PHYSICAL-LAYER SECURITY

A Dissertation
Presented to
The Academic Faculty

By

Matthieu Bloch

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
in
Electrical and Computer Engineering

School of Electrical and Computer Engineering
Georgia Institute of Technology
August 2008

PHYSICAL-LAYER SECURITY

Approved by:

Dr. Steven W. McLaughlin, Advisor
School of ECE
Georgia Institute of Technology

Dr. Aaron D. Lanterman
School of ECE
Georgia Institute of Technology

Dr. João Barros
Department of Computer Science
University of Porto

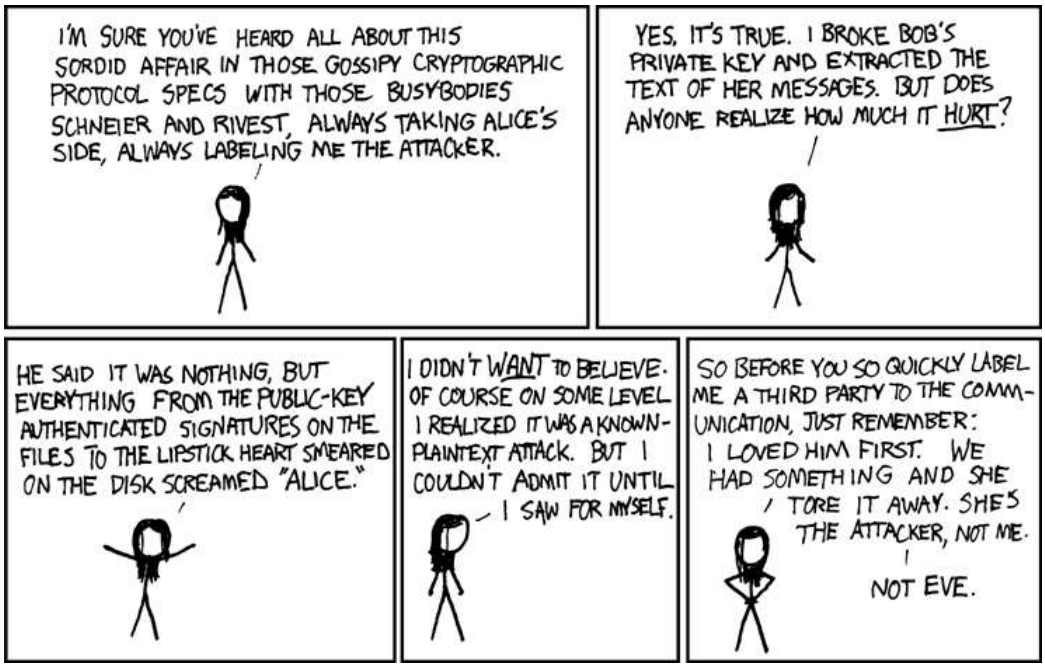
Dr. Faramarz Fekri
School of ECE
Georgia Institute of Technology

Dr. Jean Bellissard
School of Mathematics
Georgia Institute of Technology

Date Approved: April 16th, 2008

DEDICATION

To Alice, Bob, and Eve.



Original picture: http://xkcd.com/177/

ACKNOWLEDGMENTS

I have been extremely fortunate to work under the supervision of Dr. Steven McLaughlin. His thoughtful guidance, unconditional support, and endless patience have made my years in graduate school an enjoyable and rewarding experience. I am especially grateful for all the opportunities that he provided me. In truth, this dissertation only reflects a fraction of what I have learned as a Ph.D. student. Thanks to Steve, I have largely benefited from the globalization of education and research, and I have come to grasp what the *flat world* praised by Thomas Friedman is all about. Not only did I share my time between the campuses of Georgia Tech in France and in the United States, but I also spent three months in India at the Indian Institute of Technology Madras and three months in Portugal at the University of Porto.

I also wish to thank Dr. João Barros, Dr. Jean Bellissard, Dr. Faramarz Fekri, and Dr. Aaron Lanterman for being part of my dissertation committee, and the CNRS and Texas Instruments for supporting my research.

My survival at Georgia Tech depended on many friends. The local French mafia definitely helped me keep a certain level of sanity. Jérôme, David, Nicolas, Arnaud, Yannick, and Xavier never really managed to drag me to a night-club, but we have had a lot of fun recreating a bit of a French atmosphere in Atlanta. I was also very fortunate to be surrounded by great labmates, who made a point proving that working and having fun was indeed compatible. Many thanks to the Indian click, Badri, Rajesh, Aru, Shayan, Yogesh, and Arun, the American gang, Tim and his family, David, Will, Martin, and Kevin, and the lone Slovenian, Demijan, who managed to bear with me both as labmate and roommate.

