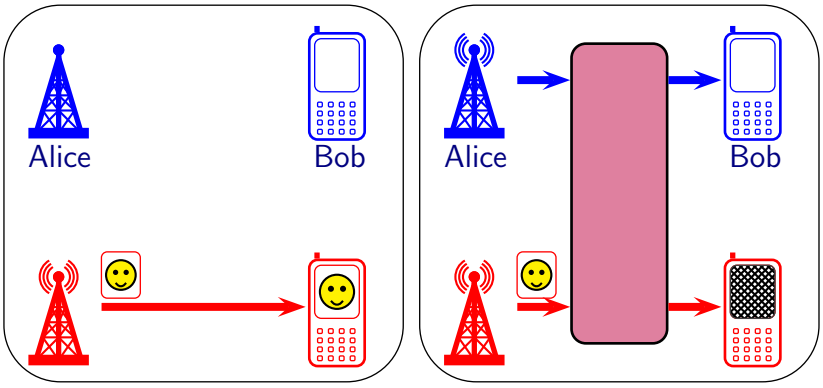

Figure 1.1: The figure on the left depicts a system successfully transmitting the image of a smiley. This legacy system works under the assumption that no other systems will be active. If Alice starts transmitting on the same spectrum, however, her signal can interfere with the legacy system, as seen in the figure on the right. Chapter 2 models and analyzes a scenario in which Alice and Bob must simultaneously communicate while controlling the interference for a specific type of legacy system.

Now suppose Alice and Bob correspond to the transmitter and receiver, respectively, of one of these unlicensed devices, which are sometimes called *cognitive radios*. Alice and Bob must coexist with a legacy system with which their transmissions can potentially interfere, a scenario is depicted in Figure 1.1. The challenge for Alice

[1] "FCC Adopts Rules for Unlicensed Use of Television White Spaces", FCC News, November 4, 2008:
`http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-286566A1.pdf`

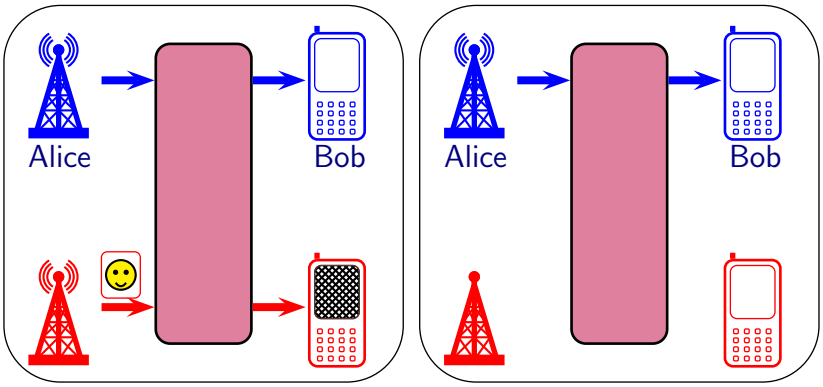

Figure 1.2: The figure on the left depicts Alice and Bob interfering with a system that is trying to transmit a smiley. In the figure on the right, this system stops transmitting in response to the interference, which adds a layer of complexity not captured by the model in Chapter 2. Chapter 3 develops a general model to study how Alice and Bob balance communication with control under such dynamics.

and Bob is that the legacy systems were designed without the expectation that they would be sharing the spectrum with other devices. Thus, without a redesign, the legacy system may not even realize that Alice and Bob are responsible for the interference or indicate to them how much interference they cause. At the concert, this would correspond to a situation in which Candice would not whisper, "Shhh!" to Alice even if Alice were to talk too loudly. However, audience members could still signal that they are unable to hear the performance, which is precisely what several wireless communication systems already do via feedback signals. In Chapter 2, we consider a model for a particular type of legacy system that uses such feedback for which Alice, by overhearing this feedback, can estimate and control how much interference she generates to the legacy system.

The setting in Chapter 2 develops a model to study a special case that may arise in cognitive radio, but it lacks generality. At the concert, recall that Candice and Damon switch seats in response to Alice and Bob. If Candice and Damon correspond

to wireless devices, the model in Chapter 2 is too limited to capture any interference generated by their dynamics. Alternatively, what if the musicians were to respond to Alice and Bob's interference by stopping the concert? If the musicians correspond to the transmitter of a legacy system, which, as depicted in Figure 1.2, shuts off in response to interference, the model in Chapter 2 is again too limited to capture the dynamics. As a step towards addressing these issues, in Chapter 3, we generalize one of the ideas from that case study: namely, the constraint Alice must satisfy while communicating to Bob. This generalization allows us to consider the richer dynamics that the model in Chapter 2 cannot.

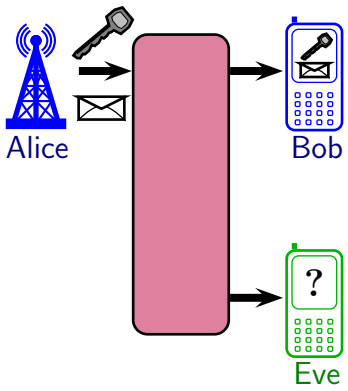

Figure 1.3: Chapter 4 considers how much secret information Alice and Bob can share in the presence of Eve, an eavesdropper. It turns out the answer depends not only the resources available at Alice, Bob, and Eve, but also on the type of information that Alice and Bob plan to share.

The above settings draw on the fact that wireless systems by nature affect unintended recipients in a surrounding area, and some of these recipients may consider the effect to be interference. However, interference to one party may be valuable to another, and Alice and Bob may intend the information they share to be secret from an eavesdropper Eve. Chapter 4 explores a model in which Alice, Bob, and Eve have access to correlated sensor observations in addition to the wireless channel and given

these resources, considers how Alice and Bob may share secret information. It turns out the nature of the information that Alice and Bob intend to share plays a role in answering this question. For instance, one may take the aphorism that "a bit is a bit" seriously, and it does not matter what bits Alice and Bob share so long as they are secret from Eve. This scenario may arise practically when Alice and Bob want to share a *key* that can later be used for encryption. On the other hand, Alice may have a particular *message* to send to Bob, and thus Alice and Bob want to keep a specific sequence of bits secret from Eve. A schematic figure of this setting is given in Figure 1.3.

1.2 Methodology

Within the context of each problem stated above, our goal in this thesis is to answer the following question: is it feasible for Alice to convey the Big News to Bob before the concert ends, and if so, how can she achieve this? To make progress on this question, we develop models that derive from information theory, a field of applied mathematics for which there exists a significant body of literature on how to characterize and achieve the fundamental limits of data transmission over noisy communication channels.

Many of these models are generalizations of Claude Shannon's seminal work, "A Mathematical Theory of Communication," which introduces the problem of sending data reliably over a noisy communication channel [78]. For this problem, which is depicted in Figure 1.4, Shannon proves that given any message in a finite set and sufficient uses of a noisy channel, one can design an encoder that transforms the given message into the channel's inputs and a decoder that transforms the channel's outputs into a reconstructed message such that the reconstructed message matches the

Figure 1.4: Shannon's original problem considers how to maximize the size of the set of possible transmitted messages to transform a transmitted message M into a sequence of channel inputs $(X_1, X_2, \ldots, X_N)$ and a sequence of channel outputs $(Y_1, Y_2, \ldots, Y_N)$ into a reconstructed message $\hat{M}$ such that the transmitted and reconstructed messages match with high probability as $N \to \infty$.

given message with probability arbitrarily close to one.[2] Furthermore, the cardinality of the message set can grow exponentially in the number of channel uses if and only if the exponent is less than the *capacity*, a nonnegative number that depends solely on the channel's statistics. Motivated by scenarios of physical interest, in the same work Shannon proves an extension of the result for an analog channel subject to a power constraint on the channel inputs. For this setting, Shannon shows the capacity depends on both the channel's statistics and the power constraint.

In the original problem, the source of interest has a single point of origin and a single destination (i.e., *point-to-point*), but generalizations of the problem have been studied for networks with multiple sources, destinations, and intermediate nodes, sometimes called *multiterminal* problems (see e.g. [1; 86]). The rate region of the multiple-access channel, in which there are multiple sources and a single destination, was characterized by Liao [54] as well as Slepian and Wolf [79]. Unlike the multiple-access channel, the rate region is generally unknown for most networks. The relay

[2]While Shannon proves the existence of such encoders and decoders, the work does not give explicit constructions for them. Indeed, an intense focus since the work has been to produce explicit constructions that approach the capacity bound (see e.g. [6; 15; 55]).

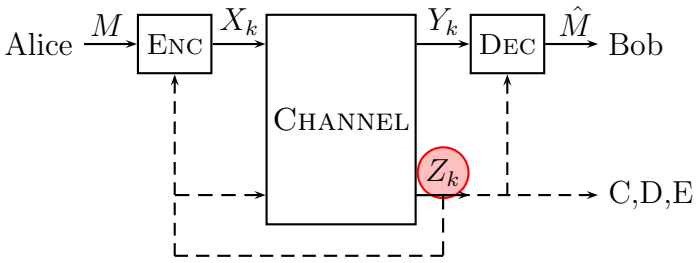

Figure 1.5: This is a generic block diagram to characterize the problems studied in this thesis. In these problems, Alice wants to send Bob a message while satisfying some constraint imposed by a third-party C,D,E. The dashed lines indicate that depending on the model, Alice or Bob may be aware of the channel outputs at the third-party.

channel, introduced by Van Der Meulen [85], considers a single source and destination with an intermediate helper node, and partial results for the rate region have been given by Cover and El Gamal [19], Kramer, Gastpar, and Gupta [52], and Avestimehr, Diggavi, and Tse [4], among others [31; 41; 94]. The broadcast channel, introduced by Cover [17], considers a single source and multiple destinations. While the capacity region is known for some special cases, it remains open in general [50; 59; 30], and many of the exact results are restricted to channels with additive Gaussian noise [18; 88; 90]. Finally, the interference channel, introduced by Ahlswede [2], considers two sources and two destinations in its simplest form. Despite a sophisticated achievable strategy introduced by Han and Kobayashi [45], the capacity region remains open with only an approximate result in the Gaussian case [37].

The models developed in this thesis are most closely related to the broadcast channel. For the classical broadcast channel, Alice has messages for Bob and Betty, who are located at separate destinations, and the three of them must devise a coding

strategy such that Bob and Betty can decode their intended messages. The work in this thesis replaces Betty with an unintended recipient, a third party like Candice, Damon, or Eve, whose channel outputs introduce an additional constraint on the communication between Alice and Bob. Figure 1.5 provides a generic block diagram to capture the several variations of this setting.

1.3 Contributions

As stated above, this thesis focuses on a series of theoretical studies to gain insights into a class of communication problems. Before delving into these theoretical studies, we briefly summarize the main contributions of the thesis:

- In Chapter 2, we develop a communication model that derives from a case study of cognitive radio. In the case study, a cognitive radio Alice must communicate to Bob while simultaneously guaranteeing a target rate to a legacy communication system over whose spectrum she transmits. We provide coding strategies that adapt the cognitive radio's *duty cycle*[3] and prove that for these strategies, the legacy system achieves its target rate. Furthermore, under certain interference conditions, we prove these strategies are rate-optimal for the cognitive radio, as well.

- In Chapter 3, we introduce a feedback communication model that generalizes aspects of the cognitive radio model from Chapter 2. For this model, Alice must communicate to Bob while simultaneously controlling a state process to satisfy a cost constraint. For this setting, there is a tension between communication and control: namely, if communication is ignored, the setup reduces to a classical stochastic control problem, and if satisfying the cost constraint is

[3]A duty cycle refers to the fraction of time a system is active.

ignored, it reduces to a well-understood feedback communication problem. We find an expression that characterizes the tradeoff between these two extremes and demonstrate how to evaluate it via a series of examples.

- In Chapter 4, we provide an achievable tradeoff for sending a secret message and secret key when there are correlated source observations at Alice, Bob, and Eve, as well as a one-way broadcast channel from Alice to Bob and Eve. The strategy has an interesting *separation* architecture that highlights the interplay between a secret key and a secret message. Furthermore, for a class of sources and channels, we show this strategy is optimal and establish a formal relationship between these different notions of secret information sharing.

These contributions are stated precisely in Chapters 2, 3, and 4, which assume the reader has some knowledge of information theory. For readers uncomfortable with information theory, Appendix A should provide the background necessary to understand the contents of these chapters. For everyone else, we now proceed with the technical portion of this thesis.

Chapter 2

A Case Study from Cognitive Radio

2.1 Introduction

In this chapter, we consider a problem motivated by cognitive radio to address the case in which a cognitive radio must communicate while simultaneously satisfying a random cost constraint to control the amount of interference it generates on a legacy system. In this setting, we demonstrate how feedback enables the cognitive radio to dynamically adapt its channel inputs and rate by monitoring the empirical cost.

2.1.1 Problem Motivation

Systems often need to be designed so that they do not disrupt pre-existing systems with which they interact. This backwards compatibility problem is a central issue in the study of cognitive radio systems. A cognitive radio is a device that can sense and adjust its power, frequency band, etc. to peacefully coexist with other radios with which it shares spectrum [61]. At the time of writing, the FCC and interna-

tional regulatory bodies have started to modify their rules to allow for such systems to occupy unlicensed bands or to share bands with licensed, predesigned communication systems. These licensed users are often called *primaries*, *legacy systems*, or *incumbents*.

The aim of this chapter is to study sharing spectrum with legacy systems, in which the backwards compatibility problem arises. One potential solution is to transmit on a band that is currently unoccupied and to leave that band once a primary is detected. For these "detect-and-avoid" systems, one research aim is to understand the feasibility of detecting the presence of a primary system subject to noise uncertainty and quantization effects [81; 72; 82].

A different approach is for the cognitive radio to occupy bands on which the primary is already active but in such a way as to mitigate the interference generated on the primary system. Two such information-theoretic models have been introduced to study cognitive radio and spectrum sharing systems. The first is sometimes called the cognitive radio channel [26; 46; 56; 57; 58; 43; 27]. This channel is a variation on the two-user interference channel [2; 74; 9; 16] with the modification that the cognitive radio (one of the transmitters) knows the message that the primary (the other transmitter) will send. Among these papers, Devroye, Mitran, and Tarokh [26] as well as Jovičić and Viswanath [46] consider a Gaussian scenario in which the primary's strategy can be thought of as a fixed, predesigned legacy system. Specifically, they show that for their setup, there is an optimal achievable strategy that enables the primary to continue using a point to point Gaussian codebook. The result highlights the fact that in cognitive radio problems, one may not have the flexibility to modify the primary's design and must instead design the cognitive radio in such a way that the primary continues to meet its target performance. The second approach is to consider the capacity of systems with a constraint on the interference power generated at certain locations. The assumption is that the primary

systems that occupy these locations will be able to handle this level of interference [39; 42].

We take inspiration from these two models in the following example, which forms the starting point of the current investigation. It is a model of the practically most interesting case, where the cognitive transmitter is close to the primary receiver, thus creating substantial interference. For purposes of illustration, this example assumes a noiseless channel for the cognitive radio. Indeed, the more general model will consider a noisy channel, where the noise may be the result of numerous factors, including interference from the primary system.

Example 2.1. Suppose the primary transmitter Candice sends packets across an erasure channel to her receiver Damon, who responds with feedback to retransmit the packet or send the next one. We call these feedback signals ARQs[1]. The cognitive radio transmitter Alice, on the other hand, has a noiseless channel to her receiver Bob with $P+1$ channel inputs divided into two classes: a *silent symbol* x_{off}, for which Damon successfully receives Candice's transmitted packet, and the remaining P *transmit symbols*, which cause Candice's packet to be erased before reaching Damon. Suppose Candice and Damon want a guaranteed packet rate of $\frac{1}{2}$; that is, Damon should successfully receive one packet per two transmissions on average. By simply alternating channel uses between the silent symbol and sending information with the P transmit symbols, Alice guarantees this packet rate 1/2 to Candice and Damon. Furthermore, Alice and Bob and achieve a rate of

$$\frac{1}{2}\log P\ . \tag{2.1}$$

In the spirit of the previous work, Example 2.1 considers a primary that is unaware

[1]ARQ is an acronym for an *automatic repeat request*, which is feedback that asks a sender to retransmit a packet.

of the cognitive radio. However, Example 2.1 makes the more dubious assumption that the cognitive radio knows what the primary's erasure probabilities are for its two classes of inputs. As a result, the strategy presented is not robust to deviations from the erasure probabilities provided in the example. For instance, if the primary's erasure probability for a silent symbol x_{off} is $\epsilon_0 > 0$, the strategy outlined will not allow the primary to meet its rate $1/2$ target. The issue is that the cognitive radio will not be able to directly estimate the interference it creates for the primary. Such estimates are generally obtained by training via pilot symbols, but the primary receiver is unlikely to train with the cognitive radio transmitter. However, certain kinds of 1-bit feedback have been shown to be sufficient for beamforming [64; 63]. We adopt this insight as we build on Example 2.1 by introducing both an uncertainty and sensing component to the problem.

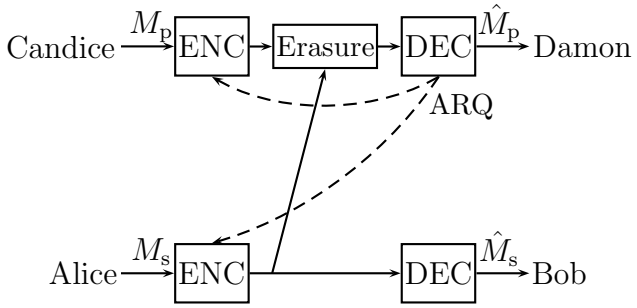

Figure 2.1: An example of the type of channel model for our cognitive radio system (Alice and Bob) in which the primary Candice has message M_{p} for Damon, and the cognitive radio Alice has message M_{s} for Bob. Alice can overhear the ARQ feedback that Damon sends to Candice and adapts her channel inputs to reduce interference to Damon.

Example 2.1, continued: Alice's silent symbol now induces an erasure probability $\epsilon_0 < 1/2$, and her transmit symbols induce an erasure probability of $\epsilon_1 > 1/2$, but Alice does not know either ϵ_0 or ϵ_1. However, Alice can overhear the Damon's ARQ feedback, which we will denote with indicator random variables A_k that test whether

Candice's k-th transmission is received at Damon. Figure 2.1 shows a schematic block diagram of this setup.

Alice's strategy is as follows: if the primaries Candice and Damon exceed their rate target at time k, then Alice sends one of the transmit symbols based on her message. This happens if $\frac{1}{k}\sum_{i=1}^{k} A_k \geq \frac{1}{2}$. Otherwise, Alice sends the silent symbol. Let τ_k be the indicator function that Alice sends a transmit symbol at time k. Thus, $\mathbb{P}(A_k = 1|\tau_k) = 1 - \epsilon_{\tau_k}$. Then at time n, Alice and Bob achieve the rate

$$\frac{1}{n}\sum_{k=1}^{n} \tau_k \log P \ . \tag{2.2}$$

Note that for $\epsilon_0 = 0$, $\epsilon_1 = 1$, this strategy achieves the same rates as the one outlined in Example 2.1.

What can we say about the rate for the primary (Candice and Damon) and cognitive radio (Alice and Bob) in Example 2.1? Let $S_0 = 0$ and $S_k = S_{k-1} + (A_k - 1/2)$ represent the difference between the number of packets the primary has received by time k and its targeted number of packets by time k based on a target packet rate of $\frac{1}{2}$. Suppose the A_k are independent in k. Then S_k is a positive recurrent Markov chain and is nonnegative if and only if $\tau_k = 1$, which can be verified by confirming that its stationary distribution S_k, denoted $\pi_{i/2} = \mathbb{P}(S_k = i/2)$, is

$$\pi_{i/2} = \begin{cases} \frac{(2\epsilon_1 - 1)(1-2\epsilon_0)}{2\epsilon_1(\epsilon_1 - \epsilon_0)} \left(\frac{1-\epsilon_1}{\epsilon_1}\right)^{i} & i \geq 0 \\ \frac{(2\epsilon_1 - 1)(1-2\epsilon_0)}{2(1-\epsilon_0)(\epsilon_1 - \epsilon_0)} \left(\frac{\epsilon_0}{1-\epsilon_0}\right)^{-i+1} & i < 0 \end{cases} \ . \tag{2.3}$$

We can make the following statement.

Fact: Suppose S_0 is distributed according to π. Then for all $k \geq 1$,

$$\mathbb{P}(\tau_k = 1) = \sum_{i=0}^{\infty} \pi_{i/2} = \frac{1/2 - \epsilon_0}{\epsilon_1 - \epsilon_0} \ . \tag{2.4}$$

The fact allows us to get a handle on the cognitive radio's rate. Furthermore, the primary's expected rate is

$$k^{-1} \sum_{i=1}^{k} \sum_{j=0,1} \mathbb{P}(A_i = 1 | \tau_i = j) \mathbb{P}(\tau_i = j) = \frac{1}{2} \ . \tag{2.5}$$

Note that this strategy does not depend on the cognitive radio knowing the values ϵ_0 and ϵ_1 a priori. However, the cognitive radio does know the primary's rate target, which is $1/2$ in this example. In the remainder of the chapter, we assume the primary's rate target is known in advance to the cognitive radio, but the primary's erasure probabilities are unknown.

2.1.2 Bits through ARQs

In the rest of this chapter, we consider optimal coding strategies for the case in which the primary is a packet erasure system as described in Example 2.1.[2] For the channel of the cognitive radio, we consider a more general class of (noisy) channels. As we show, the primary can meet its rate target even if the cognitive radio is active for a certain fraction of channel uses. This *interference budget* available to the cognitive radio, while unknown a priori, can be estimated via the primary ARQs and rate target, which are known at the cognitive radio encoder. One can determine the capacity of the cognitive radio in terms of this interference budget, which we call the

[2]This formulation lends itself well to many spectrum sharing problems in which the primary is a separately designed system and whose exact implementation is partially obscured from the cognitive radio.

rate-interference budget (RIB) tradeoff function. We show an achievable strategy for the general case in which the primary's packet erasure probabilities can fluctuate and find a matching converse for the RIB function when they do not.

In Section 2.2, we define the problem we are considering precisely, including the channel model for the cognitive radio and the allowable coding strategies that the cognitive radio can adopt. These strategies force the cognitive radio to provide guarantees about the primary's rate that do not depend on the time horizon that the cognitive radio uses to measure its own rate (*horzion-independence* condition) and force it to be robust to fluctuations in the primary's packet erasure probabilities (*robustness* condition).

In Section 2.3, we show how to refine the strategy from Example 2.1 to provide such guarantees that also allow positive rate for the cognitive radio, which leads to two new strategies: the *fixed-codebook protocol* and the *codebook-adaptive protocol*. In Section 2.4, we present a converse when the erasure probabilities are time-invariant, which matches the rates achievable by the codebook-adaptive protocol proposed in Section 2.3. Section 2.5 revisits Example 2.1 in the introduction and considers new ones. Section 2.6 concludes the chapter with a discussion of our contributions and future work.

2.2 Problem Setup and Main Result

Capital letters X, Y, Z represent random variables and calligraphic letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ denote finite sets. We will focus on discrete memoryless channels in this work, but potential extensions to Gaussian channels will be discussed in Section 2.6. For convenience, $p(x)$ is the probability distribution of X at x. Similarly, $p(y|x)$ is the conditional probability distribution of Y at y given $X = x$. Notation for entropy $H(X)$, mutual information $I(X;Y)$, etc. are consistent with the notation of Cover

and Thomas [21].

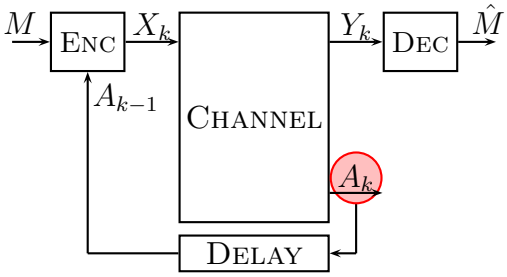

Figure 2.2: Equivalent channel model from the cognitive radio's perspective. A_k is an indicator random variable: $A_k = 1$ means that the packet sent by the primary at time k was successfully received. The relationship between X_k and A_k is unknown a priori and varies in time.

2.2.1 Equivalent Channel Model

As a legacy ARQ system, the primary is assumed to have the following fixed strategy. At time i, it sends a packet to its receiver and receives feedback A_i to indicate whether the packet was erased or successfully received ($A_i = 0$ or 1, respectively). If the packet is erased, the primary retransmits the same packet at time $i+1$. If the packet is successfully received, the primary transmits a new packet at time $i+1$. Thus, we will refer to A_i as the *primary's ARQ feedback*.

Since the primary's strategy is fixed, we now have to design the cognitive radio's strategy. Figure 2.2 illustrates this problem; the primary merely appears as a constraint on the cognitive radio in the shape of A_i. That is, in addition to communicating, the cognitive radio must also control its channel inputs to guarantee the

primary's rate, i.e. such that to the first-order[3],

$$k^{-1}\sum_{i=1}^{k} A_i \geq R_{\mathrm{p}} \ , \tag{2.6}$$

where R_{p} is the desired and prespecified performance of the primary system. Furthermore, this control must be robust to fluctuations in the channel between the cognitive radio transmitter and primary receiver. Thus, the primary's ARQ feedback provides a means for the cognitive radio to apply this control.

2.2.2 Channel Model and Coding

We now consider the DMC with feedback from Figure 2.2 in more detail. Let $\mathcal{X} = \{x_{\mathrm{off}}, 1, \ldots, P\}$ be the channel inputs so $|\mathcal{X}| = P+1$. Then at time i, the conditional distribution of the channel output Y_i and primary's ARQ A_i given $X_i = x$ can be expressed as

$$p(y_i, a_i | x_i = x) = p(y_i | x_i = x) \cdot \epsilon_{x,i} \cdot \exp\left(a_i \cdot \log \frac{1-\epsilon_{x,i}}{\epsilon_{x,i}}\right) \ . \tag{2.7}$$

We assume that the sequences $\{\epsilon_{x,i}\}_{i=1}^{\infty}$, for $x \in \mathcal{X}$ are unknown at the encoder and decoder. For notational convenience, we denote

$$\epsilon_{x_{\mathrm{off}},i} = \epsilon_{0,i} \ . \tag{2.8}$$

Note that allowing $\epsilon_{x,i}$ to vary with i reflects uncertainty about the amount of interference the cognitive radio is generating on the primary. We will assume that for

[3]Second order issues and tight delay constraints are discussed in Section 2.6.

all $x \neq x_{\text{off}}$ and $i = 1, 2, \ldots,$

$$\epsilon_{x,i} > \epsilon_{0,i} \ . \tag{2.9}$$

For notational convenience, we will assume without loss of generality there exists a channel input $x_{\text{rep}} \neq x_{\text{off}}$ and $y \in \mathcal{Y}$ such that

$$\mathbb{P}(Y = y | X = x_{\text{off}}) \neq \mathbb{P}(Y = y | X = x_{\text{rep}}) \ . \tag{2.10}$$

Note that we can assume this without loss of generality since if it were not true for any symbol, the cognitive radio's channel inputs X_i would be independent of the channel outputs Y_i, and the channel could not be used for communicating in the first place.

In order for the channel to be useful to the cognitive radio, i.e. in order for the cognitive radio to achieve a positive rate, the cognitive radio should be able to choose input symbols other than x_{off} and still have the primary meet its rate target. Thus, we make the following technical assumption.

Assumption 2.1. There is a known constant $\nu > 0$ such that for all i, $R_{\text{p}} + \nu < 1 - \epsilon_{0,i}$.

Assumption 2.1 enables the primary to tolerate some interference from the cognitive radio while guaranteeing the cognitive radio achieves a positive rate.

The definition of the rate and capacity for the secondary are complicated by the fact that the number of channel uses depends on the realizations of $\epsilon_{x,i}$. Therefore, we need to be precise on what is meant by messages. We define the set of possible messages to be the set of binary sequences $\{0,1\}^{nC_{\max}}$, where $C_{\max} = \log \min\{|\mathcal{X}|, |\mathcal{Y}|\}$. Let M_k be the first k bits of the message and $M = M_{nC_{\max}}$.

Definition 2.1. An $(n, f^n, g, \nu, R_{\text{p}})$ *code* (we call n the blocklength) consists of a

sequence of encoding functions $f_{i,\nu,R_\mathrm{p}} : \{0,1\}^{i-1} \times \{0,1\}^{nC_{\max}} \to \mathcal{X}$ for $i = 1, 2, \ldots, n$,

$$X_i = f_{i,\nu,R_\mathrm{p}}(A^{i-1}, M) \ , \tag{2.11}$$

and decoding function $g_{\nu,R_\mathrm{p}} : \mathcal{Y}^n \to \{0,1\}^{nC_{\max}}$

$$\hat{M} = g_{\nu,R_\mathrm{p}}(Y^n) \ . \tag{2.12}$$

Definition 2.2. Given a rate target R_p and $\nu > 0$, a *strategy* is a sequence of $(n, f^n, g, \nu, R_\mathrm{p})$ codes indexed by n on the positive integers $n = 1, 2, \ldots$.

Strategies must respect the primary's rate target, so the following definition restricts the type of strategies we allow.

Definition 2.3. Given a rate target R_p and $\nu > 0$, a strategy is *valid* if for all $\{\epsilon_{x,i}\}_{i=1}^{\infty}$ satisfying Assumption 2.1, for every $(n, f^n, g, \nu, R_\mathrm{p})$ code in the strategy and $k \leq n$,

$$\mathbb{P}_{\{\epsilon_{x,i}\}_{i=1}^{\infty}} \left(k^{-1} \sum_{i=1}^{k} A_i \leq R_\mathrm{p} \right) \leq K_{1,R_\mathrm{p},\nu,k} \cdot e^{-k \cdot K_{2,R_\mathrm{p},\nu}} \ , \tag{2.13}$$

where the constants $K_{1,R_\mathrm{p},\nu,k} < \infty$, $0 < K_{2,R_\mathrm{p},\nu} < \infty$ depend only on ν and rate target R_p, and for all $r > 0$, as $k \to \infty$,

$$K_{1,R_\mathrm{p},\nu,k} \cdot e^{-k \cdot r} \to 0 \ .$$

Note that a valid strategy imposes two restrictions. First, the primary's rate should meet its rate target with a failure probability that decays exponentially in time and with the same exponent for each code within a strategy (*horizon-independence* condition). The horizon-independence condition gets its name from the fact that a

rate guarantee to the primary should not depend on the blocklength of the cognitive radio's code. Second, this failure probability should not depend on the sequences $\{\epsilon_{x,i}\}_{i=1}^{\infty}$ (*robustness* condition). For a given valid strategy, we will use the notation $\hat{M}^n$ to denote the decoded output for its code of blocklength n.

Definition 2.4. Given a rate target R_{p} and $\nu > 0$, a valid strategy *achieves rate R for the sequence* $\{\epsilon_{x,i}\}_{i=1}^{\infty}$ if for all $\delta > 0$, there exists $n_0\big(\delta, \{\epsilon_{x,i}\}_{i=1}^{\infty}, R_{\mathrm{p}}\big)$ such that for every code in the strategy with blocklength $n \geq n_0$,

$$\mathbb{P}_{\{\epsilon_{x,i}\}_{i=1}^{\infty}}(\hat{M}^n_{\lfloor n(R-\delta)\rfloor} \neq M_{\lfloor n(R-\delta)\rfloor}) \leq \delta \ . \tag{2.14}$$

Given a rate target R_{p} and $\nu > 0$, if there exists a valid strategy that achieves rate R for the sequence $\{\epsilon_{x,i}\}_{i=1}^{\infty}$, then that rate R is *achievable for the sequence* $\{\epsilon_{x,i}\}_{i=1}^{\infty}$. The set of achievable R is denoted as $\mathcal{R}(\{\epsilon_{x,i}\}_{i=1}^{\infty}, R_{\mathrm{p}})$.

Definition 2.5. The *rate-interference budget (RIB) function* $R_{\mathrm{IB}}(\{\epsilon_{x,i}\}_{i=1}^{\infty}, R_{\mathrm{p}})$ is defined as

$$R_{\mathrm{IB}}\left(\{\epsilon_{x,i}\}_{i=1}^{\infty}, R_{\mathrm{p}}\right) = \sup_{R \in \mathcal{R}(\epsilon_0, \{\epsilon_{x,i}\}_{i=1}^{\infty}, R_{\mathrm{p}})} R. \tag{2.15}$$

For the special case in which $\epsilon_{x,i} = \epsilon_x$ for all i, it will be convenient to use the shorthand $\vec{\epsilon}$, where $\vec{\epsilon}$ is a length $|\mathcal{X}|$ vector, and we will use the shorthand $R_{\mathrm{IB}}(\vec{\epsilon}, R_{\mathrm{p}})$.

2.2.3 Contributions

We now state the main contributions of this chapter. First, we find a valid strategy that achieves positive rates for the cognitive radio. From the definition of a valid strategy, this implies that there exists a sequence of codes such that the primary meets its rate target irrespective of $\{\epsilon_{x,i}\}_{i=1}^{\infty}$.

Proposition 2.1. *Given any rate target R_p and $\nu > 0$, there exists a valid strategy (i.e. a strategy that under Assumption 2.1 satisfies* (2.13) *for all $\{\epsilon_{x,i}\}_{i=1}^{\infty}$) that achieves the following rates for $\{\epsilon_{x,i}\}_{i=1}^{\infty}$ and thereby provides a corresponding lower-bound to the RIB function:*

$$R_{IB}\left(\{\epsilon_{x,i}\}_{i=1}^{\infty}, R_p\right) \geq \begin{cases} C^* \ , & \forall\ i\ , 1 - \sum_x p^*(x) \cdot \epsilon_{x,i} \geq R_p + \nu \\ \frac{1-R_p-\hat{\epsilon}_0^+}{\Delta\epsilon^+} \cdot C^* \ , & \forall\ i\ , 1 - R_p < \min\{\epsilon_{x_{rep},i}, \sum_x p^*(x)\epsilon_{x,i}\} \\ \frac{1-R_p-\hat{\epsilon}_0^+}{1-\epsilon_0^-} \cdot C^* \ , & \textit{otherwise} \end{cases} , \tag{2.16}$$

where $C^* = \max_{p(x)} I(X;Y)$, $p^* = \arg\max_{p(x)} I(X;Y)$, $\Delta\epsilon^+ = \sup_{x,i} \epsilon_{x,i} - \epsilon_{0,i}$, $\hat{\epsilon}_0^+ = \limsup_k k^{-1} \sum_{i=1}^{k} \epsilon_{0,i}$, and $\epsilon_0^- = \inf_i \epsilon_{0,i}$.

Proposition 2.1 follows immediately from Theorem 2.3. Furthermore, we can precisely characterize the capacity of the cognitive radio for the case of time-invariant interference on the primary, in which $\epsilon_{x,i} = \epsilon_x$ for all $x \in \mathcal{X}$.

Proposition 2.2. *Given any rate target R_p and $\nu > 0$, there exists a valid strategy (i.e. a strategy that under Assumption 2.1 satisfies* (2.13) *for all $\{\epsilon_{x,i}\}_{i=1}^{\infty}$) that achieves rates equal to the RIB function on the subset of $\{\epsilon_{x,i}\}_{i=1}^{\infty}$ for which $\epsilon_{x,i} = \epsilon_x$ for all i:*

$$R_{IB}(\vec{\epsilon}, R_p) = \max_{\substack{p(x): \\ \sum_x \epsilon_x p(x) \leq 1 - R_p}} I(X;Y) \ . \tag{2.17}$$

For this setting, we will refer to the constraint $\sum_x \epsilon_x p(x) \leq 1 - R_{\mathrm{p}}$ as the *interference budget*. Note that the constraint is based on how much interference each of the cognitive radio's channel inputs generates on the primary compared to how much is tolerable for the primary's desired performance.

Proposition 2.2 follows immediately from Theorem 2.4, which provides achievability, and Theorem 2.5, which provides the converse. These are stated in Sections III and IV, respectively. We note that Theorem 2.4 relies on a more intricate valid strategy than the one in the proof of Theorem 2.3.

2.3 Achievable Strategies

In this section, we present two achievable strategies and state results on the rates the cognitive radio can achieve while guaranteeing rate to the primary under various interference conditions. The first of these– the fixed-codebook protocol– is a generalization of the approach considered in Example 2.1, in which the cognitive radio becomes active only when the primary is meeting its rate target. We show that this strategy is valid, i.e. the primary meets its rate target under unknown time-varying interference characteristics, and can give equally general rate guarantees for the cognitive radio. The second strategy– the codebook-adaptive protocol– builds on the first strategy to predict the amount of interference the cognitive radio will generate on the primary and optimize its codebook to maximize its own rate. Like the first strategy, this strategy is also valid, so the primary meets its rate target under unknown time-varying interference characteristics. We provide rate guarantees for the cognitive radio under the more limited set of unknown time-invariant interference characteristics, and in Section 2.4, we show that the codebook-adaptive protocol provides the optimum rate for the cognitive radio within this set.

2.3.1 Fixed-Codebook Protocol

Recall the approach considered in Example 2.1 over the noiseless channel. The silent symbol x_{off} is used for each channel use when the primary is not meeting its rate

target. Otherwise, one of the remaining P symbols is used to send information about the message. As demonstrated in that example, this leads to a rate proportional to $\log P$. However, this strategy appears to be wasteful in that x_{off} is not being used to send information about the message.

One way to overcome this limitation is to group multiple channel uses into *frames*. Each frame is either silent – consisting of only silent symbols x_{off} – or active – consisting of any combinations of *all* $P+1$ symbols, including x_{off}. Clearly, over the active frames, this increases the rate since the available channel input alphabet is larger. The main issues are:

- To find a rule by which the cognitive transmitter decides before each frame whether the frame will be silent or active. The cognitive transmitter then also needs some way of indicating its choice to the cognitive receiver.
- To appropriately select the frame length. If the frame length is too short, then no rate gain is attained. Conversely, if the frame length is too large, then the non-interference guarantee given in (2.13) can no longer be respected.

We now illustrate the approach in the context of Example 2.1. For the sake of concreteness, consider the case in which the frame length $K_n = 3$ channel uses. For this illustration, we will assume the decision to become active is governed by the threshold rule $\sum_{j=1}^{3\lfloor (i-1)/3 \rfloor}(A_j - \frac{1}{2}) > 0$. Then a sample run may look as follows:

i	$\mathbf{1}$	2	3	$\mathbf{4}$	5	6	$\mathbf{7}$	$\cdots$
$\sum_{j=1}^{i-1}\left(A_j - \frac{1}{2}\right)$	$\mathbf{0}$	$\frac{1}{2}$	1	$\mathbf{\frac{1}{2}}$	0	$-\frac{1}{2}$	$\mathbf{-1}$	$\cdots$
X_i	$\mathbf{x}_{\text{off}}$	x_{off}	x_{off}	$\mathbf{x}_{\text{on,1}}$	x_{off}	$x_{\text{on,P}}$	$\mathbf{x}_{\text{off}}$	$\cdots$

Times $i = 4, 5, 6$ represent an active frame, where the channel input at time $i = 4$ is simply a beacon to indicate to the decoder that the frame is active; the message

information is sent over $i = 5, 6$. Despite the fact that the primary meets the rate target $R_{\mathrm{p}} = \frac{1}{2}$ over channel uses 2 and 3, the cognitive radio sends the silent symbol x_{off} for the duration of the frame. Thus, one has to be careful to set the frame length K_n and transmission threshold to make sure the cognitive radio can achieve a significant rate. Likewise, the cognitive radio sends the message information $(x_{\mathrm{off}}, x_{\mathrm{on,P}})$ over channel uses 5 and 6 even though the primary no longer exceeds the rate target $\frac{1}{2}$. Thus, one has to be careful to set the frame length K_n and transmission threshold so that the primary's rate satisfies (2.13), so the strategy is valid.

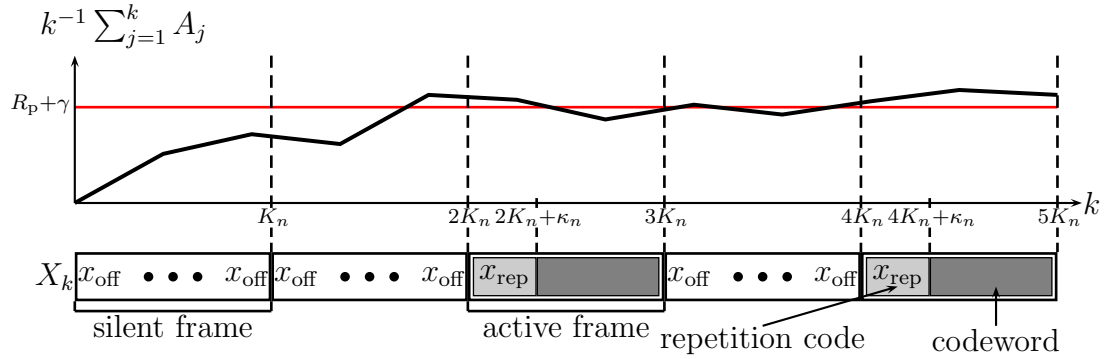

Figure 2.3: In the fixed-codebook protocol, channel uses are grouped into units known as frames. At the start of a frame, the cognitive radio encoder chooses to become active if the primary's packet rate $k^{-1}\sum_{j=1}^{k} A_j$ is above a threshold $R_{\mathbf{p}} + \gamma + o(1)$. Otherwise, it stays silent for the frame, i.e. sends the symbol $x_{\mathbf{off}}$. On an active frame, the encoder uses a length κ_n repetition code to signal to the decoder that it is active and sends a codeword over the remaining channel uses to convey additional bits of the message.

We now use the intuition from the illustration above to construct the *fixed-codebook protocol*, which we will then prove is a valid strategy, as defined in Section 2.2. Figure 2.3 provides an illustration of the fixed-codebook protocol. For convenience, we define

for all $k \geq 1$,

$$S_k = S_{k-1} + (A_k - R_{\mathrm{p}}) \ , \tag{2.18}$$

$$S_0 = s \ , \tag{2.19}$$

where $s \leq 0$. Note that for $s = 0$, we have that $S_k = \sum_{i=1}^{k}(A_i - R_{\mathrm{p}})$, which is positive at time k if and only if the primary is exceeding its rate target. Thus, $s < 0$ will be an initial penalty term that can effectively delay the start of an active frame.

2.3.1.1 Determining Silent Frames

As before, the cognitive radio makes a decision to be silent or active over frames of length K_n channel uses. Specifically, the following condition specifies the frames over which the cognitive radio is silent:

$$X_j = x_{\mathrm{off}} \ \text{ if } \ S_{i(j)} - i(j) \cdot \gamma < K_n \ , \tag{2.20}$$

where $i(j) = \lfloor (j-1)/K_n \rfloor \cdot K_n$, and γ is an additional parameter for setting the threshold along with K_n to satisfy condition (2.13).

2.3.1.2 Active Frames

It remains to define what the cognitive transmitter does over an active frame. As in the noiseless case, we want to inform the decoder that the frame is active, but in the noisy case, it cannot be done with a single channel use.

2.3.1.2.1 Repetition Coding The cognitive transmitter uses a length κ_n repetition code to inform the cognitive receiver that the frame is active. While a repetition code may appear suboptimal, κ_n will be chosen to be small relative to the blocklength

n and thus have no impact on the rate asymptotically. However, one may in practice replace this with a more sophisticated code. Recall our assumption in (2.10). Then the repetition code over the first κ_n channel uses of an active frame is specified by the following condition:

$$X_j = x_{\text{rep}} \text{ if } S_{i_j} - i_j\gamma \geq K_n,\ i_j < j \leq i_j + \kappa_n\ , \tag{2.21}$$

where $i_j = \lfloor (j-1)/K_n \rfloor \cdot K_n$.

2.3.1.2.2 Message Information For the remaining channel uses of an active frame, the encoder sends information about the message to the decoder. It does so with a blocklength $K_n - \kappa_n$ codebook $\mathcal{C}_{\text{fixed}}$ of rate $C^* - \tilde{\delta}$, where $C^* = \max_{p(x)} I(X;Y)$. To show the existence of a valid strategy that achieves the rates claimed in Proposition 2.1, we specifically will rely on an $(K_n - \kappa_n, C^* - \tilde{\delta}, p^*)$ random codebook with a maximum likelihood codebook (see Definition A.13), where p^* is chosen to be the input distribution that yields C^*. We will denote codeword m as $\tilde{X}^{K_n-\kappa_n}(m)$, where $m \in \{1, \ldots, \exp\{(K_n - \kappa_n)(C^* - \tilde{\delta})\}\}$.

The following notation will be useful for understanding the channel inputs during the remainder of an active frame. Let V_1 denote the channel index preceding the start of the first active frame, V_2 the second, V_3 the third, and so on. That is,

$$V_1 = \inf\{i \geq 0 : S_i - i\gamma \geq K_n\ , i = mK_n \text{ for some } m \in \mathbb{Z}\}\ , \tag{2.22}$$

$$V_k = \inf\{i > V_{k-1} : S_i - i\gamma \geq K_n\ , i = mK_n \text{ for some } m \in \mathbb{Z}\}\ . \tag{2.23}$$

We now characterize the remaining channel inputs. For the ℓ-th active frame and letting m_ℓ be bits $\ell(K_n - \kappa_n)(C^* - \tilde{\delta})\log_2 e + 1$ through $(\ell+1)(K_n - \kappa_n)(C^* - \tilde{\delta})\log_2 e$

of message M,

$$X_j = \tilde{X}_{j-V_\ell-\kappa_n}(m_\ell) \;\text{ if }\; V_\ell + K_n \geq j > V_\ell + \kappa_n \;. \tag{2.24}$$

A summary of the fixed-codebook protocol is given in Table 2.1.

Table 2.1: Summary of the fixed-codebook protocol.

Equation	Description
(2.20)	Silent frame
(2.21)	Active frame: repetition code to synchronize decoder.
(2.24)	Active frame: send next unsent message fragment.

2.3.1.3 Performance of the Fixed-Codebook Protocol

For this strategy, we have the following result.

Theorem 2.3. *Given rate target R_p and $\nu > 0$, there exist choices of $\kappa_n, K_n, \gamma, \tilde{\delta}$ such that the fixed-codebook protocol is a valid strategy, i.e. primary's packet rate satisfies the condition in* (2.13) *for all $\{\epsilon_{x,i}\}_{i=1}^{\infty}$ satisfying Assumption 2.1. Furthermore, for these parameter choices, the strategy achieves rates of at least the following:*

$$\begin{cases} C^* \;, & \forall\; i \;, 1 - \sum_x p^*(x) \cdot \epsilon_{x,i} \geq R_p + \nu \\ \frac{1-R_p-\hat{\epsilon}_0^+}{\Delta\epsilon^+} \cdot C^* \;, & \forall\; i \;, 1 - R_p < \min\{\epsilon_{x_{rep},i}, \sum_x p^*(x)\epsilon_{x,i}\} \\ \frac{1-R_p-\hat{\epsilon}_0^+}{1-\epsilon_0^-} \cdot C^* \;, & \textit{otherwise} \end{cases} , \tag{2.25}$$

where $C^* = \max_{p(x)} I(X;Y)$, $p^* = \arg\max_{p(x)} I(X;Y)$, $\Delta\epsilon^+ = \sup_{x,i} \epsilon_{x,i} - \epsilon_{0,i}$, $\hat{\epsilon}_0^+ = \limsup_k k^{-1} \sum_{i=1}^k \epsilon_{0,i}$, and $\epsilon_0^- = \inf_i \epsilon_{0,i}$.

Proof. While other choices will work, for the purposes of the proof, we will let

$$K_n = \lfloor n^{1/8} \rfloor \ , \kappa_n = \lfloor n^{1/16} \rfloor \ , \tag{2.26}$$

and any γ satisfying $0 < \gamma < \min\{\tilde{\delta}/2, \nu/2\}$. We will assume that $\tilde{\delta} > 0$, but a detailed prescription is given in Lemma B.8 below to allow the rate loss to become arbitrarily small.

The proof of the theorem is divided into three parts.

1. The primary's rate satisfies condition (2.13), so the strategy is valid. (Lemma B.3)

2. There exists a codebook such that cognitive radio decoder error probability is small, thus satisfying (2.14) for some R. (Lemma B.4)

3. By appropriately choosing $\tilde{\delta}$, R in (2.14) can be made arbitrarily close to (2.25). (Lemma B.8)

These results are proved in the Appendix B.1. □

2.3.2 Codebook-Adaptive Protocol

Let us return to Example 2.1, in which the erasure probability for x_{off} is fixed at $\epsilon_0 < 1 - R_{\text{p}}$ for all time, all other inputs have a fixed erasure probability $\epsilon_1 > 1 - R_{\text{p}}$, and the channel is noiseless. Theorem 2.3 implies that when the fixed-codebook protocol is applied to the noiseless channel, the cognitive radio is guaranteed to achieve rates

$$R \geq \frac{1 - \epsilon_0 - R_{\text{p}}}{\epsilon_1 - \epsilon_0} \cdot \log(1 + P) \ . \tag{2.27}$$

Hence, the fixed-codebook protocol only adapts the duty-cycle to the actual degree of interference. Note that the fixed codebook protocol, the strategy is chosen so that when a frame is active, all inputs are uniform, and the rate loss comes from the fraction of silent frames that occur. However, it may be preferable to adapt the codebook so that the input x_{off} is chosen more frequently in an active frame. If this can be done so that the fraction of silent frames has a negligible impact on the rate, the increased *duty-cycle* may provide a net benefit to the rate. In this section, we propose such a strategy, which we refer to as the *codebook-adaptive protocol*. The strategy functions almost identically to the fixed-codebook protocol except that it tries to estimate the amount of interference it causes and then chooses a codebook that will meet this interference constraint such that if the interference characteristics stay fixed, the fraction of silent frames becomes negligible. To achieve this, the strategy consists of three phases:

- Phase I: estimate the amount of interference generated during the first active frame,
- Phase II: select a codebook that satisfies the interference constraints based on the estimates in Phase I, and notify the decoder of this choice during the second active frame.
- Phase III: run the fixed-codebook protocol described in Section 2.3.1 with the modification that the codebook used is the one selected in Phase II.