My short stay in India was a tremendously rewarding experience, and Andrew Thangaraj deserves most of the praise for making it a success. I also extend my sincere gratitude to Anil, Namita, Shanti, and Subu for their kind welcome and many games of statistical basketball. I am also very grateful to Sunil and Pradeep, without whom I would not have dared wandering out of the IIT Madras campus much, and to Yogesh and his family for hosting me in Mysore.

I experienced European cooperation at its best in Porto, where João Barros and Miguel

Rodrigues somehow managed to have England, France, Germany and Portugal represented in the same office. Many thanks to Rui(s), Luisa, João(s), Paulo, Gerhard, William, and Miguel(s) for all the fun both inside and outside the lab, and my heartfelt thanks to João, Ana, Daniel, and Andre.

I have lived in far too many different place to list them all, but I still think of my parents' house as home. I can probably count the number of days I spent there over the last 5 years (if not 10 years), but it would not be fair not to acknowledge the constant support of my family. Skype and the Internet definitely helped making distances seems a little shorter. I suspect that they sometimes seriously wonder why I have studied nonsensical things for so long, but the truth is I am not sure I know myself...

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

As wireless networks continue to flourish worldwide and play an increasingly prominent role, it has become crucial to provide effective solutions to the inherent security issues associated with a wireless transmission medium. Unlike traditional solutions, which usually handle security at the application layer, the primary concern of this thesis is to analyze and develop solutions based on coding techniques at the physical layer.

First, an information-theoretically secure communication protocol for quasi-static fading channels was developed and its performance with respect to theoretical limits was analyzed. A key element of the protocol is a reconciliation scheme for secret-key agreement based on low-density parity-check codes, which is specifically designed to operate on non-binary random variables and offers high reconciliation efficiency.

Second, the fundamental trade-offs between cooperation and security were analyzed by investigating the transmission of confidential messages to cooperative relays. This information-theoretic study highlighted the importance of jamming as a means to increase secrecy and confirmed the importance of carefully chosen relaying strategies.

Third, other applications of physical-layer security were investigated. Specifically, the use of secret-key agreement techniques for alternative cryptographic purposes was analyzed, and a framework for the design of practical information-theoretic commitment protocols over noisy channels was proposed.

Finally, the benefit of using physical-layer coding techniques beyond the physical layer was illustrated by studying security issues in client-server networks. A coding scheme exploiting packet losses at the network layer was proposed to ensure reliable communication between clients and servers and security against colluding attackers.

CHAPTER 1
INTRODUCTION

The advent and success of the Internet, together with the large-scale deployment of wireless networks, now allows ubiquitous access to communication networks; however, the pervasive access to online services often comes at the expense of security. For instance, the broadcast nature of wireless communications makes them particularly sensitive to eavesdropping. Given our increased dependency on network services, the interception and malicious use of data could have a tremendous societal cost, and consequently, there is an increasing need for secure communication solutions. Unlike traditional approaches, which handle security at the application layer, *physical-layer security* aims at developing effective secure communication schemes exploiting the properties of the physical layer. As we discuss in this dissertation, this new paradigm has the potential of strengthening the security of existing systems by introducing a level of information-theoretic security, now widely accepted as a stronger notion than computational security.

1.1 Motivating example: security issues in wireless communications

In addition to standard security issues, wireless systems face very specific security vulnerabilities caused by the inherent openness of wireless media. First, wireless channels are susceptible to *channel jamming*. An attacker can easily jam physical communication channels and prevent legitimate users from accessing a network. This threat is all the more difficult to counter as it aims at disrupting traffic and not intercepting information. Second, without proper authentication mechanisms, an attacker can gain *unauthorized access* to network resources and bypass security infrastructures such as firewalls. Finally, because of the open nature of wireless media, *eavesdropping* can be performed without resorting to advanced technological devices. In principle, even legitimate users in a network could be regarded as potential eavesdroppers.

Solutions for the aforementioned security issues have been engineered using a layered approach. Historically, this approach has been used to simplify the design of communication

protocols – with little consideration for security. Figure 1 illustrates the various layers considered in a typical wireless communication protocol, and indicates their specific purposes. For instance, channel coding is implemented at the Physical (PHY) layer, which ensures that all above layers operate essentially on error-free information, and admission control is handled at the Medium Access Control (MAC) layer. Although the design of modern communication protocols does not follow a strict layered approach and considers cross-layer aspects, layering remains a convenient conceptual representation that we use in the rest of this dissertation.