The codebook-adaptive protocol is summarized in Figure 2.4. As noted above, the codebook-adaptive protocol can be thought of as an adaption of the fixed-codebook protocol. Indeed, the codebook-adaptive protocol uses the same threshold rule and repetition code to signify an active frame. That is, the codebook-adaptive protocol

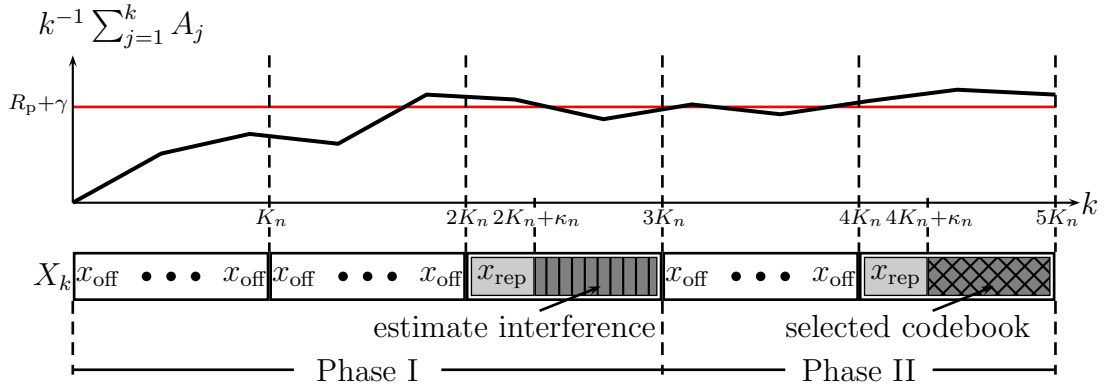

Figure 2.4: The codebook-adaptive protocol is like the fixed-codebook protocol except the first two active frames are used to select a codebook to use and inform the decoder about it. In Phase I, the cognitive radio sends pilots of each of its channel inputs and uses the ARQs to create estimates of the interference it generates on the primary. In Phase II, the cognitive radio notifies the decoder which among a polynomial sized set of codebooks it has selected based on its estimates from Phase I. Phase III, which immediately follows the end of Phase II above, is almost identical to the fixed-codebook protocol, except the codewords are now from the codebook selected during Phase I and Phase II.

follows the rules:

$$X_j = x_{\mathrm{off}} \text{ if } S_{i_j} - i_j\gamma < K_n \ , \tag{2.28}$$

$$X_j = x_{\mathrm{rep}} \text{ if } S_{i_j} - i_j\gamma \geq K_n,\ i_j < j \leq i_j + \kappa_n \ , \tag{2.29}$$

where $i_j = \lfloor (j-1)/K_n \rfloor \cdot K_n$, are identical to conditions (2.20) and (2.21) in the fixed-codebook protocol.

The difference between the two strategies is thus in what follows the repetition code in an active frame. In particular, the encoder uses the first active frame to estimate the channel, the second to inform the decoder which codebook it will use based on those rates, and the third and greater active frames to send message information using the selected codebook.

As before, let V_1 denote the channel index preceding the start of the first active frame, V_2 the second, V_3 the third, and so on. That is, $V_1 = \inf\{i \geq 0 : S_i - i\gamma \geq K_n \ , i = mK_n \text{ for some } m \in \mathbb{Z}\}$ and $V_k = \inf\{i > V_{k-1} : S_i - i\gamma \geq K_n \ , i = mK_n \text{ for some } m \in \mathbb{Z}\}$.

2.3.2.1 Phase I

During the first active frame, the cognitive radio estimates the interference produced by each channel input. Let $\mu = \lfloor \frac{K_n - \kappa_n}{|\mathcal{X}|} \rfloor$. Then for $x \in \{0, \ldots, |\mathcal{X}| - 1\}$, the channel inputs for the first frame can be described as

$$X_j = x \ , \quad \text{if} \ \ V_1 + \kappa_n + (x+1)\mu \geq j > V_1 + \kappa_n + x\mu \ . \tag{2.30}$$

Using these channel inputs, the encoder can use the ARQs to estimate the primary's erasure probabilities.

$$\hat{\epsilon}_x = \mu^{-1} \sum_{i=V_1+\kappa_n+x\mu+1}^{V_1+\kappa_n+(x+1)\mu} A_i \ . \tag{2.31}$$

With these estimates, the end of this first active frame marks the end of Phase I.

2.3.2.2 Phase II

Based on the estimates $\hat{\epsilon}_x$, the encoder chooses a codebook among a set of codebooks; it informs the decoder of this choice in Phase II.

Each codebook in the set has a different input distribution corresponding uniquely to each length-C_n type $p_{x^{C_n}}$ of $\mathcal{X}$, i.e. $p_{x^{C_n}}$ is a probability distribution with the property that for all $x \in \mathcal{X}$, $p_{x^{C_n}}(x) = n_x/C_n$ such that n_x is a nonnegative integer and $\sum_{x \in \mathcal{X}} n_x = C_n$. Thus, there are at most $(C_n + 1)^{|\mathcal{X}|}$ codebooks in the set. The codebook $\mathcal{C}_{x^{C_n}}$ of type $p_{x^{C_n}}$ is a random codebook with codewords generated i.i.d.

according to $\prod_{k=1}^{K_n-\kappa_n} p_{x^{C_n}}(x_k)$ and has

$$\tilde{M}_{x^{C_n}} = \exp\{(K_n - \kappa_n)(R_{x^{C_n}} - \tilde{\delta})^+\} \ , \tag{2.32}$$

where $R_{x^{C_n}}$ is the mutual information $I(X;Y)$ with X having input probability distribution $p_{x^{C_n}}(x)$. One then selects the codebook according to the following rule:

$$\chi = \underset{\substack{x^{C_n}: \\ \sum_x \hat{\epsilon}_x p_{x^{C_n}}(x) \leq 1 - R_{\mathrm{p}} - 2\gamma - \tilde{\delta}}}{\operatorname{argmax}} \tilde{M}_{x^{C_n}} \ . \tag{2.33}$$

The encoder uses the codebook from the fixed-codebook protocol in the second active frame to inform the decoder of its codebook selected codebook. (Note: Based on the parameter choices considered in this work, $(C_n + 1)^{|\mathcal{X}|}$ is small enough so that the fixed-codebook protocol's codebook contains sufficiently many codewords for sending this information. Thus, the encoder simply uses the codewords that result in the lowest probability of error.) Suppose the selected codebook corresponds to message m^{χ}. Then for the second active frame,

$$X_j = \tilde{X}_{j-V_2-\kappa_n}(m^{\chi}) \ \text{ if } \ V_2 + K_n \geq j > V_2 + \kappa_n \ . \tag{2.34}$$

With the decoder informed of which codebook has been selected, the end of this active frame marks the end of Phase II.

2.3.2.3 Phase III

In Phase III, the active frames are now used to send message information. Thus, they resemble the active frames in the fixed-codebook protocol, with the main difference that the codebook χ is used.

Let m_ℓ be bits $\ell(K_n - \kappa_n)(R_\chi - \tilde{\delta}) \log_2 e + 1$ through $(\ell+1)(K_n - \kappa_n)(R_\chi - \tilde{\delta}) \log_2 e$

of message M. For the $(\ell+2)$-th active frame, we can express the message information segment of the frame as

$$X_j = \tilde{X}_{j-V_{\ell+2}-\kappa_n}(m_\ell) \;\text{ if }\; V_{\ell+2}+K_n \geq j > V_{\ell+2}+\kappa_n \;, \tag{2.35}$$

where $\tilde{X}^{K_n-\kappa_n}(m_\ell) \in \mathcal{C}_\chi$.

A summary of the codebook-adaptive protocol is given in Table 2.2.

Table 2.2: Summary of the codebook-adaptive protocol.

Equation	Description
(2.28)	Silent frame
(2.29)	Active frame: repetition code to synchronize decoder
(2.30)	Active frame (Phase I): estimate interference with primary's ARQs.
(2.34)	Active frame (Phase II): notify decoder of selected codebook.
(2.35)	Active frame (Phase III): send next unsent message fragment.

We now state the main result for the codebook-adaptive protocol, which confirms that it is a valid strategy and gives a bound on the rates it achieves.

Theorem 2.4. *Given a rate target R_p and $\nu > 0$, there exists a choice of $K_n, \kappa_n, C_n, \gamma, \tilde{\delta}$ such that the codebook-adaptive protocol is a valid strategy, i.e. the primary's packet rate satisfies the condition* (2.13) *for all $\{\epsilon_{x,i}\}_{i=1}^{\infty}$ satisfying Assumption 2.1. Furthermore, when the interference on the primary is time-invariant, i.e. $\epsilon_{x,i} = \epsilon_x$, the codebook-adaptive protocol, under these parameter settings, achieves the rate*

$$\max_{\substack{p(x):\\ \sum_x \epsilon_x p(x) \leq 1-R_p}} I(X;Y) \;. \tag{2.36}$$

Proof. The parameters K_n, κ_n will be set as in (2.26), $C_n = \sqrt{\kappa_n}$, and any γ satisfying $0 < \gamma < \min\{\tilde{\delta}/2, \nu/2\}$. We will assume $\tilde{\delta} > 0$, but a detailed prescription is given in

Lemma B.14 below to get arbitrarily close to the rate in the statement of the theorem.

The proof of the theorem is divided into three parts.

1. As in the fixed-codebook protocol, we can apply Lemma B.3 since (2.20) and (2.28) are identical conditions. Thus, the primary's rate satisfies the condition (2.13), so the strategy is valid.

2. The cognitive radio decoder error probability is small, thus satisfying (2.14) for some R. (Lemma B.9)

3. By appropriately choosing $\tilde{\delta}$, the R in (2.14) can be made arbitrarily close to $R_{\mathrm{IB}}(\vec{\epsilon}, R_{\mathrm{p}})$ with probability going to 1 as $n \to \infty$. (Lemma B.14)

With the exception of Lemma B.3, these results are proved in the Appendix B.2. □

2.4 Converse

To show the converse, we will relax the conditions stipulated in the problem setup, thereby allowing a larger class of strategies. It turns out that in some cases, this larger class does not increase the rate region.

Theorem 2.5. *Given a rate target R_p and $\nu > 0$, if $\epsilon_{x,i} = \epsilon_x$ for all i, then*

$$R_{IB}(\vec{\epsilon}, R_p) \leq \max_{\substack{p(x): \\ \sum_x \epsilon_x p(x) \leq 1 - R_p}} I(X;Y) \ . \tag{2.37}$$

Proof. From the definition of achievable rate,

$$\begin{aligned} nR &\leq H(M_{\lfloor n(R-\delta)\rfloor}) + n\delta + 1 && (2.38)\\ &\leq I(M_{\lfloor n(R-\delta)\rfloor}; Y^n) + 2n\delta + 1 && (2.39)\\ &= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, M_{\lfloor n(R-\delta)\rfloor}) + 2n\delta + 1 && (2.40)\\ &\leq \sum_{i=1}^{n} H(Y_i) - H(Y_i|Y^{i-1}, M, A^{i-1}, X^i) + 2n\delta + 1 && (2.41)\\ &= \sum_{i=1}^{n} H(Y_i) - H(Y_i|X_i) + 2n\delta + 1 && (2.42)\\ &= \sum_{i=1}^{n} I(X_i; Y_i) + 2n\delta + 1 \ , && (2.43) \end{aligned}$$

where (2.39) follows from Fano's inequality, (2.40) from the chain rule, (2.41) since conditioning cannot increase entropy, (2.42) by the Markov chain $(M, A^{i-1}, Y^{i-1}, X^{i-1}) - X_i - Y_i$, and (2.43) by definition.

We have yet to place a restriction on the strategies. Recall that valid strategies need to satisfy the condition in (2.13). If condition (2.13) is satisfied, then the code of blocklength n satisfies

$$n^{-1} \sum_{i=1}^{n} \mathbb{E}\left[A_i\right] \geq R_{\mathrm{p}} - K_{1,R_{\mathrm{p}},\nu,n} e^{-n \cdot K_{2,R_{\mathrm{p}},\nu}} \ . \tag{2.44}$$

We now consider only this weaker condition on the channel inputs as opposed to the stronger one given by (2.13). By the concavity of mutual information with respect to its input distribution, we can combine (2.43) and (2.44) to yield that for all $\delta > 0$,

there exists large enough n such that

$$R \leq \max_{\substack{p(x): \\ \sum_x \epsilon_x p(x) \leq 1 - R_{\mathrm{p}} + \delta}} I(X;Y) + 3\delta \tag{2.45}$$

Since δ can be made arbitrarily small, we can conclude the result. □

2.5 Examples

Propositions 2.1 and 2.2 provide a lower bound and an exact result for the RIB function under different interference conditions, respectively. In this section, we evaluate the RIB function given in Proposition 2.2 for cases in which the interference characteristics on the primary are time-invariant. We then evaluate the RIB function lower bound given in Proposition 2.1 for these examples when the interference characteristics are time-varying.

2.5.1 Evaluation of the RIB Function for Time-Invariant Interference Characteristics

We first explore the setting in which the interference parameters $\epsilon_{x,i}$ are time-invariant, i.e. $\epsilon_{x,i} = \epsilon_x$ for all i, x. In this setting, Proposition 2.2 gives an exact expression for the RIB function $R_{\mathrm{IB}}(\vec{\epsilon}, R_{\mathrm{p}})$.

We first evaluate the RIB function for Example 2.1. We first rewrite the expression

in Proposition 2.2 as

$$R_{\mathrm{IB}}(\vec{\epsilon}, R_{\mathrm{p}}) = \max_{\substack{p(x): \\ \sum_x \epsilon_x p(x) \le 1-R_{\mathrm{p}}}} I(X;Y) \tag{2.46}$$

$$= \max_{\substack{p(x): \\ p(x\neq x_{\mathrm{off}})\epsilon_1 + p(x=x_{\mathrm{off}})\epsilon_0 \le 1-R_{\mathrm{p}}}} H(X)$$

$$= \max_{p \le \frac{1-R_{\mathrm{p}}-\epsilon_0}{\epsilon_1-\epsilon_0}} h(p) + p\log P$$

$$= \begin{cases} \log(P+1)\,, & b \ge \frac{P}{P+1} \\ h(b) + b\log P\,, & \text{otherwise} \end{cases}, \tag{2.47}$$

where $b = \frac{1-R_{\mathrm{p}}-\epsilon_0}{\epsilon_1-\epsilon_0}$. Figure 2.5 shows (2.47) in terms of b, which we can think of as a summary of the interference budget.

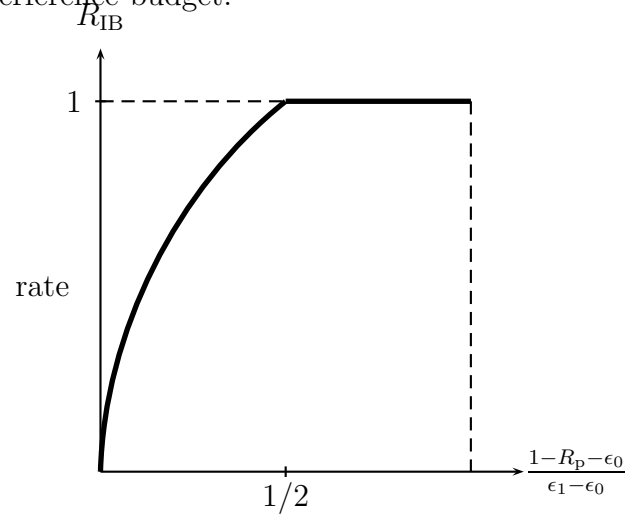

Figure 2.5: A schematic plot of the RIB function for Example 2.1 when $P = 1$.

Example 2.2. Consider a DMC with $|\mathcal{X}| = 1+P$ channel input symbols and $|\mathcal{Y}| = P$

output symbols with the following property:

$$\mathbb{P}(Y = y|X = x) = \begin{cases} 1, & y = x, \quad x \in \{1, \ldots, P\} \\ \frac{1}{P}, & y \in \mathcal{Y}, \quad x = x_{\text{off}} \ . \end{cases} \ . \tag{2.48}$$

If $\epsilon_0 = 0$, $\epsilon_x = 1$ for $x \in \{1, \ldots, P\}$, then evaluating the RIB function from Proposition 2.2 yields

$$R_{\text{IB}}(\vec{\epsilon}, R_{\text{p}}) = (1 - R_{\text{p}}) \log_2 P \ , \tag{2.49}$$

where the units are in bits per channel use.

We now consider a case in which the secondary has an alternative to x_{off} to control interference. The channel model resembles the one in Example 2.2, except there are now additional channel inputs.

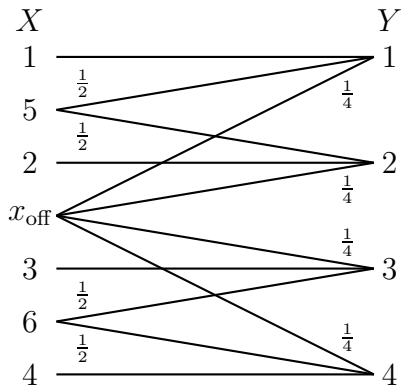

Figure 2.6: Illustration of transition probabilities for Example 2.3 when $P = 4$.

Example 2.3. Let P be even and consider a DMC with $|\mathcal{X}| = 1 + 3P/2$ channel

input symbols and $|\mathcal{Y}| = P$ output symbols with the following property:

$$\mathbb{P}(Y = y|X = x) = \begin{cases} 1, & y = x, \quad x \in \{1, \ldots, P\} \\ \frac{1}{P}, & y \in \mathcal{Y}, \quad x = x_{\text{off}} \\ \frac{1}{2}, & y = 2(x - P) - 1, \quad x \in \{P + 1, \ldots, P + P/2\} \\ \frac{1}{2}, & y = 2(x - P), \quad x \in \{P + 1, \ldots, P + P/2\} \end{cases} . \quad (2.50)$$

An illustration of these transition probabilities are given in Figure 2.6. We now consider the case in which $\epsilon_0 = 0$, $0 < R_{\text{p}} < 1$, $\epsilon_x = 1$ for $x \in \{1, \ldots, P\}$, and $\epsilon_x = \epsilon_{1/2} < 1 - R_{\text{p}}$ for $x \in \{P+1, \ldots, P+P/2\}$. Under these assumptions, evaluating the RIB function from Proposition 2.2 yields

$$R_{\text{IB}}(\vec{\epsilon}, R_{\text{p}}) = \max_{p(x):\sum_x \epsilon_x p(x) \leq 1 - R_{\text{p}}} I(X;Y) \quad (2.51)$$

$$= \frac{1 - R_{\text{p}} - \epsilon_{1/2}}{1 - \epsilon_{1/2}} \log_2 P + \frac{R_{\text{p}}}{1 - \epsilon_{1/2}} \log_2(P/2) \quad (2.52)$$

$$= \log_2 P - \frac{R_{\text{p}}}{1 - \epsilon_{1/2}} , \quad (2.53)$$

where the units are in bits per channel use. Note that the rate loss due to the primary can be at most 1 bit in this setting. Moreover, this can be arbitrarily better than the case in Example 2.2 by making P large and R_{p} close to 1, for which the rate target R_{p} induced a multiplicative penalty on the $\log_2 P$ term in (2.49).

2.5.2 Further Considerations for Time-Varying Interference Characteristics

The most interesting and realistic scenarios concern the case when the interference characteristics are time-varying. The codebook-adaptive protocol introduced in Section III can deal with this as long as it is well behaved. However, for some "maliciously

chosen" time-varying characteristics, the proposed startegy can be fooled into choosing a low rate codebook in Phase II when the interference conditions are less severe in Phase III. The effect of such a possibility is illustrated in Figure 2.7. One option might be to consider a strategy that periodically readapts the codebook, which, while potentially beneficial, is outside the scope of this work. Instead, we consider the simpler strategy given by the fixed-codebook protocol.

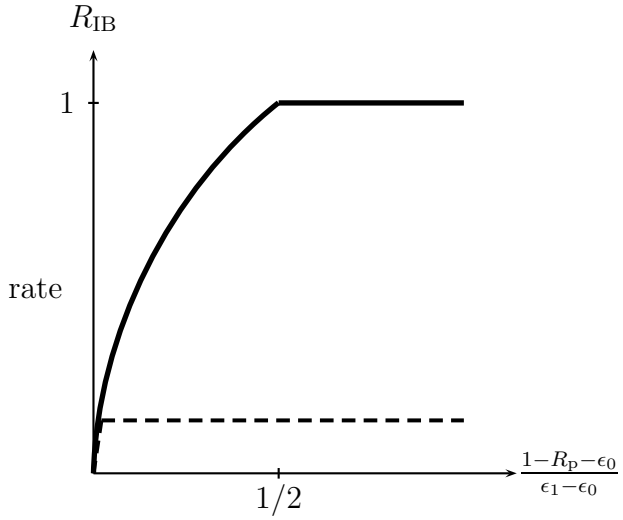

Figure 2.7: A schematic plot of the RIB function for Example 2.1 when $P = 1$. The dashed line suggests how maliciously chosen time-varying characteristics can cause the encoder to select a "low rate" codebook, which saturates well below the actual RIB function when the interference budget is large.

For Example 2.1, Proposition 2.1 implies the fixed-codebook protocol lets the cognitive radio achieve the rate

$$R_{\text{IB}}\left(\epsilon_0, \{\epsilon_{x,i}\}_{i=1}^{\infty}, R_{\text{p}}\right) \geq \left(1 - R_{\text{p}}/(1-\epsilon_0)\right) \log(1+P) \tag{2.54}$$

for all $\{\epsilon_{x,i}\}_{i=1}^{\infty}$, $x \neq x_{\text{off}}$. For the restricted time-invariant interference setting of Example 2.1, can use $b = \frac{1-R_{\text{p}}-\epsilon_0}{\epsilon_1-\epsilon_0}$ to compare its performance against the RIB func-

tion. It turns out that for $b > b^*$, the codebook chosen by the codebook-adaptive protocol has the same asymptotic rate and produces the same interference on the primary as that in the fixed-codebook protocol. Thus, depending on one's assumptions about the interference environment, there are instances in which the fixed-codebook protocol may be preferable to the codebook-adaptive protocol.

Despite these guarantees, there are situations in which the fixed-codebook protocol can be arbitrarily worse. Recall Examples 2.2 and 2.3. It turns out that in both cases when $\epsilon_0 = 0$, Proposition 2.1 implies the fixed-codebook protocol guarantees rates given by

$$R_{\mathrm{IB}}\left(0, \{\epsilon_{x,i}\}_{i=1}^{\infty}, R_{\mathrm{p}}\right) \geq (1 - R_{\mathrm{p}}) \log_2 P \ , \tag{2.55}$$

which matches the RIB function in (2.49) for Example 2.2. However, as already illustrated, by making P large and R_{p} close to 1, the RIB function in Example 2.3, given in (2.53), can be made arbitrarily larger than the one in Example 2.2. This implies that the loss for applying the fixed-codebook protocol can be significant. Thus, one's choice between these two protocols depends jointly on the cognitive radio's channel and the interference generated on the primary. Indeed, there may exist strategies that can trade off the competing desires of optimality and robustness better than the ones proposed. These are discussed further in the next section.

2.6 Discussion

In this chapter, a novel model was proposed for a cognitive radio problem. The basic problem is that the cognitive radio must not disturb the primary user (i.e., the license holder). The specific aspect of our model is that the cognitive radio is ignorant of the channel characteristics according to which it interferes with the primary. To

mitigate this uncertainty, the cognitive radio may eavesdrop on the primary system's ARQ feedback signal. We show how this can be exploited to design two adaptive cognitive radio strategies, each of which provides a fixed rate guarantee to the primary and variable rate guarantee to the cognitive radio that depends on its interference budget, the amount of interference it is allowed to generate on the primary user. The problem statement and results provide a starting point for new research directions and problems, some of which we briefly outline in the sequel.

2.6.1 Gaussian Channels

In this work, the cognitive radio's channel is a DMC with each symbol affecting the primary's erasure probability. An analogous model and result for the Gaussian setting would be desirable to gain further intuitions about the design of a cognitive radio system. For instance, if the primary employs a Gaussian codebook that assumes a certain level of interference, the cognitive radio may use the ARQs to choose the highest power codebook that maintains that level of interference on the primary.

2.6.2 Primary with a Fixed Delay Constraint

In our model, the cognitive radio must operate such that eventually, the primary attains its prespecfied rate target. A more restrictive setting would be to also enforce a delay constraint. That is, the cognitive radio must operate such as to not delay packets by more than a certain prespecified bound. Alternatively, this can be formulated as a "sliding window" rate constraint: over any window of a prespecified length, the primary must attain its prespecified rate. It would be interesting to understand by how much this lowers the "interference budget" of the cognitive radio, and thus, its capacity.

2.6.3 Improved Strategies

The cognitive radio's rate guarantees for the fixed-codebook protocol are somewhat pessimistic, and the rate guarantees for the codebook-adaptive protocol are restricted to the smaller class of time-invariant interference parameters on the primary. The problem is that since the codebook-adaptive protocol only selects the codebook once, varying the interference conditions in the time can lead to suboptimal performance. For instance, the interference conditions in Phase I can be such that the primary selects a codebook with negligible rate in Phase II only to discover that there is no interference to the primary in Phase III. Thus, its performance can be significantly worse than the fixed-codebook protocol in the time-varying setting.

An obvious alternative would be a strategy that periodically readapts the codebook, which, if done properly, may be able to provide stronger rate guarantees than those already provided in the time-varying setting. One may also wish to restrict the set of codebooks so that all codebooks have a rate above a certain threshold. Then, arguments similar to those used for the fixed-codebook protocol can provide rate guarantees for the time-varying case, and one can also exploit the advantage afforded by adapting one's codebook for the time-invariant case.

2.6.4 Multiple Cognitive Radios

In our model, there is only a single cognitive radio interfering with the primary. A more interesting situation will involve multiple cognitive radios all competing for the same interference budget. Clearly, this significantly changes the dynamics of the problem. Are there efficient strategies that give good rates for the cognitive radios while respecting the primary user? First of all, if all the cognitive radios have access to A_k with different delays, then the arguments in this work would need to be extended. The existence of multiple users also leads to the issue that any individual

cognitive radio may not cause significant interference to the primary by itself, but the aggregate interference from all cognitive radios can still be quite large. Another issue to consider is how the cognitive radios might divide their rate in an equitable way based not only on their own channels but also on how much interference each generates on the primary.

2.6.5 Noisy feedback

In our model, the cognitive radio has a perfect observation of the ARQ signal of the primary, i.e., of the values of A_k. However, in practice there may be noise that corrupts the encoder's knowledge of A_k. This may also play a crucial role for the case of multiple cognitive radios, in which the noise may be different for different terminals in the system.

Chapter 3

Towards a Theory: Communication with a Dynamic Cost

3.1 Introduction

In the previous chapter, we studied a particular situation that could arise for a cognitive radio that needs to communicate in the presence of a primary and modeled several aspects of the problem, including the nature of interference the cognitive radio generates on the primary, uncertainty at the cognitive radio about the statistics of this interference, and the type of constraint the cognitive radio must satisfy.

In this chapter, we focus on one aspect of the model from Chapter 2 and generalize it: the cost constraint. In Chapter 2, the cost is the number of dropped packets at the primary. It is additive and random: namely, the cost at each time is a random variable with a distribution that is a function of the cognitive radio's current channel input. However, such a cost does not capture dynamics in a real system over multiple time steps. For instance, if the primary's packet were dropped, it could choose to modify its own duty cycle, and thus the cognitive radio's input could also affect the

long-term behavior. Furthermore, such a cost does not capture the fact that a burst of dropped packets could be worse than the same number of dropped packets sparsely distributed over time. Our goal in this chapter is to model such dynamics generally and understand the structure of optimal strategies for such models.

The following tongue-in-cheek story will help illustrate the issues that arise in this more general setting. It is the story of General Alice S. Grant (Alice) and General Bob E. Lee (Bob), old friends whose armies have fallen into a Civil War reenactment against each other. The reenactment is divided into a series of battles. While there is room for improvisation in the battles, to maintain some historical accuracy, the reenactment requires that Alice's army win enough battles to claim victory in the war. Alice has a dual objective with her battle tactics: the first is to win the war, and the second is to send Bob a message. Bob, for his part, wants to decode the message from Alice, but he must also follow military protocol and respond to Alice's tactics by preparing his troops for future battles. Thus, if Alice is not careful in how she chooses her tactics, it may cost her army in battle losses.

Example 3.1 (Alice, Bob, and Fibonacci)**.** For battle k, we model Alice's tactics as an input X_k over a channel, where Bob observes the output Y_k. Alice can use this channel for n battles. There are two possible inputs to the channel: attack ($X_k = 1$) and retreat ($X_k = 0$). Bob observes these perfectly in his output ($Y_k = X_k$) and responds by updating the state (S_k) of his troops. If Alice attacks ($X_k = 1$), then Bob prepares his troops for the next battle ($S_{k+1} = 1$). If Alice retreats ($X_k = 0$), then Bob does not prepare his troops for the next battle ($S_{k+1} = 0$). Before choosing a tactic for battle k, Alice can observe the state of Bob's troops (S_k). If Alice attacks when Bob's troops are prepared ($X_k = 1, S_k = 1$), then Alice loses the battle. Otherwise, Alice does not lose. Alice can the win the war if she loses no more than a fraction α of the n battles, and we wish to find rates R at which Alice can send a

message $M \in \{1, \ldots, 2^{nR}\}$ reliably to Bob while simultaneously winning the war.

Here is a strategy for $\alpha = 0$: Alice maps each possible message into a string of bits. If Bob's troops are prepared for battle ($S_k = 1$), Alice orders a retreat ($X_k = 0$). However, if Bob's troops are unprepared ($S_k = 0$), Alice sends the next bit in the bit string for the selected message. To decode, Bob looks at the outputs for which his troops are unprepared (Y_k such that $S_k = 0$) and inverts the mapping to determine the selected message. Thus, we can count the total number of messages simply by counting the number of uniquely distinguishable bit strings.

This can be done inductively. Note trivially that if no battles are fought ($n = 0$), there can only be $N_0 = 1$ message. Let us assume Bob's troops are initially unprepared ($S_1 = 0$). For $n = 1$, there are $N_1 = 2$ messages with bit strings 0 and 1. For $n = 2$, there are $N_2 = N_1 + N_0 = 3$ messages, which can be seen from the following. For messages with 0 in the first position of the bit string, Alice retreats ($X_1 = 0$), so Bob's troops are unprepared for the second battle ($S_2 = 0$), and we are back to the case of $n = 1$, for which are $N_1 = 2$ such distinguishable messages. For messages with 1 in the first position of the bit string, Alice attacks ($X_1 = 1$), so Bob prepares his troops for the second battle ($S_2 = 1$). Thus, Alice retreats in the second battle ($X_2 = 0$), and we are back to the case of $n = 0$, for which there is only $N_0 = 1$ message. By applying this logic to general n, Alice can send $N_n = N_{n-1} + N_{n-2}$ messages at time n, which is the Fibonacci sequence $(1, 2, 3, 5, 8, \ldots)$, and the rate for cost $\alpha = 0$ converges to $\lim_{n\to\infty} n^{-1} \log N_n = \log \phi$, where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio.

In the remainder of this chapter, we build on Example 3.1 to study the capacity of a channel with feedback and a cost constraint that depends on both the state of the channel and the channel input, where the state depends on both the previous state and channel output. The formulation is related to problems considered by Chen and Berger [12], Yang, Kavcic, and Tatikonda [96], as well as Permuter, Weissman, and

Goldsmith [67] and Permuter, Cuff, Van Roy, and Weissman [66]. The key difference in the current work is that the objective is not only to communicate reliably, but also to control the channel state to satisfy a cost constraint. As an aside, we also note an interesting problem in the nonfeedback setting by Koch et al. [49] to model on-chip communication.

In this and related problems [12; 96; 67; 66], ideas from stochastic control are leveraged to find the capacity. One should note that connections between communication and control have also been observed in other settings. For instance, work by Sahai [70; 71; 73] has found that reliable communication plays a critical role in controlling a plant when there is noise in the feedback loop. Similarly, work by Mitter [62], Elia [33; 32], and Tatikonda [83; 84] have considered control problems in which the control channel has been constrained in the amount of information it can convey. Recently, Grover and Sahai [44] have shown how one can improve on the performance of a linear control system control problem, which is closely related to a well-known counterexample proposed by Witsenhausen in stochastic optimal control [92].

In Section 3.2, we provide an explicit problem setup and useful definitions for characterizing the capacity. In Section 3.3, we provide the achievable strategy, which generalizes the achievable strategy considered in Example 3.1. Section 3.4 provides the converse, which relies on the solution to the Bellman equation for an infinite-horizon dynamic programming problem. In Section 3.5, we revisit Example 3.1 to find its capacity and consider other examples, including one that demonstrates that the gap between the feedback and non-feedback capacity can be unbounded. Section 3.6 concludes the chapter with a discussion about the results and directions for future research.

3.2 Problem Setup

Before proceeding, we explain the notation used in this chapter. Let capital letters X, Y, Z denote random variables, lower-case letters x, y, z real-valued numbers, and calligraphic letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ sets. Two forms of vector notation are used in this chapter: $\vec{x}$ and x^n. The first notation $\vec{x}$ specifies a vector without directly referencing its dimension while the second notation x^n uses the superscript to denote an n-dimensional vector. The subscripts x_k denotes the k-th component of x^n. Subscript are also used to denote time.

3.2.1 Channel Model and Coding Strategies

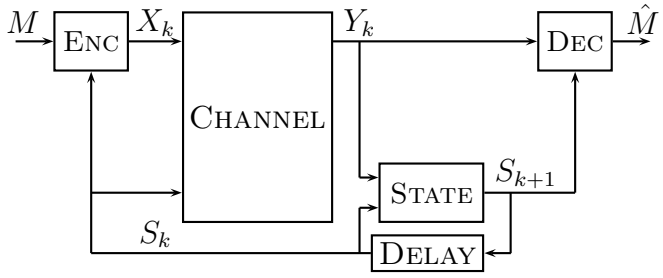

Figure 3.1: At time k, the channel assesses a cost as a function of the channel input X_k and the state S_k. The next state S_{k+1} depends on the current state S_k and the channel output Y_k.

We assume that at time k, the channel takes the channel input $X_k \in \mathcal{X}$ and a state $S_k \in \mathcal{S}$ to generate the output $Y_k \in \mathcal{Y}$ and the next channel state $S_{k+1} \in \mathcal{S}$. We make the following two assumptions about the channel behavior: *(i)* given the channel input X_k and state S_k at time k, the channel output Y_k and state S_{k+1} are conditionally independent of message $M \in \mathcal{M}$, channel inputs X^k, states S^k, and channel outputs Y^{k-1}; and *(ii)* given the channel output Y_k and state S_k, the next

state S_{k+1} is conditionally independent of the channel input X_k. Mathematically, we can express these assumptions as follows:

$$\begin{aligned}
&\mathbb{P}(S_{k+1} = s_{k+1}, Y_k = y_k | M = m, X^k = x^k, Y^{k-1} = y^{k-1}, S^k = s^k) \\
&\quad = \mathbb{P}(S_{k+1} = s_{k+1} | Y_k = y_k, S_k = s_k) \cdot \mathbb{P}(Y_k = y_k | X_k = x_k, S_k = s_k) \qquad (3.1) \\
&\quad = q(s_{k+1} | y_k, s_k) \cdot w(y_k | x_k, s_k) \ , \qquad (3.2)
\end{aligned}$$

where q and w are convenient shorthand to denote $\mathbb{P}(S_{k+1} = s_{k+1} | Y_k = y_k, S_k = s_k)$ and $\mathbb{P}(Y_k = y_k | X_k = x_k, S_k = s_k)$, respectively. We note a correspondence between this channel and finite state channels (see e.g. [67; 97]).

The cost function $\Gamma : \mathcal{S} \times \mathcal{X} \to \Re^+$ depends on both the channel input and the state. Given the cost function, the total cost at time n is

$$\sum_{k=1}^{n} \mathbb{E}[\Gamma(S_k, X_k)] \ . \qquad (3.3)$$

We are now in a position to define what we mean by a coding strategy and achievable rate. Our definitions factor the need for the encoder to both communicate over the channel and satisfy the cost constraint.

Definition 3.1. An (n, R) *coding strategy* with blocklength n and rate R consists of a sequence of encoding functions $\mu_k : \mathcal{M} \times \mathcal{S}^k \to \mathcal{X}$ for $k = 1, \ldots, n$, and a decoding function $\nu : \mathcal{Y}^n \times \mathcal{S}^n \to \mathcal{M}$, where $\mathcal{M} = \{1, 2, \ldots, 2^{nR}\}$. Given an (n, R) coding strategy and a message $M \in \mathcal{M}$, the channel input is given by $X_k = \mu_k(M, S^k)$.

Definition 3.2. Given a function $\Gamma : \mathcal{S} \times \mathcal{X} \to \Re^+$, a rate R is *achievable* with cost constraint α if for all $\delta > 0$, there exists for n sufficiently large an $(n, R - \delta)$ coding

strategy such that for all messages $m \in \mathcal{M}$ and for all initial states $s \in \mathcal{S}$,

$$\mathbb{P}(\nu(Y^n, S^n) \neq m | M = m) \leq \eta(\delta) \ , \tag{3.4}$$

$$n^{-1} \sum_{k=1}^{n} \mathbb{E}[\Gamma(S_k, X_k) | M = m, S_1 = s] \leq \alpha + \gamma(\delta) \ , \tag{3.5}$$

where $\gamma(\delta) \to 0$ and $\eta(\delta) \to 0$ as $\delta \to 0$. The set of achievable rates is denoted $\mathcal{R}(\alpha)$, and the *capacity* is written as $C(\alpha) = \sup \mathcal{R}(\alpha)$.

Remark: The cost function $\Gamma : \mathcal{S} \times \mathcal{X} \to \Re^+$ is a bounded function that depends both on the channel input and the state. Note that if one's only concern is minimizing the cost function and not communication (i.e. communicating at rate $R = 0$), the problem reduces to an infinite-horizon stochastic control problem (see, e.g. [53, Ch. 8]). If $S_k = Y_k$ and Γ depends only on the state S_k and not X_k, the constraint mirrors the output cost constraint considered in [40].

3.2.2 Properties of the State Process

We can characterize the capacity when the state process $\{S_k\}_{k \geq 1}$ satisfies an irreducibility condition that we now define formally. The definition is based on one given in [53, Definition (8.5.14), p. 159].

Definition 3.3. Consider a state process $\{S_k\}_{k \geq 1}$ defined according to (3.2). For a transition probability function p on $\mathcal{X}$ given $\mathcal{S}$, let $[P^p_{u,v}]$ be a transition probability matrix, where row $u \in \mathcal{S}$ and column $v \in \mathcal{S}$ in the matrix are defined as follows:

$$P^p_{u,v} = \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} q(v|y,u) \cdot w(y|x,u) \cdot p(x|u) \ .$$

Then $\{S_k\}_{k \geq 1}$ is *irreducible* if for all transition probability functions p on $\mathcal{X}$ given $\mathcal{S}$, the Markov chain with transition probability matrix $[P^p_{u,v}]$ is irreducible.

The following notion of a stationary distribution for $\{S_k\}_{k\geq 1}$ will also be convenient to characterize the capacity.

Definition 3.4. Consider a state process $\{S_k\}_{k\geq 1}$ defined according to (3.2). For a transition probability function p on $\mathcal{X}$ given $\mathcal{S}$, let $[P^p_{u,v}]$ be a transition probability matrix, where row $u \in \mathcal{S}$ and column $v \in \mathcal{S}$ in the matrix are defined as follows:

$$P^p_{u,v} = \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} q(v|y,u)\cdot w(y|x,u)\cdot p(x|u) \ .$$

If a probability distribution $\pi(\cdot)$ satisfies, for all $v \in \mathcal{S}$,

$$\pi(v) = \sum_{u\in\mathcal{S}} P^p_{u,v}\cdot \pi(u) \ ,$$

then $\pi(\cdot)$ is a *stationary distribution* of $\{S_k\}_{k\geq 1}$ induced by p.

Remark: Note that if the state process $\{S_k\}_{k\geq 1}$ is irreducible, then given any transition probability function p on $\mathcal{X}$ given $\mathcal{S}$, there exists a unique stationary distribution of $\{S_k\}_{k\geq 1}$ induced by p.

3.3 Achievable Strategy

We consider a simple coding strategy that works as follows: for each state $s \in \mathcal{S}$, we construct a random codebook according to the distribution $p(x|s)$, where the length of each codebook corresponds to the number of times we expect to see that particular state over the course of the blocklength. That is, the length of each of these codebooks is roughly proportional to the probability of being in the corresponding state. The encoder then splits the message by state to select a codeword from each codebook. If at time k, the state $S_k = s$, the encoder chooses the next unsent symbol for the

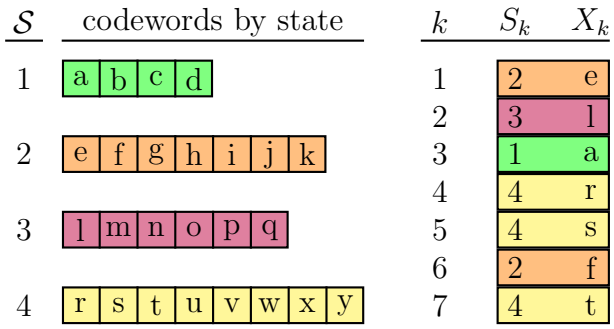

Figure 3.2: The coding strategy in Section 3.3 for a channel with state space $\mathcal{S} = \{1, 2, 3, 4\}$. For each state, there is a codebook, possibly of different lengths. The encoder splits the message to choose codewords by state; it chooses the channel input to be the next unsent symbol from the codeword of the current state. The decoder unscrambles the channel outputs by state, determines each codeword, and thereby reconstructs the message.

codeword corresponding to state s. The decoder, which has access to the state vector S^n, separates the channel outputs by state and for each state, decodes the correspond codeword. From these codewords, the decoder reconstructs the original message. A schematic description of this strategy is provided in Figure 3.2. We are now in a position to state the rates achieved by this strategy.

Theorem 3.1. *Suppose $\pi(s)$ is the stationary distribution of the state process $\{S_k\}_{k\geq 1}$ induced by some $p(x|s)$ and $\sum_{x,s} \Gamma(s,x) \cdot p(x|s) \cdot \pi(s) \leq \alpha$. Then the following rate is achievable:*

$$\sum_s I(X;Y|S=s)\pi(s) \ , \tag{3.6}$$

where the random variables X, Y, S have joint distribution $w(y|x,s) \cdot p(x|s) \cdot \pi(s)$.

Furthermore, if the state process $\{S_k\}_{k\geq 1}$ *is irreducible, then the capacity is at least*

$$C(\alpha) \geq \max_{\substack{p(x|s):\\ \mathbb{E}[\Gamma(S,X)]\leq\alpha}} \sum_s I(X;Y|S=s)\pi(s) \ , \tag{3.7}$$

where $\pi(s)$ *is the stationary distribution of the state process* $\{S_k\}_{k\geq 1}$*, defined in* (3.2)*, induced by* $p(x|s)$*, and the distribution of* Y *given* (X,S) *is* $w(y|x,s)$*.*

Remark: One may note that this is reminiscent of the achievable strategy given in [89], which considers a channel model in which the channel state is also available at both the encoder and decoder. The key difference in the current setting is that channel inputs may also affect future states.

The proof of Theorem 3.1 is given in Appendix C.1. While we consider a different coding strategy, one could attain the same result by extending the feedback strategy considered in [67].

3.4 Converse

In this section, we present a partial converse. We start by defining the key quantity that will permit us to characterize the class for which our converse applies. In particular, our converse hinges on a solution to Bellman's equation (see, e.g. [53, Ch. 8]) for an infinite-horizon dynamic programing problem that we derive in its proof. Specifically, it requires the existence of $|\mathcal{S}|+1$ real numbers

$$(J^*_{\lambda,\alpha}, \ell(1), \ldots, \ell(|\mathcal{S}|))$$

such that for all $i \in \mathcal{S}$,

$$J^*_{\lambda,\alpha} + \ell(i) = \max_{p(x|i)} \left\{ I(X;Y|S=i) + \lambda(\alpha - \mathbb{E}\left[\Gamma(S,X)|S=i\right]) + \sum_{j\in\mathcal{S}} P^p_{i,j} \cdot \ell(j) \right\} . \tag{3.8}$$

Let $p^*_{\lambda,\alpha}(x|i)$ be the maximizing distribution for each $i \in \mathcal{S}$. Then define

$$\alpha^*_\lambda = \sum_{s,x} \Gamma(s,x) \cdot p^*_{\lambda,\alpha}(x|s) \cdot \pi^*_{\lambda,\alpha}(s) \ , \tag{3.9}$$

where $\pi^*_{\lambda,\alpha}(s)$ is the stationary distribution induced by $p^*_{\lambda,\alpha}(x|s)$ on the state process.

Theorem 3.2. *The statement of the result is divided into two parts:*

(i), *If, for some $\lambda \geq 0$, there exist $|\mathcal{S}|+1$ real numbers*

$$(J^*_{\lambda,\alpha}, \ell(1), \ldots, \ell(|\mathcal{S}|))$$

that satisfy (3.8) *for all $i \in \mathcal{S}$, then*

$$C(\alpha) \leq J^*_{\lambda,\alpha} \ , \tag{3.10}$$

and the existence is guaranteed for all $\lambda \geq 0$ if the state process $\{S_k\}_{k\geq 1}$ is irreducible.

(ii), *Suppose the state process $\{S_k\}_{k\geq 1}$ is irreducible. If for some $\lambda' \geq 0$, either $(\lambda' \geq 0, \alpha = \alpha^*_{\lambda'})$ or $(\lambda' = 0, \alpha \geq \alpha^*_{\lambda'})$, then*

$$\begin{aligned} C(\alpha) &= J^*_{\lambda',\alpha} \\ &= \max_{\substack{p(x|s):\\ \mathbb{E}[\Gamma(S,X)]\leq\alpha}} \sum_s I(X;Y|S=s)\pi(s) \ , \end{aligned} \tag{3.11}$$

where $\pi(s)$ is the stationary distribution of the state process $\{S_k\}_{k\geq 1}$, defined in (3.2)*, induced by $p(x|s)$, and the distribution of Y given (X, S) is $w(y|x, s)$.*

Remark: The technical conditions in Part *(ii)* of this theorem are related to a concept in convex optimization known as Slater's condition, which, under suitable regularity conditions, can be used to establish that a convex problem and its Lagrange dual are equal [8, pp. 226-227].

The proof of Theorem 3.2 is given in Appendix C.2.

3.5 Examples

While the technical conditions in Theorem 3.2 might seem daunting at first, the following examples provide insights into when these conditions hold and how they can be used to find the capacity. Let us start by revisiting Example 3.1. Note that the problem can be modeled as follows:

$$\Gamma(s_k, x_k) = \begin{cases} 1, & s_k = 1, x_k = 1 \\ 0, & \text{otherwise} \end{cases} \tag{3.12}$$

$$w(y_k|x_k, s_k) = \begin{cases} 1, & y_k = x_k \\ 0, & \text{otherwise} \end{cases} \tag{3.13}$$

$$q(s_{k+1} = 1|y_k, s_k) = \begin{cases} 0, & y_k = 0 \\ 1, & y_k = 1 \end{cases}. \tag{3.14}$$

It turns out the state process is not irreducible, which can be seen if for the choice

$$p(x_k = 1|s_k) = \begin{cases} 0, & s_k = 0 \\ 1, & s_k = 1 \end{cases}, \tag{3.15}$$

for which there is no path from state 0 to state 1 or vice versa, and the Markov chain is not irreducible, so neither is the state process is not irreducible. However, irreducibility is not necessary for either Theorem 3.1 or Theorem 3.2, and indeed, we can apply both to determine the capacity for $\alpha = 0$.

Proposition 3.3. *The capacity in Example 3.1 with cost constraint $\alpha = 0$ is*

$$\log \phi \ , \tag{3.16}$$

where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio, which is the solution to the equation $\frac{1}{\phi} = \phi - 1$.

Indeed, the achievable strategy presented in the introduction is consistent with the strategy presented in Theorem 3.1, so proving the result simply verifies trying to find an upper bound by solving the Bellman equation in Theorem 3.2. The proof of this is given in Appendix C.3.1.

The examples considered in the sequel all assume an irreducible state process.

3.5.1 Cost function depends only on the state

Our next example considers the case in which the cost depends only on the current channel state.

Example 3.2. Let $\mathcal{X} = \{0, 1\}$, $\mathcal{S} = \{0, 1\}$, and $\mathcal{Y} = \{0, 1\}$. Finally, for $0 < \epsilon_0 <$

$\epsilon_1 < 1$ and $0 < \beta < 1$, we let

$$\Gamma(s_k, x_k) = \begin{cases} 1, & s_k = 1 \\ 0, & \text{otherwise} \end{cases} \tag{3.17}$$

$$w(y_k|x_k, s_k) = \begin{cases} 1, & y_k = x_k \\ 0, & \text{otherwise} \end{cases} \tag{3.18}$$

$$q(s_{k+1} = 1|y_k, s_k) = \begin{cases} \epsilon_0, & y_k = 0, s_k = 0 \\ \epsilon_1, & y_k = 1, s_k = 0 \\ 1 - \beta, & s_k = 1 \end{cases} . \tag{3.19}$$

Proposition 3.4. *The capacity in Example 3.2 with cost constraint α is given by the following expression:*

$$C(\alpha) = \begin{cases} \alpha \cdot \log 2 + (1 - \alpha) \cdot h_b(p_\alpha), & \frac{\epsilon_0}{\epsilon_0+\beta} < \alpha \leq \frac{\epsilon_0+\epsilon_1}{\epsilon_0+\epsilon_1+2\beta} \\ \log 2, & \alpha > \frac{\epsilon_0+\epsilon_1}{\epsilon_0+\epsilon_1+2\beta} \end{cases} , \tag{3.20}$$

where $p_\alpha = \frac{(\epsilon_0+\beta)\alpha-\epsilon_0}{(1-\alpha)\cdot(\epsilon_1-\epsilon_0)}$.

Remark: Alice's achievable strategy here is quite simple: for the state 1, Alice's codebook consists of codewords with roughly half of their inputs as 1 and the other half to 0. For the state 0, Alice's codebook consists of codewords with roughly p_α of their inputs as 1 and the rest as 0, where p_α increases to $\frac{1}{2}$ as α increases to $\frac{\epsilon_0+\epsilon_1}{\epsilon_0+\epsilon_1+2\beta}$.

The proof of Proposition 3.4 is given in Appendix C.3.2. A plot of these rates is given in Example 3.2 under the parameters $\epsilon_0 = 0, \epsilon_1 = \frac{3}{4}, \beta = \frac{1}{2}$.

3.5.2 Cost function depends only on the state and the input

Our next example, an extension of the previous one, examines a case in which the cost function Γ depends upon both the state and the channel input.

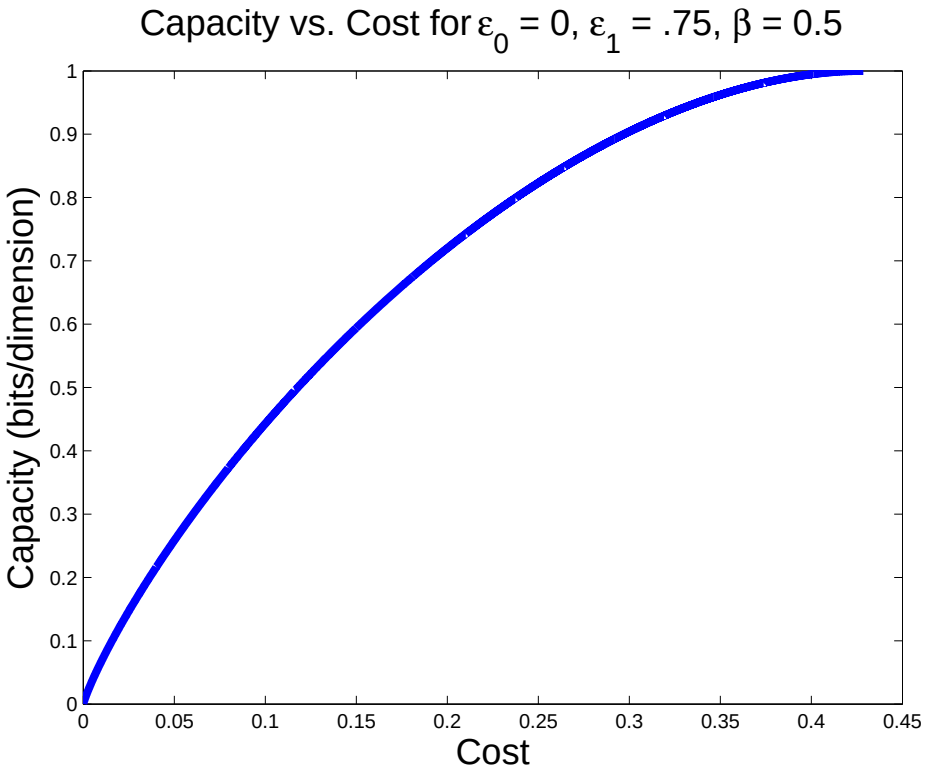

Figure 3.3: The capacity in Example 3.2 is a concave function of the cost, suggesting that without using the state feedback, a naive timesharing of codebooks optimal for the extreme is suboptimal.

Example 3.3. Let $\mathcal{X} = \{0, 1, \ldots, L\}$, $\mathcal{S} = \{0, 1\}$, and $\mathcal{Y} = \mathcal{X}$. Furthermore, we let

$$\Gamma(s_k, x_k) = \begin{cases} 1 & x_k \neq 0, s_k = 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.21}$$

$$w(y_k|x_k, s_k) = \begin{cases} 1 & y_k = x_k \\ 0 & \text{otherwise} \end{cases} \tag{3.22}$$

$$q(s_{k+1} = 0|y_k, s_k) = \begin{cases} 1 & y_k = 0, s_k = 1 \\ \frac{1}{2} & \text{otherwise} \end{cases} . \tag{3.23}$$

Proposition 3.5. *The capacity in Example 3.3 with cost constraint α is given by the*

following expression:

$$C(\alpha) = \begin{cases} \frac{2-\alpha}{3} \cdot \log(L+1) + \alpha \cdot \log L + \frac{1+\alpha}{3} \cdot h_b\left(\frac{1-2\alpha}{1+\alpha}\right), & 0 \leq \alpha \leq \frac{L}{2L+3} \\ \log(L+1), & \alpha > \frac{L}{2L+3} \end{cases}, \quad (3.24)$$

where $h_b(\cdot)$ is the binary entropy function.

Remark: Note that at $\alpha = 0$, $C(0) = \frac{2}{3}\log(L+1)$. By contrast, it can be shown by a type counting argument that the capacity without feedback for this case is 0. Since L can be arbitrary large, the gap between feedback and nonfeedback capacity can be arbitrarily large for an appropriately chosen example. The interested reader is referred to [34] for further discussion on such examples.

The proof of Proposition 3.5 is given in C.3.3.

3.5.3 Noisy channel

Note that above, we have considered Alice's actions to have deterministic consequences with respect to victory. However, the victories of Alice's army may not be so predictable, which the next example explores.

Example 3.4. Let $\mathcal{X} = \{0, 1, \ldots, L\}$, $\mathcal{S} = \{e, c\}$, and $\mathcal{Y} = \{e, 0, 1, \ldots, L\}$. Furthermore, we let

$$\Gamma(s_k, x_k) = \begin{cases} 1 & x_k \neq 0, s_k = e \\ 0 & \text{otherwise} \end{cases} \quad (3.25)$$

$$w(y_k|x_k, s_k) = \begin{cases} 1-\epsilon & y_k = x_k \\ \epsilon & y_k = e \end{cases} \quad (3.26)$$

$$q(s_{k+1} = e|y_k, s_k) = \begin{cases} 1 & y_k = e \\ 0 & \text{otherwise} \end{cases}. \quad (3.27)$$

Proposition 3.6. *The capacity in Example 3.4 with cost constraint α is given by the following expression:*

$$C(\alpha) = \begin{cases} (1-\epsilon)\cdot\left((1-\epsilon)\log(L+1) + \epsilon\cdot h_b\left(\frac{\epsilon-\alpha}{\epsilon}\right) + \alpha\log L\right), & 0 \le \alpha \le \frac{\epsilon L}{L+1} \\ (1-\epsilon)\cdot\log(L+1), & \alpha > \frac{\epsilon L}{L+1} \end{cases}, \tag{3.28}$$

where $h_b(\cdot)$ is the binary entropy function.

The proof of Proposition 3.6 is given in Appendix C.3.4.

3.6 Discussion

In this chapter, we have explored a tradeoff between communicating over a channel and simultaneously controlling a channel state to satisfy a cost constraint. We have presented an achievable strategy and provided a partial converse under which it yields the optimal tradeoff between communication and control. Using these results, we presented a series of examples in which we could evaluate the capacity expression. As a by-product of one example, we found that the gap between non-feedback and feedback capacity can be arbitrarily large within our framework.

This study was inspired by the problem of a cognitive radio sharing spectrum with other users, and in such a setting, the cost represents both the physical limitations of the cognitive radio, which is reflected in the channel input, and how much interference it inflicts on other users, which is reflected in the channel state. This state was known causally at both the encoder and decoder, which resulted in an achievable strategy whereby the encoder and decoder synchronize over parallel codebooks that depend on their knowledge of the current channel state. As noted in [12], because the channel state is available at the decoder, an equivalent model is to collapse the state into the

channel output, where the cost is then a function of the channel input and output.

However, if the state actually reflects the interference to a third-party user, there could be uncertainty about the current channel state at both the encoder and decoder, which could be introduced either by a *delay* in receiving this information or *noise* in the state observations. If there were a delay, using the approach in [89], the encoder and decoder could still synchronize across parallel codebooks based on a common channel state estimate constructed from the states available to both. On the other hand, if there were noise in the state observations, such synchronization might not be possible, and tree-structured codes as in [96; 67] would have to be adapted.

The current study was limited to discrete channels, and within this class, it was shown by an example that the gap between feedback and non-feedback capacity can be unbounded. While this observation may not be too surprising, one should perhaps recall that for a class of Gaussian channels with memory, feedback can increase the capacity by at most the smaller of $\frac{1}{2}$ a bit and twice the non-feedback capacity [20]. This class, however, does not assume that the cost function can depend on the state of the channel. If the current study were extended to the Gaussian setting, it could give further insights into the gap between the feedback and non-feedback capacity.

Our problem was point-to-point, consisting of a single encoder and decoder. Again, inspired by spectrum sharing, one could consider what would happen if multiple users needed to satisfy a common cost constraint, as in [40]. In such a setting, feedback could serve two roles. First, the encoders could control the state and thus satisfy the cost constraint. Second, the encoders could correlate their channel inputs, thereby increasing their rate region. Indeed, this latter approach has been leveraged in other problems with multiple transmitters or receivers [65; 75; 76; 51; 91]. Such studies could shed further light on the role of feedback in communication.

Chapter 4

Secrecy via Sources and Channels

4.1 Introduction

4.1.1 Secrecy Constraints and Private Key Cryptography

On October 21, 2007 the canton of Geneva used quantum cryptography to secure the wired link used to count ballots during the Swiss national election.[1] The codes used were dubbed "unbreakable" because of the secrecy constraint it must satisfy [14, p. 584]. The constraints are based on a classical problem considered by Shannon [77] for private key cryptosystems. Shannon's results gave a pessimistic requirement that the entropy of a private key should be at least as large as the entropy of the message to be sent, a result that motivated the development of public key cryptography [28; 69]. This chapter focuses on a variation of the private key cryptography problem.

[1] "Geneva is counting on Quantum Cryptography as it counts its votes," Press release of Geneva State Chancellery:
`http://www.idquantique.com/news/files/com_swissquantum_ang.pdf`

4.1.2 Secrecy via Sources and Channels

Alice has a secret message she wants to send Bob, but unfortunately, she must do so in the presence of Eve, an eavesdropper. This chapter explores a new dimension of this familiar problem: how can Alice efficiently utilize two disparate resources in order to keep this message secret from Eve? The first resource is a one-way noisy broadcast channel from Alice to Bob and Eve, and the second resource is the presence of correlated source observations at Alice, Bob, and Eve. Specifically, we are interested in understanding how to design achievable strategies that combine these resources optimally in order to support secure communication between Alice and Bob.

There already exists a body of literature for cases in which only one of these resources is available. Wyner's seminal work, "The Wire-tap Channel" [93] considered secure communication over degraded broadcast channels [17] and was later generalized by Csiszár and Körner [22] to cover all broadcast channels. Consider the following example, which highlights one of the key principles presented in these works. This will be the first in a sequence of three examples, the third of which builds on the first two to illustrate the main ideas in this chapter.

Example 4.1. Suppose Alice has a three-bit noiseless channel to Bob with each bit denoted as 0 or 1. Eve can observe only two of the three bits sent by Alice, but not all of them. Alice can use this advantage Bob has over Eve to send a one-bit secret message $M \in \{\mathsf{a}, \mathsf{b}\}$ to Bob such that Eve will consider both outcomes to be equally likely. In order to do this, Alice may make use of two fair coin tosses, for which `T` or `H` are equiprobable. Specifically, Alice chooses her channel input according to the

following table:

	coin toss			
	TT	TH	HT	HH
$M = \mathsf{a}$	000	011	101	110
$M = \mathsf{b}$	001	010	100	111

Note that in the top row, 1 occurs an even number of times (even parity) whereas in the bottom row, 1 occurs an odd number of times (odd parity). Since Bob can observe Alice's channel input noiselessly, Bob can calculate its parity to determine the row and thus the message M. Eve, on the other hand, will see only two of the bits and because of the fair coin toss, the remaining bit from Eve's perspective is equally likely to be 0 or 1. Thus, from Eve's perspective, it is equally likely that the output came from either row of the table, and thus each message is equally likely.

Analogously, Ahlswede and Csiszár [3] and Maurer [60] recognized that dependent source observations available at the terminals can be used as a resource for generating a secret-key (a uniform random variable shared by Alice and Bob which Eve is oblivious to) if the terminals can communicate over a noiseless public channel (which delivers all its input faithfully to all the terminals including the eavesdropper). In [3], the secret-key capacity of dependent sources was characterized if a one-way noiseless public channel from Alice to Bob and Eve of unconstrained capacity is available. The characterization for the case when there is a constraint on the capacity of the public channel was later found by Csiszár and Narayan [24] as a special case of their results on a class of common randomness generation problems using a helper. As in the channel setting, one can also exploit distributed sources for sending a secret message. The next example highlights this principle in the source setting.

Example 4.2. (a) Consider the setting in which Alice is allowed to transmit one

bit across a noiseless public channel to Bob and Eve. Furthermore, Alice observes a two-bit string uniformly distributed over all of these strings. Bob observes either the first or the second bit of Alice's string, but not both, and Alice does not learn which of the two bits Bob observed. Thus, Bob can narrow down Alice's two-bit string to one of two possible strings. Then Alice and Bob can agree on a common random bit, a secret key, as follows. Let the secret key be the first bit of Alice's two-bit string. Alice simply transmits the XOR of her two-bit string:

Alice's 2-bit sequence		
key=0	00	01
key=1	11	10
bit-pipe input:	0	1

.

Given the XOR bit, Bob can recover Alice's two-bit string and thus determine the first bit. Furthermore, regardless of what the first bit of Alice's string is, Eve is equally likely to see a 0 or 1 from the public channel. Thus, Alice and Bob share a secret bit that is hidden from Eve.

(b) Suppose that Alice is allowed to transmit an additional bit across the noiseless public channel. Note given a secret message $M \in \{\mathsf{a}, \mathsf{b}\}$, Alice can simply transmit the XOR of the message ($\mathsf{a} = 0, \mathsf{b} = 1$) and the key, so Bob can decode the message. To see that Eve will be confused about which message is sent, consider Alice's overall

strategy for selecting the public channel input:

	Alice's 2-bit string			
	key = 0		key = 1	
	00	01	11	10
$M = \mathsf{a}$	00	10	01	11
$M = \mathsf{b}$	01	11	00	10

Note that regardless of which message is selected, Eve is equally likely to see all four possible public channel outputs, and thus for her, both messages appear equally likely.

We now provide an example to illustrate the key features that occur when both resources are available as in this chapter.

Example 4.3. Suppose Alice has a three-bit noiseless channel to Bob, and Eve can observe only two of the three bits as in Example 4.1. Additionally, Alice and Bob have source observations as in Example 4.2. The key idea is to combine the strategies used above, except to replace Alice's coin tosses in Example 4.1 with the input to the public channel from Example 4.2(b). With this combined strategy, Alice can send a two-bit secret message $M \in \{\mathsf{ca}, \mathsf{cb}, \mathsf{da}, \mathsf{db}\}$ to Bob. To make this explicit, Alice sends the first bit of the message over the channel using a strategy analogous to Example 4.1, swapping only the coin tosses with the public channel input from Example 4.2(b), which encodes the second bit of the message:

	public channel input from Example 4.2(b)			
	00	01	10	11
$M = \mathsf{c}*$	000	011	101	110
$M = \mathsf{d}*$	001	010	100	111

Because Bob can see the channel output noiselessly, he can determine both the first

bit of Alice's two-bit string and the corresponding input to the public channel input from Example 4.2(b) which is encoded as before:

	Alice's 2-bit string			
	key = 0		key = 1	
	00	01	11	10
$M = *\mathsf{a}$	00	10	01	11
$M = *\mathsf{b}$	01	11	00	10

Again, as in Example 4.2(b), Bob can decode. Since the public channel input in Example 4.2(b) is uniform, the channel advantage enables the first bit of the message to be kept secret from Eve as in Example 4.1. Note that even if Eve were given the first bit of the message, and thereby were able to determine the 2-bit input to the public channel, by Example 4.2(b), this alone is insufficient for Eve to determine the second bit of the message. Thus, with this joint strategy, Alice can keep both bits of the message secret from Eve.

A more realistic scenario than Example 4.3 arises in wireless sensor networks, in which sensors have access to both a wireless channel and their correlated sensor readings. Note that in such situations, fading can cause the channel characteristics to be more or less favorable to secrecy at different points in time. Thus, when the channel characteristics are favorable, it can be advantageous for Alice and Bob, instead of (or in addition to) sending a specific secret message, to simply agree on a sequence of private common random bits (a secret key) to be used later when the characteristics are unfavorable. See Khalil et al. [47], for an example of how this can enable a form of secure communication with delay constraints under such conditions. Not surprisingly, it turns out that in some settings, one can achieve higher rates for the secret key than the more restrictive secret message. The general problem we consider abstracts this

issue into considering a tradeoff between transmitting a uniform source privately (a secret message) and generating private common randomness (a secret key). Another related recent work is [48] which independently investigates secret key generation in a similar setting.

Section 4.2 gives a formal description of the problem setup, and Section 4.3 describes the main results presented in this chapter. Section 4.4 considers examples to illustrate the essence of our scheme. Here we show how the scheme can be interpreted as one which separates the source and channel as resources for aquiring secrecy. Section 4.5 gives a formal proof of the coding scheme. The chapter concludes with a discussion and directions for future work in Section 4.6.

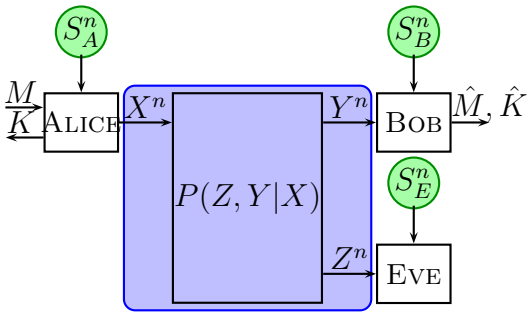

Figure 4.1: Problem setup: Alice and Bob want to share a key K and independent message M, both of which they want to be kept secret from Eve. Alice has a memoryless broadcast to Bob and Eve. Additionally, Alice, Bob, and Eve have three have correlated memoryless sources.

4.2 Problem Setup

Notation: We denote random variables by upper-case letters (*e.g.*, X), their realizations by lower-case letters (*e.g.*, x), and the alphabets over which they take values by calligraphic letters (*e.g.*, $\mathcal{X}$). A vector $(X_k, X_{k+1}, \ldots, X_n)$ will be denoted by X_k^n. When $k = 1$, the subscript will be dropped as in $X^n = (X_1, X_2, \ldots, X^n)$.

We make the following assumptions on the channel and sources:

- The channel and the sources are memoryless.
- The channel is independent of the sources.
- The number of source observations is the same as the number of channel uses available. Note that we allow all the sources to be observed ahead of time in the sense that the input to the channel may depend on the block of source observations.

The assumption on the number of source observations and channel uses can be easily relaxed and is made here only for reducing the notation. However, the independence assumption is critical to the results we present here. The memoryless assumption is useful for getting simple closed-form single-letter expressions.

We consider the following model. Alice, Bob and Eve observe, respectively, the dependent memoryless processes (sources) $S_{A,k}, S_{B,k}, S_{E,k}$, where $k = 1, 2, \ldots$ is the time index. They have a joint distribution p_{S_A,S_B,S_E} over the alphabet $\mathcal{S}_A \times \mathcal{S}_B \times \mathcal{S}_E$. Independent of these sources, there is a memoryless broadcast channel from Alice to Bob and Eve given by $p_{Y,Z|X}$, where X_k is the input to the channel, Y_k is Bob's output, and Z_k Eve's. We will also allow Alice to have access to a private random variable Φ_A uniformly distributed over the unit interval $[0, 1]$, which is not available to Bob and Eve and which is independent of all other random variables. Alice may use this private random variable for purposes of randomization. Finally, there is a secret message M, which is uniformly distributed over its alphabet $\mathcal{M}$ and independent of the sources and channel

Definition 4.1. An $(n, R_{\mathsf{SK}}, R_{\mathsf{SM}})$ *secrecy codebook* with blocklength n, secret key rate R_{SK} and secret message rate R_{SM}, consists of an encoding function $\mu : \mathcal{M} \times \mathcal{S}_A^n \times [0, 1] \to \mathcal{X}^n$, key function $g : \mathcal{S}_A^n \times [0, 1] \to \mathcal{K}$, and a decoding function $\nu : \mathcal{Y}^n \times \mathcal{S}_B^n \to$

$\mathcal{M} \times \mathcal{K}$, where $R_{\mathsf{SK}} = \frac{1}{n} \log |\mathcal{K}|$ and $R_{\mathsf{SM}} = \frac{1}{n} \log |\mathcal{M}|$. Given an $(n, R_{\mathsf{SK}}, R_{\mathsf{SM}})$ secrecy codebook, the channel inputs are given by $X^n = \mu(M, S_A^n, \Phi_A)$, and the secret key is given by $K = g(S_A^n, \Phi_A)$.

From Definition 4.1, the following distinction between secret key and message is apparent: the secret key is constructed by the secrecy codebook, but the message is prespecified.

Definition 4.2. A secret key and secret message rate pair $(R_{\mathsf{SK}}, R_{\mathsf{SM}})$ is ϵ-*achievable* with blocklength n if there exists an $(n, R_{\mathsf{SK}} - \epsilon, R_{\mathsf{SM}} - \epsilon)$ secrecy codebook such that the following conditions hold:

$$\mathbb{P}\left(\nu(Y^n, S_B^n) \neq (K, M)\right) \leq \epsilon \tag{4.1}$$

$$\frac{1}{n} \log |\mathcal{K}| - \frac{1}{n} H(K) \leq \epsilon \tag{4.2}$$

$$\frac{1}{n} I(M, K; Z^n, S_E^n) \leq \epsilon \ . \tag{4.3}$$

Definition 4.3. A secret key and secret message rate pair $(R_{\mathsf{SK}}, R_{\mathsf{SM}})$ is *achievable* if for all $\epsilon > 0$ and sufficiently large n, $(R_{\mathsf{SK}}, R_{\mathsf{SM}})$ is ϵ-achievable with blocklength n.

Definition 4.4. We define the rate region $\mathcal{R}$ to be the set of all achievable rate pairs.

4.3 Results

Let $\mathcal{P}$ be the set of all joint distributions p of random variables U_1, V_1, V_2, X, Y, Z, S_A, S_B, S_E such that U_1 and (V_1, V_2) are independent, the following two Markov chains holds:

$$U_1 - S_A - (S_B, S_E),$$
$$V_2 - V_1 - X - (Y, Z),$$

the joint distribution of (S_A, S_B, S_E) and the joint conditional distribution of (Y, Z) given X are consistent with the given source and channel respectively, and

$$I(V_1; Y) \geq I(U_1; S_A|S_B).$$

For $p \in \mathcal{P}$, let $\mathcal{R}(p)$ be the set of all non-negative pairs $(R_{\mathsf{SK}}, R_{\mathsf{SM}})$ which satisfy the following two inequalities

$$R_{\mathsf{SM}} \leq I(V_1; Y) + I(U_1; S_B) - I(U_1; S_A), \tag{4.4}$$

$$R_{\mathsf{SK}} + R_{\mathsf{SM}} \leq [I(V_1; Y|V_2) - I(V_1; Z|V_2)]_+ + [I(U_1; S_B) - I(U_1; S_E)]_+, \tag{4.5}$$

where $[x]_+ \stackrel{\text{def}}{=} \max(0, x)$. The next theorem states that all pairs of rates belonging to $\mathcal{R}(p)$ are achievable. The complete proof is provided in Section 4.5.

Theorem 4.1.

$$\mathcal{R} \supseteq \bigcup_{p \in \mathcal{P}} \mathcal{R}(p).$$

Remark: It can be shown that in taking the union above, it suffices to consider auxiliary random variables with a sufficiently large, but finite cardinality. This can be shown using Carathéodery's theorem (see [22], for instance).

The next theorem states that the above inner bound to the trade-off region can be used to derive a tight innerbound for the parallel channels and sources case when each sub-channel and source component satisfies a degradation order (either in favor of the legitimate receiver or the eavesdropper). Figure 4.2 depicts this scenario.

Theorem 4.2. *Consider the following:*

(i) The channel has two independent components[2] *denoted by F and R: $\mathbf{X} = (X_F, X_R)$, $\mathbf{Y} = (Y_F, Y_R)$, and $\mathbf{Z} = (Z_F, Z_R)$ such that $p_{Y_F, Y_R, Z_F, Z_R|X_F, X_R} =$*

[2]We denote the channel input, outputs, and the sources using bold letters to make this explicit.

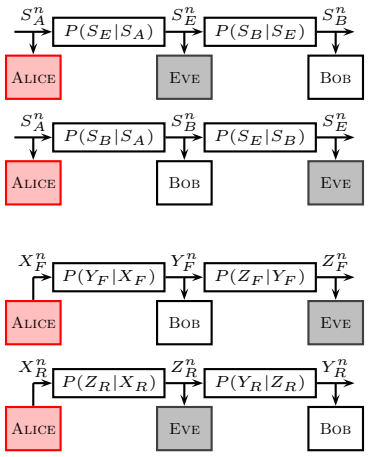

Figure 4.2: Theorem 4.2 states that the separation strategies used to establish Theorem 4.1 are optimal if the sources and channels can be decomposed to satisfy a degradation order, either in favor of Bob or Eve.

$p_{Y_F,Z_F|X_F} p_{Y_R,Z_R|X_R}$. *Moreover, the first sub-channel F is degraded in favor of Bob, which we call forwardly degraded, and the second sub-channel R is degraded in favor of Eve, which we call reversely degraded;* i.e., $X_F - Y_F - Z_F$ *and* $X_R - Z_R - Y_R$ *are Markov chains.*

(ii) The sources also have two independent components, again denoted by F and R: $\mathbf{S}_A = (S_{A,F}, S_{A,R})$, $\mathbf{S}_B = (S_{B,F}, S_{B,R})$, *and* $\mathbf{S}_E = (S_{E,F}, S_{E,R})$ *with* $p_{\mathbf{S}_A,\mathbf{S}_B,\mathbf{S}_E} = p_{S_{A,F},S_{B,F},S_{E,F}} p_{S_{A,R},S_{B,R}S_{E,R}}$. *The first component is degraded in favor of Bob and the second in favor of Eve;* i.e., $S_{A,F} - S_{B,F} - S_{E,F}$ *and* $S_{A,R} - S_{E,R} - S_{B,R}$ *are Markov chains.*

In this case,

$$\mathcal{R} = \bigcup_{p \in \tilde{\mathcal{P}}} \tilde{\mathcal{R}}(p),$$

where $\tilde{\mathcal{P}}$ is the set of joint distributions of the form

$$p_{V_{2,F},X_F} \cdot p_{Y_F,Z_F|X_F} \cdot p_{X_R} \cdot p_{Y_R,Z_R|X_R} \cdot p_{U_{1,F}|S_{A,F}} \cdot p_{S_{A,F},S_{B,R},S_{E,R}} \cdot p_{S_{A,R},S_{B,R},S_{E,R}}$$

and $\tilde{\mathcal{R}}(p)$ *is the set of non-negative pairs of* $(R_{\mathsf{SK}}, R_{\mathsf{SM}})$ *satisfying*

$$R_{\mathsf{SM}} \leq I(X_F; Y_F) + I(X_R; Y_R) - I(U_{1,F}; S_{A,F}|S_{B,F}), \text{ and} \tag{4.6}$$

$$R_{\mathsf{SK}} + R_{\mathsf{SM}} \leq I(X_F; Y_F|V_{2,F}) - I(X_F; Z_F|V_{2,F}) + I(U_{1,F}; S_{B,F}|S_{E,F}). \tag{4.7}$$

We prove this theorem in Appendix D.2. It turns out the result is more general than the form presented above, but these extensions are omitted to be able to state the result cleanly. These extensions are discussed in greater detail in Section 4.6.

4.4 Achievability as a Separation Strategy

In this section we will sketch informally the achievable scheme behind Theorem 4.1. The sketch of strategy follows the spirit of Examples 4.1, 4.2, and 4.3 from the introduction to this chapter and provide an interpretation of the result as a separation strategy. A formal proof is provided in Section 4.5. The section concludes with a Gaussian example.

4.4.1 Case of no sources: secrecy via the channel

Consider the case in which there is a noisy broadcast channel from Alice to Bob and Eve but no sources. Note that this resembles the cases studied in Example 4.1 with the added wrinkle that the channel to Bob may also be noisy. Recall that in Example 4.1, given sufficiently many fair coin tosses, Alice uses the channel to send a message secretly to Bob. The work of Csiszár and Körner [22] generalizes this approach as a means of providing secrecy for all noisy broadcast channels. The following proposition can be proved (adapted from [95]) if we assume the messages and coin tosses and independent and uniform over their alphabets.

Proposition 4.3. *For any given joint distribution of random variables V_1, V_2, X, Y, Z such that $V_2 - V_1 - X - (Y, Z)$ is a Markov chain and the joint conditional distribution of (Y, Z) given X is consistent with the given channel, then using the channel and R_{public} tosses of a fair coin, the secret message rate of R_{private} is achievable, where*

$$R_{\text{private}} = [I(V_1; Y|V_2) - I(V_1; Z|V_2)]_+$$
$$R_{\text{public}} = I(V_1; Y) - R_{\text{private}} \ .$$

Furthermore, Bob can determine the outcome of the R_{public} coin tosses with high probability.

4.4.2 Case of two noiseless bit pipes: private and public

Now consider the setting in which the channel is deterministic. In particular, the channel is made up of two bit-pipes: (1) a *private* bit-pipe of rate R_{private} which delivers its input bits from Alice faithfully and only to Bob, and (2) a *public* bit-pipe of rate R_{public} which delivers faithfully its input bits from Alice to both Bob and Eve.

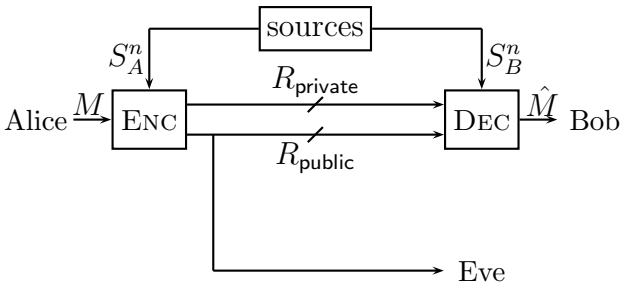

Figure 4.3: Consider the case in which Alice and Bob share correlated source observations, and there is both a private bit-pipe from Alice to Bob and a public bit-pipe from Alice to Bob and Eve. This generalizes the problem considered in Example 4.2, and the strategy considered in that setting generalizes naturally, as well.

4.4.2.1 Secret-key only; no source observation at Eve

Consider the goal of generating the largest secret-key rate possible when there is no source observation at Eve.[3] This is reminiscent of Example 4.2(a) but with two added dimensions not present in that setting. First, there may not be enough rate on the public bit-pipe for Bob to determine Alice's source observation perfectly to generate a secret key. A modified form of Wyner-Ziv's source coding strategy can be employed to handle this, which simply involves quantizing Alice's source and using that to generate the secret key. Second, in addition to the public bit-pipe, there is also a private bit-pipe. Note that any component sent on the private bit-pipe is automatically a secret key, as well. Thus, if part of the bin index is sent on the private bit-pipe, it is also secret from Eve.

Then, given an auxiliary random variable U which satisfies the Markov chain $U - S_A - S_B$, a secret-key rate of $(I(U;S_B) + R) + R_{\mathsf{private}} - R$ if

$$I(U;S_A) - I(U;S_B) \leq R_{\mathsf{public}} + R, \text{ and}$$
$$R \leq R_{\mathsf{private}}.$$

Again, the work of Ahlswede-Csiszár [3] can be used to show that this has the required secrecy and uniformity properties, and thus, the resulting secret-key rate is

$$R_{\mathsf{SK}} = (I(U;S_B) + R) + (R_{\mathsf{private}} - R) = I(U;S_B) + R_{\mathsf{private}},$$

[3] When Eve has a dependent source observation S_E, a further binning of the codebook described in this section can be used to get a secret-key rate of

$$R_{\mathsf{SK}} = I(U;S_B) - I(U;S_E) + R_{\mathsf{private}},$$

where we restrict U to those which satisfy

$$I(U;S_A) - I(U;S_B) < R_{\mathsf{public}} + R_{\mathsf{private}},$$

and the Markov chain $U - S_A - (S_B, S_E)$.

where we restrict U to those which satisfy

$$I(U;S_A) - I(U;S_B) < R_{\mathsf{public}} + R_{\mathsf{private}},$$

and the Markov chain $U - S_A - S_B$.

4.4.2.2 Secret message only; no source observation at Eve

Consider the case in which Alice desires to communicate a message secretly at the largest possible rate when there is no source observation at Eve.[4] This scenario, depicted in Figure 4.3, is a straightforward generalization of Example 4.2(b) from the introduction. In that example, Alice achieves secrecy across a public bit-pipe by binning her source observation based on Bob's side information to generate a shared secret key. On the rest of the public bit-pipe, Alice uses this key as a one-time pad to send the secret message.

As earlier, there are two added dimensions in the current setting that are not present in Example 4.2(b). First, there may not be enough rate on the public bit-pipe for Bob to determine Alice's source observation perfectly and thus generate a secret key. Again, Alice simply quantizes the source observation and applies the binning strategy as before, which corresponds to Wyner-Ziv's source coding scheme. Second, in addition to the public bit-pipe, there is also a private bit-pipe. Because there is now a secret message, we split the message into two parts: the private bit-pipe is used

[4]When Eve has a correlated source observation, a further binning of the codebook described in this section can be used to get a secret message rate of

$$R_{\mathsf{SM}} = R_{\mathsf{private}} + [I(U:S_B) - I(U;S_E)]_+,$$

where U satisfies the Markov chain $U - S_A - (S_B, S_E)$ and the condition

$$R_{\mathsf{public}} > I(U;S_A).$$

fully to send part of the secret message (at rate R_{private}), and the public bit-pipe is used as before to communicate the remaining bits secretly with the correlated sources being exploited to provide the secrecy. However, since we now have to agree on *specific* random bits instead of *any* common random bits, we have two additional restrictions, which can cause the rate of the secret message to be lower than the secret key case above. First, we have to reserve part of the public bit-pipe for sending the one-time padded secret message, which constrains part of the public bit-pipe rate R_{public} for generating the secret key from the sources. Second, sending part of the Wyner-Ziv bin index on the private bit-pipe will cost rate that can be used for sending a private message. Thus, it is better to reserve the private bit-pipe for sending a secret message, which costs $R \leq R_{\mathsf{private}}$ bits that could have been used for generating the secret key, which means the rate of the secret key used as a one-time pad, and thus the effectiveness of the public bit-pipe, is significantly limited compared with the case of the secret key.

The work of Ahlswede-Csiszár [3] can be adapted to show that this approach satisfies the required secrecy and uniformity properties. The secret-key is then used as a one-time pad to encrypt some extra messages bits. Using this approach, given any joint distribution of $U - S_A - S_B$, a secret key of $I(U;S_B)$ can be generated by consuming $I(U;S_A) - I(U;S_B)$ bits from the public bit-pipe. This secret key can then be used as a one-time pad on another $I(U;S_B)$ bits of the public bit-pipe to send a secret message of that rate. Hence, we must choose auxiliary random variable U such that

$$R_{\mathsf{public}} > (I(U;S_A) - I(U;S_B)) + I(U;S_B) = I(U;S_A),$$

and the total secret message rate obtained is

$$R_{\mathsf{SM}} = R_{\mathsf{private}} + I(U : S_B).$$

Unlike in the work of Csiszar-Narayan [24], in which Alice and Bob only need to agree on *any* common random bits to construct a secret key, for a secret message, we have the added constraint that they must agree on *specific* random bits. Thus, the rates achievable for secret message are less than those achievable for secret key.

4.4.2.3 Secret message – secret-key tradeoff; no source observation at Eve

A secret-message – secret-key tradeoff optimal strategy here[5] turns out to be a natural combination of the above two: If (1) $R_{\mathsf{SM}} \leq R_{\mathsf{private}}$, the secret-message is sent entirely over the private bit-pipe, and the left-over rate $(R_{\mathsf{private}} - R_{\mathsf{SM}})$ of the private bit-pipe rate along with the public bit-pipe is used for agreeing on a secret-key from the correlated sources. This secret-key step is essentially the *secret-key only* case discussed above. Otherwise, *i.e.*, if (2) $R_{\mathsf{SM}} \geq R_{\mathsf{private}}$, all of the private bit-pipe is used to carry a part of the secret message. For communicating the rest of the secret message, at a rate of $R_{\mathsf{SM}} - R_{\mathsf{private}}$, and for agreeing on a secret-key, the public bit-pipe and the sources are made use of. The way the public bit-pipe is used is essentially the same as in the *secret message only* case above. The only difference is that instead of utilizing all of the secret-key generated from the sources as a one-time

[5] When Eve has a correlated source observation S_E, the tradeoff becomes

$$\begin{aligned} R_{\mathsf{SM}} &\leq R_{\mathsf{public}} + R_{\mathsf{private}} - (I(U;S_A) - I(U;S_B)), \text{ and} \\ R_{\mathsf{SM}} + R_{\mathsf{SK}} &\leq [I(U;S_B) - I(U;S_E)]_+ + R_{\mathsf{private}}, \end{aligned}$$

where U satisfies the Markov chain $U - S_A - S_B$ and the condition

$$I(U;S_A) - I(U;S_B) \leq R_{\mathsf{public}} + R_{\mathsf{private}}.$$

pad to secure communication of a message over the public bit-pipe, here, only a part of the secret-key is used for this purpose. The rate of the unused part of the secret-key is R_{SK}.

The resulting tradeoff is given by

$$R_{\mathsf{SM}} \leq R_{\mathsf{public}} + R_{\mathsf{private}} - (I(U;S_A) - I(U;S_B)), \text{ and}$$
$$R_{\mathsf{SM}} + R_{\mathsf{SK}} \leq I(U;S_B) + R_{\mathsf{private}},$$

where U satisfies the Markov chain $U - S_A - S_B$ and the condition

$$I(U;S_A) - I(U;S_B) \leq R_{\mathsf{public}} + R_{\mathsf{private}}.$$

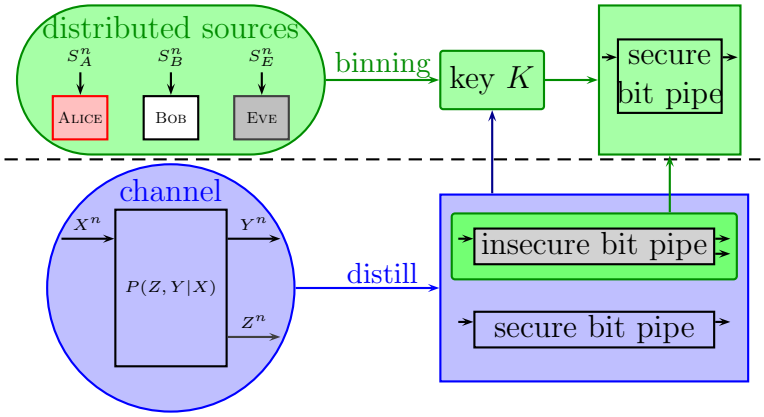

Figure 4.4: In the separation strategy, the channel is distilled into a public and private bit pipe. The sources take advantage of part of the rate from each bit pipe to generate a secret key. This key is divided into the final secret key and a one-time pad that is used to secure the remainder of the public bit pipe for sending part of the secret message. The remainder of the private bit pipe is used to send the remainder of the secret message.

4.4.3 The General case

Now let us turn to the general case with sources in which the channel is not necessarily deterministic. This resembles Example 4.3, and as in that case, we can apply a combination of the strategies in Section 4.4.1 and 4.4.2. Indeed, by treating the random coin tosses Alice uses in Proposition 4.3 as a public bit-pipe, we can construct a public and private bit-pipe from the channel and can leverage the source strategy from Section 4.4.2. This approach enables us to obtain the rates in (4.4) and (4.5). However, we should note that neither the independence requirement nor the uniformity requirement in Proposition 4.3 hold for the messages sent over the bit-pipes in 4.4.2, though they may hold approximately. Hence, this discussion does not constitute a proof of Theorem 4.1. We prove rigorously in Section 4.5. A schematic interpretation of the discussion in this section is shown in Figure 4.4.

4.4.4 A Gaussian example

Let us consider a scalar Gaussian example. Suppose the observations of Alice and Bob are jointly Gaussian. Then, without loss of generality, we can model them as

$$S_B = S_A + N_{\text{source}},$$

where S_A and N_{source} are zero mean Gaussian. Let N_{source} be unit variance, and let the variance of S_A be SNR_{src}. Let Eve have no source observation. Suppose that the broadcast channel has additive Gaussian noise with a power constraint on X of SNR_{Bob}. Let

$$Y = X + N_{\text{Bob}}, \text{ and}$$
$$Z = X + N_{\text{Eve}},$$

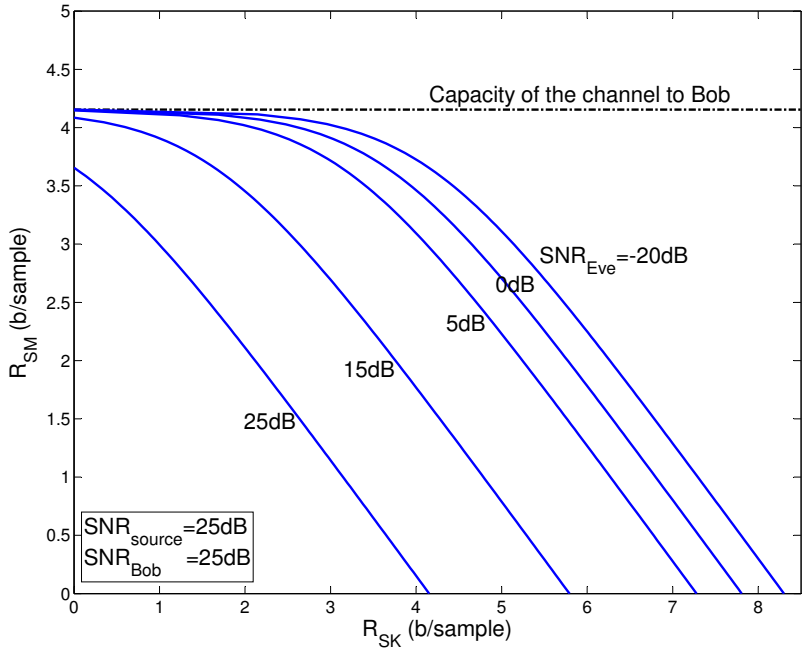

Figure 4.5: The tradeoff between key and message in the Gaussian setting. Note that this tradeoff is not simply linear.

where N_{Bob} and N_{Eve} are Gaussians independent of X, and such that N_{Bob} has unit variance and N_{Eve} has a variance $\text{SNR}_{\text{Bob}}/\text{SNR}_{\text{Eve}}$. We have the following proposition which is proved in Appendix D.3.

Proposition 4.4. *The rate region $\mathcal{R}$ for this problem is set of all non-negative $(R_{\mathsf{SK}}, R_{\mathsf{SM}})$ pairs satisfying*

$$R_{\mathsf{SM}} \leq \frac{1}{2}\log\frac{(1+SNR_{src})(1+SNR_{Bob})}{1+SNR_{src}+\min(SNR_{Bob},SNR_{Eve})},$$
$$R_{\mathsf{SK}} \leq \frac{1}{2}\log\frac{(1+SNR_{src})(1+SNR_{Bob})\exp(-2R_{\mathsf{SM}})-SNR_{src}}{1+\min(SNR_{Bob},SNR_{Eve})}$$

Remark: When Eve also has a source observation jointly Gaussian with the observa-

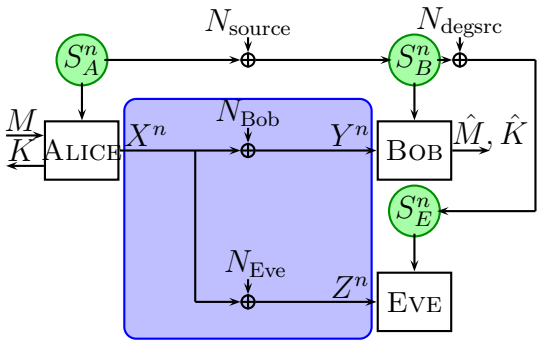

Figure 4.6: When the noises are additive Gaussian, and the sources are jointly Gaussian, Theorem 4.2 applies, and the rate region can achieved by a separation strategy.

tions of Alice and Bob, the problem is covered by the cases in Theorem 4.2. However, unlike in the proposition above, we were unable to show that a Gaussian choice of the auxiliary random variables is optimal.

4.5 General Achievable Strategy

Note that in the previous section, we outlined how the rate region in Theorem 4.1 can be achieved by a separation strategy: i.e. one that considers sources and channels separately. The strategy assumed sources and channels were independent and involved Wyner-Ziv encoding of the sources and Gelfand-Pinsker type of the channel to ensure secrecy. In this section, we construct an achievable strategy for the more general case in which the sources and channel are correlated from an insight based on the duality between Wyner-Ziv and Gelfand-Pinsker. We then show that the region achieved by this general strategy reduces to the separation region when the sources and channel are independent.

Before proceeding, we will create a notational change that will greatly simplify the arguments and intuition about the proof. Let $S_k = S_{A,k}$, $\mathbf{Y}_k = (Y_k, S_{B,k})$, and $\mathbf{Z}_k =$

$(Z_k, S_{E,k})$. Then we have a memoryless broadcast channel $p_{\mathbf{Y},\mathbf{Z}|X,S}$ with non-causal state information S^n at the encoder Alice. This state sequence S^n is independent and identically distributed with a probability mass function $p_S(s)$. Note that the source observations at Bob and Eve (which have been absorbed into the channel outputs) may depend on the channel.

Let $\mathcal{P}_{\mathsf{joint}}$ be the set of all joint distributions p of random variables $\mathbf{V}, \mathbf{U}, X, S, \mathbf{Y}, \mathbf{Z}$ such that (i) the following Markov chain holds:

$$\mathbf{V} - \mathbf{U} - (X, S) - (\mathbf{Y}, \mathbf{Z}) \ ,$$

(ii) $\mathbf{V}$ is independent of S, and (iii) the joint conditional distribution of $(\mathbf{Y}, \mathbf{Z})$ given (X, S) as well as the marginal distribution of S are consistent with the given source and channel respectively.

For $p \in \mathcal{P}_{\mathsf{joint}}$, let $\mathcal{R}_{\mathsf{joint}}(p)$ be the set of all non-negative pairs $(R_{\mathsf{SK}}, R_{\mathsf{SM}})$ which satisfy the following two inequalities:

$$R_{\mathsf{SM}} \leq I(\mathbf{U}; \mathbf{Y}) - I(\mathbf{U}; S) \tag{4.8}$$

$$R_{\mathsf{SK}} + R_{\mathsf{SM}} \leq I(\mathbf{U}; \mathbf{Y}|\mathbf{V}) - I(\mathbf{U}; \mathbf{Z}|\mathbf{V}) \ . \tag{4.9}$$

Theorem 4.5.

$$\mathcal{R} \supseteq \bigcup_{p \in \mathcal{P}_{\mathsf{joint}}} \mathcal{R}_{\mathsf{joint}}(p) \ . \tag{4.10}$$

We defer the proof of Theorem 4.5 to Appendix D.1. Note that Theorem 4.1 is a special case of Theorem 4.5.

Corollary 4.6.

$$\bigcup_{p \in \mathcal{P}_{\mathsf{joint}}} \mathcal{R}_{\mathsf{joint}}(p) \supseteq \bigcup_{p \in \mathcal{P}} \mathcal{R}(p) \tag{4.11}$$

Proof. For (V_2, V_1, U_1) as defined in Theorem 4.1, set $\mathbf{V} = V_2$ and $\mathbf{U} = (U_1, V_1)$. Then we have the following:

$$\begin{aligned} I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};S) &= I(V_1;Y) + I(U_1;S_B) - I(U_1;S_A) \\ I(\mathbf{U};\mathbf{Y}|\mathbf{V}) - I(\mathbf{U};\mathbf{Z}|\mathbf{V}) &= I(V_1;Y|V_2) - I(V_1;Z|V_2) + I(U_1;S_B) - I(U_1;S_E) \end{aligned}$$

Note that if $I(U_1;S_B) - I(U_1;S_E) \leq 0$, we can increase the achievable region by making U_1 independent of S_A. Likewise, if $I(V_1;Y|V_2) - I(V_1;Z|V_2) < 0$, we can increase the region by making $V_1 = V_2$. Thus, we have established the rate region in Theorem 4.1 as a special case. □

4.6 Discussion

In this chapter, we have considered a setting in which Alice and Bob must agree on both a secret message and a secret key when there are two resources available for generating secrecy: a one-way noisy broadcast channel from Alice to Bob and Eve and dependent sources observed at Alice, Bob, and Eve. In our analysis, we presented an achievable region that trades off between communicating a secret message and generating a secret key, showed this region is optimal when Eve's source and channel are degraded versions of Bob's, as well as for cases in which either Bob's source or channel is by itself useless in generating a secret key. Finally, we evaluated this optimal trade-off in the Gaussian setting and established an achievable rate region in terms of the signal-to-noise ratio of channel at Bob and Eve, as well as the correlation

of the sources.

4.6.1 Extensions

There are several straightforward extensions of the results presented in this chapter that have been established. For instance, it turns out that the result presented in Theorem 4.2 holds more generally than the degradedness conditions outlined. First, the degradedness conditions can be relaxed to stochastically degraded conditions for both the source and channels. This simply involves a slightly more cumbersome argument in our converse proof, but no changes to the achievable strategy are necessary.

Two other extensions of the results in Theorem 4.2 were shown in [68]. In order to state the first, note that given only the sources and a public bit-pipe from Alice to Bob and Eve, the condition under which Alice and Bob cannot generate a positive rate secret-key is in fact weaker than the case where the sources are degraded in favor of Eve[6]. Under this weaker condition, it was shown in [68] that the optimal strategy involves ignoring the sources, and utilizing only the channel. In particular, $\mathcal{R}(p)$ is now the set of all non-negative rate pairs satisfying the condition

$$R_{\mathsf{SK}} + R_{\mathsf{SM}} = [I(V_1; Y) - I(V_1; Z)]_+,$$

where $V_1 - X - (Y, Z)$ is a Markov chain. Thus the optimal strategy in this case reduces to that of Csiszár and Körner [22], and there is essentially no distinction between sending a secret message and generating a secret-key.

To state the second, note that a channel degraded in favor of Eve is a condition

[6]This condition which can be inferred from [3] is that for every $\tilde{U}_1, \tilde{U}_2$ satisfying the Markov chain $\tilde{U}_2 - \tilde{U}_1 - S_A - (S_B, S_E)$,

$$I(\tilde{U}_1; S_B|\tilde{U}_2) \leq I(\tilde{U}_1; S_E|\tilde{U}_2).$$

under which the channel resource by itself cannot provide any secrecy, but the condition under which the channel resource cannot provide any secrecy is looser than this type of degradation. This condition is when the channel to Eve is 'less noisy' than the channel to Bob [22, Corollary 3, pg. 341]. Under this looser condition, but when the source component degraded in favor of Eve is absent, the optimality of turning the channel into a public bit-pipe was shown in [68] for secret-key generation. In the special case where Eve has no source observation, this optimality was shown for secret communication as well.