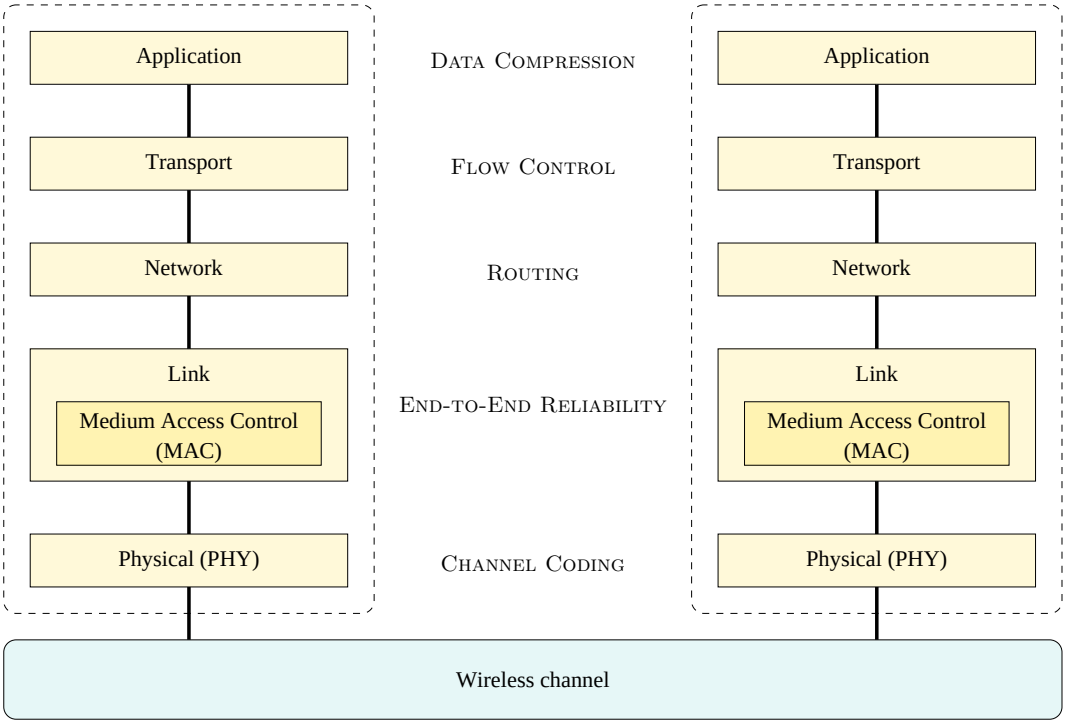


Figure 1. Layered protocol architecture.

As examples of layer-specific security solutions, spread-spectrum modulation techniques are used at the PHY layer to mitigate channel jamming, authentication mechanisms are implemented at the Link layer to prevent unauthorized access, and message encryption is performed at the Application layer to render eavesdropping useless. One can notice that channel jamming and unauthorized access, which are vulnerabilities at the PHY layer and Link layer, respectively, are handled by security solutions at their respective layers; however, eavesdropping, which is also a PHY layer vulnerability, is currently handled by a solution

at the application layer. One can naturally ask whether ignoring the physical phenomena occurring at the PHY layer is appropriate, and whether there exist security solutions against eavesdropping at the PHY layer.

1.2 Physical-layer security

To illustrate the general concept of physical layer security, consider the example of a three-node wireless network in Figure 2 where the communication between terminals T_1 and T_2 is being eavesdropped by an unauthorized terminal T_3. The communication channel between the legitimate users is called the *main channel*, whereas the communication channel between T_1 and T_3 is referred to as the *eavesdropper's channel*.

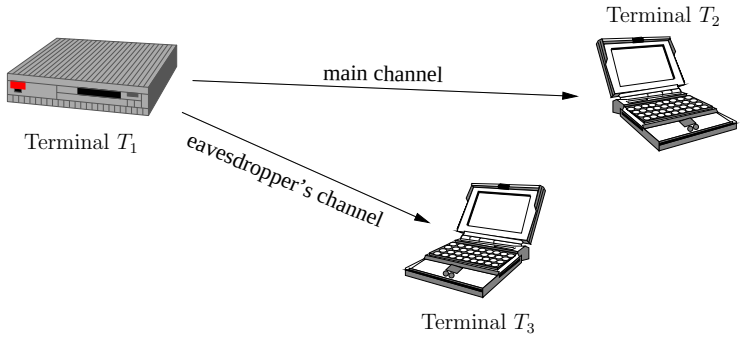


Figure 2. Illustration of eavesdropping scenario in wireless network.

When terminals T_2 and T_3 are not collocated, radiofrequency signals observed at the outputs of the main channel and eavesdropper's channel are usually different. Natural discrepancies are caused by physical phenomena, and for wireless communications, the most notable effects are *fading* and *path-loss*. Fading is a self-interference phenomenon that results from the multi-path propagation of radiofrequency waves while path-loss is simply the attenuation of wave amplitude with distance. As a consequence of these effects, if the transmission distance over the main channel is much smaller than the transmission distance over the eavesdropper's channel, one can expect the detection of signals at terminal T_3 to be much harder than at terminal T_2. For instance, if T_1 broadcasts a video stream, the signal obtained by T_3 is significantly degraded compared to the one received by T_2; this degradation can even prevent T_3 from understanding the content of the video stream. Presently, security solutions against eavesdropping totally disregard these effects and operate as if the

eavesdropper detected the same signal as the legitimate receiver. In contrast, the key idea of physical-layer security is to explicitly take into account differences at the PHY layer to better protect the messages exchanged over the main channel.

As we detail in Chapter 2, the three-node communication problem in Figure 1 can be studied from an information-theoretic perspective, and in this case, the security obtained by exploiting the PHY layer can be precisely quantified. This innate connection between physical-layer security and information-theoretic security provides the main motivation for this dissertation. Our study of physical-layer security is exclusively carried out from an information-theoretic and coding perspective but we acknowledge that the scope of physical-layer security goes well beyond these considerations. In particular, we do not consider a large class of techniques that aim to modify the PHY layer to impair potential eavesdroppers. Examples of such techniques are coded-division multiple-access signaling, which gives signals a noise-like appearance, and beamforming with smart antennas, which essentially prevents an eavesdropper located away from the legitimate receiver to detect signals.

1.3 Outline of the dissertation

This dissertation is organized as follows. Chapter 2 introduces fundamental concepts of information-theoretic security, summarizes the state of the art, and sets the notation used in subsequent chapters. Our main discussion and results on physical-layer security for wireless channels are contained in Chapters 3-5. Specifically, Chapter 3 introduces an efficient Slepian-Wolf compression algorithm for continuous random variables, which is a key element in the design of secure communication schemes for wireless channels. Chapter 4 presents a practical key agreement protocol for quasi-static fading wireless channels and analyzes its performance. Chapter 5 treats the more theoretical problem of trade-offs between cooperation and security in wireless environments. Chapter 6 and 7 discuss applications of physical-layer security beyond wireless communications. Chapter 6 treats the problem of information-theoretic commitment over noisy channels, and Chapter 7 discusses the design of network architectures exploiting ideas borrowed from physical-layer security. Finally, Chapter 8 summarizes our conclusions and points to areas for future research.

CHAPTER 2
FUNDAMENTALS OF INFORMATION-THEORETIC SECURITY

This chapter summarizes fundamental information-theoretic security results, which are key tools used for studying physical-layer security in this dissertation. To date, these results have been confined within the information theory community, arguably because the assumptions required are often judged impractical from a cryptographic perspective. We believe that information-theoretic security has the potential to significantly strengthen the security level of current systems, but we acknowledge that some claims found in the literature regarding the pertinence of information-theoretic results can be misleading. We attempt to avoid that pitfall and provide a fair comparison of computational security and information-theoretic security by highlighting the strengths and weaknesses of both approaches. In particular, we clearly state the assumptions used in information-theoretic models.

2.1 Notation and basic definitions

In the rest of this dissertation, scalars are denoted by normal letters (x), vectors of length n are denoted by boldface letters ($\mathbf{x}^n = (x_1, \ldots, x_n)$), and random variables are denoted by capital letters (X). The entropy of discrete random variables and the differential entropy of continuous random variables are denoted by $H(\cdot)$ and $h(\cdot)$, respectively, and the mutual information between two random variables, as defined in [1], is denoted by $I(\cdot;\cdot)$. The probability distribution of a random variable X taking values in a set $\mathcal{X}$ is denoted by $p_X(x)$, and the conditional distribution of $X \in \mathcal{X}$ given $Y \in \mathcal{Y}$ is denoted by $p_{X|Y}(x|y)$. Subscripts may be omitted to simplify notation, in which case the random variable considered should be inferred from the context.

The main focus of the present work is a three-party cryptographic scenario similar to the one illustrated in Figure 2, where a transceiver attempts to communicate with a legitimate receiver while being spied on by an eavesdropper. Following the tradition, the transceiver, receiver, and eavesdropper are named Alice, Bob, and Eve, respectively. The information sent by Alice consists of a set of *messages*, which are encoded into *codewords* to ensure reliable or secure transmissions. In cryptography, messages and codewords are often

referred to as *plaintexts* and *cyphertexts*, respectively. A *code* is the set of mechanisms by which messages are encoded into and retrieved from codewords. In all scenarios considered thereafter, it is assumed that the algorithms used to code messages and decode codewords are publicly known.

2.2 Principles and fundamental limits of modern cryptography

Modern cryptography is not limited to the analysis and design of encryption schemes, but also tackles issues such as data signature, message authentication, data integrity, etc. A detailed presentation of all these aspects goes well beyond the scope of this dissertation, and we restrict ourselves to a succinct description of *secret key* and *public key* encryption schemes. The objective of this section is merely to highlight the salient features of these systems and to provide a reasonable basis for comparison with information-theoretic schemes.