Note that sending a secret message is equivalent to the case in which Alice must send a discrete uniform source losslessly to Bob that must be kept secret from Eve. A straightforward extension of our result for the secret message case, shown in [36], demonstrates that optimality continues to hold if one is interested in reconstructing any discrete memoryless source, both for the lossless and lossy cases. In this situation, an additional layer of separation between the private bit pipes and the compression of the source can be used to establish the result.

4.6.2 Open problems

The above extensions do not close the door on this problem, and there are several considerations that currently warrant further research. Indeed, the general rate region and structure of optimal strategies are still open problems. One avenue is to consider extensions of the result beyond the degraded case and beyond some of the extensions discussed above.

Another avenue to consider is the setting in which the sources and channel are correlated. Note that in such a setting, there may not be a clean distinction between a source observation and a channel output at either Bob or Eve, which resembles the setup for Theorem 4.5 in Section 4.5. Indeed, the strategy and proof presented for

Theorem 4.5 continues to hold if the sources and channels are correlated.

Such a setting bears a resemblance to a problem studied by Chen and Vinck [13] in which Alice must send Bob a secret message, where Alice has non-causal state information about the channel, and Eve observes degraded versions of Bob's channel outputs. If one considers the non-causal state information to be the source observations at Alice, and the channel outputs at Bob and Eve to be a mixture of their source and channel outputs, then their setting considers sending secret message when sources and channels are correlated and degraded favors Bob, and the rate region and strategy in Theorem 4.5 is equivalent to theirs for this setting.

While Chen and Vinck present an upper bound on the secret message rate, their bound is not tight for their achievable strategy nor Theorem 4.5 in this chapter. The work in [35] provides a marginal improvement to the upper bound presented by Chen and Vinck, but the problems of characterizing the rate region and optimal strategies remain open. Indeed, this region may also be tightened by improving upon the achievable strategy in Theorem 4.5.

Even if one were to characterize the rate region, lingering issues still prevent the private key cryptography promised by these strategies from being adopted. Namely, the channel to the eavesdropper would likely be unknown. The problem is then to find ways to characterize the statistics of an eavesdropper or applications in which such knowledge might be available. Another potential avenue is to consider places in which such a secrecy constraint might be useful. One should note that a variation on the secrecy condition has been used to study anonymous routing protocols in networks [87], such as Chaum mixing [11].

Progress on any of these fronts could lead to new insights on designing strategies to optimally combine source and channel resources for secrecy, as well as understanding the interplay between secret keys and messages.

Chapter 5

Conclusions

Back at the concert, Alice turns to Bob, takes a deep breath, and tells him the Big News: "Bob, I want to take things to the next level, so I'm taking you to meet my family: the March Hare and the Mad Hatter. We'll follow the White Rabbit into Wonderland tomorrow morning. The Cheshire Cat will pick us up on the other side of the looking glass ..."

This thesis has developed new models for communication between Alice and Bob in the presence of various third parties. In Chapter 2, Alice and Bob represented the transmitter and receiver of an adaptable, unlicensed wireless device, and the third party represented the rigid, licensed legacy system with which Alice and Bob must share spectrum. While the legacy system's strategy was fixed, given a target rate for that system, we presented adaptive strategies for Alice and Bob whereby Alice could reliably send a message to Bob such that the legacy system could also meet its target rate. In Chapter 3, we introduced a new model to handle more dynamic third parties. In this model, Alice's inputs to the channel also control a state process, and we considered a problem in which Alice must send a message to Bob and simultaneously control the state process to satisfy a cost constraint. Our analysis

led to an interesting achievable strategy, and under some technical conditions, we showed that this strategy achieves the optimal tradeoff between communication and control. In Chapter 4, we showed how Alice and Bob can keep a secret from Eve by efficiently utilizing two resources: a noisy channel and correlated source observations. Furthermore, we showed a relationship between two different notions of secrecy: in the first, Alice and Bob can agree on any common sequence of bits, and in the second, Alice and Bob must agree on a specific sequence of bits.

While the concert may be over, Alice and Bob have left us with unanswered questions. We now state some of these questions and suggest their relevance to the communication problem:

- Why did Alice choose to tell Bob the Big News at a crowded concert? Could Alice have been wiser about where to tell Bob? In the wireless setting, this corresponds to the case in which the wireless devices can switch among different frequency bands.

- Alice shared the Big News with Bob, but what if she also had Other News for Betty, who was also at the concert? What if Andrew, seated one row behind Alice, had a message for Bob? In the communication context, one may consider how the strategies developed in this thesis can be adapted to scale in multi-terminal settings, in which information can have multiple sources intended for multiple destinations.

- How would Alice have broached the Big News at a restaurant? What about a library? In these environments, there might be different rules of etiquette than at a concert. How do these rules of etiquette affect the strategies used to communicate and the rates they can achieve?

There are likely several other questions one can ask. If there are ways to address any

of them, the answers could provide further insights and enable the development of new communication technology.

Appendices

Appendix A

Technical Preliminaries

This appendix contains technical definitions and preliminary results that will be useful in understanding the technical results of this thesis. We begin by describing our notation. Unless stated otherwise, all random variables are represented by capital letters X, Y, Z, their realizations by lowercase letters x, y, z, and sets by calligraphic letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$. The cardinality of a set $\mathcal{X}$ is denoted $|\mathcal{X}|$. The symbol $\mathbb{P}$ denotes the probability of an event, and $\mathbb{E}$ denotes the expectation of a random variable.

A.1 Entropy, Mutual Information, and Divergence

We start by defining the notion of entropy, which will be useful in the sequel.

Definition A.1. Let X, Y be discrete random variables with joint distribution $p(x, y)$. Then the *entropy* of X $H(X)$ is a real number given by the expression

$$H(X) = -\mathbb{E}[\log p(X)] = -\sum_{x} p(x) \log p(x) \ ,$$

the *joint entropy* of X and Y $H(X,Y)$ is a real number given by the expression

$$H(X,Y) = -\mathbb{E}[\log p(X,Y)] = -\sum_{x,y} p(x,y) \log p(x,y) \ ,$$

and the *conditional entropy* of X given Y $H(X|Y)$ is a real number given by the expression

$$H(X|Y) = -\mathbb{E}[\log p(X|Y)] = H(X,Y) - H(Y) \ .$$

The *binary entropy function* $h_b(p)$ is

$$h_b(p) = -p \log p - (1-p) \log(1-p) \ .$$

With a definition for entropy, we can now define mutual information.

Definition A.2. Let X, Y be discrete random variables with joint distribution $p(x,y)$. Then the mutual information $I(X;Y)$ between X and Y is a real number given by the expression

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \ .$$

Another useful definition is the Kullback-Leibler (KL) divergence.

Definition A.3. Given distributions $p(\cdot)$ and $q(\cdot)$ on the alphabet $\mathcal{X}$, the *Kullback-Leibler (KL) divergence*, denoted $D(p(\cdot)\|q(\cdot))$, is

$$D(p(\cdot)\|q(\cdot)) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \ .$$

The *binary KL divergence* $D(p\|q)$ is

$$D(p\|q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}\ .$$

Similar definitions exist in the continuous setting.

Definition A.4. Let X, Y be continuous random variables with joint density $f(x, y)$. Then we define the *differential entropy*

$$h(X) = -\mathbb{E}[\log f(X)],$$

joint differential entropy

$$h(X, Y) = -\mathbb{E}[\log f(X, Y)],$$

conditional differential entropy

$$h(X|Y) = -\mathbb{E}[\log f(X|Y)],$$

and *mutual information*

$$I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X).$$

Definition A.5. Let X be a continuous random variable with density $f(x, y)$. Then its *entropy power* Q_X is given by

$$Q_X = \frac{1}{2\pi e}\exp\{2h(X)\}\ . \tag{A.1}$$

Theorem A.1 (Entropy Power Inequality [7; 80])**.** *Let X and Y be two independent random variables. Then*

$$Q_{X+Y} \geq Q_X + Q_Y \ .$$

The central concept is that of an *information source*, which is merely a discrete-time random process $\{S_k\}_{k>0}$. For the purposes of this chapter, we assume that all information sources are stationary and ergodic. When the random variables $\{S_k\}_{k>0}$ take values in a *discrete* set, we will refer to the source as a *discrete information source*. For simplicity, we will refer to a source $\{S_k\}_{k>0}$ with the abbreviated notation S.

Definition A.6. The entropy rate of a stationary and ergodic discrete information source S is

$$H_\infty(S) = \lim_{n\to\infty} \frac{1}{n} H(S_1, S_2, \ldots, S_n) = \lim_{k\to\infty} H(S_k|S_{k-1}, S_{k-2}, \ldots, S_1) \ . \tag{A.2}$$

In many of the concrete examples discussed in this chapter, S will be assumed to be a sequence of independent and identically distributed (i.i.d.) random variables. This is usually referred to as a *memoryless* source, and one can show that $H_\infty(S) = H(S_1)$. For notational convenience, we will denote this simply as $H(S)$.

A.2 Typical Sequences

For $x \in \mathcal{X}$ and a random vector X^n, $N(x; X^n)$ refers to the number of times the symbol x appears in the sequence. That is $N(x; X^n) = \sum_i \mathbf{1}_{\{X_i = x\}}$.

Definition A.7. Consider a distribution $p_X(\cdot)$ on the alphabet $\mathcal{X}$. A sequence x^n is *δ-typical* with respect to p_X (denoted $x^n \in \mathcal{T}_\delta^{*(n)}(p_X)$) if for all $a \in \mathcal{X}$,

$$\left|\frac{1}{n}N(a; x^n) - p_X(a)\right| < \frac{\delta}{|\mathcal{X}|} \ . \tag{A.3}$$

This is often called strong typicality in the literature [5].

Definition A.8. Consider a transition probability function $p_{Y|X}$. A sequence y^n is *conditionally δ-typical* on a sequence x^n with respect to $p_{Y|X}$ (denoted $y^n \in \mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))$) if for all $(a, b) \in \mathcal{X} \times \mathcal{Y}$, they satisfy

$$\left|\frac{1}{n}N(a, b; x^n, y^n) - \frac{1}{n}N(a; x^n) \cdot p_{Y|X}(b|a)\right| < \frac{\delta}{|\mathcal{X}| \cdot |\mathcal{Y}|} \ . \tag{A.4}$$

This is often called strong typicality in the literature [5].

Definition A.9. Consider a joint distribution $p_{X,Y}(\cdot, \cdot)$ on the product alphabet $\mathcal{X} \times \mathcal{Y}$. A pair of sequences (x^n, y^n) are *jointly δ-typical* with respect to $p_{X,Y}$ (denoted $(x^n, y^n) \in \mathcal{T}_\delta^{*(n)}(p_{X,Y})$) if for all $(a, b) \in \mathcal{X} \times \mathcal{Y}$, they satisfy

$$\left|\frac{1}{n}N(a, b; x^n, y^n) - p_{X,Y}(a, b)\right| < \frac{\delta}{|\mathcal{X}| \cdot |\mathcal{Y}|} \ . \tag{A.5}$$

Remark: We will suppress the dependence on $p_{X,Y}$ and denote the set of jointly typical sequences as $\mathcal{T}_\delta^{*(n)}$ when it is clear from context.

The following properties about jointly δ-typical sequences, taken from the text of Csiszar and Korner [23, pp. 34– 35], are included here for completeness.

Lemma A.2. *[23, p. 34, Lemma 1.2.10] If $x^n \in \mathcal{T}_{\delta/|\mathcal{Y}|}^{*(n)}(p_X)$, $y^n \in \mathcal{T}_{\delta'}^{*(n)}(p_{Y|X}(\cdot|x^n))$, then $(x^n, y^n) \in \mathcal{T}_{\delta+\delta'}^{*(n)}(p_{X,Y})$, and consequently, $y^n \in \mathcal{T}_{\delta+\delta'}^{*(n)}(p_Y)$.*

Lemma A.3 (Asymptotic Equipartition Property (AEP))**.** *[23, p. 34, Lemma 1.2.12] Let $(X_k, Y_k)_{k\geq 1}$ be a sequence drawn i.i.d. from the joint distribution $p_{X,Y}$. Then, for*

all $\delta, \epsilon > 0$, there exists $n_0(\delta, \epsilon)$ such that for all $n \geq n_0$,

$$\mathbb{P}\left(X^n \in \mathcal{T}_\delta^{*(n)}(p_X)\right) \geq 1 - \epsilon \ , \tag{A.6}$$

$$\mathbb{P}\left(Y^n \in \mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))\middle| X^n = x^n\right) \geq 1 - \epsilon \ . \tag{A.7}$$

The final two results in this section are proved for completeness using techniques that can be found in Berger's lecture notes [5].

Lemma A.4 (Typical Sequences Almost Equiprobable)**.** *Let X, Y be random variables with joint distribution $p_{X,Y}$, which has the property that for all $(a, b) \in \mathcal{X} \times \mathcal{Y}$, $p_{X,Y}(a, b) > 0$. If $x^n \in \mathcal{T}_\delta^{*(n)}(p_X)$ and $y^n \in \mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))$, then for the product distributions $p_{X^n}(x^n) = \prod p_X(x_k)$ and $p_{Y^n|X^n}(y^n|x^n) = \prod p_{Y|X}(y_k|x_k)$,*

$$\begin{array}{rcl} 2^{-n(H(X)+\delta')} \leq & p_{X^n}(x^n) & \leq 2^{-n(H(X)-\delta')} \\ 2^{-n(H(Y|X)+\delta'')} \leq & p_{Y^n|X^n}(y^n|x^n) & \leq 2^{-n(H(Y|X)-\delta'')} \end{array} \ ,$$

where $\delta' = \delta \max_a \log \frac{1}{p_X(a)}$ and $\delta'' = 2\delta \max_{a,b} \log \frac{1}{p_{Y|X}(b|a)}$.

Proof. If $x^n \in \mathcal{T}_\delta^{*(n)}(p_X)$, the definition of a typical sequence implies that

$$\begin{aligned} \prod_{k=1}^n p_X(x_k) &= \prod_{a \in \mathcal{X}} p_X(a)^{N(a;x^n)} \\ &\leq \prod_{a \in \mathcal{X}} p_X(a)^{n(p_X(a) - \delta/|\mathcal{X}|)} \\ &= 2^{n \sum_{a \in \mathcal{X}} (p_X(a) - \delta/|\mathcal{X}|) \log p_X(a)} \\ &= 2^{-nH(X) - n \sum_{a \in \mathcal{X}} \delta/|\mathcal{X}| \log p_X(a)} \\ &\leq 2^{-nH(X) + n\delta \max_a \log \frac{1}{p_X(a)}} \ . \end{aligned}$$

By a symmetric argument,

$$\begin{aligned}
\prod_{k=1}^{n} p_X(x_k) &= \prod_{a\in\mathcal{X}} p_X(a)^{N(a;x^n)} \\
&\geq \prod_{a\in\mathcal{X}} p_X(a)^{n(p_X(a)+\delta\cdot|\mathcal{X}|^{-1})} \\
&= 2^{n\sum_{a\in\mathcal{X}}(p_X(a)+\delta\cdot|\mathcal{X}|^{-1})\log p_X(a)} \\
&= 2^{-nH(X)+n\sum_{a\in\mathcal{X}}\delta/|\mathcal{X}|\log p_X(a)} \\
&\geq 2^{-nH(X)-n\delta\max_a \log\frac{1}{p_X(a)}}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\prod_{k=1}^{n} p_{Y|X}(y_k|x_k) &= \prod_{(a,b)\in\mathcal{X}\times\mathcal{Y}} p_{Y|X}(b|a)^{N(a,b;x^n,y^n)} \\
&\leq \prod_{(a,b)\in\mathcal{X}\times\mathcal{Y}} p_{Y|X}(b|a)^{N(a;x^n)\cdot p_{Y|X}(b|a)-n\delta\cdot|\mathcal{X}|^{-1}\cdot|\mathcal{Y}|^{-1}} \\
&= 2^{n\sum_{(a,b)\in\mathcal{X}\times\mathcal{Y}}(n(p_X(a)-\delta\cdot|\mathcal{X}|^{-1})\cdot p_{Y|X}(b|a)-\delta\cdot|\mathcal{X}|^{-1}\cdot|\mathcal{Y}|^{-1})\log p_{Y|X}(b|a)} \\
&= 2^{-nH(Y|X)-n\sum_{(a,b)\in\mathcal{X}\times\mathcal{Y}}\delta\cdot|\mathcal{X}|^{-1}\cdot(p_{Y|X}(b|a)+|\mathcal{Y}|^{-1})\log p_{Y|X}(b|a)} \\
&\leq 2^{-nH(Y|X)+2n\delta\max_{a,b}\log\frac{1}{p_{Y|X}(b|a)}}
\end{aligned}$$

By a symmetric argument, one can show that

$$\prod_{k=1}^{n} p_{Y|X}(y_k|x_k) \geq 2^{-nH(Y|X)-2n\delta\max_{a,b}\log\frac{1}{p_{Y|X}(b|a)}}$$

□

Lemma A.5 (Number of Typical Sequences)**.** *Suppose there exist two random variables X, Y with joint distribution $p_{Y|X}\cdot p_X$. Then for all $\delta > 0$, there exists $n_0(\delta)$*

such that for all $n \geq n_0$,

$$\left| \frac{1}{n} \log |\mathcal{T}_\delta^{*(n)}(p_X)| - H(X) \right| \leq 2\delta' \tag{A.8}$$

$$\left| \frac{1}{n} \log |\mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))| - H(Y|X) \right| \leq 2\delta'' \ , \tag{A.9}$$

where $\delta' = \delta \max_a \log \frac{1}{p_X(a)}$ *and* $\delta'' = 2\delta \max_{a,b} \log \frac{1}{p_{Y|X}(b|a)}$.

Remark: A variation of this result exists in the text of Csiszar and Korner [23, p. 34, Lemma 1.2.13].

Proof. Since the probability of all typical sequences can be no more 1, we have the following bound:

$$\begin{aligned} 1 &\geq \sum_{\substack{x^n \in \mathcal{T}_\delta^{*(n)}(p_X) \\ y^n \in \mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))}} p_{Y^n|X^n}(y^n|x^n) \\ &\geq |\mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))| 2^{-n(H(Y|X)+\delta'')} \ , \end{aligned}$$

where the last line follows from Lemma A.4. Rearranging terms, we have that

$$|\mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))| \geq 2^{n(H(Y|X)+\delta'')}$$

Similarly, for all $\epsilon > 0$ and n sufficiently large, the AEP (Lemma A.3) implies that

$$\begin{aligned} 1 - \epsilon &\leq \sum_{\substack{x^n \in \mathcal{T}_\delta^{*(n)}(p_X) \\ y^n \in \mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))}} p_{Y^n|X^n}(y^n|x^n) \\ &\leq |\mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))| 2^{-n(H(Y|X)-\delta'')} \ , \end{aligned}$$

where the last line follows from Lemma A.4. Rearranging terms and a judicious choice

of ϵ gives that for sufficiently large n,

$$|\mathcal{T}_\delta^{*(n)}(p_{Y|X}(\cdot|x^n))| \geq 2^{n(H(Y|X)-2\delta'')} \ .$$

A similar argument gives the bounds for $|\mathcal{T}_\delta^{*(n)}(p_X)|$. □

A.3 Codebooks and Hypothesis Testing

Definition A.10 (Codebook)**.** Given an input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, an (n, R) *codebook* consists of an encoding function $f : \mathcal{M} \to \mathcal{X}^n$ and a decoding function $g : \mathcal{Y}^n \to \mathcal{M}$, where the set $\mathcal{M}$ is referred to as the message set and has cardinality $|\mathcal{M}| = 2^{nR}$. For each message $m \in \mathcal{M}$, we say $f(m)$ is the m-th codeword of the codebook, n is called the *blocklength*, and R is called the *rate* of the codebook.

Definition A.11 (Discrete Memoryless Channel)**.** Given an input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, a *discrete memoryless channel (DMC)* $p_{Y|X}$ is a transition probability function from $\mathcal{X}$ to $\mathcal{Y}$ such that

$$\mathbb{P}(Y_k = y_k|Y^{k-1} = y^{k-1}, X^k = x^k) = p_{Y|X}(y_k|x_k), \qquad k = 1, 2, \ldots, n \ .$$

Definition A.12 (Random Codebook with Jointly Typical Decoding)**.** Given an input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and a discrete memoryless channel $p_{Y|X}$, an (n, R, p_X) *random codebook with a jointly δ-typical decoder* is the distribution of (n, R) codebooks for which each codeword is drawn independently at random according to the distribution

$$p_{X^n}(x^n) = \prod_{k=1}^{n} p_X(x_k) \ , \tag{A.10}$$

and the decoder, given Y^n, performs jointly typical decoding for the message, i.e. determines the codeword such that are jointly δ-typical with respect to the joint distribution $p_{Y|X} \cdot p_X$. If no such codeword exists or is not unique, the decoder can select an arbitrary message.

Lemma A.6 (Performance of a Random Codebook with Jointly Typical Decoding [5])**.** *Suppose there exist two random variables X, Y with joint distribution $p_{Y|X} \cdot p_X$. Then given an input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and a discrete memoryless channel $p_{Y|X}$, for all $R < I(X;Y)$, $\delta > 0$, there exists a δ' and n sufficiently large such that an (n, R, p_X) random codebook with a jointly δ'-typical decoder (encoder-decoder pair f, g)such that for all messages $m \in \mathcal{M}$, if $X^n = f(m)$, then the decoding error probability is*

$$\mathbb{P}(g(Y^n) \neq m) \leq \delta \ . \tag{A.11}$$

Remark: Similar results are contained in the textbooks of Csiszar and Korner [23, Corollary 2.1.5, p. 102] as well as Cover and Thomas [21].

Sometimes a tighter bound on the error probability of a codebook may be required. In these instances, we perform maximum likelihood decoding instead and rely on known results bounding the error probability.

Definition A.13 (Random Codebook with Maximum Likelihood Decoding)**.** Given an input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and a discrete memoryless channel $p_{Y|X}$, an (n, R, p_X) *random codebook with a maximum likelihood (ML) decoder* is the distribution of (n, R) codebooks for which each codeword is drawn independently at random according to the distribution

$$p_{X^n}(x^n) = \prod_{k=1}^{n} p_X(x_k) \ , \tag{A.12}$$

and the decoder, given Y^n, performs maximum likelihood decoding for the message, i.e. select $g(Y^n) = \arg\max_m \mathbb{P}(Y^n = y^n | X^n = f(m))$.

Lemma A.7 (Performance of a Random Codebook with ML Decoding)**.** *Suppose there exist two random variables X, Y with joint distribution $p_{Y|X} \cdot p_X$. Then given an input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and a discrete memoryless channel $p_{Y|X}$. Let $C = I(X;Y)$. Then for all $R < I(X;Y)$ and all messages $m \in \mathcal{M}$, for an (n, R, p_X) random codebook with an ML decoder (encoder-decoder pair f, g) with $X^n = f(m)$, the decoding error probability is*

$$\mathbb{P}(g(Y^n) \neq m) \leq \exp\{-n \cdot r \cdot (C - R)^2\} \ , \tag{A.13}$$

where $r > 0$ is some constant that does not depend on n.

Remark: The result is based on an exercise in Gallager's book [38, p. 539, Problem 5.23], which in turn derives from a result in that text [38, p. 138, Theorem 5.6.2]. While there is a small error in the derivation outlined in that exercise, a corrected proof is given in [10].

Lemma A.8 (Stein's Lemma)**.** *Let $p_1(y)$ and $p_2(y)$ be two probability distributions such that $p_1(y) \neq p_2(y)$ for at least one $y \in \mathcal{Y}$. Given ℓ samples of one of these distributions, then there exists a (random) hypothesis test U such that*

$$\mathbb{P}(U(Y^\ell) \neq i | Y^\ell \ \ \textit{selected i.i.d. from } p_i) \leq e^{-\ell \cdot r} \ , \tag{A.14}$$

where $r > 0$ is some constant that does not depend on ℓ.

Remark: Stein's Lemma is a well-known result that can be found in several texts, including Csiszar and Korner [23, Corollary 1.1.2, p. 19], Cover and Thomas [21], and Dembo and Zeitouni [25].

A.4 Dynamic Programming

In this section, we consider the problem maximizing a reward function, which is sometimes referred to as infinite horizon dynamic programming (see e.g. [53, Ch. 8]). We begin by setting up the problem and then stating a well known result associated with it. Before proceeding, we define the following notation: $\mathcal{U} \subseteq \Re^d$ is the control space, where d is a positive integer, and $\mathcal{S}$ the state space.

1. Consider a state process $\{S_k\}_{k\geq 1}$, with state transition probability matrix $[P^u_{i,j}]$, where the entry on row $i \in \mathcal{S}$ and column $j \in \mathcal{S}$ is defined as

$$P^u_{i,j} = \mathbb{P}(S_{k+1} = j | S_k = i, U_k = u) \ , \tag{A.15}$$

   where U_k is the *control*.

2. A *control law* is any sequence $g = \{g_0, g_1, \ldots\}$ with

$$U_k = g_k(S^k) \in \mathcal{U} \ , \tag{A.16}$$

   where $\mathcal{U} \subseteq \Re^d$ for some d.

3. Given a *reward function* $r : \mathcal{S} \times \mathcal{U} \to \Re$, a control sequence g, and an initial state $S_1 = s$, the *average reward* is defined as

$$J_s(g) = \liminf_n \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[r(S_k, U_k)] \ . \tag{A.17}$$

Given items 1), 2), and 3) above, the *maximum average reward* is

$$J^*_s = \max_g J_s(g) \ . \tag{A.18}$$

Theorem A.9 (Bellman's Equation)**.** *[53, Lemma (8.5.8) and Theorem (8.5.21), pp. 158-160] Given items 1), 2), and 3) above, if there exists $|\mathcal{S}|+1$ real numbers*

$$(J^*, \ell(1), \ldots, \ell(|\mathcal{S}|))$$

such that for all $i \in \mathcal{S}$,

$$J^* + \ell(i) = \max_{u \in \mathcal{U}} \left\{ r(i,u) + \sum_{j=1}^{\mathcal{S}} P_{i,j}^u \cdot \ell(j) \right\} , \tag{A.19}$$

then for all $s \in \mathcal{S}$, J^* is the maximum average reward:

$$J^* = \max_g J_s(g) . \tag{A.20}$$

Furthermore, if for all $u \in \mathcal{U}$, a Markov chain with transition probability matrix $[P_{i,j}^u]$ over the state space $\mathcal{S}$ is irreducible, then a solution to (A.19) *exists, so J^* is the maximum average reward.*

A.5 Miscellaneous Results

A.5.1 A Markov Property

Lemma A.10 (Stopping Times and the Strong Markov Property)**.** *Let A_j be i.i.d. Bernoulli-p_j random variables such that $p \in (0,1)$. Define stopping times*

$$\tilde{N}_{2k-1} = r \cdot \inf\{ i > r^{-1} \tilde{N}_{2k-2} : \sum_{j=1}^{i} A_j - i \cdot q \geq \tau \} , \tag{A.21}$$

$$\tilde{N}_{2k} = r \cdot \inf\{ i > r^{-1} \tilde{N}_{2k-1} : \sum_{j=1}^{i} A_j - i \cdot q < \tau \} , \tag{A.22}$$

where $\tilde{N}_0 = 0$. *Then for all real* τ, q *and integers* r, *and on the event* $\{\tilde{N}_{2k} < \infty\}$,

$$\begin{aligned}&\mathbb{P}_{(p_j)_{j\geq 1}}(\tilde{N}_{2k+1} - \tilde{N}_{2k} > \ell \cdot r | \tilde{N}_{2k} < \infty) \\ &\qquad = \mathbb{P}_{(p_{j+\tilde{N}_{2k}})_{j\geq 1}}(\tilde{N}_1 \geq \ell \cdot r | \tau - (q+1)\cdot r \leq \tilde{S}_0 < \tau)\ . \end{aligned} \tag{A.23}$$

Proof. Define $\tilde{S}_i = \sum_{j=1}^{i} A_i - i \cdot q$ and note that it is Markov.

$$\begin{aligned}&\mathbb{P}_{(p_j)_{j\geq 1}}(\tilde{N}_{2k+1} - \tilde{N}_{2k} > \ell \cdot r | \tilde{N}_{2k} < \infty) \\ &\qquad = \mathbb{P}_{(p_j)_{j\geq 1}}(\tilde{S}_{\tilde{N}_{2k}+1} \geq \tau | \tau - (q+1)\cdot r \leq \tilde{S}_{\tilde{N}_{2k}} < \tau, \tilde{N}_{2k} < \infty) \\ &\qquad\quad \cdot \prod_{m=1}^{\ell\cdot r - 1} \mathbb{P}_{(p_j)_{j\geq 1}}(\tilde{S}_{\tilde{N}_{2k}+m+1} < \tau | \tilde{S}_{\tilde{N}_{2k}+m} < \tau, \tilde{N}_{2k} < \infty) \qquad \text{(A.24)} \\ &\qquad = \mathbb{P}_{(p_{j+\tilde{N}_{2k}})_{j\geq 1}}(\tilde{S}_1 \geq \tau | \tau - (q+1)\cdot r \leq \tilde{S}_0 < \tau, \tilde{N}_{2k} < \infty) \\ &\qquad\quad \cdot \prod_{m=1}^{\ell\cdot r - 1} \mathbb{P}_{(p_{j-\tilde{N}_{2k}})_{j\geq 1}}(\tilde{S}_{m+1} < \tau | \tilde{S}_m < \tau) \qquad \text{(A.25)} \\ &\qquad = \mathbb{P}_{(p_{j+\tilde{N}_{2k}})_{j\geq 1}}(\tilde{N}_1 \geq \ell \cdot r | \tau - (q+1)\cdot r \leq \tilde{S}_0 < \tau) \qquad \text{(A.26)}\end{aligned}$$

where (A.24) follows from the definition of the stopping times and the fact that $\tilde{S}_i$ is Markov, (A.25) from the strong Markov property [29, p. 285, Theorem 5.2.4], and (A.26) by the definition of the stopping time. □

Lemma A.11. *Suppose* $\{S_k\}_{k>0}$ *is an irreducible, positive recurrent Markov chain with stationary distribution* $\pi(s)$. *Then for all* $\delta > 0$, $s, s' \in \mathcal{S}$ *such that* $S_1 = s'$,

$$\mathbb{P}\left(\left|\frac{N_s^n}{n} - \pi(s)\right| > \delta\right) \to 0, \tag{A.27}$$

where $N_s^n = \sum_{k=1}^{n} \mathbf{1}_{\{i:S_i = s\}}$.

Proof. This result is simply a combination of [29, p. 303, Theorem 5.4.6] and [29, p. 308, Theorem 5.5.1]. Note that said results prove the stronger result of almost sure

convergence, whereas we only require convergence in probability. □

A.5.2 Monotonicity, Concavity, and Continuity of a Cost-Constrained Capacity

Lemma A.12. *Let $0 \leq \epsilon_x \leq 1$ for all x and define $\epsilon_0 = \min_x \epsilon_x$, which is achieved uniquely by some x. Then*

$$C(\vec{\epsilon}, \lambda) = \max_{\substack{p(x): \\ \sum_x \epsilon_x p(x) \leq \lambda}} I(X;Y) \ , \tag{A.28}$$

is nondecreasing concave in λ on the interval $[\epsilon_0, 1]$.

Proof. Since increasing λ increases the set of channel input distributions over which to maximize, it is clear that $C(\vec{\epsilon}, \lambda)$ is nondecreasing. As a convenient shorthand, let $I_p = I(X;Y)$ denote the mutual information with the input distribution p. Let p_1 be the maximizing input distribution for $C(\vec{\epsilon}, \lambda_1)$ and p_2 the maximizing input distribution for $C(\vec{\epsilon}, \lambda_2)$, both of which are guaranteed to exist for $\lambda_i \in [\epsilon_0, 1]$, $i \in \{1, 2\}$. Then

$$(1-\rho)C(\vec{\epsilon}, \lambda_1) + \rho C(\vec{\epsilon}, \lambda_2) = (1-\rho)I_{p_1} + \rho I_{p_2} \tag{A.29}$$

$$\leq I_{(1-\rho)p_1 + \rho p_2} \ , \tag{A.30}$$

$$\leq C(\vec{\epsilon}, (1-\rho)\lambda_1 - \rho\lambda_2) \ , \tag{A.31}$$

where (A.30) follows from the concavity of mutual information with respect to its input distribution and (A.31) by definition. Thus, the function is concave. □

Lemma A.13. *Let $0 \leq \epsilon_x \leq 1$ for all x and define $\epsilon_0 = \min_x \epsilon_x$, which is achieved*

uniquely by some x. Consider

$$C(\vec{\epsilon}, \lambda) = \max_{\substack{p(x): \\ \sum_x \epsilon_x p(x) \leq \lambda}} I(X;Y) \ , \tag{A.32}$$

where $\lambda \in [\epsilon_0, 1]$ and $C(\vec{\epsilon}, \lambda) = 0$ for $\lambda < \epsilon_0$. Then for $0 < \Delta \leq \frac{1}{4}$,

$$0 \leq C(\vec{\epsilon}, \lambda + \Delta) - C(\vec{\epsilon}, \lambda) \leq -6\Delta \log \frac{2\Delta}{|\mathcal{X}| \cdot |\mathcal{Y}|} \ . \tag{A.33}$$

Proof. The lower bound follows immediately from Lemma A.12. For the upper bound, note that $C(\vec{\epsilon}, \epsilon_0) = 0$ since the constraint can only be met by applying all the probability to a single choice of x. Let $\lambda \geq \epsilon_0$. Let $p_1(x)$ be the maximizing input distribution for $C(\vec{\epsilon}, \lambda + \Delta)$. If $p_1(x)$ is found in the set of valid input distributions for $C(\vec{\epsilon}, \lambda)$, then $C(\vec{\epsilon}, \lambda + \Delta) = C(\vec{\epsilon}, \lambda)$. Otherwise, we can define $p_2(x) = \sum_x \epsilon_x p_2(x) = \lambda$ and observe that

$$\lambda < \sum_x \epsilon_x p_1(x) \leq \lambda + \Delta \ , \tag{A.34}$$

which implies that

$$0 < \sum_x \epsilon_x (p_1(x) - p_2(x)) \leq \Delta \tag{A.35}$$

$$0 < \sum_x (1 - \epsilon_x)(p_2(x) - p_1(x)) \leq \Delta \ . \tag{A.36}$$

Thus,

$$\sum_x |p_2(x) - p_1(x)| = \sum_x \epsilon_x |p_2(x) - p_1(x)| + \sum_x (1 - \epsilon_x)|p_2(x) - p_1(x)| \tag{A.37}$$

$$\leq 2\Delta \tag{A.38}$$

Let $I_p = I(X;Y)$ when the input distribution for X is p. Then by the continuity of entropy [23, Lemma 1.2.7, p. 33],

$$C(\vec{\epsilon}, \lambda + \Delta) = I_{p_1} \tag{A.39}$$

$$= I_{p_2} + (I_{p_1} - I_{p_2}) \tag{A.40}$$

$$\leq C(\vec{\epsilon}, \lambda) - 3 \cdot 2\Delta \log \frac{2\Delta}{|\mathcal{X}| \cdot |\mathcal{Y}|} \tag{A.41}$$

This is still valid as an upper bound for $\lambda < \epsilon_0$ because of the monotonicity of $C(\vec{\epsilon}, \lambda)$ from Lemma A.12. □

Lemma A.14. *Let $0 \leq \epsilon_x \leq 1$ for all x and define $\epsilon_0 = \min_x \epsilon_x$, which is achieved uniquely by some x, so $\epsilon_0 < \epsilon_1 = \max_x \epsilon_x$. Consider*

$$C(\vec{\epsilon}, \lambda) = \max_{\substack{p(x): \\ \sum_x \epsilon_x p(x) \leq \lambda}} I(X;Y) \ , \tag{A.42}$$

where $\lambda \in [\epsilon_0, 1]$ and $C(\vec{\epsilon}, \lambda) = 0$ for $\lambda < \epsilon_0$. Furthermore,

$$\sum_x |\epsilon_x - \tilde{\epsilon}_x| \leq \Delta \leq \frac{1}{4} \ . \tag{A.43}$$

Then

$$|C(\vec{\epsilon}, \lambda) - C(\vec{\tilde{\epsilon}}, \lambda)| \leq -6\Delta \log \frac{2\Delta}{|\mathcal{X}| \cdot |\mathcal{Y}|} \ . \tag{A.44}$$

Proof. The constraint on $p(x)$ in $C(\vec{\tilde{\epsilon}}, \lambda)$ can be rewritten as

$$\sum_x \epsilon_x p(x) \leq \lambda - \sum_x (\tilde{\epsilon}_x - \epsilon_x) p(x) \ . \tag{A.45}$$

Since $\sum_x (\tilde{\epsilon}_x - \epsilon_x) p(x) \leq \sum_x |\tilde{\epsilon}_x - \epsilon_x|$, a tighter constraint on $p(x)$ is $\sum_x \epsilon_x p(x) \leq$

$\lambda - \Delta$, and since $\sum_x (\tilde{\epsilon}_x - \epsilon_x) p(x) \geq -\sum_x |\tilde{\epsilon}_x - \epsilon_x|$, a looser constraint on $p(x)$ is $\sum_x \epsilon_x p(x) \leq \lambda + \Delta$. Thus

$$C(\vec{\epsilon}, \lambda - \Delta) \leq C(\vec{\tilde{\epsilon}}, \lambda) \leq C(\vec{\epsilon}, \lambda + \Delta) \ . \tag{A.46}$$

Then one can write

$$\begin{aligned} & C(\vec{\epsilon}, \lambda) - C(\vec{\tilde{\epsilon}}, \lambda) \\ & \qquad = C(\vec{\epsilon}, \lambda) - C(\vec{\epsilon}, \lambda - \Delta) + C(\vec{\epsilon}, \lambda - \Delta) - C(\vec{\tilde{\epsilon}}, \lambda) && \text{(A.47)} \\ & \qquad \leq C(\vec{\epsilon}, \lambda) - C(\vec{\epsilon}, \lambda - \Delta) \ , && \text{(A.48)} \end{aligned}$$

and similarly,

$$\begin{aligned} & C(\vec{\epsilon}, \lambda) - C(\vec{\tilde{\epsilon}}, \lambda) \\ & \qquad = C(\vec{\epsilon}, \lambda) - C(\vec{\epsilon}, \lambda + \Delta) + C(\vec{\epsilon}, \lambda + \Delta) - C(\vec{\tilde{\epsilon}}, \lambda) && \text{(A.49)} \\ & \qquad \geq C(\vec{\epsilon}, \lambda) - C(\vec{\epsilon}, \lambda + \Delta) \ , && \text{(A.50)} \end{aligned}$$

Lemma A.13 then implies that

$$|C(\vec{\epsilon}, \lambda) - C(\vec{\tilde{\epsilon}}, \lambda)| \leq -6\Delta \log \frac{2\Delta}{|\mathcal{X}| \cdot |\mathcal{Y}|} \ . \tag{A.51}$$

□

Appendix B

Proofs for Chapter 2

B.1 Proof of Theorem 2.3

B.1.1 Properties of the Thresholding Strategy

In this section, we consider a couple simple properties about the thresholding strategy that will be useful in showing the strategy is valid (i.e. the primary meets its rate target) and also analyzing the rate of the cognitive radio. The first lemma enables the former.

Lemma B.1. *Let S_k be defined as in* (2.18)*, S_0 as in* (2.19)*, and channel inputs satisfy the condition* (2.20)*. Let q and r be positive integers such that $q \cdot r = K_n$, and define the stopping time*

$$N = \inf\{m > 0 : \exists\ i \in \mathbb{Z}^+ \ s.t.\ m = i \cdot r,\ S_m - m \cdot \gamma - r \geq 0\}\ . \tag{B.1}$$

If $R_p + \gamma < 1 - \epsilon_0^+$*, where* $\epsilon_0^+ = \sup_i \epsilon_{0,i}$*, then*

$$\mathbb{P}(N \geq t | S_0 = s) \leq \left(\frac{(1 - R_p - \gamma)(1 - \epsilon_0^+)}{(R_p + \gamma)\epsilon_0^+} \right)^{2r-s} e^{-t \cdot D(1 - R_p - \gamma \| \epsilon_0^+)} \ . \tag{B.2}$$

Proof. The binary KL divergence $D(\cdot\|\cdot)$ (see Definition A.3) is nonnegative, so we can apply a Chernoff bound to get that

$$\mathbb{P}(N \geq t | S_0 = s) \leq \mathbb{E}[e^{N \cdot D(1 - R_\mathrm{p} - \gamma \| \epsilon_0^+)} | S_0 = s] \cdot e^{-t \cdot D(1 - R_\mathrm{p} - \gamma \| \epsilon_0^+)} \ . \tag{B.3}$$

Thus, it suffices to show that

$$\mathbb{E}[e^{N \cdot D(1 - R_\mathrm{p} - \gamma \| \epsilon_0^+)} | S_0 = s] \leq \left(\frac{(1 - R_\mathrm{p} - \gamma)(1 - \epsilon_0^+)}{(R_\mathrm{p} + \gamma)\epsilon_0^+} \right)^{2r-s} \ .$$

From the definition of S_k in (2.18), it is clear that $S_k - S_{k-1} \leq 1$, and since the stopping time N is a multiple of r, we have the following inequality:

$$S_{N \wedge k} \leq (N \wedge k) \cdot \gamma + 2r \ , \tag{B.4}$$

where we use the notation $N \wedge k$ to denotes the minimum of N and k. In the remainder of the proof, we will use (B.4) and a martingale argument to bound the expectation in (B.3), thereby concluding the result.

Define the sequence of random variables

$$M_k = e^{\lambda S_k - \sum_{i=1}^k (f_\lambda(\epsilon_{0,i}) - \lambda R_\mathrm{p})} \tag{B.5}$$

$$= e^{\lambda s + \sum_{i=1}^k (\lambda A_i - f_\lambda(\epsilon_{0,i}))} \ , \tag{B.6}$$

where $f_\lambda(\epsilon) = \log((1 - \epsilon)e^\lambda + \epsilon)$, and (B.6) follows from (2.18) and (2.19). First

observe that we can express M_k in terms of the recurrence equation

$$M_k = M_{k-1} \cdot e^{\lambda A_k - f_\lambda(\epsilon_{0,k})} \ , \ k \geq 1 \qquad\qquad M_0 = e^{\lambda s} \ .$$

Then conditioned on the event $\{N > k\}$, M_k is a martingale:

$$\begin{aligned}
\mathbb{E}[M_k | M_0, \ldots, M_{k-1}, N > k] &= \mathbb{E}[M_k | S_0, \ldots, S_{k-1}, N > k] \\
&= M_{k-1} e^{-f_\lambda(\epsilon_{0,k})} \mathbb{E}[e^{\lambda A_k} | S_0, \ldots, S_{k-1}, N > k] \\
&= M_{k-1} \ .
\end{aligned}$$

Thus, the optional stopping theorem [29, Thm. 4.7.4, p. 270] implies

$$\begin{aligned}
e^{\lambda s} &= \mathbb{E}[M_{N \wedge m} | S_0 = s] \\
&= \mathbb{E}[e^{\lambda S_{N\wedge m} - \sum_{i=1}^{N\wedge m} (f_\lambda(\epsilon_{0,i}) - \lambda R_\mathrm{p})} | S_0 = s] \ .
\end{aligned}$$

If we select $\lambda < 0$, then $f_\lambda(\epsilon)$ is monotonically increasing in ϵ, so for all i, $f_\lambda(\epsilon_{0,i}) \leq f_\lambda(\epsilon_0^+)$. Thus, for $\lambda < 0$,

$$e^{\lambda s} \geq \mathbb{E}[e^{\lambda S_{N\wedge m} - (N \wedge m) \cdot (f_\lambda(\epsilon_0^+) - \lambda R_\mathrm{p})} | S_0 = s] \ . \tag{B.7}$$

Let $\lambda = \log \frac{(R_\mathrm{p}+\gamma)\epsilon_0^+}{(1-R_\mathrm{p}-\gamma)(1-\epsilon_0^+)}$, where log is the natural logarithm. This λ is negative, which can be seen by applying the inequality $\log x \leq x - 1$ and our assumption

$R_{\mathrm{p}} + \gamma < 1 - \epsilon_0^+$:

$$\begin{aligned}
\log \frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)} &\leq \frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)} - 1 \\
&= \frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+ - (1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)} \\
&= \frac{(R_{\mathrm{p}}+\gamma) - (1-\epsilon_0^+)}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)} \\
&< 0 \ .
\end{aligned}$$

We can substitute this $\lambda < 0$ into equation (B.7) to get that

$$\begin{aligned}
& e^{s \cdot \log \frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)}} \\
&\quad \geq \mathbb{E}\Big[e^{((N\wedge m)\cdot\gamma + 2r)\cdot \log \frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)} - (N\wedge m)\cdot(\log \frac{\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)} - R_{\mathrm{p}} \cdot \log \frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)})} \Big| S_0 = s \Big] \qquad \text{(B.8)} \\
&\quad = e^{2r \cdot \log \frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)}} \cdot \mathbb{E}[e^{(N\wedge m) D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+)} | S_0 = s] \ , \qquad \text{(B.9)} \\
&\quad = \left(\frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)} \right)^{2r} \mathbb{E}\big[e^{(N\wedge m) D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+)} \big| S_0 = s \big] \ , \qquad \text{(B.10)}
\end{aligned}$$

where (B.8) follows from (B.4), and (B.9) follows from the following:

$$\begin{aligned}
& \gamma \cdot \log \frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)} - \log \frac{\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)} + R_{\mathrm{p}} \cdot \log \frac{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)} \\
&\quad = (R_{\mathrm{p}}+\gamma) \cdot \log \frac{R_{\mathrm{p}}+\gamma}{1-\epsilon_0^+} + (1-R_{\mathrm{p}}-\gamma) \cdot \log \frac{(1-R_{\mathrm{p}}-\gamma)}{\epsilon_0^+} \\
&\quad = D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+) \ .
\end{aligned}$$

By the monotone convergence [29, p. 15, Theorem 1.3.6], letting $m \to \infty$ in (B.10)

and rearranging terms gives

$$\mathbb{E}[e^{N\cdot D(1-R_{\mathrm{p}}-\gamma\|\epsilon_0^+)}|S_0=s]\leq\left(\frac{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)}{(R_{\mathrm{p}}+\gamma)\epsilon_0^+}\right)^{2r-s}\ . \tag{B.11}$$

By applying this to (B.3), we have proved the result. □

Our next lemma, which holds for a special class of $(\epsilon x,i)_{i=1}^{\infty}$, will help us provide a guarantee about the cognitive radio's rate over this class.

Lemma B.2. *Let S_k be defined as in* (2.18)*, and channel inputs satisfy the conditions* (2.20)*,* (2.21)*, and* (2.24)*. Let $p^*=\arg\max_{p(x)}I(X;Y)$. If $\max\{1-\epsilon_{x_{rep},i},1-\sum_x p^*(x)\epsilon_{x,i}\}<R_p$ for all i, then*

$$\mathbb{P}(S_m\geq t+m\cdot\gamma+K_n)\leq\left(\frac{(R_p+\gamma)(1-R_p)}{(1-R_p-\gamma)R_p}\right)^{2K_n}\cdot e^{-t\cdot D(R_p+\gamma\|R_p)}\ . \tag{B.12}$$

Proof. Define stopping times for $k\geq 1$,

$$N_{2k-1}=\inf\{m>N_{2k-2}:\exists\ i\in\mathbb{Z}^+\text{ s.t. }m=i\cdot K_n,\ S_m-m\cdot\gamma\geq K_n\}\ , \tag{B.13}$$

$$N_{2k}=\inf\{m>N_{2k-1}:\exists\ i\in\mathbb{Z}^+\text{ s.t. }m=i\cdot K_n,\ S_m-m\cdot\gamma< K_n\}\ , \tag{B.14}$$

where $N_0=0$. We seek an upper bound on S_m, defined in (2.18). Since $|S_m-S_{m-1}|\leq 1$, the following observation will enable a simple upper bound on S_m is the following:

$$\begin{aligned} &S_m<m\cdot\gamma+K_n, && N_{2k}<m\leq N_{2k+1}\\ &S_m<m\cdot\gamma+K_n+N_{2k}-N_{2k-1}, && N_{2k-1}<m\leq N_{2k}\end{aligned}$$

Thus, a simple upper bound to for $t > 0$ is

$$\begin{aligned}
&\mathbb{P}(S_m \geq t + m \cdot \gamma + K_n) \\
&\quad = \sum_{k=1}^{\infty} \mathbb{P}(S_m \geq t + m \cdot \gamma + K_n | N_{2k-1} < m \leq N_{2k}) \cdot \mathbb{P}(N_{2k-1} < m \leq N_{2k}) \\
&\quad \leq \max_k \mathbb{P}_{(\epsilon_{x,i})_{i=1}^{\infty}}(N_{2k} - N_{2k-1} \geq t | N_{2k-1} < \infty) \\
&\quad \leq \max_k \mathbb{P}_{(\epsilon_{x,i+N_{2k-1}})_{i=1}^{\infty}}(N_2 - N_1 \geq t | N_1 < \infty) \ ,
\end{aligned}$$

where the last line follows from the strong Markov property (Lemma A.10). We will provide a bound that is unaffected by the shift $(\epsilon_{x,N_{2k-1}+1})_{i=1}^{\infty}$, so we suppress the dependence in the sequel. The binary KL divergence $D(\cdot\|\cdot)$ (see Definition A.3) is nonnegative, so we can apply a Chernoff bound to get that

$$\begin{aligned}
&\mathbb{P}(N_2 - N_1 \geq t | S_0 = s, N_1 < \infty) \\
&\qquad \leq \mathbb{E}[e^{(N_2 - N_1) \cdot D(R_\mathrm{p} + \gamma \| R_\mathrm{p})} | S_0 = s, N_1 < \infty] \cdot e^{-t \cdot D(R_\mathrm{p} + \gamma \| R_\mathrm{p})} \ . \qquad \text{(B.15)}
\end{aligned}$$

Thus, it suffices to show that

$$\mathbb{E}[e^{(N_2 - N_1) \cdot D(R_\mathrm{p} + \gamma \| R_\mathrm{p})} | S_0 = s, N_1 < \infty] \leq \left(\frac{(R_\mathrm{p} + \gamma)(1 - R_\mathrm{p})}{(1 - R_\mathrm{p} - \gamma) R_\mathrm{p}} \right)^{2K_n} \ .$$

From the definition of S_k in (2.18), it is clear that $|S_k - S_{k-1}| \leq 1$, and since the stopping times N_1, N_2 are multiples of K_n, we have the following inequalities:

$$S_{N_1 \wedge k} \leq (N_1 \wedge k) \cdot \gamma + 2K_n \ , \qquad \text{(B.16)}$$

$$S_{N_2 \wedge k} \geq (N_2 \wedge k) \cdot \gamma \ , \qquad \text{(B.17)}$$

where we use the notation $N_2 \wedge k$ to denotes the minimum of N and k. In the

remainder of the proof, we will use (B.16), (B.17) and a martingale argument to bound the expectation in (B.15), thereby concluding the result.