To guarantee the confidentiality of messages transmitted by Alice, codes are based on *keys*, which are secret sequences of bits only known to Alice or Bob. The goal of Eve is to *break* the codes used by Alice and Bob, that is, to retrieve messages from codewords without having knowledge of the keys. The security of encryption schemes is traditionally assessed in terms of *computational security*[1] and relies on assumptions limiting the computing resources of eavesdroppers. Essentially, computational security ensures that the amount of computing time or memory required to break a code is "unreasonable" with today's technology. Usually, a code is regarded as secure if the computational complexity of an eavesdropper's decoding algorithm is equivalent to that required for solving "hard" mathematical problems (NP-hard problems for example). This notion of security is widely used in current cryptographic protocols, but despite being satisfactory in many situations, this notion fails to guarantee security in the long term. For instance, many codes that were regarded as secure twenty years ago are now easily breakable with off-the-shelf computers. Consequently, encoding algorithms have to be regularly updated to face the increasing power of computers.

Figure 3 illustrates the principle of secret key encryption schemes (also called *symmetric* schemes). Alice and Bob are assumed to share a secret key $\mathbf{k}$ that is used to encode

[1] This type of security is called provable security in [2]

messages **m** or decode codewords **c**. The eavesdropper Eve has knowledge of the encoding and decoding algorithms and intercepts the codewords **c**, but does not know the key **k**.

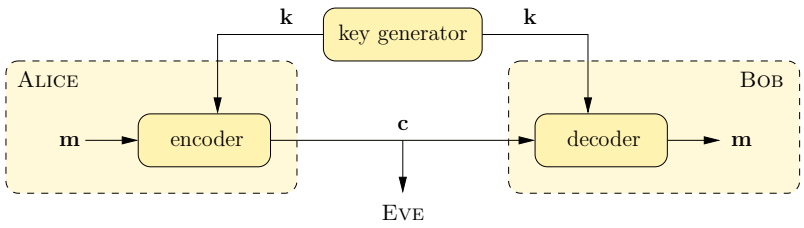


Figure 3. Principle of symmetric encryption.

Secret-key schemes offer the advantage of encrypting messages with relatively short keys and of operating at high rates. For instance, hardware implementations of the Advanced Encryption Standard (AES) can yield encryption rates on the order of gigabytes per second [3]; however, security relies not only on the existence of hardly breakable algorithms, but also on the ability of distributing secure keys efficiently between Alice and Bob.

Public-key encryption schemes (also called *asymmetric* schemes) were first proposed in the late 1970s as a solution to the key distribution problem. As shown in Figure 4, their principle is fundamentally different from that of symmetric schemes since Alice and Bob own distinct keys. The *public* key $\mathbf{k}_{pub}$ is publicly available and used by Alice for message encryption. The *private* key $\mathbf{k}_{priv}$ is kept secret and is only used for message decryption. In other words, the public key plays the role of an open vault that anyone can close but that cannot be opened by anybody but Bob.

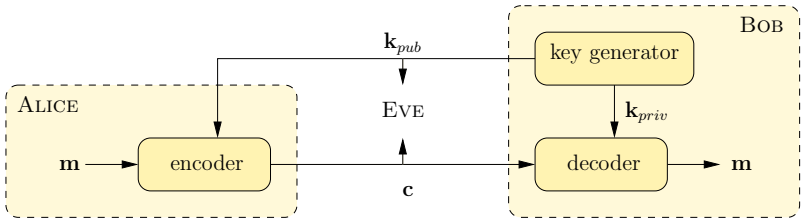


Figure 4. Principle of asymmetric encryption.

Evidently, this scheme is useful, provided that the knowledge of $\mathbf{k}_{pub}$ does not allow one to recover the key $\mathbf{k}_{priv}$. The keys are not independent in practice, but are usually constructed based on mathematical conjectures suggesting that recovering $\mathbf{k}_{priv}$ from $\mathbf{k}_{pub}$

cannot be done in a reasonable time. For instance, the security of the infamous RSA protocol is linked to the intractability of the decomposition of large integers into prime factors. Unfortunately, popular public-key schemes suffer from low encryption rates, typically several orders of magnitude slower than those of symmetric schemes.

To avoid the limitations associated with a computational measure of secrecy, one can consider a more stringent criterion based on an information-theoretic measure. Specifically, one can treat messages and codewords as random variables denoted by M and C, respectively, and decide that a codeword is secure if the Shannon uncertainty of the message after observing the codeword $H(M|C)$ is equal to the *a priori* uncertainty of the message $H(M)$. This definition of security is referred to as *unconditional security* and is now widely accepted as the strictest notion of security. This definition does not place any restriction on the resources of the eavesdropper, but there exist few practical methods satisfying the above criterion.

Before discussing unconditional security any further, it is worth clarifying the practical meaning of this notion. First of all, one should notice that there exist no unbreakable encryption schemes. In fact, if a message contains k bits of information, trying to guess every bit at random is a poor but valid eavesdropping strategy, whose probability of success is 2^{-k}. This result does not imply that securing data is impossible, but only points out that assessing the security of a system should ultimately be done with a probabilistic measure. The unconditional security criterion, which ensures that messages and codewords are statistically independent, essentially means that the aforementioned guessing strategy is the best strategy that an eavesdropper can implement to retrieve messages. In particular, there are no correlations between messages and codewords that could be exploited, and unconditionally secure schemes are immune to cryptanalysis techniques.

Analyzing modern cryptography schemes from the perspective of unconditional security yields a rather surprising and disappointing answer. In fact, Shannon proved that the only encryption scheme satisfying the unconditional security criterion is the so-called *one-time pad* [4] illustrated in Figure 5. Alice and Bob are assumed to share perfectly random secret keys whose size is at least as long as the messages that they wish to exchange, and they encrypt or decrypt messages by summing key bits and message bits modulo two.

Since each key bit can be used only once, Shannon's result states that it is impossible to ensure unconditional security with modern cryptographic techniques based on the repeated use of small secret keys. Actually, the result is even more disappointing since a one-time pad encryption is impractical unless there exist efficient means of distributing secret keys between Alice and Bob.

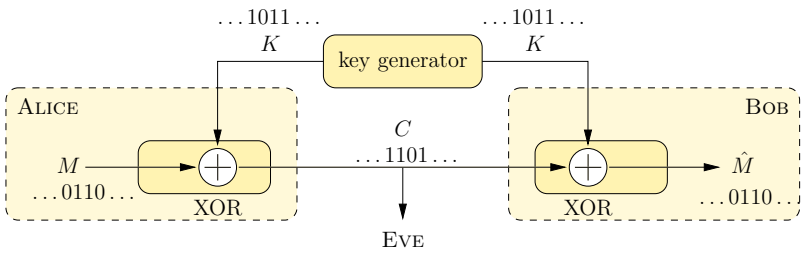


Figure 5. One-time pad encryption scheme.

2.3 The wiretap channel

Based on Shannon's result, one may believe that unconditional security is not achievable with practical systems; however, as we already pointed out in Chapter 1, the secure communication framework investigated in Figures 3-5 is overly pessimistic since it does not account for the physical reality of communication channels. Especially, it does not consider the degradation of signals because of noise. This observation naturally leads to the introduction of a more realistic communication model, now known as the *wiretap channel*, where noise in the main channel and eavesdropper's channel is explicitly introduced.

2.3.1 Wiretap channel model

The wiretap channel model was initially introduced by Wyner [5] and later refined by Csiszár and Körner [6]. Figure 6 illustrates the latter model, which is also called a broadcast channel with confidential messages.

Alice and Bob communicate over a *discrete broadcast channel* characterized by a discrete input alphabet $\mathcal{X}$, two discrete output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, and a probability transition function $p_{YZ|X}(y, z|x)$. The channel is also assumed to be memoryless, that is the transition

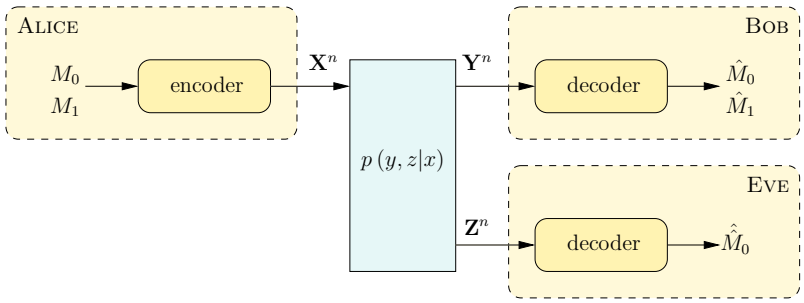


Figure 6. Broadcast channel with confidential messages (wiretap channel).

probability of a sequence of n symbols is given by

$$p\left(\mathbf{y}^n, \mathbf{z}^n | \mathbf{x}^n\right) = \prod_{i=1}^{n} p_{YZ|X}\left(y_i, z_i | x_i\right).$$

It is also assumed that Alice wishes to send a common message M_0 to both Bob and Eve and a private message M_1 to Bob only.

Definition 2.1. *A* $\left(2^{nR_0}, 2^{nR_1}, n\right)$ *code for the broadcast channel with confidential messages consists of the following.*

- *Two message sets* $\mathcal{M}_0 = \left\{1, 2, \ldots, 2^{nR_0}\right\}$ *and* $\mathcal{M}_1 = \left\{1, 2, \ldots, 2^{nR_1}\right\}$.
- *An encoding function (possibly stochastic)* $f_n : \mathcal{M}_0 \times \mathcal{M}_1 \to \mathcal{X}^n$, *which maps each message pair* $(m_0, m_1) \in \mathcal{M}_0 \times \mathcal{M}_1$ *to a codeword* $\mathbf{x}^n \in \mathcal{X}^n$.
- *Two decoding functions* $g_n : \mathcal{Y}^n \to \mathcal{M}_0 \times \mathcal{M}_1$ *and* $h_n : \mathcal{Z}^n \to \mathcal{M}_0$, *which map an observation* $\mathbf{y}^n$ *to a message pair* $(\hat{m}_0, \hat{m}_1)$ *and an observation* $\mathbf{z}^n$ *to a message* $\hat{\hat{m}}_0$.