Define $\bar{\epsilon}_j = 1 - \mathbb{E}[A_j | S_0, \ldots, S_{j-1}]$ and the sequence of random variables

$$M_k = e^{\lambda S_k - \sum_{i=1}^{k} (f_\lambda(\bar{\epsilon}_i) - \lambda R_{\mathrm{p}})} \tag{B.18}$$

$$= e^{\lambda s + \sum_{i=1}^{k} (\lambda A_i - f_\lambda(\bar{\epsilon}_i))} \ , \tag{B.19}$$

where $f_\lambda(\epsilon) = \log((1-\epsilon)e^\lambda + \epsilon)$, and (B.19) follows from (2.18) and (2.19). First observe that we can express M_k in terms of the recurrence equation

$$M_k = M_{k-1} \cdot e^{\lambda A_k - f_\lambda(\bar{\epsilon}_k)} \ , \ k \geq 1 \qquad\qquad M_0 = e^{\lambda s} \ .$$

Then $\tilde{M}_k = M_k \cdot (M_{N_1 \wedge k})^{-1}$ is a martingale:

$$\begin{aligned}
\mathbb{E}[\tilde{M}_k | \tilde{M}_0, \ldots, \tilde{M}_{k-1}] &= \mathbb{E}[M_k \cdot (M_{N_1 \wedge k})^{-1} | S_0, \ldots, S_{k-1}] \\
&= \tilde{M}_{k-1} \cdot \mathbb{E}[e^{\lambda A_k - \lambda A_{N_1 \wedge k}} \cdot e^{f_\lambda(\bar{\epsilon}_{N_1 \wedge k}) - f_\lambda(\bar{\epsilon}_k)} | S_0, \ldots, S_{k-1}] \\
&= \tilde{M}_{k-1} \ .
\end{aligned}$$

Thus, the optional stopping theorem [29, Thm. 4.7.4, p. 270] implies

$$\begin{aligned}
1 &= \mathbb{E}[\tilde{M}_{N_2 \wedge m} | S_0 = s] \\
&= \mathbb{E}\left[\exp\left\{ \lambda S_{N_2 \wedge m} - \lambda S_{N_1 \wedge m} - \sum_{i=N_1 \wedge m + 1}^{N_2 \wedge m} (f_\lambda(\bar{\epsilon}_i) - \lambda R_{\mathrm{p}}) \right\} \middle| S_0 = s \right] \ .
\end{aligned}$$

If we select $\lambda > 0$, then $f_\lambda(\epsilon)$ is monotonically decreasing in ϵ, so for all $N_1 \wedge m + 1 \leq i \leq N_2 \wedge m$, our assumption $\max\{1 - \epsilon_{x_{\mathrm{rep}},i}, 1 - \sum_x p^*(x)\epsilon_{x,i}\} < R_{\mathrm{p}}$ implies that

$f_\lambda(\bar{\epsilon}_i) \le f_\lambda(1-R_\mathrm{p})$. Thus, for $\lambda > 0$,

$$\begin{aligned} 1 &\ge \mathbb{E}[e^{\lambda(S_{N_2\wedge m}-S_{N_1\wedge m})-(N_2\wedge m-N_1\wedge m)\cdot(f_\lambda(1-R_\mathrm{p})-\lambda R_\mathrm{p})}|S_0=s] \\ &\ge \mathbb{E}[e^{-2K_n\cdot\lambda-(N_2\wedge m-N_1\wedge m)\cdot(f_\lambda(1-R_\mathrm{p})-\lambda(R_\mathrm{p}+\gamma))}|S_0=s]\ , \end{aligned} \tag{B.20}$$

where (B.20) follows from the bounds in (B.16) and (B.17). Let $\lambda = \log\frac{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})}{(1-R_\mathrm{p}-\gamma)R_\mathrm{p}}$, where log is the natural logarithm. This λ is positive, which can be seen by applying the inequality $\log x \le x-1$ and our assumption $\gamma > 0$:

$$\begin{aligned} -\log\frac{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})}{(1-R_\mathrm{p}-\gamma)R_\mathrm{p}} &\le \frac{(1-R_\mathrm{p}-\gamma)R_\mathrm{p}}{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})} - 1 \\ &= \frac{(1-R_\mathrm{p}-\gamma)R_\mathrm{p}-(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})}{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})} \\ &= \frac{R_\mathrm{p}-(R_\mathrm{p}+\gamma)}{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})} \\ &< 0\ . \end{aligned}$$

We can substitute this $\lambda > 0$ into equation (B.20) to get that

$$\begin{aligned} 1 &\ge \mathbb{E}\Big[e^{-2K_n\cdot\log\frac{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})}{(1-R_\mathrm{p}-\gamma)R_\mathrm{p}}-(N_2\wedge m-N_1\wedge m)\cdot(\log\frac{1-R_\mathrm{p}}{(1-R_\mathrm{p}-\gamma)}-(R_\mathrm{p}+\gamma)\cdot\log\frac{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})}{(1-R_\mathrm{p}-\gamma)R_\mathrm{p}})}\Big|S_0=s\Big] \\ &= e^{-2K_n\cdot\log\frac{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})}{(1-R_\mathrm{p}-\gamma)R_\mathrm{p}}}\cdot\mathbb{E}[e^{(N_2\wedge m-N_1\wedge m)\cdot D(R_\mathrm{p}+\gamma\|R_\mathrm{p})}|S_0=s]\ , \end{aligned} \tag{B.21}$$

$$= \left(\frac{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})}{(1-R_\mathrm{p}-\gamma)R_\mathrm{p}}\right)^{-2K_n}\mathbb{E}[e^{(N_2\wedge m-N_1\wedge m)\cdot D(R_\mathrm{p}+\gamma\|R_\mathrm{p})}|S_0=s]\ , \tag{B.22}$$

where (B.21) follows from the following:

$$\begin{aligned}
&-\log\frac{1-R_{\mathrm{p}}}{(1-R_{\mathrm{p}}-\gamma)}+(R_{\mathrm{p}}+\gamma)\cdot\log\frac{(R_{\mathrm{p}}+\gamma)(1-R_{\mathrm{p}})}{(1-R_{\mathrm{p}}-\gamma)R_{\mathrm{p}}}\\
&\quad=(R_{\mathrm{p}}+\gamma)\cdot\log\frac{R_{\mathrm{p}}+\gamma}{R_{\mathrm{p}}}+(1-R_{\mathrm{p}}-\gamma)\cdot\log\frac{(1-R_{\mathrm{p}}-\gamma)}{1-R_{\mathrm{p}}}\\
&\quad=D(R_{\mathrm{p}}+\gamma\|R_{\mathrm{p}})\ .
\end{aligned}$$

By the monotone convergence [29, p. 15, Theorem 1.3.6], letting $m\to\infty$ in (B.22) and rearranging terms gives

$$\mathbb{E}[e^{(N_2-N_1)\cdot D(R_{\mathrm{p}}+\gamma\|R_{\mathrm{p}})}|S_0=s,N_1<\infty]\le\left(\frac{(R_{\mathrm{p}}+\gamma)(1-R_{\mathrm{p}})}{(1-R_{\mathrm{p}}-\gamma)R_{\mathrm{p}}}\right)^{2K_n}\ . \tag{B.23}$$

By applying this to (B.15), we have proved the result. □

B.1.2 Primary Meets Rate Target

Lemma B.3. *Given a rate target R_p and $\nu>0$, if for every code of blocklength n satisfying* (2.20) *on page 27, $\gamma>0$ is chosen so that $\gamma<\frac{\nu}{2}$, a strategy composed of these codes is valid. That is, such a strategy satisfies, for all $\{\epsilon_{x,i}\}_{i=1}^{\infty}$ under Assumption 2.1,*

$$\mathbb{P}_{\{\epsilon_{x,i}\}_{i=1}^{\infty}}\left(\ell^{-1}\sum_{k=1}^{\ell}A_k<R_p\right)\le K_{1,R_p,\nu,\ell}\cdot e^{-\ell K_{2,R_p,\nu}}\ , \tag{B.24}$$

where $0<K_{2,R_p,\nu}<\infty$, $K_{1,R_p,\nu,\ell}<\infty$, and for all $r>0$, $K_{1,R_p,\nu,\ell}\cdot e^{-\ell\cdot r}\to 0$ as $\ell\to\infty$.

Proof. When $S_\ell>0$ (see (2.18) for the definition), the primary is meeting its rate target. Furthermore, if $S_{iK_n}-K_n-iK_n\gamma>0$, then the primary will be guaranteed to meet its rate target over the next primary frame. Thus, it suffices to consider frames

when $S_{iK_n} - K_n - iK_n\gamma \leq 0$, which correspond directly to silent frames. To consider what happens in these settings, we define stopping times to threshold $S_\ell - \ell\gamma$.

$$N_{2k-1} = \inf\{\ell > N_{2k-2} : S_\ell - \ell\gamma \geq 1\}, \tag{B.25}$$

$$N_{2k} = \inf\{\ell > N_{2k-1} : S_\ell - \ell\gamma < 1\}. \tag{B.26}$$

Negative deviations occur only when $N_{2k} \leq \ell < N_{2k+1}$, so

$$\begin{aligned}
&\mathbb{P}(S_\ell \leq 0) \\
&\quad \leq \sum_k \mathbb{P}(-S_\ell + \ell\gamma \geq \ell\gamma | N_{2k} \leq \ell < N_{2k+1}) \mathbb{P}(N_{2k} \leq \ell < N_{2k+1}) && \text{(B.27)} \\
&\quad \leq \sum_k \mathbb{P}(N_{2k+1} - N_{2k} \geq \ell\gamma | N_{2k} \leq \ell < N_{2k+1}) \\
&\qquad \cdot \mathbb{P}(N_{2k} \leq \ell < N_{2k+1}) && \text{(B.28)} \\
&\quad \leq \sum_k \mathbb{P}(N_{2k+1} - N_{2k} \geq \ell\gamma | N_{2k} < \infty) \\
&\qquad \cdot \mathbb{P}(N_{2k} \leq \ell < N_{2k+1} | N_{2k+1} - N_{2k} \geq \ell\gamma, N_{2k} < \infty) && \text{(B.29)} \\
&\quad \leq \frac{\ell}{2} \max_{1 \leq 2k \leq \ell} \mathbb{P}(N_{2k+1} - N_{2k} \geq \ell\gamma | N_{2k} < \infty) \ , && \text{(B.30)}
\end{aligned}$$

where (B.27) follows from the law of total probability; (B.28) since the bounded increments of S_ℓ imply that $-(S_\ell - \ell\gamma) \leq N_{2k+1} - N_{2k}$ given $N_{2k} \leq \ell < N_{2k+1}$; (B.29) from Bayes theorem and that probabilities are bounded from above by 1; and (B.30) since probabilities are bounded from above by 1 and for $N_{2k} \leq \ell$, it is necessary for $2k \leq \ell$ by the definition of N_{2k}, N_{2k+1}. Since S_n is a Markov chain, then for all k,

Lemma A.10 implies that we only need to consider

$$\mathbb{P}(N_{2k+1} - N_{2k} \geq \ell\gamma | N_{2k} < \infty) \leq \max_{s\in[-1,0]} \mathbb{P}(N_1 \geq \ell\gamma | S_0 = s) \; . \tag{B.31}$$

$$\leq \left(\frac{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)}{(R_{\mathrm{p}}+\gamma)\epsilon_0^+} \right)^3 e^{-\ell\gamma \cdot D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+)} \; , \tag{B.32}$$

where (B.32) follows from Lemma B.1. Since the above holds for all k, substituting it into (B.30) gives

$$\mathbb{P}(S_\ell \leq 0)$$

$$\leq \frac{\ell}{2} \left(\frac{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)}{(R_{\mathrm{p}}+\gamma)\epsilon_0^+} \right)^3 e^{-\ell\gamma \cdot D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+)} \tag{B.33}$$

$$= \frac{\ell}{2} \cdot \exp\left(\frac{3 \cdot \left(D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+) - \log \frac{R_{\mathrm{p}}+\gamma}{1-\epsilon_0^+} \right)}{1-R_{\mathrm{p}}-\gamma} - \ell\gamma \cdot D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+) \right) \tag{B.34}$$

$$= \frac{\ell}{2} \cdot \exp\left(-\frac{3 \cdot \log \frac{R_{\mathrm{p}}+\gamma}{1-\epsilon_0^+}}{1-R_{\mathrm{p}}-\gamma} - (\ell\gamma - 3/(1-R_{\mathrm{p}}-\gamma)) D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+) \right) \tag{B.35}$$

$$\leq \frac{\ell}{2} \cdot \exp\left(-\frac{3 \cdot \log(R_{\mathrm{p}}+\gamma)}{1-R_{\mathrm{p}}-\gamma} - (\ell\gamma - 3/(1-R_{\mathrm{p}}-\gamma)) D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+) \right) \; , \tag{B.36}$$

where the last line follows since $1 - \epsilon_0^+ \leq 1$, and (B.34) follows from the following

observation:

$$\begin{aligned}
&3 \cdot \log \frac{(1-R_{\mathrm{p}}-\gamma)(1-\epsilon_0^+)}{(R_{\mathrm{p}}+\gamma)\epsilon_0^+} \\
&\quad = 3 \cdot \left(\log \frac{1-R_{\mathrm{p}}-\gamma}{\epsilon_0^+} - \log \frac{R_{\mathrm{p}}+\gamma}{1-\epsilon_0^+} \right) \\
&\quad = 3 \cdot \left(\log \frac{1-R_{\mathrm{p}}-\gamma}{\epsilon_0^+} - \log \frac{R_{\mathrm{p}}+\gamma}{1-\epsilon_0^+} \right) \\
&\quad = \frac{3 \cdot \left((1-R_{\mathrm{p}}-\gamma) \cdot \log \frac{1-R_{\mathrm{p}}-\gamma}{\epsilon_0^+} - (1-R_{\mathrm{p}}-\gamma) \cdot \log \frac{R_{\mathrm{p}}+\gamma}{1-\epsilon_0^+} \right)}{1-R_{\mathrm{p}}-\gamma} \\
&\quad = \frac{3 \cdot \left(D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+) - \log \frac{R_{\mathrm{p}}+\gamma}{1-\epsilon_0^+} \right)}{1-R_{\mathrm{p}}-\gamma} \ .
\end{aligned}$$

For $\ell\gamma > 3/(1-R_{\mathrm{p}}-\gamma)$ in (B.36), we use the fact that $D(1-R_{\mathrm{p}}-\gamma \| \epsilon_0^+) \geq \frac{(1-R_{\mathrm{p}}-\gamma-\epsilon_0^+)^2}{2}$ to get that

$$\begin{aligned}
&\mathbb{P}(S_\ell \leq 0) \\
&\quad \leq \frac{\ell}{2} \cdot \exp\left(-\frac{3 \cdot \log(R_{\mathrm{p}}+\gamma)}{1-R_{\mathrm{p}}-\gamma} - (\ell\gamma - 2/(1-R_{\mathrm{p}}-\gamma)) \frac{(1-R_{\mathrm{p}}-\gamma-\epsilon_0^+)^2}{2} \right) && \text{(B.37)} \\
&\quad \leq \frac{\ell}{2} \cdot \exp\left(-\frac{3 \cdot \log(R_{\mathrm{p}}+\gamma)}{1-R_{\mathrm{p}}-\gamma} - (\ell\gamma - 3/(1-R_{\mathrm{p}}-\gamma)) \frac{\nu^2}{8} \right) \ , && \text{(B.38)}
\end{aligned}$$

where the last line follows by Assumption 2.1 and the assumption in this lemma that $\gamma < \frac{\nu}{2}$. Note that this expression goes to 0 as $\ell \to \infty$. Let $K_{2,R_{\mathrm{p}},\nu} = \frac{\gamma \cdot \nu^2}{8}$. Then for $\ell\gamma > 3/(1-R_{\mathrm{p}}-\gamma)$, let $K_{1,R_{\mathrm{p}},\nu,\ell}$ (B.38) divided by $e^{-\ell K_{2,R_{\mathrm{p}},\nu}}$ and write

$$\mathbb{P}(S_\ell \leq 0) \leq K_{1,R_{\mathrm{p}},\nu,\ell} e^{-\ell K_{2,R_{\mathrm{p}},\nu}} \ . \tag{B.39}$$

For $\ell\gamma \leq 3/(1-R_{\mathrm{p}}-\gamma)$, we simply choose $K_{1,R_{\mathrm{p}},\nu,\ell}$ to make the probability upper bound 1. Then we can conclude our result by simply recalling the definition of S_ℓ in

(2.18). □

B.1.3 Decoder Error

Lemma B.4. *Define error events as follows:*

1. E_1*: $\exists$ frame in which the decoder misidentifies it as active or silent.*
2. E_2*: $\exists$ frame in which the decoder misidentifies the codeword.*

Then, for $\tilde{\delta} > 0, \kappa_n = \lfloor n^{1/16} \rfloor, K_n = \lfloor n^{1/8} \rfloor$ *in the fixed-codebook protocol, as* $n \to \infty$*,*

$$\mathbb{P}(E_1 \cup E_2) \to 0 \ . \tag{B.40}$$

Proof. Note that we can bound the error as

$$\mathbb{P}(E_1 \cup E_2) \leq \mathbb{P}(E_1) + \mathbb{P}(E_2 | E_1^c) \ . \tag{B.41}$$

First consider the event of misidentifying whether transmission is taking place over a frame. There are $\frac{n}{K_n}$ frames, and an error occurs if misidenification happens over any one of them. Thus, by taking a union bound over all frames and applying Stein's Lemma (Lemma A.8) for the error of the repetition code used to distinguish these frames, the error probability is bounded by

$$\mathbb{P}(E_1) \leq \frac{n}{K_n} e^{-\kappa_n \cdot r} \ , \tag{B.42}$$

where $r > 0$ is independent of n.

Finally, we can consider the error corresponding to misidentifying the codewords sent in each frame. By Lemma A.7 and a union bound per frame, we also also have

that

$$\mathbb{P}(E_2|E_1^c) \leq \frac{n}{K_n} \cdot e^{-(K_n-\kappa_n)\cdot r'\cdot\delta^2} \ . \tag{B.43}$$

Combining (B.42) and (B.43) with (B.41), we complete the proof. □

B.1.4 Rate Analysis

The rate achievable by the cognitive radio is directly proportional to the fraction of frames in which it is active. Thus, we consider a bound on the number of frames the cognitive radio is guaranteed to be active. Our first bound holds quite generally for sequences $(\epsilon_{x,i})_{i=1}^{\infty}$.

Lemma B.5. *For all $\delta > 0$ and $\gamma < \delta \cdot (1-\epsilon_0^-)/2$, there exists an $n_0(\delta)$ such that for $n \geq n_0(\delta)$,*

$$\begin{aligned}\mathbb{P}\left(n^{-1}\sum_{k=1}^{n}\tau_k \leq \frac{1-R_p-\epsilon_0^{(n)}}{1-\epsilon_0^-} - \delta\right) \\ \leq \exp\left\{-nD\left(\frac{1+\delta\cdot(1-\epsilon_0^-)/2}{2} - n^{-1}K_n \middle\| \frac{1}{2}\right)\right\} \ ,\end{aligned} \tag{B.44}$$

where $\epsilon_0^{(n)} = n^{-1}\sum_{k=1}^{n}\epsilon_{0,k}$, $\epsilon_0^- = \inf_k \epsilon_{0,k}$, and τ_k is an indicator random variable to denote that k is in an active frame. That is,

$$\tau_k = \mathbb{I}_{\{\exists \ell \in \mathbb{Z}^+ \ such\ that\ V_\ell < k \leq V_\ell + K_n\}} \ ,$$

where V_ℓ is defined in (2.23)*. Furthermore, if $K_n = o(n)$,* (B.44) *goes to* 0 *as $n \to \infty$.*

Proof. Note that if the sequence $\epsilon_{x,k} = 1$ for all k, $x \neq x_{\text{off}}$, and the transmitter does not use the symbol x_{off} during active frames, all channel uses in an active frame result in an erased packet for the primary. Thus, given ϵ_0, this case provides a convenient

albeit conservative way to lower bound the fraction of active frames. We will assume it in the sequel.

Recalling the definition of S_n in (2.18) and the definition of τ_k in (2.20), S_n can be no more than $n\gamma + 2K_n$ during the course of a silent frame. Furthermore, it can only decrease from this during an active frame since $\epsilon_{x,k} = 1$, $x \neq x_{\text{off}}$. Thus, we are guaranteed almost surely that

$$S_n \leq n\gamma + 2K_n \ . \tag{B.45}$$

Now, a simple upper bound on the

$$\begin{aligned}
&\mathbb{P}\left(n^{-1}\sum_{k=1}^{n} \tau_k \leq \frac{1 - R_{\text{p}} - \epsilon_0^{(n)}}{1 - \epsilon_0^-} - \delta \right) \\
&\qquad = \mathbb{P}\left(\sum_{k=1}^{n} \Big(\tau_k(1 - \epsilon_0^-) - (1 - R_{\text{p}} - \epsilon_{0,k}) \Big) \leq -n\delta \cdot (1 - \epsilon_0^-) \right) \\
&\qquad \leq \mathbb{P}\left(\sum_{k=1}^{n} \Big(\tau_k(1 - \epsilon_{0,k}) - (1 - R_{\text{p}} - \epsilon_{0,k}) \Big) \leq -n\delta \cdot (1 - \epsilon_0^-) \right)
\end{aligned} \tag{B.46}$$

Thus, it suffices to find an upper bound on (B.46). Recalling the definition of S_n in (2.18), we can define

$$\tilde{S}_n = S_n + \sum_{k=1}^{n} \Big(\tau_k(1 - \epsilon_{0,k}) - (1 - R_{\text{p}} - \epsilon_{0,k}) \Big) \tag{B.47}$$

Thus, we can rewrite the upper bound in (B.46) as

$$\begin{aligned}
&\mathbb{P}\left(\sum_{k=1}^{n}\Big(\tau_k(1-\epsilon_{0,k})-(1-R_{\mathrm{p}}-\epsilon_{0,k})\Big)\leq -n\delta\cdot(1-\epsilon_0^-)\right)\\
&\qquad=\mathbb{P}(\tilde{S}_n-S_n\leq -n\delta\cdot(1-\epsilon_0^-)) &\text{(B.48)}\\
&\qquad=\mathbb{P}(\tilde{S}_n\leq S_n-n\delta\cdot(1-\epsilon_0^-)) &\text{(B.49)}\\
&\qquad\leq\mathbb{P}(\tilde{S}_n\leq -n\delta\cdot(1-\epsilon_0^-)/2+2K_n) &\text{(B.50)}
\end{aligned}$$

is small, where the last line follows from (B.45) and the assumption that $\gamma\cdot(1-\epsilon_0^-)<\delta/2$.

Under our conservative assumption at the beginning of the proof that $\epsilon_{x,k}=1$, $x\neq x_{\mathrm{off}}$, and that x_{off} is unused on active frames, it is straightforward to verify that $\tilde{S}_n$ is a martingale by checking that $\mathbb{E}[\tilde{S}_n|\tilde{S}_0,\ldots,\tilde{S}_{n-1}]=\tilde{S}_{n-1}$. Furthermore, it has bounded increments. Thus, a bounded martingale concentration inequality [25, p. 57, Corollary 2.4.7] implies that for large enough n,

$$\mathbb{P}(\tilde{S}_n\leq -n\delta\cdot(1-\epsilon_0^-)/2+2K_n)\leq\exp\left\{-nD\left(\frac{1+\delta\cdot(1-\epsilon_0^-)/2}{2}-n^{-1}K_n\middle\|\frac{1}{2}\right)\right\}. \tag{B.51}$$

The result follows immediately. □

Our next bound holds quite for a special class of sequences $(\epsilon_{x,i})_{i=1}^{\infty}$ and will be used to bound the rate achieved by the cognitive radio.

Lemma B.6. *Let* $p^*=\arg\max_{p(x)}I(X;Y)$. *If* $\max\{1-\epsilon_{x_{rep},i},1-\sum_x p^*(x)\epsilon_{x,i}\}<R_p$ *for all* i, *then for all* $\delta>0$ *and* $\gamma<\delta\cdot\Delta\epsilon^+/2$, *where* $\Delta\epsilon^+=\sup_{x,i}\epsilon_{x,i}-\epsilon_{0,i}$, *there*

exists an $n_0(\delta)$ *such that for* $n \geq n_0(\delta)$,

$$\begin{aligned}\mathbb{P}\left(n^{-1}\sum_{k=1}^{n}\tau_k \leq \frac{1-R_p-\epsilon_0^{(n)}}{\Delta\epsilon^+} - \delta\right) \\ \leq \exp\left\{-nD\left(\frac{1+\delta\cdot\Delta\epsilon^+/4}{2} - n^{-1}K_n\middle\|\frac{1}{2}\right)\right\} \\ + \left(\frac{(R_p+\gamma)(1-R_p)}{(1-R_p-\gamma)R_p}\right)^{2K_n}\cdot e^{-n\frac{\delta}{4}\cdot D(R_p+\gamma\|R_p)}\ , \end{aligned} \tag{B.52}$$

where $\epsilon_0^{(n)} = n^{-1}\sum_{k=1}^{n}\epsilon_{0,k}$, *and* τ_k *is an indicator random variable to denote that* k *is in an active frame. That is,*

$$\tau_k = \mathbb{I}_{\{\exists \ell\in\mathbb{Z}^+ \text{ such that } V_\ell < k \leq V_\ell + K_n\}}\ ,$$

where V_ℓ *is defined in* (2.23). *Furthermore, if* $K_n = o(n)$, (B.52) *goes to* 0 *as* $n\to\infty$.

Proof. Recalling the definition of S_n in (2.18) and the definition of τ_k in (2.20), Lemma B.2 yields the following bound on the positive drift of S_n:

$$\mathbb{P}(S_n \geq t + n\cdot\gamma + K_n) \leq \left(\frac{(R_\mathrm{p}+\gamma)(1-R_\mathrm{p})}{(1-R_\mathrm{p}-\gamma)R_\mathrm{p}}\right)^{2K_n}\cdot e^{-t\cdot D(R_\mathrm{p}+\gamma\|R_\mathrm{p})}\ . \tag{B.53}$$

Now, a simple upper bound on the

$$\begin{aligned}\mathbb{P}\left(n^{-1}\sum_{k=1}^{n}\tau_k \leq \frac{1-R_\mathrm{p}-\epsilon_0^{(n)}}{\Delta\epsilon^+} - \delta\right) \\ = \mathbb{P}\left(\sum_{k=1}^{n}\Big(\tau_k\cdot\Delta\epsilon^+ - (1-R_\mathrm{p}-\epsilon_{0,k})\Big) \leq -n\delta\cdot\Delta\epsilon^+\right) \\ \leq \mathbb{P}\left(\sum_{k=1}^{n}\Big(\tau_k(\bar\epsilon_k-\epsilon_{0,k}) - (1-R_\mathrm{p}-\epsilon_{0,k})\Big) \leq -n\delta\cdot\Delta\epsilon^+\right)\ , \end{aligned} \tag{B.54}$$

where $\bar\epsilon_k = 1 - \mathbb{E}[A_k|S_0,\ldots,S_k-1]$. Thus, it suffices to find an upper bound on

(B.54). Recalling the definition of S_n in (2.18), we can define

$$\tilde{S}_n = S_n + \sum_{k=1}^{n} \Big(\tau_k(\bar{\epsilon}_k - \epsilon_{0,k}) - (1 - R_{\mathrm{p}} - \epsilon_{0,k}) \Big) \tag{B.55}$$

Thus, we can rewrite the upper bound in (B.54) as

$$\begin{aligned}
&\mathbb{P}\left(\sum_{k=1}^{n} \Big(\tau_k(1 - \epsilon_{0,k}) - (1 - R_{\mathrm{p}} - \epsilon_{0,k}) \Big) \leq -n\delta \cdot \Delta\epsilon^{+} \right) \\
&\quad = \mathbb{P}(\tilde{S}_n - S_n \leq -n\delta \cdot \Delta\epsilon^{+}) && \text{(B.56)} \\
&\quad = \mathbb{P}(\tilde{S}_n \leq S_n - n\delta \cdot \Delta\epsilon^{+}) && \text{(B.57)} \\
&\quad \leq \mathbb{P}(\tilde{S}_n \leq -n\delta \cdot \Delta\epsilon^{+}/4 + 2K_n) \\
&\qquad + \left(\frac{(R_{\mathrm{p}} + \gamma)(1 - R_{\mathrm{p}})}{(1 - R_{\mathrm{p}} - \gamma)R_{\mathrm{p}}} \right)^{2K_n} \cdot e^{-n\frac{\delta \cdot \Delta\epsilon^{+}}{4} \cdot D(R_{\mathrm{p}} + \gamma \| R_{\mathrm{p}})} && \text{(B.58)}
\end{aligned}$$

is small, where the last line follows from applying (B.53) at $t = n\delta \cdot \Delta\epsilon^{+}/4$ and the assumption that $\gamma < \delta \cdot \Delta\epsilon^{+}/2$.

It is straightforward to verify that $\tilde{S}_n$ is a martingale by checking that

$$\mathbb{E}[\tilde{S}_n | \tilde{S}_0, \ldots, \tilde{S}_{n-1}] = \tilde{S}_{n-1}.$$

Furthermore, it has bounded increments. Thus, a bounded martingale concentration inequality [25, p. 57, Corollary 2.4.7] implies that for large enough n,

$$\mathbb{P}(\tilde{S}_n \leq -n\delta/4 + 2K_n) \leq \exp\left\{ -nD\left(\frac{1 + \delta \cdot \Delta\epsilon^{+}/4}{2} - n^{-1}K_n \middle\| \frac{1}{2} \right) \right\} . \tag{B.59}$$

The result follows immediately. □

Lemma B.7. *Let $p^* = \arg\max_{p(x)} I(X;Y)$. Then for all $\nu > 0$, if $1 - \sum_x p^*(x) \cdot \epsilon_{x,i} \geq R_p + \nu$ for all i, the fixed-codebook protocol with parameters $(\gamma, \kappa_n, K_n, \tilde{\delta})$ satisfying*

$0 < \gamma < \nu/2$, $\kappa_n = o(K_n)$, $K_n = o(\sqrt{n})$, $1 > \tilde{\delta} > 0$, *has the property that as* $n \to \infty$,

$$\mathbb{P}\left(\exists\ j > \sqrt{n},\ \frac{j}{K_n} \in \mathbb{Z},\ \textit{such that } S_j - j \cdot \gamma < K_n\right) \to 0\ . \tag{B.60}$$

Recall the condition for a silent frame in (2.20)*, this implies that after time* $\sqrt{n}$*, the probability of a silent frame is negligible.*

Proof. We start by defining

$$\tilde{S}_j = \tilde{S}_{j-1} + A_j - \mathbb{E}[A_j | \tilde{S}_0, \ldots, \tilde{S}_{j-1}], \quad \tilde{S}_0 = 0, \tag{B.61}$$

and it is easy to verify that $\tilde{S}_j$ is a bounded martingale. By our assumption about $\sum_x p^*(x) \cdot \epsilon_{x,i}$ for all i, (2.20) and (2.24) imply that if $\ell K_n \geq j > (\ell - 1)K_n + \kappa_n$ for some integer $\ell \geq 1$,

$$\mathbb{E}[A_j | \tilde{S}_0, \ldots, \tilde{S}_{j-1}] \geq R_{\mathrm{p}} + \nu\ . \tag{B.62}$$

Then for $k \geq n^{1/2}$,

$$\begin{aligned}
&\mathbb{P}\left(k^{-1}\sum_{i=1}^{k} A_i < R_{\mathrm{p}} + \gamma + k^{-1}K_n\right) \\
&\quad \leq \mathbb{P}\left(k^{-1}\tilde{S}_k < R_{\mathrm{p}} + \gamma + k^{-1}K_n - k^{-1}\beta_{k,n}(R_{\mathrm{p}} + \nu)\right) &\text{(B.63)}\\
&\quad \leq \mathbb{P}\left(k^{-1}\tilde{S}_k < -\nu/2 + o(1)\right)\ , &\text{(B.64)}
\end{aligned}$$

where $\beta_{k,n} = (k - n^{1/4}) \cdot \frac{K_n - \kappa_n}{K_n}$, $o(1)$ is notational convenience for $\lim_{n\to\infty} o(1) = 0$, and (B.63) follows from (B.61), (B.62), and since for all i, $A_i \geq 0$ almost surely. Since $\tilde{S}_k$ is a zero-mean bounded martingale, for $k \geq n^{1/2}$ and large enough n, we can apply

a bounded martingale concentration inequality [25, p. 57, Corollary 2.4.7] to yield

$$\begin{aligned}
&\mathbb{P}\left(k^{-1}\sum_{i=1}^{k} A_k < R_{\mathrm{p}} + \gamma + k^{-1}\cdot K_n\right) \\
&\quad \le \exp\left(-k(\nu/2+o(1))^2/2\right) && \text{(B.65)}\\
&\quad = \exp\left(-\lceil\sqrt{n}\rceil(\nu/2+o(1))^2/2\right)\cdot\exp\left(-(k-\lceil\sqrt{n}\rceil)(\nu/2+o(1))^2/2\right) && \text{(B.66)}
\end{aligned}$$

From the above result and a union bound,

$$\begin{aligned}
&\mathbb{P}(\tilde{E}) \\
&\quad \le \exp\left(-\lceil\sqrt{n}\rceil(\nu/2+o(1))^2/2\right)\cdot\sum_{k=\lceil\sqrt{n}\rceil}^{n}\exp\left(-(k-\lceil\sqrt{n}\rceil)(\nu/2+o(1))^2/2\right) && \text{(B.67)}\\
&\quad \le \exp\left(-\lceil\sqrt{n}\rceil(\nu/2+o(1))^2/2\right)\cdot\sum_{m=0}^{\infty}\exp\left(-m(\nu/2+o(1))^2/2\right) && \text{(B.68)}
\end{aligned}$$

However, the geometric series does not affect the error probability by more than a constant asymptotically, so taking the limit above completes the result. □

Lemma B.8. *Given rate target R_p and $\nu > 0$, for all $\delta > 0$ and $\{\epsilon_{x,k}\}_{k=1}^{\infty}$ satisfying Assumption 2.1, consider the fixed-codebook protocol with $\kappa_n = o(K_n), K_n = o(\sqrt{n}), K_n \to \infty$ as $n \to \infty$, and $\gamma < \tilde{\delta}/2$. Then there exists a choice of $\tilde{\delta}$ such that this strategy achieves rates of at least*

$$\begin{cases} C^* - \delta\ , & \forall\ i\ , 1 - \sum_x p^*(x)\cdot\epsilon_{x,i} \ge R_p + \nu \\ \frac{1-R_p-\hat{\epsilon}_0^+}{\Delta\epsilon^+}\cdot C^* - \delta\ , & \forall\ i\ , 1 - R_p < \min\{\epsilon_{x_{rep},i}, \sum_x p^*(x)\epsilon_{x,i}\} \\ \frac{1-R_p-\hat{\epsilon}_0^+}{1-\epsilon_0^-}\cdot C^* - \delta\ , & \textit{otherwise} \end{cases}\ , \qquad \text{(B.69)}$$

where $C^* = \max_{p(x)} I(X;Y)$, $p^* = \arg\max_{p(x)} I(X;Y)$, $\Delta\epsilon^+ = \sup_{x,i}\epsilon_{x,i} - \epsilon_{0,i}$, $\hat{\epsilon}_0^+ =$

$\limsup_k k^{-1}\sum_{i=1}^{k}\epsilon_{0,i}$, *and* $\epsilon_0^- = \inf_i \epsilon_{0,i}$.

Proof. By Lemma B.5, we know that with probability going to 1 as $n \to \infty$, a fraction of at least

$$\liminf_k \frac{1-R_{\mathrm{p}}-k^{-1}\sum_{i=1}^{k}\epsilon_{0,k}}{1-\epsilon_0^-} - \tilde{\delta} = \frac{1-R_{\mathrm{p}}-\hat{\epsilon}_0^+}{1-\epsilon_0^-} - \tilde{\delta} \tag{B.70}$$

of the frames will be active frames. By similar arguments with Lemmas B.6, and B.7, we can state the summarize the fraction of frames that will be active for different cases as follows:

$$\begin{cases} \frac{n-\sqrt{n}}{n} - \tilde{\delta}\ , & \forall\ i\ , 1-\sum_x p^*(x)\cdot\epsilon_{x,i} \geq R_{\mathrm{p}}+\nu \\ \frac{1-R_{\mathrm{p}}-\hat{\epsilon}_0^+}{\Delta\epsilon^+} - \tilde{\delta}\ , & \forall\ i\ , 1-R_{\mathrm{p}} < \min\{\epsilon_{x_{\mathrm{rep}},i}, \sum_x p^*(x)\epsilon_{x,i}\} \\ \frac{1-R_{\mathrm{p}}-\hat{\epsilon}_0^+}{1-\epsilon_0^-} - \tilde{\delta}\ , & \text{otherwise} \end{cases} \tag{B.71}$$

The rate for each active frame is

$$\frac{K_n-\kappa_n}{K_n}(C^*-\tilde{\delta}) \geq C^* - \frac{\kappa_n}{K_n}\cdot C^* - \tilde{\delta}\ , \tag{B.72}$$

so we have that the following rates are achievable for sequences such that for all i, $1-\sum_x p^*(x)\cdot\epsilon_{x,i} \geq R_{\mathrm{p}}+\nu$:

$$\begin{aligned} &\left(\frac{n-\sqrt{n}}{n} - \tilde{\delta}\right)\cdot\left(C^* - \frac{\kappa_n}{K_n}\cdot C^* - \tilde{\delta}\right) \\ &\qquad \geq C^* - (\kappa_n/K_n + \sqrt{n}/n + \tilde{\delta})\cdot C^* - \tilde{\delta}\ , \end{aligned} \tag{B.73}$$

the following for sequences such that for all i, $1 - R_{\mathrm{p}} < \min\{\epsilon_{x_{\mathrm{rep}},i}, \sum_x p^*(x)\epsilon_{x,i}\}$:

$$\left(\frac{1 - R_{\mathrm{p}} - \hat{\epsilon}_0^+}{\Delta\epsilon^+} - \tilde{\delta}\right) \cdot (C^* - \frac{\kappa_n}{K_n} \cdot C^* - \tilde{\delta}) \\ \geq \frac{1 - R_{\mathrm{p}} - \hat{\epsilon}_0^+}{\Delta\epsilon^+} \cdot C^* - (\kappa_n/K_n + \tilde{\delta}) \cdot C^* - \tilde{\delta} \ , \tag{B.74}$$

and the following for all other sequences:

$$\left(\frac{1 - R_{\mathrm{p}} - \hat{\epsilon}_0^+}{1 - \epsilon_0^-} - \tilde{\delta}\right) \cdot \frac{K_n - \kappa_n}{K_n}(C^* - \tilde{\delta}) \\ \geq \left(\frac{1 - R_{\mathrm{p}} - \hat{\epsilon}_0^+}{1 - \epsilon_0^-}\right) C^* - \tilde{\delta} - (\kappa_n/K_n + \tilde{\delta}) \cdot C^* \ . \tag{B.75}$$

For large enough n, by assumption $\kappa_n/K_n + \sqrt{n}/n \leq \tilde{\delta}/C^*$, so that rates of at least the following are achievable:

$$\begin{cases} C^* - 2\tilde{\delta} - \tilde{\delta} \cdot C^* \ , & \forall\ i\ , \sum_x p^*(x) \cdot \epsilon_{x,i} \geq R_{\mathrm{p}} + \nu \\ \frac{1-R_{\mathrm{p}}-\hat{\epsilon}_0^+}{\Delta\epsilon^+} \cdot C^* - 2\tilde{\delta} - \tilde{\delta} \cdot C^* \ , & \forall\ i\ , 1 - R_{\mathrm{p}} < \min\{\epsilon_{x_{\mathrm{rep}},i}, \sum_x p^*(x)\epsilon_{x,i}\} \\ \frac{1-R_{\mathrm{p}}-\hat{\epsilon}_0^+}{1-\epsilon_0^-} \cdot C^* - 2\tilde{\delta} - \tilde{\delta} \cdot C^* \ , & \text{otherwise} \end{cases} \tag{B.76}$$

are achievable with probability going to 1 as $n \to \infty$. By choosing $\tilde{\delta} = \frac{1}{3}\min\{\delta, \delta/C^*\}$, we can conclude the result. □

B.2 Proof of Theorem 2.4

B.2.1 Decoder Error

Lemma B.9. *Define error events as follows:*

1. E_1*: $\exists$ frame in which the decoder misidentifies whether a frame is active or silent.*

2. *E_2: the decoder misidentifies the selected codebook.*

3. *E_3: $\exists$ an active frame in which decoder misidentifies the codeword.*

Then, for $\tilde{\delta} > 0, C_n = \lfloor n^{1/32} \rfloor, \kappa_n = \lfloor n^{1/16} \rfloor, K_n = \lfloor n^{1/8} \rfloor$ in the codebook-adaptive protocol, as $n \to \infty$,

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) \to 0 \ . \tag{B.77}$$

Proof. Note that we can bound the error as

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) \leq \mathbb{P}(E_1) + \mathbb{P}(E_2|E_1^c) + \mathbb{P}(E_3|E_1^c, E_2^c) \ . \tag{B.78}$$

First consider the event of misidentifying whether transmission is taking place over a frame. There are $\frac{n}{K_n}$ frames, and an error occurs if misidenification happens over any one of them. Thus, by taking a union bound over all frames and applying Stein's Lemma (Lemma A.8) for the error of the repetition code used to distinguish these frames, the error probability is bounded by

$$\mathbb{P}(E_1) \leq \frac{n}{K_n} e^{-\kappa_n \cdot r} \ , \tag{B.79}$$

where $r > 0$ is independent of n. Another source of error is misidentifying the codebook. For large enough n, $(C_n + 1)^{|\mathcal{X}|}$ does not exceed $C - \tilde{\delta}$, and we can apply Lemma A.7 to get the error probability

$$\mathbb{P}(E_2|E_1^c) \leq e^{-(K_n - \kappa_n) \cdot r' \cdot \tilde{\delta}^2} \ . \tag{B.80}$$

Finally, we can consider the error corresponding to misidentifying the codewords sent

in each frame. By Lemma A.7 and a union bound per frame, we also also have that

$$\mathbb{P}(E_3|E_2^c, E_1^c) \leq \frac{n}{K_n} \cdot e^{-(K_n-\kappa_n)\cdot r'\cdot\tilde{\delta}^2} \ . \tag{B.81}$$

Combining (B.79), (B.80), and (B.81) with (B.78), we complete the proof. □

B.2.2 Rate Analysis

The rate loss argument is the most tedious because one must account for a variety of factors: the length of the first two phases of transmission, the gap between the rates of quantized set of codebooks and points on the RIB function, and the number of active frames in Phase III. We therefore subdivide the result into several lemmas.

B.2.2.1 Phase I and II are short

Because the encoder does not send message information in Phase I and II, we want the length of these phases to be sublinear in n to guarantee negligible rate loss.

Lemma B.10. *For all $\nu > 0$, let $\gamma < \nu/2$, $\kappa_n = \lfloor n^{1/16} \rfloor, K_n = \lfloor n^{1/8} \rfloor$ in the codebook-adaptive protocol. Furthermore, let T be the length of Phases I and II and $\tilde{E}_1 = \{T \geq n^{1/4}\}$. Then*

$$\mathbb{P}(\tilde{E}_1) \to 0 \tag{B.82}$$

as $n \to \infty$.

Proof. Consider the transition times from silent frames to active frames and vice

versa. To do this, define the stopping times for $k \geq 1$,

$$\tilde{N}_{2k-1} = K_n \cdot \inf\{i > K_n^{-1}\tilde{N}_{2k-2} : S_{i \cdot K_n} - i \cdot K_n \cdot \gamma - K_n \geq 0\} \ , \tag{B.83}$$

$$\tilde{N}_{2k} = K_n \cdot \inf\{i > K_n^{-1}\tilde{N}_{2k-1} : S_{i \cdot K_n} - i \cdot K_n \cdot \gamma - K_n < 0\} \ , \tag{B.84}$$

where $\tilde{N}_0 = 0$. Phases I and II end after the first two active frames. We can get a bound on the start of the first active frame immediately from Lemma B.1, which implies

$$\mathbb{P}(\tilde{N}_1 \geq t | S_0 = 0) \leq \left(\frac{(1 - R_{\mathrm{p}} - \gamma)(1 - \epsilon_0)}{(R_{\mathrm{p}} + \gamma)\epsilon_0} \right)^{2K_n} e^{-tD(1 - R_{\mathrm{p}} - \gamma \| \epsilon_0)} \ . \tag{B.85}$$

Thus, if $\tilde{N}_2 > \tilde{N}_1 + K_n$, then

$$T = \tilde{N}_1 + 2K_n. \tag{B.86}$$

Together with (B.85), this implies

$$\mathbb{P}(T \geq n^{1/4} | \tilde{N}_2 > \tilde{N}_1 + K_n) \leq \left(\frac{(1 - R_{\mathrm{p}} - \gamma)(1 - \epsilon_0)}{(R_{\mathrm{p}} + \gamma)\epsilon_0} \right)^{2K_n} e^{-(n^{1/4} - K_n) \cdot D(1 - R_{\mathrm{p}} - \gamma \| \epsilon_0)} \tag{B.87}$$

The remaining case to consider is if $\tilde{N}_2 = \tilde{N}_1 + K_n$. If this happens, then Phases I and II end at

$$T = \tilde{N}_3 + K_n. \tag{B.88}$$

Then (B.88) implies

$$\begin{aligned}
&\mathbb{P}(T \geq n^{1/4} | \tilde{N}_2 = \tilde{N}_1 + K_n) \\
&\quad = \mathbb{P}(\tilde{N}_3 \geq n^{1/4} - K_n | \tilde{N}_2 = \tilde{N}_1 + K_n) && \text{(B.89)} \\
&\quad = \mathbb{P}(\tilde{N}_3 - \tilde{N}_2 \geq n^{1/4} - 2K_n - \tilde{N}_1 | \tilde{N}_2 = \tilde{N}_1 + K_n) && \text{(B.90)} \\
&\quad \leq \mathbb{P}(\tilde{N}_3 - \tilde{N}_2 \geq n^{1/4} - 2K_n - \tilde{N}_1 \text{ or } \tilde{N}_1 \geq n^{1/4}/2 | \tilde{N}_2 = \tilde{N}_1 + K_n) && \text{(B.91)} \\
&\quad \leq \mathbb{P}(\tilde{N}_1 \geq n^{1/4}/2 | \tilde{N}_2 = \tilde{N}_1 + K_n) \\
&\quad + \mathbb{P}(\tilde{N}_3 - \tilde{N}_2 \geq n^{1/4}/2 - 2K_n | \tilde{N}_2 = \tilde{N}_1 + K_n, \tilde{N}_1 < n^{1/4}/2) \ , && \text{(B.92)}
\end{aligned}$$

where (B.89) follows from (B.88), (B.90) follows by our conditioning, (B.91) follows since we are increasing the possible events over which we are taking the probability, and (B.92) follows from $P(A \text{ or } B) = P(A) + P(A^c)P(B|A^c)$.

By Lemma A.10 and Lemma B.1,

$$\begin{aligned}
&\mathbb{P}(\tilde{N}_3 - \tilde{N}_2 \geq n^{1/4}/2 - 2K_n | \tilde{N}_2 = \tilde{N}_1 + K_n, \tilde{N}_1 < n^{1/4}/2) \\
&\qquad \leq \left(\frac{(1 - R_{\mathrm{p}} - \gamma)(1 - \epsilon_0)}{(R_{\mathrm{p}} + \gamma)\epsilon_0} \right)^{2K_n} e^{-(n^{1/4}/2 - 2K_n) \cdot D(1 - R_{\mathrm{p}} - \gamma \| \epsilon_0)} \ . && \text{(B.93)}
\end{aligned}$$

By combining (B.85), and (B.92), and (B.93),

$$\begin{aligned}
&\mathbb{P}(T \geq n^{1/4} | \tilde{N}_2 = \tilde{N}_1 + K_n) \\
&\qquad \leq \left(\frac{(1 - R_{\mathrm{p}} - \gamma)(1 - \epsilon_0)}{(R_{\mathrm{p}} + \gamma)\epsilon_0} \right)^{2K_n} e^{-(n^{1/4}/2 - 2K_n) \cdot D(1 - R_{\mathrm{p}} - \gamma \| \epsilon_0)} \\
&\qquad + \left(\frac{(1 - R_{\mathrm{p}} - \gamma)(1 - \epsilon_0)}{(R_{\mathrm{p}} + \gamma)\epsilon_0} \right)^{2K_n} e^{-\frac{n^{1/4}}{2} D(1 - R_{\mathrm{p}} - \gamma \| \epsilon_0)} \ . && \text{(B.94)}
\end{aligned}$$

The result follows immediately from (B.87) and (B.94). □

B.2.2.2 Codebook Quantization

To account for the error in the interference estimates and that the encoder must inform the decoder of which rate it will be targetting, we only have a limited number of codebooks to choose from at the start of Phase II. Thus, in general there will be a gap between the rate of a selected codebook and an actual point on the RIB function. In this subsection, we ensure that this gap is small.

We first provide guarantees on accurate interference estimates.

Lemma B.11. *Let $\mathcal{D}_\ell$ be the event that Phase I of the codebook-adaptive protocol terminates at frame ℓ. Then*

$$\mathbb{P}(\max_x |\hat{\epsilon}_x - \epsilon_x| > \delta | \mathcal{D}_\ell) \leq 2|\mathcal{X}| e^{-\mu\delta^2/2} \ , \tag{B.95}$$

where $\mu = \lfloor \frac{K_n - \kappa_n}{|\mathcal{X}|} \rfloor$.