The secrecy of message M_1 with respect to the eavesdropper is measured in terms of the *equivocation rate*

$$\frac{1}{n} H(M_1 | \mathbf{Z}^n),$$

and a rate tuple (R_0, R_1, R_e) is achievable for a broadcast channel with confidential messages if and only if, for any $\epsilon > 0$, there exists a $\left(2^{nR_0}, 2^{nR_1}, n\right)$ code such that

$$\begin{aligned} \mathrm{P}[g_n(\mathbf{Y}^n) \neq (M_0, M_1) \text{ or } h_n(\mathbf{Z}^n) \neq M_0] < \epsilon \quad & \text{(reliability condition)}, \\ \frac{1}{n} H(M_1 | \mathbf{Z}^n) \geq R_e - \epsilon \quad & \text{(secrecy condition)}. \end{aligned}$$

It is not *a priori* obvious whether the conditions defined above can be satisfied simultaneously. In fact, the reliability condition calls for an increased redundancy in the coding scheme, while the secrecy condition tends to limit this redundancy. Surprisingly, the trade-off between reliability and secrecy can be characterized exactly, as shown by the following theorem.

Theorem 2.1 ([6] Theorem 1)**.** *The set of achievable rate tuples* (R_0, R_1, R_e) *is given by*

$$\mathcal{C} = \bigcup_{U \to V \to X \to YZ} \left\{ \begin{array}{l} 0 \leq R_e \leq R_1 \\ R_e \leq I(V;Y|U) - I(V;Z|U) \\ R_1 + R_0 \leq I(V;Y|U) + \min\left(I(U;Y), I(U;Z)\right) \\ 0 \leq R_0 \leq \min\left(I(U;Y), I(U;Z)\right) \end{array} \right\}$$

It is also convenient to define a scalar metric characterizing the inherent security that a channel can provide. The *secrecy capacity* of a broadcast channel with confidential messages is defined as the supremum of all rates R_1 such that the tuple $(0, R_1, R_1)$ is achievable. This metric provides a counterpart to the usual channel capacity, which considers only reliable communications without secrecy constraints. Based on Theorem 2.1, the following result can be proven.

Corollary 2.1 ([6], Corollary 2)**.** *The secrecy capacity of a broadcast channel with confidential messages is given by*

$$C_s = \max_{V \to X \to YZ} \left[I(V;Y) - I(V;Z) \right]. \tag{2.1}$$

Corollary 2.1 provides a formula that allows, in principle, to compute the secrecy capacity of any discrete memoryless channel, and it can also be shown that the result holds for continuous memoryless channels; however, Equation (2.1) involves a maximization over random variables satisfying a Markov chain condition, which provides little insight in practical situations. Nevertheless, Equation (2.1) shows that the secrecy capacity depends on the channel transition probability only through the marginal probabilities $p_{Y|X}(y|x)$ and $p_{Z|X}(z|x)$.

It is instructive to consider the case where the main channel and eavesdropper's channel are noiseless. Clearly, the secrecy capacity of a noiseless broadcast channel is zero, which

confirms that information-theoretic security cannot be obtained within the framework of traditional cryptography.

For certain channels, a closed-form expression of the secrecy capacity can be obtained. The most useful model for practical purposes is probably the Gaussian wiretap channel illustrated in Figure 7.

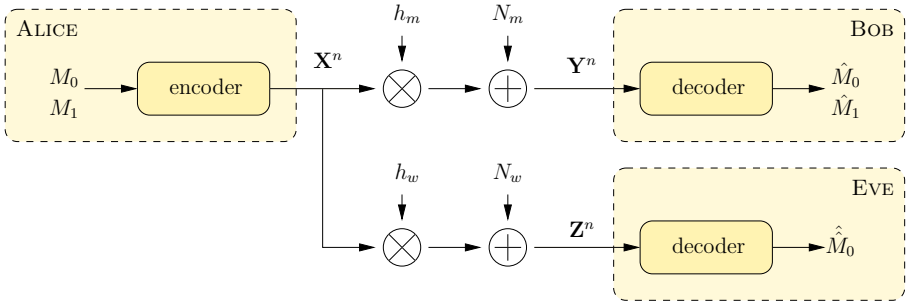


Figure 7. Gaussian wiretap channel.