Proof. Recall the definition of the estimates given in (2.31). Then $\mathcal{D}_\ell = \{V_1 = (\ell-1)K_n\}$ is an equivalent expression for the event. By Hoeffding's inequality [25, p. 57, Corollary 2.4.7], we have for each $x \in \mathcal{X}$

$$\mathbb{P}(|\hat{\epsilon}_x - \epsilon_x| > \delta | V_1 = (\ell-1)K_n) \leq 2e^{-\mu\delta^2/2} \ . \tag{B.96}$$

The result then follows from a union bound on $\mathbb{P}(\max_x |\hat{\epsilon}_x - \epsilon_x| > \delta | \mathcal{D}_\ell)$. □

Lemma B.12. *For the selected codebook χ given in (2.33), define $\tilde{E}_2^c$ as the event where the following two conditions are met:*

$$\sum_x \epsilon_x p_\chi(x) \leq 1 - R_p - 2\gamma \tag{B.97}$$

$$6\frac{|\mathcal{X}|}{C_n}\log\frac{2}{C_n\cdot|\mathcal{Y}|}+\frac{3\tilde{\delta}}{2|\mathcal{X}|}\log\frac{\tilde{\delta}}{2|\mathcal{X}|^2\cdot|\mathcal{Y}|}$$
$$\leq R_\chi - R_{IB}(\vec{\epsilon}, R_p+2\gamma+\tilde{\delta}) \leq \tag{B.98}$$
$$-\frac{3\tilde{\delta}}{2|\mathcal{X}|}\log\frac{\tilde{\delta}}{2|\mathcal{X}|^2\cdot|\mathcal{Y}|}.$$

Then for all $1>\tilde{\delta}>0$, $C_n\geq 4|\mathcal{X}|$, *and as* $n\to\infty$,

$$\mathbb{P}(\tilde{E}_2|\tilde{E}_1^c)\to 0\ , \tag{B.99}$$

where $\tilde{E}_1$ *is defined in Lemma B.10.*

Proof. Given $\tilde{E}_1^c$, we can guarantee by setting $\delta=\frac{\tilde{\delta}}{4|\mathcal{X}|}$ in Lemma B.11 that

$$\sum_x|\hat{\epsilon}_x-\epsilon_x|\leq\frac{\tilde{\delta}}{4} \tag{B.100}$$

with probability going to 1 as $n\to\infty$. Furthermore, we know that by definition

$$\sum_x\hat{\epsilon}_x p_\chi(x)\leq 1-R_{\mathrm{p}}-2\gamma-\tilde{\delta} \tag{B.101}$$

From (B.100) and (B.101), we have that

$$\sum_x\epsilon_x p_\chi(x)\leq\sum_x\hat{\epsilon}_x p_\chi(x)+\tilde{\delta} \tag{B.102}$$
$$\leq 1-R_{\mathrm{p}}-2\gamma. \tag{B.103}$$

It remains to verify the other condition. Note that for any $p(x)$, there is a codebook in the set with input distribution type $p_{x^{C_n}}(x)$ such that $\sum_x|p(x)-p_{x^{C_n}}(x)|\leq\frac{|\mathcal{X}|}{C_n}$.

Then by the continuity of entropy [23, Lemma 2.7, p. 33], we know that

$$R_{\mathrm{IB}}(\vec{\tilde{\epsilon}}, R_{\mathrm{p}} + 2\gamma + \tilde{\delta}) + 6\frac{|\mathcal{X}|}{C_n} \log \frac{2}{C_n \cdot |\mathcal{Y}|} \leq R_\chi \leq R_{\mathrm{IB}}(\vec{\tilde{\epsilon}}, R_{\mathrm{p}} + 2\gamma + \tilde{\delta}) \ , \tag{B.104}$$

where the inequality on the right follows from (B.101) and the definition of the R_{IB} function. Lemma A.14 and (B.100) imply that

$$|R_{\mathrm{IB}}(\vec{\epsilon}, R_{\mathrm{p}} + 2\gamma + \tilde{\delta}) - R_{\mathrm{IB}}(\vec{\tilde{\epsilon}}, R_{\mathrm{p}} + 2\gamma + \tilde{\delta})| \leq -\frac{3\tilde{\delta}}{2|\mathcal{X}|} \log \frac{\tilde{\delta}}{2|\mathcal{X}|^2 \cdot |\mathcal{Y}|} \ . \tag{B.105}$$

Combining (B.104) and (B.105) yield the result. □

B.2.2.3 Always On

Our next lemma shows that all frames are active after time $\sqrt{n}$ with probability going to 1 as $n \to \infty$.

Lemma B.13. *Let $\tilde{E}_1$ and $\tilde{E}_2$ be defined as in Lemmas B.10 and B.12, respectively. Define $\tilde{E}_3$ to be the event that for some $j > \sqrt{n}$, the condition in (2.28) is met, resulting in a silent frame. Then for all $\nu > 0$ and $\epsilon_{x,i} = \epsilon_x$ for $x \neq x_{\mathit{off}}$, the codebook-adaptive protocol with parameters $(\gamma, C_n, \kappa_n, K_n, \tilde{\delta})$ satisfying $0 < \gamma < \nu/2, C_n = \lfloor n^{1/32} \rfloor, \kappa_n = \lfloor n^{1/16} \rfloor, K_n = \lfloor n^{1/8} \rfloor$, $1 > \tilde{\delta} > 0$, has the property that as $n \to \infty$,*

$$\mathbb{P}(\tilde{E}_3 | \tilde{E}_1^c, \tilde{E}_2^c) \to 0 \ . \tag{B.106}$$

Proof. We start by defining

$$\tilde{S}_j = \tilde{S}_{j-1} + A_j - \mathbb{E}[A_j | \tilde{S}_0, \ldots, \tilde{S}_{j-1}], \quad \tilde{S}_0 = 0, \tag{B.107}$$

and it is easy to verify that $\tilde{S}_j$ is a bounded martingale. From the definition of $\tilde{E}_2^c$ in Lemma B.12 and for $\nu > 0$, (B.97), (2.28), and (2.24) imply that for j satisfying

$j > V_2 + K_n$ and $\ell K_n \geq j > (\ell - 1)K_n + \kappa_n$ for some integer $\ell \geq 1$,

$$\mathbb{E}[A_j | \tilde{S}_0, \ldots, \tilde{S}_{j-1}] \geq R_{\mathrm{p}} + 2\gamma \ . \tag{B.108}$$

Then for $k \geq n^{1/2}$ and under $\tilde{E}_1^c$,

$$\begin{aligned}
&\mathbb{P}\left(k^{-1} \sum_{i=1}^{k} A_i < R_{\mathrm{p}} + \gamma + k^{-1} K_n \middle| \tilde{E}_1^c, \tilde{E}_2^c \right) \\
&\leq \mathbb{P}\left(k^{-1} \tilde{S}_k < R_{\mathrm{p}} + \gamma + k^{-1} K_n - k^{-1} \beta_{k,n} (R_{\mathrm{p}} + 2\gamma) \middle| \tilde{E}_1^c, \tilde{E}_2^c \right) &\text{(B.109)} \\
&\leq \mathbb{P}\left(k^{-1} \tilde{S}_k < -\gamma + o(1) \middle| \tilde{E}_1^c, \tilde{E}_2^c \right) \ , &\text{(B.110)}
\end{aligned}$$

where $\beta_{k,n} = (k - n^{1/4}) \cdot \frac{K_n - \kappa_n}{K_n}$, $o(1)$ is notational convenience for $\lim_{n\to\infty} o(1) = 0$, and (B.109) follows from (B.107), (B.108), and since for all i, $A_i \geq 0$ almost surely. Since $\tilde{S}_k$ is a zero-mean bounded martingale, for $k \geq n^{1/2}$ and large enough n, we can apply a bounded martingale concentration inequality [25, p. 57, Corollary 2.4.7] to yield

$$\begin{aligned}
&\mathbb{P}\left(k^{-1} \sum_{i=1}^{k} A_k < R_{\mathrm{p}} + \gamma + k^{-1} \cdot K_n \middle| \tilde{E}_1^c, \tilde{E}_2^c \right) \\
&\leq \exp\left(-k(\gamma + o(1))^2 / 2 \right) &\text{(B.111)} \\
&= \exp\left(-\lceil \sqrt{n} \rceil (\gamma + o(1))^2 / 2 \right) \cdot \exp\left(-(k - \lceil \sqrt{n} \rceil)(\gamma + o(1))^2 / 2 \right) &\text{(B.112)}
\end{aligned}$$

From the above result and a union bound,

$$\mathbb{P}(\tilde{E}_3|\tilde{E}_1^c, \tilde{E}_2^c)$$
$$\leq \exp\left(-\lceil\sqrt{n}\rceil(\gamma + o(1))^2/2\right) \cdot \sum_{k=\lceil\sqrt{n}\rceil}^{n} \exp\left(-(k - \lceil\sqrt{n}\rceil)(\gamma + o(1))^2/2\right) \tag{B.113}$$
$$\leq \exp\left(-\lceil\sqrt{n}\rceil(\gamma + o(1))^2/2\right) \cdot \sum_{m=0}^{\infty} \exp\left(-m(\gamma + o(1))^2/2\right) \tag{B.114}$$

However, the geometric series does not affect the error probability by more than a constant asymptotically, so taking the limit above completes the result. □

B.2.2.4 Overall Rate Loss

Lemma B.14. *For all $\nu > 0$, $\epsilon_{x,i} = \epsilon_x$ for $x \neq x_{off}$ and given any $\delta > 0$, consider the codebook-adaptive protocol with parameters $(\gamma, C_n, \kappa_n, K_n)$ satisfying $0 < \gamma < \min\{\nu/2, \tilde{\delta}/2\}, C_n = \lfloor n^{1/32} \rfloor, \kappa_n = \lfloor n^{1/16} \rfloor, K_n = \lfloor n^{1/8} \rfloor$. Then there exists a choice of the parameter $\tilde{\delta} \in (0, 1/8)$ so that with probability going to 1 as $n \to \infty$, the cognitive radio achieves rates*

$$R \geq R_{IB}(\vec{\epsilon}, R_p) - \delta \ . \tag{B.115}$$

Proof. Let $\tilde{E}_1, \tilde{E}_2, \tilde{E}_3$ be defined as in Lemmas B.10, B.12, and B.13 respectively. From these results, we know that

$$\mathbb{P}(\tilde{E}_1 \cup \tilde{E}_2 \cup \tilde{E}_3) \to 0 \tag{B.116}$$

as $n \to \infty$ and thus with high probability,

1. Phases I and II are short, ending by $n^{1/4}$ (Lemma B.10).

2. The gap between the codebook's rate and the RIB function is small (Lemma B.12).

3. After time $n^{1/2}$, all frames are active frames (Lemma B.13).

Furthermore, we know that by our repetition code, there is a loss of κ_n positions for our repetition code over a frame K_n. Factoring in this source of rate loss along with the fact that we are in Phase III by time $n^{1/2}$ (Lemma B.10) and always in an active frame (Lemma B.13), the rate

$$\begin{aligned}\frac{n-\sqrt{n}}{n} \cdot \frac{K_n - \kappa_n}{K_n}(R_\chi - \tilde{\delta}) \\ \geq (R_\chi - \tilde{\delta}) - \left(n^{-1/2} + \frac{\kappa_n}{K_n}\right) \log|\mathcal{X}|\end{aligned} \tag{B.117}$$

is achievable for the cognitive radio with probability going to 1 as $n \to \infty$. Finally, we know that the gap between the codebook's rate and the RIB function is small (Lemma B.12), so

$$R_\chi \geq R_{\mathrm{IB}}(\vec{\epsilon}, R_{\mathrm{p}} + 2\gamma + \tilde{\delta}) + 6\frac{|\mathcal{X}|}{C_n} \log \frac{2}{C_n \cdot |\mathcal{Y}|} + \frac{3\tilde{\delta}}{2|\mathcal{X}|} \log \frac{\tilde{\delta}}{2|\mathcal{X}|^2 \cdot |\mathcal{Y}|} \tag{B.118}$$

$$\geq R_{\mathrm{IB}}(\vec{\epsilon}, R_{\mathrm{p}} + 2\tilde{\delta}) + 6\frac{|\mathcal{X}|}{C_n} \log \frac{2}{C_n \cdot |\mathcal{Y}|} + \frac{3\tilde{\delta}}{2|\mathcal{X}|} \log \frac{\tilde{\delta}}{2|\mathcal{X}|^2 \cdot |\mathcal{Y}|}\ , \tag{B.119}$$

$$\geq R_{\mathrm{IB}}(\vec{\epsilon}, R_{\mathrm{p}}) + 6\frac{|\mathcal{X}|}{C_n} \log \frac{2}{C_n \cdot |\mathcal{Y}|} + \frac{3\tilde{\delta}}{2|\mathcal{X}|} \log \frac{\tilde{\delta}}{2|\mathcal{X}|^2 \cdot |\mathcal{Y}|} + 12\tilde{\delta} \log \frac{4\tilde{\delta}}{|\mathcal{X}| \cdot |\mathcal{Y}|}\ , \tag{B.120}$$

where (B.119) follows by our assumption about γ and (B.120) from Lemma A.13 given our assumption about $\tilde{\delta}$. Combining (B.117) with (B.120), our assumptions

about (C_n, κ_n, K_n) imply that for large enough n, the rate

$$R_{\mathrm{IB}}(\vec{\epsilon}, R_{\mathrm{p}}) - \frac{\delta}{2} - \left(\tilde{\delta} + \frac{3\tilde{\delta}}{2|\mathcal{X}|} \log \frac{2|\mathcal{X}|^2 \cdot |\mathcal{Y}|}{\tilde{\delta}} + 12\tilde{\delta} \log \frac{|\mathcal{X}| \cdot |\mathcal{Y}|}{4\tilde{\delta}} \right) \tag{B.121}$$

is achievable for the cognitive radio with probability going to 1 as $n \to \infty$. One can now observe that the parenthetical term in (B.121) vanishes as $\tilde{\delta}$ goes to 0, so choosing $\tilde{\delta} \in (0, 1/8)$ such that this parenthetical term is less than $\frac{\delta}{2}$ completes the proof. □

Appendix C

Proofs for Chaper 3

C.1 Proof of Theorem 3.1 (Achievability)

The strategy we outline is reminiscent of an achievable strategy of Viswanathan [89]. Because the encoder and decoder both have access to the channel state process $\{S_i\}$, they can synchronize their transmissions according to the state to create $|\mathcal{S}|$ parallel channels. Let $p(x|s)$ be a transition probability function for $\mathcal{X}$ given $\mathcal{S}$ and $\pi(s)$ the stationary distribution on s determined by $p(x|s)$. If the state process $\{S_k\}_{k\geq 1}$ is irreducible, then by Definition 3.3, the existence of the stationary distribution is guaranteed for any choice of $p(x|s)$ since irreducible finite state Markov chains are positive recurrent. We say that $p(x|s) \in \mathcal{D}_{\mathsf{ach}}$ if it satisfies the following constraint:

$$\sum_{x,s} p(x|s)\pi(s)\Gamma(s,x) \leq \alpha \ . \tag{C.1}$$

We can construct a random coding strategy as follows.

1. *Codebook Generation*: Fix $\min_s \pi(s) > \delta > 0$ (Note: positive recurrence implies $\pi(s) > 0$, so such a δ is guaranteed to exist). Select a $p(x|s) \in \mathcal{D}_{\mathsf{ach}}$ and define

$n_s = n \cdot (\pi(s) - \delta)$. For each state $s \in \mathcal{S}$, construct an $(n_s, R_s, p(x|s))$ random codebook with a jointly δ'-typical decoder[1], where $R_s = |I(X;Y|S=s) - \delta|_+$, where $|\cdot|_+ = \max\{\cdot, 0\}$. We will ignore integer effects and assume that $2^{n_s R_s}$ is an integer.

2. *Encoding*: Let the message $m \in \{1, \ldots, 2^{nR}\}$, where R is defined by the equation: $nR = \sum_s n_s R_s$. Split m into components $(m_1, m_2, \ldots, m_{|\mathcal{S}|})$, where $m_s \in \{1, \ldots, 2^{n_s R_s}\}$. When $S_i = s$, the input X_i is given by the next symbol in codebook s for message m_s that has yet to be transmitted. If all symbols have been transmitted in codebook s, generate a random input according to the distribution $p(x|s)$.

3. *Decoding*: Divide the outputs according to the channel state. For each state s, decode $\hat{m}_s$ using a maximum likelihood decoder based on the first n_s channel outputs Y_i for which $S_i = s$. If there are less than n_s channel outputs for state s, declare an error. Reconstruct message $\hat{m}$ from $(\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_{|\mathcal{S}|})$.

We now proceed to bound the error of this coding strategy. We first use this random coding strategy to prove the existence of a coding strategy with a low probability of certain error events, including decoding error. We then show that such a coding strategy satisfies the cost constraint.

Random Coding Error Analysis.

There are two sources of error that motivate our error events: a decoding error and violating the cost constraint. A decoding error occurs if there are too few channel outputs for a given state, or, given this does not happen, the decoder selects the wrong message. By defining $N_s = |\{i : S_i = s\}|$, we can formally characterize these

[1]For a clarification on the meaning of a random codebook with jointly typical decoding, see Definition A.12 on page 104.

events as follows:

$$E_{\mathsf{a}} = \{\exists\ s, N_s < n_s\}\ ,$$
$$E_{\mathsf{b}} = \{\exists\ s, m_s \neq \hat{m}_s\}\ .$$

Note that the cost depends directly on the channel inputs and state. Thus, we can violate the cost constraint if a given channel state occurs either too often or rarely relative to its expectation, or if the codewords are not typical. We can define the error event as follows:

$$E_{\mathsf{c}} = \{\exists\ s, |N_s - n\pi(s)| \geq n\delta\}$$
$$E_{\mathsf{d}} = \{\exists\ s, X^{N_s}(s) \notin \mathcal{T}_\delta^{*(N_s)}\}\ ,$$

where $X^{N_s}(s)$ is shorthand for the subsequence X_i for which $S_i = s$. We will now show there exists a sequence of codebooks for which these error probabilities are small with increasing blocklength, and then show that this is sufficient to satisfy the conditions of achievability in Definition 3.2.

First, let us rewrite the event E_{a} as

$$\begin{aligned} E_{\mathsf{a}} &= \{\exists\ s, N_s < n\pi(s) - n\delta\} \\ &= \{\exists\ s, N_s - n\pi(s) < -n\delta\}\ . \end{aligned}$$

Thus, $E_{\mathsf{c}} \supseteq E_{\mathsf{a}}$, so it is sufficient to ignore the probability of E_{a} and consider only E_{c}.

Now consider the event of a decoding failure E_{b} given each state occurs sufficiently often to guarantee that its corresponding codeword is sent. Then, a bound on the error for the random codebook (Lemma A.6) for each state and a union bound over

all states imply that for sufficiently large n,

$$\mathbb{P}(E_{\mathsf{b}}|E_{\mathsf{a}}^c) \leq |\mathcal{S}| \cdot \delta \ . \tag{C.2}$$

Note that for sufficiently large n, the fraction of times each state appears converges to its stationary distribution (Lemma A.11). Thus, a union bound over all states implies that for $\epsilon = \delta \cdot |\mathcal{X}| > 0$, there exists sufficiently large n such that

$$\mathbb{P}(E_{\mathsf{c}}) \leq |\mathcal{S}| \cdot |\mathcal{X}| \cdot \delta \ . \tag{C.3}$$

Thus, we can show that for all $\delta > 0$, there exists sufficiently large n such that for the random codebook,

$$\begin{aligned} \mathbb{P}(E_{\mathsf{a}} \cup E_{\mathsf{b}} \cup E_{\mathsf{c}}) &\leq \mathbb{P}(E_{\mathsf{c}}) + \mathbb{P}(E_{\mathsf{b}}|E_{\mathsf{a}}^c) & \text{(C.4)} \\ &\leq |\mathcal{S}| \cdot (|\mathcal{X}| \cdot \delta + \delta) \ . & \text{(C.5)} \end{aligned}$$

Finally, we want to establish that every sub-sequence $X^{N_s}(s)$ is typical. By the AEP (Lemma A.3) and a union bound over all states, we have that for sufficiently large n,

$$\mathbb{P}(E_{\mathsf{d}}|E_{\mathsf{a}}^c) \leq |\mathcal{S}| \cdot \delta \ ,$$

so for sufficiently large n,

$$\mathbb{P}(E_{\mathsf{a}} \cup E_{\mathsf{b}} \cup E_{\mathsf{c}} \cup E_{\mathsf{d}}) \leq |\mathcal{S}| \cdot (|\mathcal{X}| \cdot \delta + 2\delta) \ . \tag{C.6}$$

Since this is true for a random coding scheme, there exists a deterministic coding scheme with the same average error probability. Furthermore, at least half of the codewords in such a scheme have no more than twice the error probability $\eta(\delta) =$

$2|\mathcal{S}| \cdot (2 + |\mathcal{X}|) \cdot \delta$, which goes to 0 as $\delta \to 0$.

Coding Strategy Satisfies Cost Constraint.

For this deterministic codebook at such n, we can bound the expected cost as follows:

$$
\begin{aligned}
& n^{-1} \sum_{k=1}^{n} E[\Gamma(S_k, X_k) | M = m, S_1 = s'] \\
& \qquad \leq \sum_{x,s} \Gamma(s,x)(p(x|s) + \frac{\delta}{|\mathcal{X}|})(\pi(s) + \delta) + \max_{s',x'} \Gamma(s',x') \cdot \eta(\delta) && \text{(C.7)} \\
& \qquad \leq \alpha + \max_{s',x'} \Gamma(s',x') \cdot (\eta(\delta) + |\mathcal{S}| \cdot |\mathcal{X}| \cdot \delta + |\mathcal{S}| \cdot \delta^2 + \delta) && \text{(C.8)} \\
& \qquad \leq \alpha + \gamma(\delta) \ , && \text{(C.9)}
\end{aligned}
$$

where $\gamma(\delta) = \max_{s',x'} \Gamma(s',x') \cdot (\eta(\delta) + |\mathcal{S}| \cdot |\mathcal{X}| \cdot \delta + |\mathcal{S}| \cdot \delta^2 + \delta)$. Note that as $\delta \to 0$, $\gamma \to 0$. Thus, we have shown the existence of a coding strategy that achieves the rates guaranteed in the statement of the theorem, thereby completing the proof.

C.2 Proof of Theorem 3.2 (Converse)

Key Lemmas.

Lemmas C.1, C.2, and C.3, the three lemmas proved in this subsection, will enable us to prove the main theorem. Lemma C.1 uses Fano's inequality and the Lagrange dual to give an upper bound on the capacity. Lemmas C.2 and C.3 then recast this upper bound as a control problem (see Section A.4), where Lemma C.2 converts the upper bound into one of maximizing the average reward, and Lemma C.3 gives a simpler characterization of the maximum average reward via Bellman's equation (Theorem A.9).

Lemma C.1. *For all $\lambda \geq 0$, the capacity can be bounded as follows:*

$$C(\alpha) \leq \sup_n \max_{\prod_{i=1}^n p(X_i|S^i,X^{i-1})} \min_s R_n(\alpha, \lambda, s) \ , \tag{C.10}$$

where

$$\begin{aligned} n \cdot R_n(\alpha, \lambda, s) = I(X_1; Y_1|S_1 = s) + \sum_{i=2}^{n} I(X_i; Y_i|S_i) \\ + \lambda \left(n\alpha - \sum_{i=1}^{n} \mathbb{E}\left[\Gamma(S_i, X_i)|S_1 = s\right] \right) \ . \end{aligned} \tag{C.11}$$

Proof. Suppose a rate R is achievable and M is uniformly distributed over the set of messages. By the definition of achievable rate, for all $\epsilon > 0$ and sufficiently large n, Fano's inequality implies the following:

$$\begin{aligned} nR &= H(M) \\ &\leq I(Y^n, S^n; M) + n\epsilon \ . \end{aligned} \tag{C.12}$$

We now isolate $I(Y^n, S^n; M)$ in (C.12) to get the following, for $S_1 = s$:

$$\begin{aligned}
I(Y^n, S^n; M) &\leq I(Y^n, S^{n+1}; M) \\
&\leq \sum_{i=1}^{n} I(Y_i, S_{i+1}; M | Y^{i-1}, S^i) && \text{(C.13)} \\
&= \sum_{i=1}^{n} H(Y_i, S_{i+1} | Y^{i-1}, S^i) - H(Y_i, S_{i+1} | M, Y^{i-1}, S^i) \\
&\leq \sum_{i=1}^{n} H(Y_i, S_{i+1} | S_i) - H(Y_i, S_{i+1} | M, Y^{i-1}, S^i, X_i(M, S^i)) && \text{(C.14)} \\
&= \sum_{i=1}^{n} H(Y_i, S_{i+1} | S_i) - H(Y_i, S_{i+1} | S_i, X_i) && \text{(C.15)} \\
&= \sum_{i=1}^{n} I(X_i; Y_i, S_{i+1} | S_i) \\
&= \sum_{i=1}^{n} I(X_i; Y_i | S_i) + I(X_i; S_{i+1} | Y_i, S_i) \\
&= \sum_{i=1}^{n} I(X_i; Y_i | S_i) \ , && \text{(C.16)}
\end{aligned}$$

where (C.13) follows from the chain rule and since S_1 and M are independent; (C.14) since conditioning cannot increase entropy; (C.15) from the Markov chain $(Y_i, S_{i+1}) - (X_i, S_i) - X_i, M, Y^{i-1}, S^i$; and (C.16) from the Markov chain $S_{i+1} - (Y_i, S_i) - X_i$. Note that this bound holds for all $s \in \mathcal{S}$ such that $S_1 = s$, so minimizing over all initial states gives the following:

$$nR \leq \min_s \left\{ I(X_1; Y_1 | S_1 = s) + \sum_{i=2}^{n} I(X_i; Y_i | S_i) \right\} + n\epsilon \ , \qquad \text{(C.17)}$$

Furthermore, by the definition of an achievable rate, we have an additional cost constraint, so for all $\epsilon > 0$, $s \in \mathcal{S}$, and sufficiently large n, the constraint can be

expressed as follows:

$$\sum_{i=1}^{n} \mathbb{E}\left[\Gamma(S_i, X_i)|S_1 = s\right] \leq n(\alpha + \epsilon) \ . \tag{C.18}$$

Thus, by combining (C.17), and (C.18) we can formulate the problem as follows:

$$nR \leq \max_{\prod_{i=1}^{n} p(X_i|S^i, X^{i-1}) \in \mathcal{D}_\epsilon} \left\{ \min_s \left\{ I(X_1; Y_1|S_1 = s) + \sum_{i=2}^{n} I(X_i; Y_i|S_i) \right\} \right\} + n\epsilon \ , \tag{C.19}$$

where $\mathcal{D}_\epsilon$ is the set of transition probabilities satisfying (C.18) for all $s \in \mathcal{S}$. We can replace this constraint by a Lagrange multiplier, which gives an upper bound on the value on the right side (c.f. the Lagrange dual in [8, p. 216]), so for all $\lambda \geq 0$:

$$nR \leq \max_{\prod_{i=1}^{n} p(X_i|S^i, X^{i-1})} \left\{ \min_s \left\{ n \cdot R_n(\alpha + \epsilon, \lambda, s) \right\} \right\} + n\epsilon \ , \tag{C.20}$$

and since this holds for all $\epsilon > 0$ and sufficiently large n, we can conclude the following:

$$R \leq \sup_n \max_{\prod_{i=1}^{n} p(X_i|S^i, X^{i-1})} \min_s R_n(\alpha, \lambda, s) \ . \tag{C.21}$$

□

Lemma C.2. *Let $R_n(\alpha, \lambda, s)$ be defined as in Lemma C.1. Then the following inequality holds:*

$$\sup_n \max_{\prod_{i=1}^{n} p(X_i|S^i, X^{i-1})} \min_s R_n(\alpha, \lambda, s) \leq \max_{\prod_{i=1}^{\infty} p(X_i|S^i)} \liminf_n \min_s R_n(\alpha, \lambda, s) \ . \tag{C.22}$$

Proof. The proof is divided into two parts. In the first part, we will use Fekete's

lemma to show that

$$\sup_n \max_{\prod_{i=1}^n p(X_i|S^i,X^{i-1})} \min_s R_n(\alpha,\lambda,s) \leq \max_{\prod_{i=1}^\infty p(X_i|S^i,X^{i-1})} \liminf_n \min_s R_n(\alpha,\lambda,s) \quad \text{(C.23)}$$

and in the second part, we will use an induction argument to show that

$$\max_{\prod_{i=1}^\infty p(X_i|S^i,X^{i-1})} \liminf_n \min_s R_n(\alpha,\lambda,s) = \max_{\prod_{i=1}^\infty p(X_i|S^i)} \liminf_n \min_s R_n(\alpha,\lambda,s) \ , \quad \text{(C.24)}$$

thereby concluding the result. To begin the first part of the proof, we define $C_n(\alpha,\lambda)$ as

$$C_n(\alpha,\lambda) = \max_{\prod_{i=1}^n p(X_i|S^i,X^{i-1})} \min_s R_n(\alpha,\lambda,s). \quad \text{(C.25)}$$

Let $k+\ell = n$ for nonnegative integers k,ℓ. Suppose we consider distributions

$$\prod_{i=1}^k p(X_i|S^i,X^{i-1})$$

and $\prod_{i=1}^\ell p(X_{k+i}|S_{k+1}^{k+i},X_{k+1}^{k+i-1})$ that achieve $C_k(\alpha,\lambda)$ and $C_\ell(\alpha,\lambda)$, respectively. Then

for $R_n(\alpha, \lambda)$ with these choices of distribution,

$$\begin{aligned}
&n \cdot C_n(\alpha, \lambda) \\
&\quad \geq \min_s R_n(\alpha, \lambda, s) \\
&\quad = k \cdot C_k(\alpha, \lambda) + \min_s \sum_{i=1}^{\ell} I(X_{k+i}; Y_{k+i} | S_{k+i}) + \lambda \left(\ell\alpha - \textstyle\sum_{i=1}^{\ell} \mathbb{E}[\Gamma(S_{k+i}, X_{k+i}) | S_1 = s] \right) \\
&\quad \geq k \cdot C_k(\alpha, \lambda) + \min_{\tilde{s}} I(X_{k+1}; Y_{k+1} | S_{k+1} = \tilde{s}) \\
&\qquad\qquad + \sum_{i=2}^{\ell} I(X_{k+i}; Y_{k+i} | S_{k+i}) + \lambda \left(\ell\alpha - \textstyle\sum_{i=1}^{\ell} \mathbb{E}[\Gamma(S_{k+i}, X_{k+i}) | S_{k+1} = \tilde{s}] \right) \\
&\quad = k \cdot C_k(\alpha, \lambda) + \ell \cdot C_\ell(\alpha, \lambda) \ . \qquad \text{(C.26)}
\end{aligned}$$

Thus, the sequence $n \cdot C_n(\alpha, \lambda)$ is superadditive, and we can conclude by Fekete's Lemma (see e.g. [38, p. 112, Lemma 4A.2]) that $\lim_{n\to\infty} C_n(\alpha, \lambda)$ exists and, a fortiori, $\lim_{n\to\infty} C_n(\alpha, \lambda) = \sup_n C_n(\alpha, \lambda)$. Thus,

$$\begin{aligned}
&\lim_{n\to\infty} C_n(\alpha, \lambda) \\
&\qquad = \sup_n C_n(\alpha, \lambda) \\
&\qquad = \sup_n \max_{\prod_{i=1}^n p(X_i | S^i, X^{i-1})} \min_s R_n(\alpha, \lambda, s) \\
&\qquad = \max_{\prod_{i=1}^\infty p(X_i | S^i, X^{i-1})} \sup_n \min_s R_n(\alpha, \lambda, s) \\
&\qquad \geq \max_{\prod_{i=1}^\infty p(X_i | S^i, X^{i-1})} \liminf_n \min_s R_n(\alpha, \lambda, s) \ . \qquad \text{(C.27)}
\end{aligned}$$

From the definition of limit and the above, for all $\epsilon > 0$, there exists $N(\epsilon)$ such that for $n \geq N(\epsilon)$, $C_n(\alpha, \lambda) \geq \sup_n C_n(\alpha, \lambda) - \epsilon$. Let p^* be the maximizing input distribution

for $C_{N(\epsilon)}(\alpha, \lambda)$, and for all nonnegative integers k, choose the input distribution to be

$$\prod_{i=1}^{N(\epsilon)} p(X_{k\cdot N(\epsilon)+i} | S_{k\cdot N(\epsilon)+1}^{k\cdot N(\epsilon)+i}, X_{k\cdot N(\epsilon)+1}^{k\cdot N(\epsilon)+i-1}) = p^*.$$

Then for this choice of we get that

$$\begin{aligned} \max_{\prod_{i=1}^{\infty} p(X_i|S^i, X^{i-1})} \liminf_n \min_s R_n(\alpha, \lambda, s) \\ &\geq C_{N(\epsilon)}(\alpha, \lambda) \\ &\geq \sup_n C_n(\alpha, \lambda) - \epsilon \ . \end{aligned} \tag{C.28}$$

Thus, for all $\epsilon > 0$ and n sufficiently large,

$$\sup_n C_n(\alpha, \lambda) \leq \max_{\prod_{i=1}^{\infty} p(X_i|S^i, X^{i-1})} \liminf_n \min_s R_n(\alpha, \lambda, s) + \epsilon \ . \tag{C.29}$$

Since we can make ϵ arbitrarily small, we get the following:

$$\sup_n C_n(\alpha, \lambda) \leq \max_{\prod_{i=1}^{\infty} p(X_i|S^i, X^{i-1})} \liminf_n \min_s R_n(\alpha, \lambda, s) \ , \tag{C.30}$$

which completes the first part of the proof.

To show the second part of the proof, let us define the following quantities:

$$\begin{aligned} J_k &= \max_{\prod_{i=1}^{k} p(X_i|S^i) \prod_{\ell=k+1}^{\infty} p(X_\ell|S^\ell, X^{\ell-1})} \liminf_n \min_s R_n(\alpha, \lambda, s) \\ J_\infty &= \max_{\prod_{i=1}^{\infty} p(X_i|S^i)} \liminf_n \min_s R_n(\alpha, \lambda, s) \end{aligned}$$

Note that from the first part of the proof, we have shown that

$$\sup_n \max_{\prod_{i=1}^{n} p(X_i|S^i, X^{i-1})} \min_s R_n(\alpha, \lambda, s) \leq J_1.$$

Furthermore, for all $k \geq 1$, $J_k \geq J_{k+1}$. Thus, to show the second part of the proof, is equivalent to proving that

$$J_\infty = J_1 \ . \tag{C.31}$$

We proceed by induction and argue that $J_k = J_1$ for all $k \geq 1$. The base case is trivially satisfied for $k = 1$. Suppose that $J_k = J_1$ for some $k \geq 1$. We will now show that $J_k = J_{k+1}$. Since $J_k \geq J_{k+1}$, it suffices to show that $J_k \leq J_{k+1}$. First observe that

$$\begin{aligned} &I(X_{k+1}; Y_{k+1}|S_{k+1}) - \lambda \mathbb{E}[\Gamma(S_{k+1}, X_{k+1})] \\ &\quad = \mathbb{E}\left[\mathbb{E}\left[\log \frac{w(Y_{k+1}|X_{k+1}, S_{k+1})}{\mathbb{E}[w(Y_{k+1}|X_{k+1}, S_{k+1})|Y_{k+1}]} - \lambda \Gamma(S_{k+1}, X_{k+1}) \middle| X_{k+1}, S^{k+1}\right]\right] \end{aligned} \tag{C.32}$$

Thus, under the optimizing distribution p^* for J_k, if we replace $p^*(X_{k+1} = x_{k+1}|S^{k+1} = s^{k+1}, X^k = x^k)$ with $\mathbb{E}[p^*(X_{k+1}|S^{k+1}, X^k)|X_{k+1} = x_{k+1}, S_{k+1} = s_{k+1}]$, which is in the set of valid distributions for J_{k+1}, the quantity

$$I(X_{k+1}; Y_{k+1}|S_{k+1}) + \lambda \left(\alpha - \mathbb{E}[\Gamma(S_{k+1}, X_{k+1})|S_1 = s]\right) \tag{C.33}$$

remains the same. Furthermore,

$$\sum_{i=1}^{k} I(X_i; Y_i|S_i) + \lambda \left(n\alpha - \textstyle\sum_{i=1}^{k} \mathbb{E}[\Gamma(S_i, X_i)|S_1 = s]\right) \tag{C.34}$$

remains the same because we kept the previous p^* unchanged. Finally, the distribution

of $\mathbb{P}(S_{k+2})$ remains the same by linearity of expectation, so for $\ell \geq k+1$,

$$I(X_\ell; Y_\ell | S_\ell) + \lambda \left(\alpha - \mathbb{E}[\Gamma(S_\ell, X_\ell) | S_1 = s]\right) \tag{C.35}$$

remains the same because we have not changed future p^*, and we can conclude that under this modification to p^*,

$$J_k = \liminf_n \min_s R_n(\alpha, \lambda, s) \ . \tag{C.36}$$

However, since this is a valid distribution for J_{k+1}, we have that $J_k \leq J_{k+1}$ and thus $J_k = J_{k+1}$. The result follows immediately. □

Lemma C.3. *If there exist* $|\mathcal{S}| + 1$ *real numbers*

$$(J^*_{\lambda,\alpha}, \ell(1), \ldots, \ell(|\mathcal{S}|))$$

such that for all $i \in \mathcal{S}$*,*

$$J^*_{\lambda,\alpha} + \ell(i) = \max_{p(x|i)} \left\{ I(X;Y|S=i) + \lambda(\alpha - \mathbb{E}\left[\Gamma(S,X)|S=i\right]) + \sum_{j=1}^{\mathcal{S}} P^p_{i,j} \cdot \ell(j) \right\} \ , \tag{C.37}$$

then

$$C(\alpha) \leq J^*_{\lambda,\alpha} \ . \tag{C.38}$$

Furthermore, if the state process $\{S_k\}_{k\geq 1}$ *is irreducible, then there exists a solution to* (C.37)*, and* (C.38) *holds.*

Proof. From Lemmas C.1 and C.2, we have the following upper bound to capacity:

$$C(\alpha) \leq \max_{\prod_{i=1}^{\infty} p(X_i|S^i)} \liminf_n \min_s R_n(\alpha, \lambda, s) \ , \tag{C.39}$$

where the definition of $R_n(\alpha, \lambda)$ is given in Lemma C.1. We want to apply Theorem A.9 to simplify the upper bound, which first requires that we provide an explicit connection to the control problem on page 106, specifically to items 1), 2), and 3). Item 1) is given in Definition 3.3 on page 3.3. That is, the transition probability matrix $[P^u_{i,j}]$, where the entry on row $i \in \mathcal{S}$ and column $j \in \mathcal{S}$ is defined as

$$P^u_{i,j} = \sum_{\substack{x \in \mathcal{X} \\ y \in \mathcal{Y}}} q(j|y,i) \cdot w(y|x,i) \cdot u(x) \ ,$$

where the control $u \in \mathcal{U}$, and the control space $\mathcal{U}$ is the probability simplex of $\Re^{|\mathcal{X}|}$.

Now, for item 2), the control law is

$$U_k = g_k(S^k) = \left[\ p(X_k = \cdot|S^k) \ \right] \in \mathcal{U} \ . \tag{C.40}$$

For item 3), the reward function $r : \mathcal{S} \times \mathcal{U} \to \Re$ is

$$r(s,u) = \sum_{x,y} w(y|x,s) \cdot u(x) \cdot \log \frac{w(y|x,s)}{\sum_{\tilde{x}} w(y|\tilde{x},s) \cdot u(\tilde{x})} + \lambda \left(\alpha - \sum_x \Gamma(s,x) \cdot u(x) \right) \ , \tag{C.41}$$

and thus given a control sequence and an initial state,

$$\begin{aligned} J_s(g) &= \liminf_n \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[r(S_k, U_k)] \\ &= \liminf_n R_n(\alpha, \lambda, s) \ . \end{aligned}$$

We have now established the necessary connection to the control problem to apply Theorem A.9, which states that if there exist $|\mathcal{S}| + 1$ real numbers

$$(J^*_{\lambda,\alpha}, \ell(1), \ldots, \ell(|\mathcal{S}|))$$

such that, for all $i \in \mathcal{S}$,

$$J^*_{\lambda,\alpha} + \ell(i) = \max_{u \in \mathcal{U}} \left\{ r(i,u) + \sum_{j=1}^{\mathcal{S}} P^u_{i,j} \cdot \ell(j) \right\} \ , \tag{C.42}$$

then

$$J^*_{\lambda,\alpha} = \max_{\prod_{i=1}^{\infty} p(X_i|S^i)} \liminf_n R_n(\alpha, \lambda, s) \ ,$$

and it further states that such a solution exists if the state process $\{S_k\}_{k \geq 1}$ is irreducible (see Definition 3.3). Since $J^*_{\lambda,\alpha} = \max J_s(g)$ for all s,

$$J^*_{\lambda,\alpha} = \max_{\prod_{i=1}^{\infty} p(X_i|S^i)} \liminf_n \min_s R_n(\alpha, \lambda, s) \ .$$

Thus, by (C.39), we have that

$$C(\alpha) \leq J^*_{\lambda,\alpha} \ , \tag{C.43}$$

thereby completing the proof. □

Proof of Main Theorem.

We are now in a position to prove Theorem 3.2. Observe that Lemma C.3 already gives the upper bound to capacity, which comprises the first part of the result. To show equality, we start with the upper bound in Lemma C.3 and show it matches the in achievable rates from Theorem 3.1 under our assumptions.

Let $\pi^*(\cdot)$ be the stationary distribution of the maximizing $p^*(x|\cdot)$ in the right side of (C.37) over all $i \in \{1, \ldots, |\mathcal{S}|\}$. By Lemma C.3, (C.37) implies the following:

$$\begin{aligned} J^*_{\lambda,\alpha} + \sum_s \ell(s) \cdot \pi^*(s) = \sum_s I(X;Y|S=s)\pi^*(s) \\ + \lambda(\alpha - \sum_s \mathbb{E}\left[\Gamma(S,X)|S=s\right]\pi(s)) \\ + \sum_j \ell(j) \cdot \pi^*(j) \ , \end{aligned} \tag{C.44}$$

which follows since under p^*, $\sum_{i=1}^{\mathcal{S}} P^p_{i,j} \cdot \pi^*(i) = \pi^*(j)$. This simplifies under our assumptions that either $(\lambda \geq 0, \alpha = \alpha^*_\lambda)$ or $(\lambda = 0, \alpha \geq \alpha^*_\lambda)$, where α^*_λ is defined in (3.9):

$$J^*_{\lambda,\alpha} + \sum_s \ell(s) \cdot \pi^*(s) = \sum_s I(X;Y|S=s)\pi^*(s) + \sum_j \ell(j) \cdot \pi^*(j) \ ,$$

so we can solve for $J^*_{\lambda,\alpha}$ to get the following:

$$J^*_{\lambda,\alpha} = \sum_s I(X;Y|S=s)\pi^*(s) \ , \tag{C.45}$$

where our assumptions further imply that

$$\sum_s \mathbb{E}\left[\Gamma(S,X)|S=s\right]\pi(s) \leq \alpha.$$

Thus, by Theorem 3.1 and Lemma C.3, our assumptions imply the following:

$$\begin{aligned}
&\max_{\substack{p(x|s):\\ E[\Gamma(S,X)]\leq\alpha}} \sum_s I(X;Y|S=s)\pi(s) \\
&\qquad \leq C(\alpha) && \text{(C.46)}\\
&\qquad \leq J^*_{\lambda,\alpha} && \text{(C.47)}\\
&\qquad = \sum_s I(X;Y|S=s)\pi^*(s) && \text{(C.48)}\\
&\qquad \leq \max_{\substack{p(x|s):\\ E[\Gamma(S,X)]\leq\alpha}} \sum_s I(X;Y|S=s)\pi(s)\ , && \text{(C.49)}
\end{aligned}$$

thereby proving the result.

C.3 Proof of Examples

C.3.1 Proof of Proposition 3.3

Since the achievable strategy shows the rate $\log\phi$ is achievable, where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio and satisfies $\phi^{-1} = \phi - 1$. it remains to show the converse. To do this, we will find solutions for the Bellman equation in Theorem 3.2 for $\lambda \geq 0$, and show that as $\lambda \to \infty$, these bounds converge to $\log\phi$.

Thus, we must find $(J^*_{\lambda,0}, \ell(0), \ell(1))$ that satisfy the following equations:

$$\begin{aligned}
J^*_{\lambda,0} + \ell(0) &= \max_{p_0} \{h_b(p_0) - \lambda\cdot 0 + p_0\cdot(\ell(1)-\ell(0)) + \ell(0)\} && \text{(C.50)}\\
J^*_{\lambda,0} + \ell(1) &= \max_{p_1} \{h_b(p_1) - \lambda\cdot p_1 + p_1\cdot(\ell(1)-\ell(0)) + \ell(0)\}\ . && \text{(C.51)}
\end{aligned}$$

Since the functions to be maximized are concave, by setting the derivative of each of the right sides above with respect to p_i to 0 and solving for p_i, the optimal choices

are the following:

$$p_0^* = \frac{\exp\{\ell(1) - \ell(0)\}}{1 + \exp\{\ell(1) - \ell(0)\}} \tag{C.52}$$

$$p_1^* = \frac{\exp\{\ell(0) - \ell(1) - \lambda\}}{1 + \exp\{\ell(0) - \ell(1) - \lambda\}} \ , \tag{C.53}$$

and the two expressions for $J^*_{\lambda,0}$ simplify as follows:

$$J^*_{\lambda,0} = \log(1 + \exp\{\ell(1) - \ell(0)\}) \tag{C.54}$$

$$J^*_{\lambda,0} = \ell(0) - \ell(1) + \log(1 + \exp\{\ell(1) - \ell(0) - \lambda\}) \ . \tag{C.55}$$

For simplicity, we will define $\varphi = \exp\{\ell(0) - \ell(1)\}$ to get the following:

$$J^*_{\lambda,0} = \log(1 + \varphi^{-1}) \tag{C.56}$$

$$J^*_{\lambda,0} = \log \varphi + \log(1 + \varphi^{-1} \cdot \exp\{-\lambda\}) \tag{C.57}$$

$$= \log(\varphi + \exp\{-\lambda\}) \ . \tag{C.58}$$

Note that if we can find $(\varphi, J^*_{\lambda,0})$, then we can find $(J^*_{\lambda,0}, \ell(0), \ell(1))$ that satisfy the Bellman equation, thereby giving us the upper bound $J^*_{\lambda,0}$ to the capacity via Theorem 3.2. By setting the two equations above to be equal and exponentiating both sides, we get that

$$\varphi + \exp\{-\lambda\} = 1 + \varphi^{-1}. \tag{C.59}$$

Note that for $\varphi \in [1, \phi)$, $1 + \varphi^{-1} - \varphi$ is a continuous function and ranges over $(0, 1]$. Thus, for every $\varphi \in [1, \phi)$, there exists a $\lambda \geq 0$ such that it is a solution, and thus for

all $\epsilon > 0$, there exists a $\lambda' \geq 0$ such that $\phi - \epsilon$ is a solution, and thus

$$C(0) \leq J^*_{\lambda',0} \tag{C.60}$$

$$= \log\left(1 + \frac{1}{\phi - \epsilon}\right) \tag{C.61}$$

$$= \log\phi + \log\frac{1 + \frac{1}{\phi-\epsilon}}{\phi} \tag{C.62}$$

$$= \log\phi + \log\frac{1 + \frac{1}{\phi-\epsilon}}{1 + \frac{1}{\phi}} \ . \tag{C.63}$$

Since ϵ can be made arbitrarily small, we have our converse.

C.3.2 Proof of Proposition 3.4

By Theorem 3.2, we know that $C(\alpha) = J^*_{\lambda,\alpha}$ if λ satisfies either $(\lambda \geq 0, \alpha = \alpha^*_\lambda)$ or $(\lambda = 0, \alpha \geq \alpha^*_\lambda)$, where α^*_λ is defined in (3.9). $J^*_{\lambda,\alpha}$ satisfies the following:

$$J^*_{\lambda,\alpha} = \max_{0\leq p_1\leq 1}\{h_b(p_1) + \lambda\alpha + \bar{\epsilon}(\ell(1) - \ell(0))\} \tag{C.64}$$

$$J^*_{\lambda,\alpha} = \max_{0\leq p_2\leq 1}\{h_b(p_2) + \lambda(\alpha - 1) + \beta(\ell(0) - \ell(1))\}$$

$$= \log 2 + \lambda(\alpha - 1) + \beta(\ell(0) - \ell(1)) \ , \tag{C.65}$$

where $\bar{\epsilon} = \bar{\epsilon}(p_1) = (1 - p_1)\cdot\epsilon_0 + p_1\cdot\epsilon_1$. Since the right side of (C.64) is convex-$\cap$ and continuous in p_1, by taking the derivative with respect to p_1, it is straightforward to show that the optimizing p^*_1 satisfies the following condition:

$$\ell(1) - \ell(0) = \frac{1}{\epsilon_1 - \epsilon_0}\cdot\log\frac{p^*_1}{1 - p^*_1} \ . \tag{C.66}$$

Note that by scaling (C.64) and (C.65) by $\frac{\beta}{\bar{\epsilon}^*+\beta}$ and $\frac{\bar{\epsilon}^*}{\bar{\epsilon}^*+\beta}$ and adding them, we get the following:

$$J^*_{\lambda,\alpha} = \frac{\bar{\epsilon}^*}{\bar{\epsilon}^*+\beta} \cdot \log 2 + \frac{\beta}{\bar{\epsilon}^*+\beta} \cdot h_b(p_1^*) + \lambda \left(\alpha - \frac{\bar{\epsilon}^*}{\bar{\epsilon}^*+\beta} \right) , \tag{C.67}$$

where $\bar{\epsilon}^* = \bar{\epsilon}(p_1^*)$. If $\frac{\bar{\epsilon}^*}{\bar{\epsilon}^*+\beta} = \alpha$, then we can solve for λ in (C.64) by substituting J^* for the right side of (C.67) to get that

$$\lambda = \log 2 - h_b(p_1^*) - \frac{\bar{\epsilon}^*+\beta}{\epsilon_1-\epsilon_0} \cdot \log \frac{p_1^*}{1-p_1^*} , \tag{C.68}$$

Note that $\lambda \geq 0$ and monotonically increasing over $p_1^* \in (0, 1/2)$. Thus, for $0 < p_1^* < \frac{1}{2}$, we have the following:

$$C\left(\frac{\bar{\epsilon}^*}{\bar{\epsilon}^*+\beta} \right) = \frac{\bar{\epsilon}^*}{\bar{\epsilon}^*+\beta} \cdot \log 2 + \frac{\beta}{\bar{\epsilon}^*+\beta} \cdot h_b(p_1^*) . \tag{C.69}$$

C.3.3 Proof of Proposition 3.5

By Theorem 3.2, we know that $C(\alpha) = J^*_{\lambda,\alpha}$ if λ satisfies either $(\lambda \geq 0, \alpha = \alpha^*_\lambda)$ or $(\lambda = 0, \alpha \geq \alpha^*_\lambda)$, where α^*_λ is defined in (3.9). $J^*_{\lambda,\alpha}$ satisfies the following:

$$\begin{aligned} J^*_{\lambda,\alpha} + \ell(1) &= \max_{p(x|1)} \left\{ H(X|S=1) + \lambda\alpha + \frac{1}{2}(\ell(1)+\ell(2)) \right\} \\ &= \log(L+1) + \lambda\alpha + \frac{1}{2}(\ell(1)+\ell(2)) \qquad \text{(C.70)} \\ J^*_{\lambda,\alpha} + \ell(2) &= \max_{p(x|2)} \left\{ H(X|S=2) + \lambda(\alpha - (1-p_0)) + \frac{1+p_0}{2} \cdot \ell(1) + \frac{1-p_0}{2} \cdot \ell(2) \right\} \\ &= h_b(p_0^*) + (1-p_0^*) \cdot \log L + \lambda(\alpha - 1 + p_0^*) + \frac{1+p_0^*}{2} \cdot \ell(1) + \frac{1-p_0^*}{2} \cdot \ell(2) , \qquad \text{(C.71)} \end{aligned}$$

where $p_0 = p(x = 0|s = 2)$, and $p_0^* = \frac{\exp\{\lambda + \frac{1}{2}(\ell(1)-\ell(2))\}}{L+\exp\{\lambda+\frac{1}{2}(\ell(1)-\ell(2))\}}$ is the corresponding optimal choice. By adding (C.70) and (C.71) scaled by $\frac{1+p_0^*}{2+p_0^*}$ and $\frac{1}{2+p_0^*}$, respectively, we get the following expression after solving for $J_{\lambda,\alpha}^*$:

$$J_{\lambda,\alpha}^* = \frac{1+p_0^*}{2+p_0^*} \cdot \log(L+1) + \frac{1}{2+p_0^*} \cdot (h_b(p_0^*) + (1-p_0^*) \cdot \log L) + \lambda \left(\alpha - \frac{1-p_0^*}{2+p_0^*} \right) . \tag{C.72}$$

For $\lambda = 0$, $p_0^* = \frac{1}{L+1}$, $\ell(1) - \ell(2) = 0$, then $\alpha_\lambda^* = \frac{L}{2L+3}$ in (3.9), thereby giving for $\alpha \geq \frac{L}{2L+3}$,

$$C(\alpha) = \log(L+1) . \tag{C.73}$$

For $\alpha = \alpha_\lambda^* = \frac{1-p_0^*}{2+p_0^*}$ (i.e., $p_0^* = \frac{1-2\alpha}{1+\alpha}$), we simply have to verify that for $0 \leq \alpha \leq \frac{L}{2L+3}$, $\lambda \geq 0$. Since $p_0^* = \frac{\exp\{\lambda+\frac{1}{2}(\ell(1)-\ell(2))\}}{L+\exp\{\lambda+\frac{1}{2}(\ell(1)-\ell(2))\}}$,

$$\frac{1}{2}(\ell(1) - \ell(2)) = -\lambda + \log \frac{Lp_0^*}{1-p_0^*} . \tag{C.74}$$

Thus, by (C.70) and (C.72),

$$J_{\lambda,\alpha}^* = \log(L+1) - \log L + \lambda(\alpha+1) + \log \frac{1-p_0^*}{p_0^*} \tag{C.75}$$

$$J_{\lambda,\alpha}^* = \frac{1+p_0^*}{2+p_0^*} \cdot \log(L+1) + \frac{1}{2+p_0^*} \cdot (h_b(p_0^*) + (1-p_0^*) \cdot \log L) . \tag{C.76}$$

Combining these gives the following for $0 \leq \alpha \leq \frac{L}{2L+3}$, $p_0^* = \frac{1-2\alpha}{1+\alpha}$:

$$\lambda \cdot (1+\alpha) \cdot (2+p_0^*) = \log \frac{L^3}{L+1} + h_b(p_0^*) + (2+p_0^*) \log \frac{p_0^*}{1-p_0^*} \tag{C.77}$$

$$= \log \frac{L^3}{L+1} + 2 \log p_0^* + 3 \log \frac{1}{1-p_0^*} \tag{C.78}$$

$$\geq \log \frac{L^3}{L+1} + 2 \log \frac{1}{L+1} + 3 \log \frac{L+1}{L} \tag{C.79}$$

$$= 0 \ . \tag{C.80}$$

Thus, from (C.72), and $0 \leq \alpha \leq \frac{L}{2L+3}$,

$$C(\alpha) = \frac{2-\alpha}{3} \cdot \log(L+1) + \frac{1+\alpha}{3} \cdot h_b\left(\frac{1-2\alpha}{1+\alpha}\right) + \alpha \cdot \log L \ . \tag{C.81}$$

C.3.4 Proof of Proposition 3.6

Note that $I(X;Y|S) = (1-\epsilon)H(X|S)$. By Theorem 3.2, we know that $C(\alpha) = J^*_{\lambda,\alpha}$ if λ satisfies either $(\lambda \geq 0, \alpha = \alpha^*_\lambda)$ or $(\lambda = 0, \alpha \geq \alpha^*_\lambda)$, where α^*_λ is defined in (3.9). $J^*_{\lambda,\alpha}$ satisfies the following:

$$J^*_{\lambda,\alpha} + \ell(c) = \max_{p(x|c)} \{(1-\epsilon) \cdot H(X|S=c) + \lambda\alpha + \epsilon \cdot \ell(e) + (1-\epsilon) \cdot \ell(c)\}$$

$$= (1-\epsilon) \cdot \log(L+1) + \lambda\alpha + \epsilon \cdot \ell(e) + (1-\epsilon) \cdot \ell(c) \tag{C.82}$$

$$J^*_{\lambda,\alpha} + \ell(e) = \max_{p(x|e)} \{(1-\epsilon) \cdot H(X|S=e) + \lambda\,(\alpha - (1-p_0)) + \epsilon \cdot \ell(e) + (1-\epsilon) \cdot \ell(c)\}$$

$$= (1-\epsilon) \cdot (h_b(p_0^*) + (1-p_0^*) \cdot \log L)$$

$$+ \lambda\,(\alpha - (1-p_0^*)) + \epsilon \cdot \ell(e) + (1-\epsilon) \cdot \ell(c) \ , \tag{C.83}$$

where $p_0 = p(x=0|s=e)$, and $p_0^* = \frac{\exp\{\lambda/(1-\epsilon)\}}{L+\exp\{\lambda/(1-\epsilon)\}}$ is the corresponding optimal choice. By adding (C.82) and (C.83) scaled by $1-\epsilon$ and ϵ, respectively, we get the

following expression after solving for $J^*_{\lambda,\alpha}$:

$$J^*_{\lambda,\alpha} = (1-\epsilon)\cdot((1-\epsilon)\cdot\log(L+1) + \epsilon\cdot h_b(p_0^*) + \epsilon\cdot(1-p_0^*)\cdot\log L) + \lambda(\alpha - \epsilon\cdot(1-p_0^*)) \ . \tag{C.84}$$

For $\lambda = 0$, then $p_0^* = \frac{1}{L+1}$ and $\alpha^*_\lambda = \frac{\epsilon L}{L+1}$ in (3.9), thereby giving for $\alpha \geq \frac{L}{L+1}$,

$$C(\alpha) = (1-\epsilon)\log(L+1) \ . \tag{C.85}$$

For $\alpha = \alpha^*_\lambda = \epsilon(1-p_0^*)$ (i.e., $p_0^* = \frac{\epsilon-\alpha}{\epsilon}$), we simply have to verify that for $0 \leq \alpha \leq \frac{\epsilon L}{L+1}$, $\lambda \geq 0$. This is straightforward to verify since over $0 \leq \alpha \leq \frac{\epsilon L}{L+1}$, $p_0^* \geq \frac{1}{L+1}$, and since $p_0^* = \frac{\exp\{\lambda/(1-\epsilon)\}}{L+\exp\{\lambda/(1-\epsilon)\}}$, $\lambda \geq 0$. Thus, $0 \leq \alpha \leq \frac{\epsilon L}{L+1}$,

$$C(\alpha) = (1-\epsilon)\cdot\left((1-\epsilon)\cdot\log(L+1) + \epsilon\cdot h_b\left(\frac{\epsilon-\alpha}{\epsilon}\right) + \alpha\cdot\log L\right) \ . \tag{C.86}$$

Appendix D

Proofs for Chaper 4

D.1 Proof of Theorem 4.5 (Achievable Strategy)

In this section, we give a rigorous proof of the achievable strategy for the general case in which the sources can depend on the channel, and for notational convenience, we have the following: $S_k = S_{A,k}$, $\mathbf{Y}_k = (Y_k, S_{B,k})$, and $\mathbf{Z}_k = (Z_k, S_{E,k})$.

Recall that $\mathcal{P}_{\mathsf{joint}}$ is the set of all joint distributions p of random variables $\mathbf{V}$, $\mathbf{U}$, X, S, $\mathbf{Y}$, $\mathbf{Z}$ such that (i) the following Markov chain holds:

$$\mathbf{V} - \mathbf{U} - (X, S) - (\mathbf{Y}, \mathbf{Z}) \ ,$$

(ii) $\mathbf{V}$ is independent of S, and (iii) the joint conditional distribution of $(\mathbf{Y}, \mathbf{Z})$ given (X, S) as well as the marginal distribution of S are consistent with the given source and channel respectively.

For $p \in \mathcal{P}_{\mathsf{joint}}$, recall that $\mathcal{R}_{\mathsf{joint}}(p)$ is the set of all non-negative pairs $(R_{\mathsf{SK}}, R_{\mathsf{SM}})$

which satisfy the following two inequalities:

$$R_{\mathsf{SM}} \leq I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};S) \tag{D.1}$$

$$R_{\mathsf{SK}} + R_{\mathsf{SM}} \leq I(\mathbf{U};\mathbf{Y}|\mathbf{V}) - I(\mathbf{U};\mathbf{Z}|\mathbf{V}) \ . \tag{D.2}$$

Proving the theorem is equivalent to showing that this rate region is achievable.

Proof of Theorem 4.5.

Since a secret message automatically satisfies the constraints of a secret key, it is enough to prove that the following $(R_{\mathsf{SK}}, R_{\mathsf{SM}})$ pair is achievable.

$$\begin{aligned} R_{\mathsf{SM}} &= \min\{I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};\mathbf{S}), I(\mathbf{U};\mathbf{Y}|\mathbf{V}) - I(\mathbf{U};\mathbf{Z}|\mathbf{V})\}, \text{ and} \\ R_{\mathsf{SK}} &= [I(\mathbf{U};\mathbf{Y}|\mathbf{V}) - I(\mathbf{U};\mathbf{Z}|\mathbf{V}) - (I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};\mathbf{S}))]_+ \\ &= [I(\mathbf{U};\mathbf{S}) - I(\mathbf{V};\mathbf{Y}) - I(\mathbf{U};\mathbf{Z}|\mathbf{V})]_+ \end{aligned}$$

We divide the proof into two cases. In each case, we use a random coding argument to show the existence of a codebook for which the probability of an encoding error at Alice, decoding error at Bob, and decoding error at Eve given additional side information are all small. We then show that such a code satisfies the secrecy and uniformity conditions.

Case 1: $I(\mathbf{U};S) < I(\mathbf{V};\mathbf{Y}) + I(\mathbf{U};\mathbf{Z}|\mathbf{V})$

In this case, we need only prove that the pair

$$\begin{aligned} R_{\mathsf{SM}} &= I(\mathbf{U};\mathbf{Y}|\mathbf{V}) - I(\mathbf{U};\mathbf{Z}|\mathbf{V}), \text{ and} \\ R_{\mathsf{SK}} &= 0 \end{aligned}$$

is achievable.

1. *Codebook Generation*: Draw an $(n, I(\mathbf{U};\mathbf{Y}) - 3\delta)$ codebook from the follow-

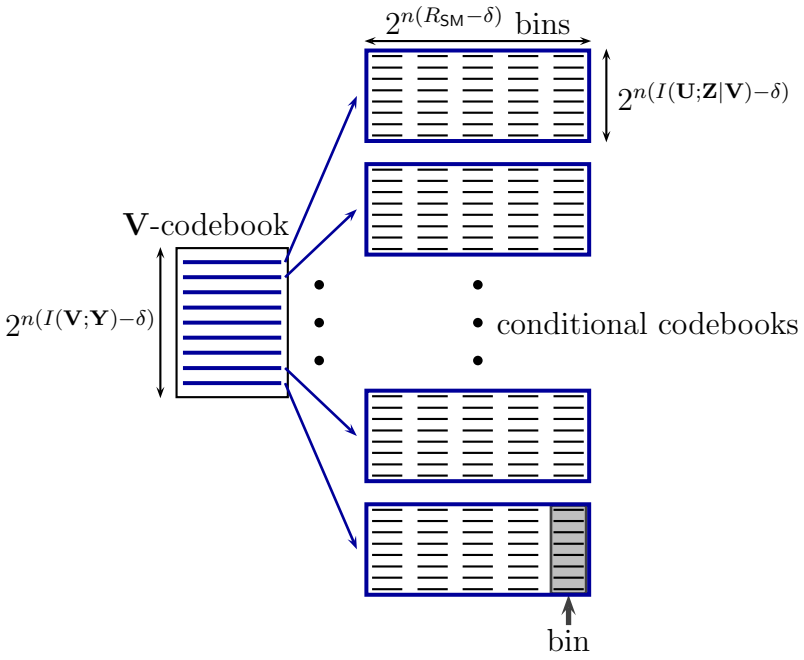

Figure D.1: The codebook used for Case 1 of the achievable strategy consists of a V-codebook, each codeword of which indexes a conditional codebook. The bins in each conditional codebook correspond directly to the private bit-pipe, and the V-codebook and codewords in each bin to the public bit-pipe. Analogously, the codewords in each conditional codebook correspond to quantization points for the source S^n.

ing distribution, which will be composed of two parts. The first part is an $(n, I(\mathbf{V};\mathbf{Y})-\delta)$ codebook with codewords drawn uniformly from the set $\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{V}})$ of ϵ-typical sequences with respect to $p_{\mathbf{V}}$ (see Definition A.7). We call this the $\mathbf{V}$-codebook and index its codewords using $i \in \{1, \dots, 2^{n(I(\mathbf{V};\mathbf{Y})-\delta)}\}$. For each such codeword $\mathbf{v}^n(i)$, we construct an $(n, I(\mathbf{U};\mathbf{Y}|\mathbf{V}) - 2\delta)$ codebook with codewords drawn uniformly from the set $\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{U}|\mathbf{V}}(\cdot|\mathbf{v}^n(i)))$ of conditionally ϵ-typical sequences on $\mathbf{v}^n(i)$ with respect to $p_{\mathbf{U}|\mathbf{V}}$ (see Definition A.8). These conditional codebooks form the second part of the codebook, and we refer to the entire codebook as a $\mathbf{U}$-codebook. For each conditional codebook, we distribute these sequences into $2^{n(R_{\mathsf{SM}}-\delta)}$ *bins* such that each bin contains $2^{n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})-\delta)}$ codewords, indexing each bin by $m \in \{1, \dots, 2^{n(R_{\mathsf{SM}}-\delta)}\}$. Let the codewords in each

bin be indexed by $j \in \{1, \ldots, 2^{n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})-\delta)}\}$.

Note that there is a direct correspondence between the bins, and the private bit-pipe, with the codewords in each bin and the $\mathbf{V}$-codebook corresponding to the public bit-pipe of the separation strategy. Furthermore, as will be seen in encoding, the $\mathbf{U}$-codewords are simply quantization points for the source S^n. A schematic of this codebook is depicted in Figure D.1.

In this separation context, Case 1 refers to the scenario in which there is insufficient randomness from the source S^n alone to determine the input to the public bit-pipe. Thus, we further divide the set of all $\mathbf{U}$-codewords into $2^{n(I(\mathbf{V};\mathbf{Y})+I(\mathbf{U};\mathbf{Z}|\mathbf{V})-I(\mathbf{U};S)-3\delta)}$ *buckets*[1] as follows: if (i) $I(\mathbf{U};S) \geq I(\mathbf{V};\mathbf{Y})$, that is, there is sufficient randomness in the source to determine the $\mathbf{V}$-codeword completely, we divide up codewords in each bin of every conditional codebook into buckets such that each bucket has the same number of codewords. Thus, given a bin, for each conditional codebook there are

$$2^{n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})+2\delta-I(\mathbf{V};\mathbf{Y})-I(\mathbf{U};\mathbf{Z}|\mathbf{V})+I(\mathbf{U};S))} = 2^{n(I(\mathbf{U};S)-I(\mathbf{V};\mathbf{Y})+2\delta)}$$

codewords in each bucket, for a total of $2^{n(I(\mathbf{U};S)+\delta)}$ codewords in each bucket over all conditional codebooks.

If (ii) $I(\mathbf{U};S) < I(\mathbf{V};\mathbf{Y})$, then, the $\mathbf{U}$-codewords are divided up among the buckets such that every bucket has no more than one codeword which belongs to the same bin of a conditional codebook. In this case, for a given bucket, there are

$$2^{n(I(\mathbf{U};\mathbf{Y})-3\delta-(R_{\mathsf{SM}}-\delta)-(I(\mathbf{V};\mathbf{Y})+I(\mathbf{U};\mathbf{Z}|\mathbf{V})-I(\mathbf{U};S)-3\delta))} = 2^{n(I(\mathbf{U};S)+\delta)}$$

[1]For there to be at least one bucket, we require that $3\delta < I(\mathbf{V};\mathbf{Y}) + I(\mathbf{U};\mathbf{Z}|\mathbf{V}) - I(\mathbf{U};S)$. However, this is not an issue since we will take $\delta \to 0$.

codewords each belonging to a different conditional codebook and holding the same bin index.

The buckets are indexed by $k \in \{1, \ldots, 2^{n(I(\mathbf{V};\mathbf{Y})+I(\mathbf{U};\mathbf{Z}|\mathbf{V})-I(\mathbf{U};S)-3\delta)}\}$. For a $\mathbf{U}$-codeword, we will explicity indicate its bucket index along with the conditional codebook it belongs to, its bin-index and its index within the bin as $\mathbf{u}^n(i, m, j, k)$.

2. *Encoding*: Let $m \in \{1, \ldots, 2^{n(R_{\mathsf{SM}}-\delta)}\}$ index the secret message. To send m, Alice selects a $\Phi_{\mathsf{bucket}} \in \{1, \ldots, 2^{n(I(\mathbf{V};\mathbf{Y})+I(\mathbf{U};\mathbf{Z}|\mathbf{V})-I(\mathbf{U};S)-3\delta)}\}$, which is uniformly distributed over its alphabet, and assigns $k = \Phi_{\mathsf{bucket}}$, and looks in bin m (of all the conditional codebooks) for a $\mathbf{V}^n(i), \mathbf{U}^n(i, m, j, k)$ such that $(\mathbf{V}^n(i), \mathbf{U}^n(i, m, j, k), S^n)$ are jointly typical. Thus, the U codeword is selected such that it belongs to bin m and bucket $k = \Phi_{\mathsf{bucket}}$ and such that it is jointly typical with the source observation S^n. If more than one choice is found, Alice chooses randomly among them. If none are found, Alice declares an error. A test channel $p_{X|\mathbf{U},S}$ stochastically generates the channel input X^n.

The probability of encoding failure can be bounded as follows. In case (i),

$$P_e \leq \mathbb{P}(S^n \notin \mathcal{T}_\epsilon^{*(n)}) + \sum_{s^n \in \mathcal{T}_\epsilon^{*(n)}} p_{S^n}(s^n) \sum_k p_{\Phi_{\mathsf{bucket}}}(k) \Bigg\{ \Bigg[\sum_{\mathbf{v}^n \in \mathcal{T}_\epsilon^{*(n)}} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n) \cdot \left[1 - \mathbb{P}((\mathbf{v}^n, \mathbf{U}^n, s^n) \in \mathcal{T}_\epsilon^{*(n)} | \mathbf{V}^n = \mathbf{v}^n)\right]^{2^{n(I(\mathbf{U};\mathbf{S})-I(\mathbf{V};\mathbf{Y})+2\delta)}} \Bigg]^{2^{n(I(\mathbf{V};\mathbf{Y})-\delta)}} \Bigg\}.$$

where the term $\mathbb{P}((v^n, \mathbf{U}^n, s^n) \in \mathcal{T}_\epsilon^{*(n)} | \mathbf{V}^n = \mathbf{v}^n)$ is evaluated with the distribution for $\mathbf{U}^n$ being given by the uniform distribution over all sequences $\mathbf{u}^n$ that are conditionally ϵ-typical on $\mathbf{v}^n$. Since $\mathbf{V}$ and S are independent, for $s^n \in \mathcal{T}_\epsilon^{*(n)}$

and $\mathbf{v}^n \in \mathcal{T}_\epsilon^{*(n)}$, this probability is

$$\mathbb{P}((\mathbf{v}^n, \mathbf{U}^n, s^n) \in \mathcal{T}_\epsilon^{*(n)} | \mathbf{V}^n = \mathbf{v}^n) \geq 2^{-n(I(\mathbf{U};S|\mathbf{V})+\epsilon_1)}, \tag{D.3}$$

where $\epsilon_1 \to 0$ as $\epsilon \to 0$. This will also be the case for any future subscripted $\epsilon_\#$ in the future. Applying (D.3) to the term within the braces in the upperbound for P_e and simplifying gives

$$\begin{aligned}
&\left[\sum_{\mathbf{v}^n \in \mathcal{T}_\epsilon^{*(n)}} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n) \cdot \left[1 - 2^{-n(I(\mathbf{U};S|\mathbf{V})+\epsilon_1)}\right]^{2^{n(I(\mathbf{U};\mathbf{S})-I(\mathbf{V};\mathbf{Y})+2\delta)}} \right]^{2^{n(I(\mathbf{V};\mathbf{Y})-\delta)}} \\
&\quad \overset{(a)}{\leq} e^{-2^{n(I(\mathbf{U};\mathbf{S})+\delta)} 2^{-n(I(\mathbf{U};S|\mathbf{V})+\epsilon_1)}} \\
&\quad \overset{(b)}{=} e^{-2^{n(\delta-\epsilon_1)}},
\end{aligned}$$

where (a) follows from $(1-x)^n \leq e^{-nx}$ and (b) from the fact that $I(\mathbf{U};S) = I(\mathbf{V},\mathbf{U};S) = I(\mathbf{U};S|\mathbf{V})$ which in turn is a consequence of the Markov chain $\mathbf{V} - \mathbf{U} - S$ and the independence of $\mathbf{V}$ and S. Substituiting this in the upperbound for P_e,

$$\begin{aligned}
P_e &\leq \mathbb{P}(S^n \notin \mathcal{T}_\epsilon^{*(n)}) + \sum_{s^n \in \mathcal{T}_\epsilon^{*(n)}} p_{S^n}(s^n) \sum_k p_{\Phi_{\mathsf{bucket}}}(k) \cdot e^{-2^{n(\delta-\epsilon_1)}} \\
&= \mathbb{P}(S^n \notin \mathcal{T}_\epsilon^{*(n)}) + (1 - \mathbb{P}(S^n \notin \mathcal{T}_\epsilon^{*(n)})) \cdot e^{-2^{n(\delta-\epsilon_1)}} \ .
\end{aligned}$$

Thus, we can make P_e as small as desired by choosing δ appropriately small, ϵ small enough such that $\epsilon_1 < \delta$, and n sufficiently large (see Lemma A.3).

Under case (ii), the probability of encoding failure can be similarly bounded.

Now, we have

$$P_e \leq \mathbb{P}(S^n \notin \mathcal{T}_\epsilon^{*(n)}) + \sum_{s^n \in \mathcal{T}_\epsilon^{*(n)}} p_{S^n}(s^n) \sum_k p_{\Phi_{\mathsf{bucket}}}(k) \cdot \left\{ \left[\sum_{\mathbf{v}^n \in \mathcal{T}_\epsilon^{*(n)}} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n) \left[1 - 2^{-n(I(\mathbf{U};S|\mathbf{V})+\epsilon_1)}\right] \right]^{2^{n(I(\mathbf{U};\mathbf{S})+\delta)}} \right\}.$$

where the term $\mathbb{P}((\mathbf{v}^n, \mathbf{U}^n, s^n) \in \mathcal{T}_\epsilon^{*(n)} | \mathbf{V}^n = \mathbf{v}^n)$ is evaluated as in (D.3). Following similar steps to those above, by choosing δ sufficiently small, ϵ small enough such that $\epsilon_1 < \delta$, and n sufficiently large, we can make P_e as small as desired.

3. *Decoding at Bob*: Bob receives $\mathbf{Y}^n$ and searches for the unique $(\mathbf{V}^n, \mathbf{U}^n)$ pair such that $(\mathbf{V}^n, \mathbf{U}^n, \mathbf{Y}^n)$ that are jointly ϵ-typical or declares an error. Bob identifies the corresponding bin-index $\hat{m}$, and declares this the secret message. Hence, conditioned on encoding being successful, a decoding error results only if there is a $\hat{m} \neq m$ such that, there are $\hat{i}$, $\hat{j}$, $\hat{k}$ such that $(\mathbf{v}^n(\hat{i}), \mathbf{u}^n(\hat{i}, \hat{m}, \hat{j}, \hat{k}), \mathbf{y}^n)$ are jointly ϵ-typical. We can upperbound the probability of this by

$$\begin{aligned}
&\sum_{\hat{i}} \sum_{\hat{m} \neq m} \sum_{\hat{j}} \mathbb{P}\left((\mathbf{V}^n(\hat{i}), \mathbf{U}^n(\hat{i}, \hat{m}, \hat{j}), \mathbf{Y}^n) \in \mathcal{T}_\epsilon^{*(n)} \right) \\
&\qquad = \sum_{\hat{i} \neq i} \sum_{\hat{m} \neq m} \sum_{\hat{j}} \mathbb{P}\left((\mathbf{V}^n(\hat{i}), \mathbf{U}^n(\hat{i}, \hat{m}, \hat{j}), \mathbf{Y}^n) \in \mathcal{T}_\epsilon^{*(n)} \right) \\
&\qquad\quad + \sum_{\hat{m} \neq m} \sum_{\hat{j}} \mathbb{P}\left((\mathbf{V}^n(i), \mathbf{U}^n(i, \hat{m}, \hat{j}), \mathbf{Y}^n) \in \mathcal{T}_\epsilon^{*(n)} \right) \\
&\qquad \leq 2^{n(I(\mathbf{U};\mathbf{Y})-3\delta)-n(I(\mathbf{U};\mathbf{Y})-\epsilon_2)} \\
&\qquad\qquad + 2^{n(R_{\mathsf{SM}}-\delta)+n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})-\delta)-n(I(\mathbf{U};\mathbf{Y}|\mathbf{V})-\epsilon_3)},
\end{aligned}$$

which can be made as small as desired by choosing δ sufficiently small, ϵ small

enough such that $\epsilon_2, \epsilon_3 < \delta$, and making n sufficiently large.

4. *Decoding at Eve with side information*: Consider Eve who has access to $M, \mathbf{V}^n$. Then, the bin in which a potential $\mathbf{U}^n$ exists is known to be at most $2^{n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})-\delta)}$. We may upperbound the probability of decoding error as we did above. Consider the jointly typical decoder for $\mathbf{U}^n$ given $\mathbf{Z}^n$ in this bin. There are two error events: E_1 is the event no sequence in the bin is jointly typical with $\mathbf{Z}^n$, and E_2 is the event a false sequence in the subbin is jointly typical with $\mathbf{Z}^n$. We have that $\mathbb{P}(E_1) \to 0$ as $n \to \infty$, and the probability a false sequence is jointly typical with $\mathbf{Z}^n$ is $2^{-n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})-\epsilon_4)}$. By a union bound, we can make the probability of error as small as desired by choosing δ sufficiently small,ϵ small enough such that $\epsilon_4 < \delta$, and taking n sufficiently large.

   By the usual random coding arguments, as in Case 1, we may now conclude that for any $\delta > 0$, for sufficiently large n, there exists a codebook such that Bob can recover the secret message with probability of error not larger than δ and Eve when provided with the message M and the $\mathbf{V}$-codeword can recover the $\mathbf{U}$-codeword with probability of error not larger than δ. We now simply have to verify that for this codebook, the secrecy condition holds: namely, Eve's information about the message (given $\mathbf{Z}^n$) is small.

Proof of Secrecy Condition.

First observe that

$$\begin{aligned} H(M|\mathbf{Z}^n) &\geq H(M|\mathbf{Z}^n, \mathbf{V}^n) \\ &= H(M, \mathbf{Z}^n|\mathbf{V}^n) - H(\mathbf{Z}^n|\mathbf{V}^n) \\ &= H(M, \mathbf{U}^n, \mathbf{Z}^n|\mathbf{V}^n) - H(\mathbf{U}^n|M, \mathbf{Z}^n, \mathbf{V}^n) - H(\mathbf{Z}^n|\mathbf{V}^n) \\ &\overset{(a)}{\geq} H(\mathbf{U}^n, \mathbf{Z}^n|\mathbf{V}^n) - H(\mathbf{U}^n|M, \mathbf{Z}^n, \mathbf{V}^n) - H(\mathbf{Z}^n|\mathbf{V}^n) \\ &= H(\mathbf{U}^n|\mathbf{V}^n) + H(\mathbf{Z}^n|\mathbf{U}^n, \mathbf{V}^n) - H(\mathbf{U}^n|M, \mathbf{Z}^n, \mathbf{V}^n) - H(\mathbf{Z}^n|\mathbf{V}^n). \quad \text{(D.4)} \end{aligned}$$

where (a) follows from non-negativity of conditional entropy. We now bound each of these terms.

Let us define, for every $\mathbf{u}^n(i, m, j, k)$ codeword

$$E = \{s^n : \exists(i, m, j, k) \text{ such that } (\mathbf{v}^n(i), \mathbf{u}^n(i, m, j, k), s^n) \in \mathcal{T}_\epsilon^{*(n)}\} \quad \text{(D.5)}$$

Recall that for all $\alpha > 0$, there exists n sufficiently large such that decoding (and hence encoding) succeeds with probability greater than $1 - \alpha$, *i.e.,*

$$\mathbb{P}(S^n \in E) \geq 1 - \alpha \ . \quad \text{(D.6)}$$

Furthermore, the probability

$$\begin{aligned}
&\mathbb{P}((\mathbf{V}^n, \mathbf{U}^n) = (\mathbf{v}^n(i), \mathbf{u}^n(i,m,j,k)), S^n \in E) \\
&\quad = \mathbb{P}(M = m, \Phi_{\mathsf{bucket}} = k, (\mathbf{V}^n, \mathbf{U}^n) = (\mathbf{v}^n(i), \mathbf{u}^n(i,m,j,k)), S^n \in E) \\
&\quad \leq \mathbb{P}(M = m) \cdot 2^{-n(I(\mathbf{V};\mathbf{Y})+I(\mathbf{U};\mathbf{Z}|\mathbf{V})-I(\mathbf{U};S)-3\delta)} \cdot \sum_{s^n:(\mathbf{v}^n(i),\mathbf{u}^n(i,m,j,k),s^n)\in\mathcal{T}_\epsilon^{*(n)}} \mathbb{P}(S^n = s^n) \qquad \text{(D.7)} \\
&\quad \leq 2^{-n(R_{\mathsf{SM}}-\delta)} \cdot 2^{-n(I(\mathbf{V};\mathbf{Y})+I(\mathbf{U};\mathbf{Z}|\mathbf{V})-I(\mathbf{U};S)-3\delta)} \cdot 2^{nH(S|\mathbf{U})+n\epsilon} \cdot 2^{-nH(S)+n\epsilon} \qquad \text{(D.8)} \\
&\quad = 2^{-n(R_{\mathsf{SM}}-\delta)} \cdot 2^{-n(I(\mathbf{V};\mathbf{Y})+I(\mathbf{U};\mathbf{Z}|\mathbf{V})-I(\mathbf{U};S)-3\delta)} \cdot 2^{-nI(S;\mathbf{U})+2n\epsilon} \qquad \text{(D.9)} \\
&\quad = 2^{-n(R_{\mathsf{SM}}-\delta)} \cdot 2^{-n(I(\mathbf{V};\mathbf{Y})+I(\mathbf{U};\mathbf{Z}|\mathbf{V})-3\delta)+2n\epsilon}, \qquad \text{(D.10)}
\end{aligned}$$

which along with the lowerbound on $\mathbb{P}(S^n \in E)$ above implies that

$$\begin{aligned}
&\mathbb{P}((\mathbf{V}^n, \mathbf{U}^n) = (\mathbf{v}^n(i), \mathbf{u}^n(i,m,j,k))|S^n \in E) \qquad \text{(D.11)} \\
&\quad \leq 2^{-nR_{\mathsf{SM}}} \cdot 2^{-n(I(\mathbf{V};\mathbf{Y})+I(\mathbf{U};\mathbf{Z}|\mathbf{V}))+n\epsilon_5} \qquad \text{(D.12)} \\
&\quad = 2^{-nI(\mathbf{U};\mathbf{Y})+n\epsilon_5}. \qquad \text{(D.13)}
\end{aligned}$$

Also, we know that the size of the codebook in which $(\mathbf{V}^n, \mathbf{U}^n)$ take values is less than $2^{nI(\mathbf{U};\mathbf{Y})}$ which implies that

$$H(\mathbf{U}^n, \mathbf{V}^n|S^n \in E) \geq nI(\mathbf{U};\mathbf{Y}) - n\epsilon_5.$$

Using this we can bound the first term in (D.4).

$$
\begin{aligned}
H(\mathbf{U}^n|\mathbf{V}^n) &= H(\mathbf{U}^n, \mathbf{V}^n) - H(\mathbf{V}^n) && \text{(D.14)}\\
&\overset{(a)}{\geq} H(\mathbf{U}^n, \mathbf{V}^n) - nI(\mathbf{V};\mathbf{Y}) && \text{(D.15)}\\
&\overset{(b)}{\geq} H(\mathbf{U}^n, \mathbf{V}^n|S^n \in E)\cdot \mathbb{P}(S^n \in E) - nI(\mathbf{V};\mathbf{Y}) && \text{(D.16)}\\
&= nI(\mathbf{U};\mathbf{Y}) - nI(\mathbf{V};\mathbf{Y}) - n\epsilon_6 && \text{(D.17)}\\
&= nI(\mathbf{U};\mathbf{Y}|\mathbf{V}) - n\epsilon_6, && \text{(D.18)}
\end{aligned}
$$

where (a) follows from the fact that $\mathbf{V}^n$ takes values in a codebook whose size is smaller than $2^{nI(\mathbf{V};\mathbf{Y})}$, and (b) follows from the fact that conditioning reduces entropy.

We bound the second term in (D.4) as follows

$$
\begin{aligned}
H(\mathbf{Z}^n|\mathbf{U}^n, \mathbf{V}^n) &= H(\mathbf{Z}^n|\mathbf{U}^n)\\
&= \sum_{\mathbf{u}^n} \mathbb{P}(\mathbf{U}^n = \mathbf{u}^n) H(\mathbf{Z}^n|\mathbf{U}^n = \mathbf{u}^n)\\
&\overset{(a)}{=} \sum_{\mathbf{u}^n} \mathbb{P}(\mathbf{U}^n = \mathbf{u}^n) \sum_{\mu\in\mathcal{U}} N(\mu|\mathbf{u}^n) H(\mathbf{Z}|\mathbf{U} = \mu)\\
&\overset{(b)}{\geq} \sum_{\mathbf{u}^n} \mathbb{P}(\mathbf{U}^n = \mathbf{u}^n) \sum_{\mu\in\mathcal{U}} n(\mathbb{P}(\mathbf{U} = \mu) - \epsilon) H(\mathbf{Z}|\mathbf{U} = \mu)\\
&= \sum_{\mathbf{u}^n} \mathbb{P}(\mathbf{U}^n = \mathbf{u}^n)(nH(\mathbf{Z}|\mathbf{U}) - n\epsilon_7)\\
&= nH(\mathbf{Z}|\mathbf{U}) - n\epsilon_7,
\end{aligned}
$$

where (a) follows from the memoryless nature of the virtual channel from $\mathbf{U}$ to $\mathbf{Z}$ and $N(\mu|\mathbf{u}^n)$ counts the number of times μ appears in the codeword $\mathbf{u}^n$, and (b) follows from the fact that all the $\mathbf{U}$-codewords belong to $\mathcal{T}_\epsilon^{*(n)}$.

The third terms can be bounded by using Fano's inequality and the fact that Eve can recover the $\mathbf{U}$-codeword with a probability of error ϵ when she has access to M

and $\mathbf{V}^n$ in addition to her observation $\mathbf{Z}^n$.

$$H(\mathbf{U}^n|M, K, \mathbf{Z}^n, \mathbf{V}^n) \leq 1 + n \cdot \epsilon \cdot I(\mathbf{U}; \mathbf{Z}|\mathbf{V}) = n\epsilon_8.$$

Finally, to bound the fourth term, let T be an indicator random variable which takes on the value 1 when $(\mathbf{V}^n, \mathbf{Z}^n) \in \mathcal{T}_\epsilon^{*(n)}$ and 0 otherwise.

$$\begin{aligned} H(\mathbf{Z}^n|\mathbf{V}^n) &\leq H(\mathbf{Z}^n, T|\mathbf{V}^n) \\ &\leq 1 + H(\mathbf{Z}^n|\mathbf{V}^n, T=1) \cdot \mathbb{P}(T=1) + n \log |\mathcal{Z}| \cdot \mathbb{P}(T=0). \end{aligned} \tag{D.19}$$

But

$$\mathbb{P}(T=0) = \mathbb{P}((\mathbf{V}^n, \mathbf{Z}^n) \notin \mathcal{T}_\epsilon^{*(n)}) \leq \epsilon_9.$$

Furthermore, we have

$$\begin{aligned} H(\mathbf{Z}^n|\mathbf{V}^n, T=1) &= \sum_{\mathbf{v}^n} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n|T=1) H(\mathbf{Z}^n|\mathbf{V}^n = \mathbf{v}^n, T=1) \\ &\overset{(a)}{\leq} \sum_{\mathbf{v}^n} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n|T=1) \log |\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{Z}|\mathbf{V}}|\mathbf{v}^n)| \\ &\leq \sum_{\mathbf{v}^n} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n|T=1)(nH(\mathbf{Z}|\mathbf{V}) + n\epsilon) \\ &= nH(\mathbf{Z}|\mathbf{V}) + n\epsilon, \end{aligned}$$

where in (a) we used $|\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{Z}|\mathbf{V}}|\mathbf{v}^n)|$ to denote the size of the set of all $\mathbf{z}^n$ such that $(\mathbf{z}^n, \mathbf{v}^n) \in \mathcal{T}_\epsilon^{*(n)}$. Thus, (D.19) becomes

$$H(\mathbf{Z}^n|\mathbf{V}^n) \leq nH(\mathbf{Z}|\mathbf{V}) + n\epsilon_{10}.$$

Hence, we may conclude from (D.4) that

$$\begin{aligned}\frac{1}{n}H(M|\mathbf{Z}^n) &\geq I(\mathbf{U};\mathbf{Y}|\mathbf{V}) + H(\mathbf{Z}|\mathbf{U}) - H(\mathbf{Z}|\mathbf{V}) + \epsilon_{11} \\ &= I(\mathbf{U};\mathbf{Y}|\mathbf{V}) - I(\mathbf{U};\mathbf{Z}|\mathbf{V}) + \epsilon_{11} \\ &= R_{\mathsf{SM}} + \epsilon_{11}.\end{aligned}$$

Thus we have shown the secrecy condition.

Case 2: $I(\mathbf{U};S) \geq I(\mathbf{V};\mathbf{Y}) + I(\mathbf{U};\mathbf{Z}|\mathbf{V})$

In this case, we only need to show the achievability of

$$\begin{aligned}R_{\mathsf{SM}} &= I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};\mathbf{S}), \text{ and} \\ R_{\mathsf{SK}} &= I(\mathbf{U};\mathbf{S}) - I(\mathbf{V};\mathbf{Y}) - I(\mathbf{U};\mathbf{Z}|\mathbf{V}).\end{aligned}$$

1. *Codebook Generation*: Draw an $(n, I(\mathbf{U};\mathbf{Y}) - 2\delta)$ codebook from the following distribution, which will be composed of two parts. The first part is an $(n, I(\mathbf{V};\mathbf{Y}) - \delta)$ codebook with codewords drawn uniformly from the set $\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{V}})$ of ϵ-typical sequences with respect to $p_{\mathbf{V}}$, which we again call the $\mathbf{V}$-codebook and index each codeword by $i \in \{1, \ldots, 2^{n(I(\mathbf{V};\mathbf{Y})-\delta)}\}$. For each such codeword $\mathbf{v}^n(i)$, we construct an $(n, I(\mathbf{U};\mathbf{Y}|\mathbf{V}) - \delta)$ codebook with codewords drawn uniformly from the set $\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{U}|\mathbf{V}}(\cdot|\mathbf{v}^n(i)))$ of conditionally ϵ-typical sequences on $\mathbf{v}^n(i)$ with respect to $p_{\mathbf{U}|\mathbf{V}}$, which forms the second part of the codebook. We call the entire codebook a $\mathbf{U}$-codebook.

   For each conditional codebook, distribute these sequences among $2^{n(R_{\mathsf{SM}}-3\delta)}$ *bins* such that each bin contains $2^{n(I(\mathbf{U};S)-I(\mathbf{V};\mathbf{Y})+2\delta)}$ codewords. We index each bin by m, where $m \in \{1, \ldots, 2^{n(R_{\mathsf{SM}}-3\delta)}\}$. Then assign the sequences in each bin

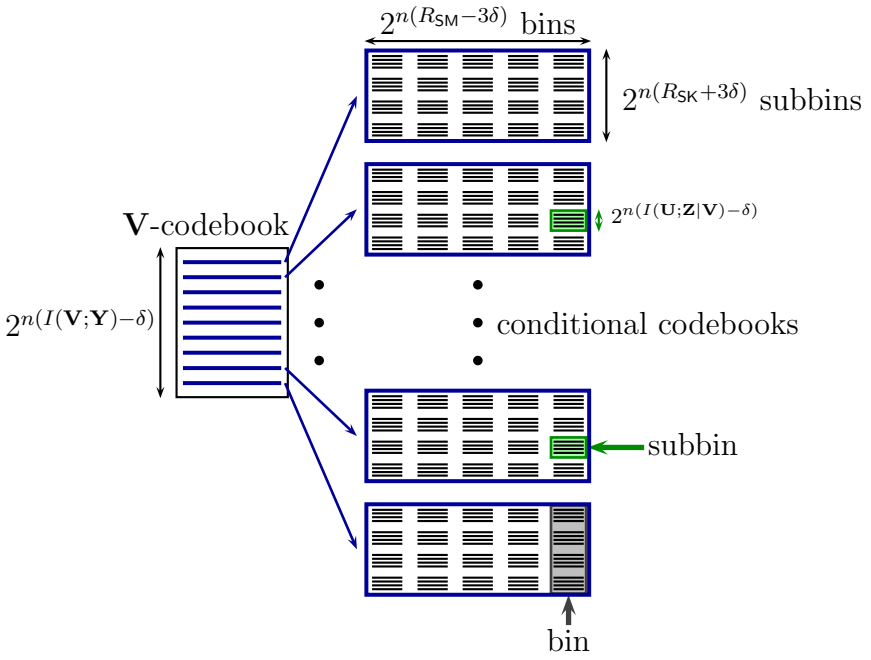

Figure D.2: The codebook used for Case 2 of the achievable strategy consists of a V-codebook, each codeword of which indexes a conditional codebook. The bins in each conditional codebook correspond directly to the private bit-pipe, and the V-codebook and codewords in each bin to the public bit-pipe. Analogously, the codewords in each conditional codebook correspond to quantization points for the source S^n.

into $2^{n(R_{\mathsf{SK}}+3\delta)}$ *subbins* so that each subbin contains $2^{n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})-\delta)}$ codewords, indexing each subbin by $k \in \{1,\ldots,2^{n(R_{\mathsf{SK}}+3\delta)}\}$. Index each of the elements in the subbin by $\ell \in \{1,\ldots,2^{n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})-\delta)}\}$, and denote the specific index as $\Phi_{\mathsf{sub-index}}$. For a $\mathbf{U}$-codeword, we will explicitly indicate its index as $\mathbf{u}^n(i,m,k,\ell)$.

2. *Encoding*: Let $m \in \{1,\ldots,2^{n(R_{\mathsf{SM}}-3\delta)}\}$ index the secret message. For this fixed m, Alice selects a $\mathbf{V}^n(i), \mathbf{U}^n(i,m,k,\ell)$ such that $(\mathbf{V}^n, \mathbf{U}^n(i,m,k,\ell), S^n)$ are jointly typical. If none are found, Alice declares an error. A test channel $p_{X|\mathbf{U},S}$ stochastically encodes the channel input X^n. The subbin k is set as the secret key. Note that the secret key is determined automatically by the $\mathbf{U}^n(i,m,k,\ell)$ selected.

Note that for fixed m, there are $2^{n(I(\mathbf{U};S)+\delta)}$ sequences among (i,k,ℓ), and the probability of an encoding failure is given by

$$P_e \leq \mathbb{P}(S^n \notin \mathcal{T}_\epsilon^{*(n)}) + \sum_{s^n \in \mathcal{T}_\epsilon^{*(n)}} p_{S^n}(s^n) \Bigg\{ \Bigg[\sum_{\mathbf{v}^n \in \mathcal{T}_\epsilon^{*(n)}} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n) \cdot \Big[1 - \mathbb{P}((\mathbf{v}^n, \mathbf{U}^n, s^n) \in \mathcal{T}_\epsilon^{*(n)} | \mathbf{V}^n = v^n) \Big]^{2^{n(I(\mathbf{U};S)-I(\mathbf{V};\mathbf{Y})+2\delta)}} \Bigg]^{2^{n(I(\mathbf{V};\mathbf{Y})-\delta)}} \Bigg\}.$$

where the term $\mathbb{P}(\mathbf{V}^n = \mathbf{v}^n)$ is uniform over the set $\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{V}})$ of ϵ-typical sequences with respect to $p_{\mathbf{V}}$, $\mathbb{P}((v^n, \mathbf{U}^n, s^n) \in \mathcal{T}_\epsilon^{*(n)} | \mathbf{V}^n = \mathbf{v}^n)$ is evaluated with $\mathbf{U}^n$ uniformly distributed over the set $\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{U}|\mathbf{V}}|\mathbf{V}^n = v^n)$ of all $\mathbf{u}^n$ sequences which are conditionally ϵ-typical on $\mathbf{v}^n$. Since $\mathbf{V}$ and S are independent, for $s^n \in \mathcal{T}_\epsilon^{*(n)}$ and $\mathbf{v}^n \in \mathcal{T}_\epsilon^{*(n)}$,

$$\begin{aligned} \mathbb{P}((\mathbf{v}^n, \mathbf{U}^n, s^n) \in \mathcal{T}_\epsilon^{*(n)} | \mathbf{V}^n = \mathbf{v}^n) &\geq 2^{-n(I(\mathbf{U};S|\mathbf{V})+\epsilon_1)} \\ &= 2^{-n(I(\mathbf{U};S)+\epsilon_1)}. \end{aligned} \tag{D.20}$$

By applying (D.20) to the term within the braces in the upperbound for P_e and simplifying, we have

$$\begin{aligned} &\Bigg[\sum_{\mathbf{v}^n \in \mathcal{T}_\epsilon^{*(n)}} |\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{V}})|^{-1} \cdot \Big[1 - 2^{-n(I(\mathbf{U};S)+\epsilon_1)} \Big]^{2^{n(I(\mathbf{U};S)-I(\mathbf{V};\mathbf{Y})+2\delta)}} \Bigg]^{2^{n(I(\mathbf{V};\mathbf{Y})-\delta)}} \\ &\quad = \Bigg[\Big[1 - 2^{-n(I(\mathbf{U};S|\mathbf{V})+\epsilon_1)} \Big]^{2^{n(I(\mathbf{U};S)-I(\mathbf{V};\mathbf{Y})+2\delta)}} \Bigg]^{2^{n(I(\mathbf{V};\mathbf{Y})-\delta)}} \\ &\quad \overset{(a)}{\leq} e^{-2^{n(\delta-\epsilon_1)}}, \end{aligned}$$

where (a) follows from the inequality $1 - x \leq e^{-x}$. Substituting this in the upper bound for P_e, as in Case 1, we can P_e as small as desired by choosing δ

sufficiently small, ϵ small enough that $\epsilon_1 < \delta$, and n sufficiently large.

3. *Decoding at Bob*: Bob receives $\mathbf{Y}^n$ and searches for a unique

$$(\mathbf{V}^n(\hat{i}), \mathbf{U}^n(\hat{i}, \hat{m}, \hat{k}, \hat{\ell}))$$

pair such that $(\mathbf{V}^n(\hat{i}), \mathbf{U}^n(\hat{i}, \hat{m}, \hat{k}, \hat{\ell}), \mathbf{Y}^n) \in \mathcal{T}_\epsilon^{*(n)}$. If not such pair exists, Bob declares an error. Otherwise Bob declares $\hat{m}$ to be the secret message and $\hat{k}$ to be the secret key. Conditioned on encoding being successful, by the AEP, the true sequence will be jointly typical with high probability. Hence, we only need to consider the decoding error resulting from a $(\hat{m}, \hat{k}) \neq (m, k)$ such that there are $\hat{i}$ and $\hat{\ell}$ and $\left(\mathbf{V}^n(\hat{i}), \mathbf{U}^n(\hat{i}, \hat{m}, \hat{k}_1, \hat{\ell}), \mathbf{Y}^n\right) \in \mathcal{T}_\epsilon^{*(n)}$. We can upperbound the probability of this by

$$\begin{aligned}
&\sum_{\hat{i}} \sum_{(\hat{m},\hat{k})\neq(m,k)} \sum_{\hat{\ell}} \mathbb{P}\left((\mathbf{V}^n(\hat{i}), \mathbf{U}^n(\hat{i}, \hat{m}, \hat{k}, \hat{\ell}), \mathbf{Y}^n) \in \mathcal{T}_\epsilon^{*(n)}\right) \\
&\quad = \sum_{\hat{i}\neq i} \sum_{(\hat{m},\hat{k})\neq(m,k)} \sum_{\hat{\ell}} \mathbb{P}\left((\mathbf{V}^n(\hat{i}), \mathbf{U}^n(\hat{i}, \hat{m}, \hat{k}, \hat{\ell}), \mathbf{Y}^n) \in \mathcal{T}_\epsilon^{*(n)}\right) \\
&\qquad + \sum_{(\hat{m},\hat{k})\neq(m,k)} \sum_{\hat{\ell}} \mathbb{P}\left((\mathbf{V}^n(i), \mathbf{U}^n(i, \hat{m}, \hat{k}, \hat{\ell}), \mathbf{Y}^n) \in \mathcal{T}_\epsilon^{*(n)}\right) \\
&\quad \leq 2^{n(I(\mathbf{U};\mathbf{Y})-2\delta)} 2^{-n(I(\mathbf{U};\mathbf{Y})-\epsilon_2)} + 2^{n(I(\mathbf{U};\mathbf{Y}|\mathbf{V})-\delta)} 2^{-n(I(\mathbf{U};\mathbf{Y}|\mathbf{V})-\epsilon_3)},
\end{aligned}$$

which can be made as small as desired by choosing δ sufficiently small, ϵ small enough such that $\epsilon_2, \epsilon_3 < \delta$, and taking n sufficiently large.

4. *Decoding at Eve with side information*: Consider Eve who has access to $M, K, \mathbf{V}^n$. Then, the subbin in which a potential $\mathbf{U}^n$ exists is known to be at most $2^{n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})-\delta)}$. The probability of error of the jointly typical decoder for $\mathbf{U}^n$ in this bin given $\mathbf{Z}^n$ can be bounded as above. There are two error events:

E_1 is the event no sequence in the bin is jointly typical with $\mathbf{Z}^n$, and E_2 is the event a false sequence in the subbin is jointly typical with $\mathbf{Z}^n$. We have, $\mathbb{P}(E_1) \to 0$ as $n \to \infty$, and the probability a false sequence is jointly typical with $\mathbf{Z}^n$ is $2^{-n(I(\mathbf{U};\mathbf{Z}|\mathbf{V})-\epsilon_4)}$. By a union bound, we can make the probability of error as small as desired by choosing δ sufficiently small,ϵ small enough such that $\epsilon_4 < \delta$, and taking n sufficiently large.

By the usual random coding arguments, we may now conclude that for any $\delta > 0$, for sufficiently large n, there exists a codebook such that Bob can recover the secret message with probability of error not larger than δ and Eve when provided with the message and the $\mathbf{V}$-codeword can recover the $\mathbf{U}$-codeword with probability of error not larger than δ. We now have to verify that this implies that (1) Eve's information about the message (given $\mathbf{Z}^n$) is small (secrecy condition) and (2) the secret key is approximately uniformly distributed over its alphabet (uniformity condition).

Proof of Secrecy Condition.
First observe that

$$\begin{aligned} H(M,K|\mathbf{Z}^n) &\geq H(M,K|\mathbf{Z}^n,\mathbf{V}^n) \\ &= H(M,K,\mathbf{Z}^n|\mathbf{V}^n) - H(\mathbf{Z}^n|\mathbf{V}^n) \\ &= H(M,K,\mathbf{U}^n,\mathbf{Z}^n|\mathbf{V}^n) - H(\mathbf{U}^n|M,K,\mathbf{Z}^n,\mathbf{V}^n) - H(\mathbf{Z}^n|\mathbf{V}^n) \\ &\overset{(a)}{=} H(\mathbf{U}^n,\mathbf{Z}^n|\mathbf{V}^n) - H(\mathbf{U}^n|M,K,\mathbf{Z}^n,\mathbf{V}^n) - H(\mathbf{Z}^n|\mathbf{V}^n) \\ &= H(\mathbf{U}^n|\mathbf{V}^n) + H(\mathbf{Z}^n|\mathbf{U}^n,\mathbf{V}^n) - H(\mathbf{U}^n|M,K,\mathbf{Z}^n,\mathbf{V}^n) - H(\mathbf{Z}^n|\mathbf{V}^n)\ , \end{aligned} \tag{D.21}$$

where (a) follows from $H(M,K|\mathbf{U}^n,\mathbf{Z}^n,\mathbf{V}^n) \geq 0$. We now bound each of these terms.

Let us define, for every $\mathbf{u}^n(i,m,k,\ell)$ codeword

$$E = \{s^n : \exists (i,m,k,\ell) \text{ such that } (\mathbf{v}^n(i), \mathbf{u}^n(i,m,k,\ell), s^n) \in \mathcal{T}_\epsilon^{*(n)}\} \tag{D.22}$$

Recall that for all $\alpha > 0$, there exists n sufficiently large such that decoding (and hence encoding) succeeds with probability greater than $1-\alpha$, *i.e.,*

$$\mathbb{P}(S^n \in E) \geq 1 - \alpha \ . \tag{D.23}$$

Furthermore, the probability

$$\begin{aligned}
&\mathbb{P}\left((\mathbf{V}^n, \mathbf{U}^n) = (\mathbf{v}^n, \mathbf{u}^n(i,m,k,\ell)), S^n \in E\right) \\
&\quad = \mathbb{P}\left(M = m, K = k, (\mathbf{V}^n, \mathbf{U}^n) = (\mathbf{v}^n, \mathbf{u}^n(i,m,k,\ell)), S^n \in E\right) && \text{(D.24)} \\
&\quad = \mathbb{P}\left(M = m\right) \cdot \sum_{s^n : (\mathbf{v}^n(i), \mathbf{u}^n(i,m,j,k), s^n) \in \mathcal{T}_\epsilon^{*(n)}} \mathbb{P}\left(S^n = s^n\right) && \text{(D.25)} \\
&\quad \leq 2^{-n(I(\mathbf{U};\mathbf{Y}) - I(\mathbf{U};S) - 3\delta)} \cdot 2^{n(H(S|\mathbf{U}) + n\epsilon} 2^{-nH(S) + n\epsilon} && \text{(D.26)} \\
&\quad = 2^{-nI(\mathbf{U};\mathbf{Y}) + 3n\delta + 2n\epsilon} \ , && \text{(D.27)}
\end{aligned}$$

which, along with the lower bound on $\mathbb{P}(S^n \in E)$ above implies that

$$\begin{aligned}
&\mathbb{P}\left((\mathbf{V}^n, \mathbf{U}^n) = (\mathbf{v}^n, \mathbf{u}^n(i,m,k,\ell)) | S^n \in E\right) \\
&\quad \leq 2^{-nI(\mathbf{U};\mathbf{Y}) + n\epsilon_{12}} \ . && \text{(D.28)}
\end{aligned}$$

Also, we know that the size of the codebook in which $(\mathbf{V}^n, \mathbf{U}^n)$ take values is less than $2^{nI(\mathbf{U};\mathbf{Y})}$, which implies that

$$H(\mathbf{U}^n, \mathbf{V}^n | S^n \in E) \geq nI(\mathbf{U};\mathbf{Y}) - n\epsilon_{12} \ .$$

Using this, we can bound the first term in (D.21):

$$H(\mathbf{U}^n|\mathbf{V}^n) = H(\mathbf{U}^n, \mathbf{V}^n) - H(\mathbf{V}^n) \tag{D.29}$$

$$\overset{(a)}{\geq} H(\mathbf{U}^n, \mathbf{V}^n) - nI(\mathbf{V};\mathbf{Y}) \tag{D.30}$$

$$\overset{(b)}{\geq} H(\mathbf{U}^n, \mathbf{V}^n|S^n \in E)\cdot \mathbb{P}(S^n \in E) - nI(\mathbf{V};\mathbf{Y}) \tag{D.31}$$

$$= nI(\mathbf{U};\mathbf{Y}) - nI(\mathbf{V};\mathbf{Y}) - n\epsilon_{13}\ , \tag{D.32}$$

where (a) follows from the fact that $\mathbf{V}^n$ takes values in a codebook whose size is smaller than $2^{nI(\mathbf{V};\mathbf{Y})}$, and (b) follows from the fact that conditioning cannot increase entropy.

We bound the second term in (D.21) as follows:

$$\begin{aligned}
H(\mathbf{Z}^n|\mathbf{U}^n, \mathbf{V}^n) &= H(\mathbf{Z}^n|\mathbf{U}^n) \\
&= \sum_{\mathbf{u}^n} Pr(\mathbf{U}^n = \mathbf{u}^n) H(\mathbf{Z}^n|\mathbf{U}^n = \mathbf{u}^n) \\
&\overset{(a)}{=} \sum_{\mathbf{u}^n} \mathbb{P}(\mathbf{U}^n = \mathbf{u}^n) \sum_{\mu\in\mathcal{U}} N(\mu|\mathbf{u}^n) H(\mathbf{Z}|\mathbf{U}=\mu) \\
&\overset{(b)}{\geq} \sum_{\mathbf{u}^n} \mathbb{P}(\mathbf{U}^n = \mathbf{u}^n) \sum_{\mu\in\mathcal{U}} n(\mathbb{P}(\mathbf{U}=\mu) - \epsilon) H(\mathbf{Z}|\mathbf{U}=\mu) \\
&= \sum_{\mathbf{u}^n} \mathbb{P}(\mathbf{U}^n = \mathbf{u}^n)(nH(\mathbf{Z}|\mathbf{U}) - n\epsilon_{14}) \\
&= nH(\mathbf{Z}|\mathbf{U}) - n\epsilon_{14},
\end{aligned}$$

where (a) follows from the memoryless nature of the virtual channel from $\mathbf{U}$ to $\mathbf{Z}$ and $N(\mu|\mathbf{u}^n)$ counts the number of times μ appears in the codeword $\mathbf{u}^n$, and (b) follows from the fact that all the $\mathbf{u}^n$ codewords belong to $\mathcal{T}_\epsilon^{*(n)}$.

The third term can be bounded by using Fano's inequality and the fact that Eve can recover the $\mathbf{U}^n$ codeword with a probability of error ϵ when she has access to M

and $\mathbf{V}^n$ in addition to her observation $\mathbf{Z}^n$.

$$H(\mathbf{U}^n|M, K, \mathbf{Z}^n, \mathbf{V}^n) \leq 1 + n \cdot \epsilon \cdot I(\mathbf{U}; \mathbf{Z}|\mathbf{V}) = n\epsilon_{15}.$$

Finally, to bound the fourth term, let T be an indicator random variable which takes on the value 1 when $(\mathbf{V}^n, \mathbf{Z}^n) \in \mathcal{T}_\epsilon^{*(n)}$ and 0 otherwise.

$$\begin{aligned} H(\mathbf{Z}^n|\mathbf{V}^n) &\leq H(\mathbf{Z}^n, T|\mathbf{V}^n) \\ &\leq 1 + H(\mathbf{Z}^n|\mathbf{V}^n, T=1)\mathbb{P}(T=1) + n\log|\mathcal{Z}|\mathbb{P}(T=0). \end{aligned} \tag{D.33}$$

But

$$\mathbb{P}(T=0) = \mathbb{P}((\mathbf{V}^n, \mathbf{Z}^n) \notin \mathcal{T}_\epsilon^{*(n)}) \leq \epsilon_{16}.$$

Furthermore, we have

$$\begin{aligned} H(\mathbf{Z}^n|\mathbf{V}^n, T=1) &= \sum_{\mathbf{v}^n} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n|T=1) H(\mathbf{Z}^n|\mathbf{V}^n = \mathbf{v}^n, T=1) \\ &\overset{(a)}{\leq} \sum_{\mathbf{v}^n} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n|T=1) \log|\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{Z}|\mathbf{V}}|\mathbf{v}^n)| \\ &\leq \sum_{\mathbf{v}^n} \mathbb{P}(\mathbf{V}^n = \mathbf{v}^n|T=1)(nH(\mathbf{Z}|\mathbf{V}) + n\epsilon) \\ &= nH(\mathbf{Z}|\mathbf{V}) + n\epsilon, \end{aligned}$$

where in (a) we used $|\mathcal{T}_\epsilon^{*(n)}(p_{\mathbf{Z}|\mathbf{V}}|\mathbf{v}^n)|$ to denote the size of the set of all $\mathbf{z}^n$ such that $(\mathbf{z}^n, \mathbf{v}^n) \in \mathcal{T}_\epsilon^{*(n)}$. Thus, (D.33) becomes

$$H(\mathbf{Z}^n|\mathbf{V}^n) \leq nH(\mathbf{Z}|\mathbf{V}) + n\epsilon_{17}.$$

Hence, we may conclude from (D.21) that

$$\begin{aligned}\frac{1}{n}H(M,K|\mathbf{Z}^n) &\geq I(\mathbf{U};\mathbf{Y}|\mathbf{V}) + H(\mathbf{Z}|\mathbf{U}) - H(\mathbf{Z}|\mathbf{V}) + \epsilon_{18} \\ &= I(\mathbf{U};\mathbf{Y}|\mathbf{V}) - I(\mathbf{U};\mathbf{Z}|\mathbf{V}) + \epsilon_{18} \\ &= R_{\mathsf{SM}} + R_{\mathsf{SK}} + \epsilon_{18}.\end{aligned}$$

Thus we have shown the secrecy condition.

Proof of Uniformity Condition.

Note that from (D.28), we have that

$$\begin{aligned}H(K) &= H(\mathbf{V}^n, M, K, \Phi_{\mathsf{sub-index}}) - H(\mathbf{V}^n, M, \Phi_{\mathsf{sub-index}}|K) \\ &\overset{(a)}{\geq} H(\mathbf{V}^n, M, K, \Phi_{\mathsf{sub-index}}) - (I(\mathbf{V};\mathbf{Y}) + nR_{\mathsf{SM}} + I(\mathbf{U};\mathbf{Z}|\mathbf{V})) \\ &\overset{(b)}{\geq} H(\mathbf{V}^n, M, K, \Phi_{\mathsf{sub-index}}|S^n \in E) \cdot \mathbb{P}(S^n \in E) \\ &\qquad - (I(\mathbf{V};\mathbf{Y}) + nR_{\mathsf{SM}} + I(\mathbf{U};\mathbf{Z}|\mathbf{V})) \\ &\overset{(c)}{\geq} nI(\mathbf{U};\mathbf{Y}) - (I(\mathbf{V};\mathbf{Y}) + nR_{\mathsf{SM}} + I(\mathbf{U};\mathbf{Z}|\mathbf{V})) - n\epsilon_{13} \\ &= R_{\mathsf{SK}} - n\epsilon_{13}\ .\end{aligned}$$

where (a) follows since $\mathbf{V}^n$ is drawn from a codebook with no more than $2^{nI(\mathbf{V};\mathbf{Y})}$ elements, M has $2^{nR_{\mathsf{SM}}}$ elements, and $\Phi_{\mathsf{sub-index}}$ has no more than $2^{nI(\mathbf{U};\mathbf{Z}|\mathbf{V})}$ elements; (b) since conditional entropy is less than or equal to entropy; and (c) from the lower bound in (D.28). Since $\epsilon_{13} \to 0$ as $\epsilon \to 0$, by appropriate choice of ϵ and sufficiently large n, we can satisfy the uniformity condition.

D.2 Proof of Theorem 4.2 (Optimality of Separation)

The achievability follows directly from Theorem 4.1 by setting the auxiliary random variables as follows.

$$V_1 = (X_F, X_R),$$
$$V_2 = (V_{2,F}, X_R),$$
$$U_1 = U_{1,F}.$$

It is easy to see that this satisfies the Markov conditions on the auxiliary random variables. Subsituting these in the expression in Theorem 4.1 shows the achievability. The interpretation is that, we have ignored the reversely degraded source component, and the reversely degraded channel is used purely as a channel for public communication.

To show the converse, let J and J' be independent random variables both uniformly distributed over $\{1, 2, \ldots, n\}$ and independent of all other random variables.

To get the first condition (ignoring $o(n)$ terms)

$$
\begin{aligned}
& n\Big(I(X_{F,J};Y_{F,J}) + I(X_{R,J};Y_{R,J})\Big) \\
& \qquad \geq nI(\mathbf{X}_J;\mathbf{Y}_J) \\
& \qquad \geq nI(\mathbf{X}_J;\mathbf{Y}_J|J) \\
& \qquad \geq I(\mathbf{X}^n;\mathbf{Y}^n) \\
& \qquad \overset{\text{(a)}}{=} I(\mathbf{X}^n;\mathbf{Y}^n,Z_F^n) \\
& \qquad = I(M,K,\mathbf{S}_A^n,\mathbf{X}^n;\mathbf{Y}^n,Z_F^n) \\
& \qquad \geq I(M,K,\mathbf{S}_A^n;\mathbf{Y}^n,Z_F^n) \\
& \qquad \geq I(M,K,\mathbf{S}_A^n;\mathbf{Y}^n,Z_F^n) - I(\mathbf{S}_B^n,\mathbf{S}_E^n;\mathbf{Y}^n,Z_F^n) \\
& \qquad \overset{\text{(b)}}{=} I(M,K,\mathbf{S}_A^n;\mathbf{Y}^n,Z_F^n|\mathbf{S}_B^n,\mathbf{S}_E^n) \\
& \qquad = I(M;\mathbf{Y}^n,Z_F^n|\mathbf{S}_B^n,\mathbf{S}_E^n) + I(K,\mathbf{S}_A^n;\mathbf{Y}^n,Z_F^n|\mathbf{S}_B^n,\mathbf{S}_E^n,M) \\
& \qquad \overset{\text{(c)}}{=} H(M|\mathbf{S}_B^n,\mathbf{S}_E^n) + I(K,\mathbf{S}_A^n;\mathbf{Y}^n,Z_F^n|\mathbf{S}_B^n,\mathbf{S}_E^n,M) \\
& \qquad = H(M) + I(K,\mathbf{S}_A^n;\mathbf{Y}^n,Z_F^n|\mathbf{S}_B^n,\mathbf{S}_E^n,M) \\
& \qquad = nR_{\mathsf{SM}} + I(K,\mathbf{S}_A^n;\mathbf{Y}^n,Z_F^n|\mathbf{S}_B^n,\mathbf{S}_E^n,M)
\end{aligned}
$$

where (a) is due to the sub-channel F to Eve being degraded w.r.t. the channel to Bob, (b) is because $(\mathbf{S}_B^n,\mathbf{S}_E^n)-\mathbf{S}_A^n-(M,K,\mathbf{Y}^n,Z_F^n)$ is a Markov chain, and (c) follows from Fano's inequality which gives $H(M|\mathbf{Y}^n,\mathbf{S}_B^n)=o(n)$. Now, to bound the second

term, we write

$$
\begin{aligned}
&I(K, \mathbf{S}_A^n; \mathbf{Y}^n, Z_F^n | \mathbf{S}_B^n, \mathbf{S}_E^n, M) \\
&\quad = H(\mathbf{Y}^n, Z_F^n | \mathbf{S}_B^n, \mathbf{S}_E^n, M) - H(\mathbf{Y}^n, Z_F s^n | K, M, \mathbf{S}_A^n, \mathbf{S}_B^n, \mathbf{S}_E^n) \\
&\quad \geq H(\mathbf{Y}^n, Z_F^n | \mathbf{S}_B^n, \mathbf{S}_E^n, M) - H(K, \mathbf{Y}^n, Z_F^n | \mathbf{S}_A^n, \mathbf{S}_B^n, \mathbf{S}_E^n, M) \\
&\quad \overset{\text{(a)}}{=} H(K, \mathbf{Y}^n, Z_F^n | \mathbf{S}_B^n, \mathbf{S}_E^n, M) - H(K, \mathbf{Y}^n, Z_F^n | \mathbf{S}_A^n, \mathbf{S}_B^n, \mathbf{S}_E^n, M) \\
&\quad = I(K, \mathbf{Y}^n, Z_F^n; \mathbf{S}_A^n | \mathbf{S}_B^n, \mathbf{S}_E^n, M) \\
&\quad \overset{\text{(b)}}{=} I(M, K, \mathbf{Y}^n, Z_F^n; \mathbf{S}_A^n | \mathbf{S}_B^n, \mathbf{S}_E^n) \\
&\quad \geq I(M, K, \mathbf{Y}^n, Z_F^n; S_{A,F}^n | S_{A,R}^n, \mathbf{S}_B^n, \mathbf{S}_E^n) \\
&\quad = I(M, K, \mathbf{Y}^n, Z_F^n; S_{A,F}^n | S_{A,R}^n, S_{B,F}^n, S_{E,F}^n) \\
&\quad = \sum_{i=1}^n I(M, K, \mathbf{Y}^n, Z_F^n; S_{A,F,i} | S_{A,F}^{i-1}, S_{A,R}^n, S_{B,F}^n, S_{E,F}^n) \\
&\quad \overset{\text{(c)}}{=} \sum_{i=1}^n I(M, K, \mathbf{Y}^n, Z_F^n, S_{A,F}^{i-1}; S_{A,F,i} | S_{A,R}^n, S_{B,F}^n, S_{E,F}^n) \\
&\quad \geq \sum_{i=1}^n I(M, K, \mathbf{Y}^n, Z_F^n; S_{A,F,i} | S_{A,R}^n, S_{B,F}^n, S_{E,F}^n) \\
&\quad = \sum_{i=1}^n I(M, K, \mathbf{Y}^n, Z_F^n, S_{B,F,\tilde{i}}, S_{E,F,\tilde{i}}, S_{A,R}^n; S_{A,F,i} | S_{B,F,i}, S_{E,F,i}) \\
&\quad = nI(U_{1,F}; S_{A,F,J'} | S_{B,F,J'}, S_{E,F,J'}),
\end{aligned}
$$

where (a) follows from Fano's inequality which implies that $H(K|\mathbf{Y}^n, \mathbf{S}_B^n) = o(n)$, (b) follows the independence of M from $(\mathbf{S}_A^n, \mathbf{S}_B^n, \mathbf{S}_C^n)$, (c) follows by the property that $S_{A,F}^n$ is memoryless, and we define $S_{B,F,\tilde{i}} \overset{\text{def}}{=} (S_{B,F}^{i-1}, S_{B,F,i+1}^n)$, $S_{E,F,\tilde{i}} \overset{\text{def}}{=} (S_{E,F}^{i-1}, S_{E,F,i+1}^n)$, and $U_{1,F} \overset{\text{def}}{=} (M, K, \mathbf{Y}^n, Z_F^n, S_{B,F,\tilde{J}'}, S_{E,F,\tilde{J}'}, S_{A,R}^n, J')$. Note that this $U_{1,F}$ does indeed

satisfy the condition $U_{1,F} - \mathbf{S}_{A,J'} - (\mathbf{S}_{B,J'}, \mathbf{S}_{E,J'})$. To get condition 2,

$$
\begin{aligned}
n(R_{\mathsf{SK}} + R_{\mathsf{SM}}) &\leq I(M,K;\mathbf{Y}^n, Z_F^n, \mathbf{S}_B^n, \mathbf{S}_E^n) \\
&\overset{\text{(a)}}{=} I(M,K;\mathbf{Y}^n, Z_F^n, \mathbf{S}_B^n, \mathbf{S}_E^n) - I(M,K;\mathbf{Z}^n, \mathbf{S}_E^n) \\
&\overset{\text{(b)}}{=} I(M,K;\mathbf{Y}^n, Z_F^n, \mathbf{S}_B^n, \mathbf{S}_E^n) - I(M,K;\mathbf{Z}^n, Y_R^n, \mathbf{S}_E^n) \\
&\leq I(M,K;\mathbf{Y}^n, Z_F^n, \mathbf{S}_B^n, \mathbf{S}_E^n) - I(M,K;Z_F^n, Y_R^n, \mathbf{S}_E^n) \\
&\overset{\text{(c)}}{=} I(M,K;Y_F^n, S_{B,F}^n | Y_R^n, Z_F^n, \mathbf{S}_E^n) \\
&= I(M,K;Y_F^n | Y_R^n, Z_F^n, \mathbf{S}_E^n) + I(M,K;S_{B,F}^n | \mathbf{Y}^n, Z_F^n, \mathbf{S}_E^n) \\
&\leq I(M,K,\mathbf{S}_E^n, Y_R^n, X_F^n; Y_F^n | Z_F^n) + I(M,K;S_{B,F}^n | \mathbf{Y}^n, Z_F^n, \mathbf{S}_E^n) \\
&= I(X_F^n; Y_F^n | Z_F^n) + \sum_{i=1}^n I(M,K;S_{B,F,i} | \mathbf{Y}^n, Z_F^n, S_{B,F}^{i-1}, \mathbf{S}_E^n) \\
&= H(Y_F^n | Z_F^n) - \sum_{i=1}^n H(Y_{F,i} | X_{F,i}, Z_{F,i}) \\
&\qquad + \sum_{i=1}^n I(M,K;S_{B,F,i} | \mathbf{Y}^n, Z_F^n, S_{B,F}^{i-1}, \mathbf{S}_E^n) \\
&\leq \sum_{i=1}^n H(Y_{F,i} | Z_{F,i}) - \sum_{i=1}^n H(Y_{F,i} | X_{F,i}, Z_{F,i}) \\
&\qquad + \sum_{i=1}^n I(M,K,\mathbf{Y}^n, Z_F^n, S_{B,F,\tilde{i}}, S_{E,F,\tilde{i}} S_{A,R}^n; S_{B,F,i} | \mathbf{S}_{E,i}) \\
&\leq nI(X_{F,J}; Y_{F,J} | Z_{F,J}, J) + nI(U_{1,F}; S_{B,F,J'} | S_{E,F,J'}) \\
&= n(I(X_{F,J}; Y_{F,J} | V_{2,F}) - I(X_{F,J}; Z_{F,J} | V_{2,F})) + nI(U_{1,F}; S_{B,F,J'} | S_{E,F,J'})
\end{aligned}
$$

where $V_{2,F} \overset{\text{def}}{=} J$, (a) follows from the hypothesis $I(M,K;\mathbf{Z}^n, \mathbf{S}_E^n) = o(n)$, (b) from the fact that $I(M,K;Y_R^n | \mathbf{S}_E^n, \mathbf{Z}^n) = 0$, which we show below, and (c) from the Markov chain $(M,K,\mathbf{Y}^n, Z_F^n, \mathbf{S}_A^n) - S_{E,R}^n - S_{B,R}^n$.

$$
0 = I(\mathbf{S}_A^n, M, K; Y_R^n | \mathbf{Z}^n) \overset{\text{(a)}}{=} I(\mathbf{S}_E^n, \mathbf{S}_A^n, M, K; Y_R^n | \mathbf{Z}^n) \geq I(M,K;Y_R^n | \mathbf{S}_E^n, \mathbf{Z}^n) \ ,
$$

where (a) follows from the Markov chain $\mathbf{S}_E^n - (\mathbf{S}_A^n, M, K) - \mathbf{Z}^n - Y_R^n$. By non-negativity of mutual information, $I(M, K; Y_R^n | \mathbf{S}_E^n, \mathbf{Z}^n) = 0$ as claimed above.

Thus, we have shown that if $(R_1, R_2) \in \mathcal{R}$, then there must exist independent random variables $U_{1,F}$ and $V_{2,F}$ such that $U_{1,F} - \mathbf{S}_A - (\mathbf{S}_B, \mathbf{S}_E)$ and $V_{2,F} - \mathbf{X} - (\mathbf{Y}, \mathbf{Z})$ are Markov chains and

$$R_{\mathsf{SM}} \leq I(X_F, Y_F) + I(X_R; Y_R) - I(U_{1,F}; S_{A,F} | S_{B,F}),$$
$$R_{\mathsf{SK}} + R_{\mathsf{SM}} \leq I(X_F; Y_F | V_{2,F}) - I(X_F; Z_F | V_{2,F}) + I(U_{1,F}; S_{B,F} | S_{E,F}).$$

The form of the right hand sides above further allow us to assert that the $U_{1,F}$ above may be independent of $S_{A,R}$. This completes the proof.

D.3 Proof of Proposition 4.4 (Gaussian Example)

While we stated the Theorems 4.1 and 4.2 only for finite alphabets, the results can be extended to continuous alphabets. We note that the scalar Gaussian problem satisfies the conditions of Theorem 4.2 (along with Remark 1 following it).

Observe that in the notation of Theorem 2, $S_{A,F} = S_A$ and $S_{B,F} = S_B$. Further, $S_{A,R}, S_{B,R}, S_{E,F}$, and $S_{E,R}$ are absent (assumed to be constants). When, $\mathrm{SNR}_{\mathrm{Eve}} \geq \mathrm{SNR}_{\mathrm{Bob}}$, we have $X_R = X, Y_R = Y$, and $Z_R = Z$, and the forwardly degraded sub-channel is absent (again, we may take the random variables of this sub-channel to be constants). When $\mathrm{SNR}_{\mathrm{Bob}} \geq \mathrm{SNR}_{\mathrm{Eve}}$, we have $X_F = X, Y_F = Y$, and $Z_F = Z$ and the reversely degraded sub-channel is absent. Hence, from theorem 2, $\mathcal{R}$ is given by

the union of $\tilde{\mathcal{R}}(p)$ over all joint distributions p. Also, $\tilde{\mathcal{R}}(p)$ is described by

$$R_{\mathsf{SM}} \leq I(X_F; Y_F) + I(X_R; Y_R) - I(U_1; S_A|S_B), \tag{D.34}$$

$$R_{\mathsf{SK}} + R_{\mathsf{SM}} \leq I(X_F; Y_F|V_2) - I(X_F; Z_F|V_2) + I(U_1; S_B). \tag{D.35}$$

When specialized to the Gaussian case above, it is easy to see that

$$I(X_F; Y_F) + I(X_R; Y_R) \leq C_Y, \text{ and}$$
$$I(X_F; Y_F|V_2) - I(X_F; Z_F|V_2) \leq [C_Y - C_Z]_+,$$

where $C_Y = \frac{1}{2}\log(1 + \mathrm{SNR}_{\mathrm{Bob}})$ and $C_Z = \frac{1}{2}\log(1 + \mathrm{SNR}_{\mathrm{Eve}})$. These bounds are simultaneously achieved when p is such that V_2 is a constant and X is Gaussian of variance $\mathrm{SNR}_{\mathrm{Bob}}$. Hence, we may rewrite, the conditions above as

$$R_{\mathsf{SM}} \leq C_Y - I(U_1; S_A) + I(U_1; S_B), \tag{D.36}$$

$$R_{\mathsf{SK}} + R_{\mathsf{SM}} \leq [C_Y - C_Z]_+ + I(U_1; S_B). \tag{D.37}$$

Now we show outerbounds to the above $\tilde{\mathcal{R}}(p)$ which match the two conditions in proposition 4.4. It will also become clear that a jointly Gaussian choice for p in fact achieves these outerbound thus completing the proof. We first derive an upperbound on R_{SM} which matches the first condition in proposition 4.4. From the two inequalities (D.36) and (D.37) above, we have

$$R_{\mathsf{SM}} \leq C_Y - I(U_1; S_A) + I(U_1; S_B), \tag{D.38}$$

$$R_{\mathsf{SM}} \leq [C_Y - C_Z]_+ + I(U_1; S_B). \tag{D.39}$$

Using the entropy power inequality (Theorem A.1),

$$\exp(2h(S_B|U)) \geq \exp(2h(S_A|U)) + \exp(2h(N_{\text{source}}))$$

Using this in (D.38), we may write

$$\begin{aligned} e^{2R_{\mathsf{SM}}} &\leq e^{2(C_Y + I(U_1;S_B) - h(S_A))} \cdot \left(e^{2(h(S_B) - I(U_1;S_B))} - e^{2h(N_{\text{source}})}\right) \\ &= e^{2(C_Y - h(S_A) + h(S_B))} - e^{2(C_Y - h(S_A) + h(N_{\text{source}}))} \cdot e^{2I(U_1;S_B)} \\ &\overset{(a)}{\leq} e^{2(C_Y - h(S_A) + h(S_B))} - e^{2R_{\mathsf{SM}}} \cdot e^{2(C_Y - [C_Y - C_Z]_+ - h(S_A) + h(N_{\text{source}}))}, \end{aligned}$$

where (a) results from (D.39). Rearranging, we have

$$\begin{aligned} R_{\mathsf{SM}} &\leq \frac{1}{2} \log \frac{\exp\left\{2(C_Y - h(S_A) + h(S_B))\right\}}{1 + \exp\left\{2(C_Y - [C_Y - C_Z]_+ - h(S_A) + h(N_{\text{source}}))\right\}} \\ &= \frac{1}{2} \log \frac{(1 + \text{SNR}_{\text{Bob}})(1 + \text{SNR}_{\text{src}})}{1 + \text{SNR}_{\text{src}} + \min(\text{SNR}_{\text{Bob}}, \text{SNR}_{\text{Eve}})}, \end{aligned}$$

which is the first condition in proposition 4.4. Now let us fix R_{SM} such that it satisfies this condition. Let us rewrite (D.36) as follows

$$h(S_A|U) \geq (R_{\mathsf{SM}} - C_Y + h(S_A) - h(S_B)) + h(S_B|U).$$

The entropy power inequality (Theorem A.1) implies that

$$\begin{aligned} \exp(2h(S_B|U)) &\geq \exp(2h(S_A|U)) + \exp(2h(N_{\text{source}})) \\ &\geq \exp(2(R_{\mathsf{SM}} - C_Y + h(S_A) - h(S_B))) \exp(2h(S_B|U)) + 1. \end{aligned}$$

Since

$$\begin{aligned} R_{\mathsf{SM}} &\leq \frac{1}{2}\log\frac{(1+\mathrm{SNR}_{\mathrm{Bob}})(1+\mathrm{SNR}_{\mathrm{src}})}{1+\mathrm{SNR}_{\mathrm{src}}+\min(\mathrm{SNR}_{\mathrm{Bob}},\mathrm{SNR}_{\mathrm{Eve}})} \\ &\leq \frac{1}{2}\log\frac{(1+\mathrm{SNR}_{\mathrm{Bob}})(1+\mathrm{SNR}_{\mathrm{src}})}{\mathrm{SNR}_{\mathrm{src}}} \\ &= C_Y - h(S_A) + h(S_B), \end{aligned}$$

we have

$$\exp(2h(S_B|U)) \geq \frac{1}{1-\exp(2(R_{\mathsf{SM}} - C_Y + h(S_A) - h(S_B)))}.$$

From (D.37),

$$\begin{aligned} \exp\left\{2R_{\mathsf{SK}}\right\} &\leq \exp\left\{2\big([C_Y - C_Z]_+ + h(S_B) - h(S_B|U) - R_{\mathsf{SM}}\big)\right\} \\ &\leq e^{2([C_Y-C_Z]_+ + h(S_B) - R_{\mathsf{SM}})} \cdot \left(1 - \exp\left\{2(R_{\mathsf{SM}} - C_Y + h(S_A) - h(S_B))\right\}\right) \\ &\leq e^{2([C_Y-C_Z]_+ - C_Y)} \cdot \left(e^{2(C_Y + h(S_B) - R_{\mathsf{SM}})} - e^{2h(S_A)}\right) , \end{aligned}$$

which evaluates to the second condition required. The inequalities used above are tight under a Gaussian choice for the auxiliary random variable, which thereby completes the proof of the proposition.

Bibliography

[1] Ahlswede, R. Multi-way communication channels. In *Proc. 2nd Int. Symp. Information Theory* (1971), pp. 103–135.

[2] Ahlswede, R. The capacity region of a channel with two senders and two receivers. *Annals of Probability 2*, 5 (1974), 805–814.

[3] Ahlswede, R., and Csiszár, I. Common randomness in information theory and cryptography – part I: Secret sharing. *IEEE Trans. Inform. Theory 39*, 4 (July 1993), 1121–1132.

[4] Avestimehr, A., Diggavi, S., and Tse, D. Approximate capacity of gaussian relay networks. In *IEEE International Symposium on Information Theory, 2008. ISIT 2008* (2008), pp. 474–478.

[5] Berger, T. Multiterminal source coding. In *Lecture Notes presented at CISM Summer School on the Information Theory Approach to Communications* (1977).

[6] Berrou, C., Glavieux, A., and Thitimajshima, P. Near Shannon limit error-correcting coding and decoding: Turbo-codes. In *IEEE International Conference on Communications, 1993. ICC 93. Geneva. Technical Program, Conference Record* (1993), vol. 2.

[7] Blachman, N. The convolution inequality for entropy powers. *IEEE Transactions on Information Theory 11* (1965), 267.

[8] Boyd, S., and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

[9] Carleial, A. Interference channels. *IEEE Transactions on Information Theory 24*, 1 (January 1978), 60–70.

[10] Chang, C. unpublished manuscript, 2006.

[11] Chaum, D. Untraceable electronic mail, return addresses, and digital pseudonyms.

[12] Chen, J., and Berger, T. The capacity of finite-state Markov channels with feedback. *IEEE Transactions on Information Theory 51*, 3 (2005), 780–798.

[13] Chen, Y., and Han Vinck, A. Wiretap channel with side information. *Information Theory, IEEE Transactions on 54*, 1 (Jan. 2008), 395–402.

[14] Chuang, I., and Nielsen, M. *Quantum Information*. Cambridge University Press, 2000.

[15] Chung, S., Forney Jr, G., Richardson, T., Urbanke, R., Inc, A., and Chelmsford, M. On the design of low-density parity-check codes within 0.0045 dB ofthe Shannon limit. *IEEE Communications letters 5*, 2 (2001), 58–60.

[16] Costa, M. H. M. On the Gaussian interference channel. *IEEE Transactions on Information Theory 31* (1985), 607–615.

[17] Cover, T. Broadcast channels. *IEEE Trans. Inform. Theory 18*, 1 (1972), 2–14.

[18] Cover, T. Comments on broadcast channels. *IEEE Transactions on Information Theory 44*, 6 (1998), 2524–2530.

[19] Cover, T., and Gamal, A. Capacity theorems for the relay channel. *IEEE Transactions on Information Theory 25*, 5 (1979), 572–584.

[20] Cover, T., and Pombra, S. Gaussian feedback capacity. *Information Theory, IEEE Transactions on 35*, 1 (1989), 37–43.

[21] Cover, T., and Thomas, J. *Elements of Information Theory*. John Wiley and Sons, 1991.

[22] Csiszár, I., and Körner, J. Broadcas channels with confidential messages. *IEEE Trans. Inform. Theory 24*, 3 (1978).

[23] Csiszár, I., and Körner, J. G. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, Inc., Orlando, FL, USA, 1982.

[24] Csiszár, I., and Narayan, P. Common randomness and secret key generation with a helper. *IEEE Trans. Inform. Theory 46*, 2 (2000), 344–366.

[25] Dembo, A., and Zeitouni, O. *Large Deviations Techniques and Applications, 2nd ed.* Stochastic Modelling and Applied Probability. Springer, 1998.

[26] Devroye, N., Mitran, P., and Tarokh, V. Achievable rates in cognitive radio channels. *IEEE Transactions on Information Theory 52*, 5 (May 2006), 1813–1827.

[27] Devroye, N., Vu, M., and Tarokh, V. Achievable rates and scaling laws in cognitive radio channels. *EURASIP Journal on Wireless Communications and Networking, special issue on Cognitive Radio and Dynamic Spectrum Sharing Systems* (2008).

[28] Diffie, W., and Hellman, M. New directions in cryptography. *IEEE Transactions on Information Theory 22*, 6 (1976), 644–654.

[29] Durrett, R. *Probability: Theory and Examples, 3rd ed.* Information and System Sciences Series. Duxbury, 2004.

[30] El Gamal, A. The capacity of a class of broadcast channels. *IEEE Trans. Inform. Theory 25*, 2 (1979), 166–169.

[31] El Gamal, A., and Aref, M. The capacity of the semideterministic relay channel. *Trans. Automat. Contr 19* (1974).

[32] Elia, N. When Bode Meets Shannon: Control-Oriented Feedback Communication Schemes. *IEEE Trans. Autom. Control 49*, 9 (2004), 1477–1487.

[33] Elia, N., and Mitter, S. Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control 46*, 9 (2001), 1384–1400.

[34] Eswaran, K., and Gastpar, M. An Observation about Feedback from Cognitive Radio. In *Proceedings of the 42nd Annual Asilomar Conference* (2008).

[35] Eswaran, K., Prabhakaran, V., and Ramchandran, K. Secret communication using sources and channels. In *Proc. 42nd Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA, 2008).

[36] Eswaran, K., and Ramchandran, K. Secrecy via sources and channels: Secure transmission of an independent source with decoder side information. In *Proceedings of the 46th Annual Allerton Conference on Communication, Control and Computation* (Monticello, IL, Sept 2008).

[37] Etkin, R., Tse, D., and Wang, H. Gaussian interference channel capacity to within one bit. *IEEE Transactions on Information Theory 54*, 12 (2008), 5534–5562.

[38] Gallager, R. *Information Theory and Reliable Communication.* John Wiley and Sons, 1968.

[39] Gastpar, M. On capacity under receive and spatial spectrum-sharing constraints. *IEEE Transactions on Information Theory 53*, 2 (February 2007), 471–487.

[40] Gastpar, M. On Capacity Under Receive and Spatial Spectrum-Sharing Constraints. *Information Theory, IEEE Transactions on 53*, 2 (2007), 471–487.

[41] Gastpar, M., and Vetterli, M. On the capacity of large Gaussian relay networks. *IEEE Transactions on Information Theory 51*, 3 (2005), 765–779.

[42] Ghasemi, A., and Sousa, E. Capacity of fading channels under spectrum-sharing constraints. In *IEEE International Conference on Communications* (June 2006), vol. 10, pp. 4373–4378.

[43] Goldsmith, A., Jafar, S., Maric, I., and Srinivasa, S. Breaking spectrum gridlock with cognitive radios: An information theoretic perspective. *Proceedings of the IEEE* (2008).

[44] Grover, P., and Sahai, A. The vector Witsenhausen problem as assisted interference cancelation. *Accepted to International Journal on Systems, Control and Communications (IJSCC)* (2008).

[45] Han, T., and Kobayashi, K. A new achievable rate region for the interference channel. *IEEE Trans. Inform. Theory 27*, 1 (1981), 49–60.

[46] Jovičić, A., and Viswanath, P. Cognitive radio: An information-theoretic perspective, 2006.

[47] Khalil, K., Youssef, M., Koyluoglu, O. O., and Gamal, H. E. On the delay limited secrecy capacity of fading channels, 2009.

[48] Khisti, A., Diggavi, S., and Wornell, G. Secret key generation using correlated sources and noisy channels. In *IEEE International Symposium on Information Theory* (Toronto, July 2008).

[49] Koch, T., Lapidoth, A., and Sotiriadis, P. A channel that heats up. In *IEEE International Symposium on Information Theory, 2007. ISIT 2007* (2007), pp. 906–910.

[50] Korner, J., and Marton, K. General broadcast channels with degraded message sets. *IEEE Transactions on Information Theory 23*, 1 (1977), 60–64.

[51] Kramer, G. Feedback strategies for white Gaussian interference networks. *Information Theory, IEEE Transactions on 48*, 6 (2002), 1423–1438.

[52] Kramer, G., Gastpar, M., and Gupta, P. Cooperative strategies and capacity theorems for relay networks. *IEEE Transactions on Information Theory 51*, 9 (2005), 3037–3063.

[53] Kumar, P., and Varaiya, P. *Stochastic systems: estimation, identification and adaptive control*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1986.

[54] Liao, H. A coding theorem for multiple access communications. In *Proceedings of IEEE International Symposium on Information Theory* (1972).

[55] MacKay, D., and Neal, R. Near Shannon limit performance of low density parity check codes. *Electronics letters 33*, 6 (1997), 457–458.

[56] Maric, I., Goldsmith, A., Kramer, G., and Shamai(Shitz), S. On the capacity of interference channels with a cognitive transmitter. In *2007 Workshop on Information Theory and Applications (ITA)* (UCSD, La Jolla, CA, January 2007).

[57] Maric, I., Goldsmith, A., Kramer, G., and Shamai(Shitz), S. On the capacity of interference channels with a partially-cognitive transmitter. In *2007 IEEE International Symposium on Information Theory* (Nice, France, June 2007).

[58] Maric, I., Yates, R., and Kramer, G. Capacity of interference channels with partial transmitter cooperation. *IEEE Transactions on Information Theory 53*, 10 (October 2007), 3536–3548.

[59] Marton, K. A coding theorem for the discrete memoryless broadcast channel. *IEEE Transactions on Information Theory 25*, 3 (1979), 306–311.

[60] Maurer, U. M. Secret key agreement by public discussion from common information. *IEEE Trans. Inform. Theory 39*, 3 (May 1993), 733–742.

[61] Mitola, J. *Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio*. PhD thesis, Royal Institute of Technology, Sweden, 2000.

[62] Mitter, S. Control with limited information. *European Journal of Control 7*, 2-3 (2001), 122–131.

[63] Mudumbai, R., Hespanha, J., Madhow, U., and Barriac, G. Scalable feedback control for distributed beamforming in sensor networks. In *International Symposium on Information Theory* (Adelaide, Australia, September 2005).

[64] Mudumbai, R., Wild, B., Madhow, U., and Ramchandran, K. Distributed beamforming using 1 bit feedback: From concept to realization. In *Allerton Conference on Communication, Control, and Computing* (Monticello, Illinois, September 2006).

[65] Ozarow, L. The capacity of the white Gaussian multiple access channel with feedback. *Information Theory, IEEE Transactions on 30*, 4 (1984), 623–629.

[66] Permuter, H., Cuff, P., Van Roy, B., and Weissman, T. Capacity of the trapdoor channel with feedback. *IEEE Transactions on Information Theory 54*, 7 (2008), 3150–3165.

[67] Permuter, H. H., Weissman, T., and Goldsmith, A. J. Finite state channels with time-invariant deterministic feedback. *Information Theory, IEEE Transactions on 55*, 2 (Feb. 2009), 644–662.

[68] Prabhakaran, V., and Ramchandran, K. A separation result for secure communication. In *Allerton Conference* (Sept 2007).

[69] Rivest, R., Shamir, A., and Adleman, L. A method for obtaining digital signatures and public-key cryptosystems.

[70] Sahai, A. *Anytime information theory*. PhD thesis, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2000.

[71] Sahai, A. Evaluating channels for control: Capacity reconsidered. In *American Control Conference, 2000. Proceedings of the 2000* (2000), vol. 4.

[72] Sahai, A., Hoven, N., and Tandra, R. Some fundamental limits on cognitive radio. In *Allerton Conference on Communication, Control, and Computing* (Monticello, Illinois, Oct. 2004).

[73] Sahai, A., and Mitter, S. The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link Part I: scalar systems. *Arxiv preprint cs/0601007* (2006).

[74] Sato, H. Two-user communication channels. Tech. Rep. B75-29, University of Hawaii, Honolulu, HI, October 1975.

[75] Schalkwijk, J. The binary multiplying channel–A coding scheme that operates beyond Shannon's inner bound region (Corresp.). *Information Theory, IEEE Transactions on 28*, 1 (1982), 107–110.

[76] Schalkwijk, J. On an extension of an achievable rate region for the binary multiplying channel (Corresp.). *Information Theory, IEEE Transactions on 29*, 3 (1983), 445–448.

[77] Shannon, C. Communication theory of secrecy systems. *Bell Systems Technical Journal 28* (1949), 656–715.

[78] Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal 27* (1948), 379–423, 623–656.

[79] Slepian, D., and Wolf, J. A coding theorem for multiple access channels with correlated sources. *Bell Syst. Tech. J 52*, 7 (1973), 1037–1076.

[80] Stam, A. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control 2* (1959), 101–112.

[81] Tandra, R., and Sahai, A. Fundamental limits on detection in low SNR under noise uncertainty. In *WirelessCom 05 Symposium on Signal Processing* (Maui, Hawaii, June 2005).

[82] Tandra, R., and Sahai, A. SNR walls for signal detection. *IEEE Journal on Selected Topics in Signal Processing 2*, 1 (February 2008), 4–17.

[83] Tatikonda, S., and Mitter, S. Control under communication constraints. *IEEE Transactions on Automatic Control 49*, 7 (2004), 1056–1068.

[84] Tatikonda, S., Sahai, A., and Mitter, S. Control of LQG systems under communication constraints. In *Decision and Control, 1998. Proceedings of the 37th IEEE Conference on* (1998), vol. 1.

[85] Van Der Meulen, E. Three-terminal communication channels. *Advances in Applied Probability* (1971), 120–154.

[86] Van der Meulen, E. A survey of multi-way channels in information theory: 1961-1976. *IEEE Transactions on Information Theory 23*, 1 (1977), 1–37.

[87] Venkitasubramaniam, P., He, T., and Tong, L. Anonymous networking amidst eavesdroppers. *IEEE Transactions on Information Theory 54*, 6 (2008), 2770–2784.

[88] Viswanath, P., and Tse, D. Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality. *IEEE Transactions on Information Theory 49*, 8 (2003), 1912–1921.

[89] Viswanathan, H. Capacity of Markov channels with receiver CSI and delayed feedback. *Information Theory, IEEE Transactions on 45*, 2 (1999), 761–771.

[90] Weingarten, H., Steinberg, Y., and Shamai, S. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Transactions on Information Theory 52*, 9 (2006), 3936–3964.

[91] Wigger, M., and Gastpar, M. The pre-log of Gaussian broadcast with feedback can be two. In *IEEE International Symposium on Information Theory, 2008. ISIT 2008* (2008), pp. 1258–1262.

[92] Witsenhausen, H. A counterexample in stochastic optimum control. *SIAM Journal on Control 6* (1968), 131.

[93] Wyner, A. The wire-tap channel. *Bell System Technical Journal 54*, 8 (1975), 1355–1387.

[94] Xie, L., and Kumar, P. An achievable rate for the multiple-level relay channel. *IEEE Transactions on Information Theory 51*, 4 (2005), 1348–1358.

[95] Xu, J., and Chen, B. Broadcast confidential and public messages. In *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on* (2008), pp. 630–635.

[96] Yang, S., Kavcic, A., and Tatikonda, S. Feedback capacity of finite-state machine channels. *IEEE Transactions on Information Theory 51*, 3 (2005), 799–810.

[97] Ziv, J. Universal decoding for finite-state channels. *IEEE Transactions on Information Theory 31*, 4 (1985), 453–460.