The main channel and eavesdropper's channel are additive white Gaussian noise channels with channel gains h_m and h_w, respectively; the Gaussian noises N_m and N_w corrupting the transmission have variance σ_m^2 and σ_w^2, respectively. It is also assumed that the codewords sent over the channels are subject to the average power constraint

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{X_i^2\} \leq P.$$

Under these assumptions, we have the following result.

Theorem 2.2 ([7, 8])**.** *The secrecy capacity of the Gaussian wiretap channel is*

$$C_s = \begin{cases} \frac{1}{2}\log_2\left(1+\frac{h_m^2 P}{\sigma_m^2}\right) - \frac{1}{2}\log_2\left(1+\frac{h_w^2 P}{\sigma_w^2}\right) & \text{if} \quad \frac{h_m^2 P}{\sigma_m^2} > \frac{h_w^2 P}{\sigma_w^2}, \\ 0 \quad \text{otherwise.} \end{cases} \tag{2.2}$$

Equation (2.2) confirms the intuition that we developed in Chapter 1. When the legitimate receiver has a better signal-to-noise ratio than the eavesdropper, there exists a coding scheme ensuring information-theoretic security, and the maximum secure communication rate is the difference between the main channel capacity and eavesdropper's channel capacity.

2.3.2 Random codes achieving secrecy capacity

Theorem 2.1 was derived in [6] using a maximal code construction argument [9]. We provide the details of another proof based on typical set decoding in Appendix A. This alternative proof follows essentially the same lines as the original proof, but the use of typical set decoding makes the combination of this random coding argument with other techniques easier. In particular, we combine wiretap coding and relay coding in Chapter 5.

To provide the reader with an intuitive understanding of the proof, we now discuss the random code construction of wiretap codes in the special case of Theorem 2.1, where $R_0 = 0$, $R_e = R_1$, $U = \emptyset$, and $V = X$. First of all, the converse part of Theorem 2.1 enforces

$$\frac{1}{n} H(M|Z^n) \leq I(X;Y) - I(X;Z) + \delta, \tag{2.3}$$

where $\delta \to 0$ as $n \to \infty$. By using basic properties of entropy, one can now bound the eavesdropper's equivocation as follows.

$$H(M|Z^n) \geq H(X^n) - I(X^n;Z^n) - H(X^n|M,Z^n)\,. \tag{2.4}$$

Since the Asymptotic Equipartition Principle [1] (AEP) guarantees that, for n large enough,

$$I(X^n;Z^n) \leq nI(X;Z) + n\pi \quad \text{with} \quad \pi \to 0 \quad \text{as} \quad n \to \infty,$$

the lower bound in Equation (2.4) can match the upper bound in Equation (2.3), provided that two conditions are met.

1. **The capacity of the main channel is exhausted**, that is, the codebook uses the maximum number of codewords that can be transmitted reliably over the main channel. Since there are roughly $2^{nI(X;Y)}$ such codewords, $H(X^n) \approx nI(X;Y)$, which is the first term in Equation (2.3).

2. **The eavesdropper is allowed to identify the transmitted codeword reliably, given the knowledge of the message**, which ensures that the term $H(X^n|M,Z^n)$ vanishes as $n \to \infty$.

It is particularly striking that security in terms of equivocation can be enforced by imposing a structure on the wiretap code, such that the eavesdropper can *decode* under certain

conditions[2].

A simple way to construct a random code satisfying the previous two conditions is to use a binning structure, as illustrated in Figure 8. Starting from a randomly generated codebook for the main channel, which contains on the order of $2^{nI(X;Y)}$ codewords, codewords are grouped at random in $2^{n(I(X;Y)-I(X;Z))}$ bins of equal size. Because each bin contains approximately $2^{nI(X;Z)}$ codewords, the eavesdropper can identify a codeword sent over the channel provided that she knows the bin to which the codeword belongs. Consequently, the index of each bin can be used as the message transmitted by Alice, and upon selection of a message, Alice should simply select a codeword uniformly at random in the corresponding bin.

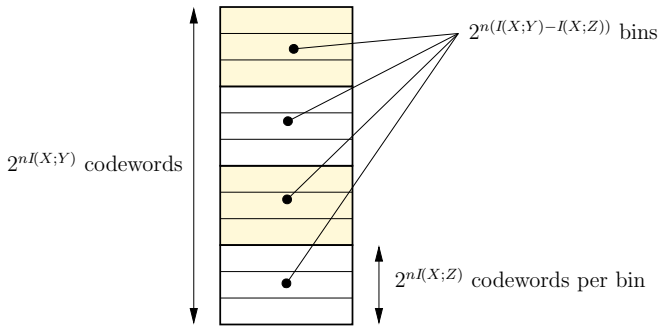


Figure 8. Binning scheme used to design wiretap codes.

2.3.3 Pertinence of wiretap channel model

In this section, we highlight the implicit assumptions inherent in the wiretap channel model.

1. **Knowledge of channel state information.** The equivocation of the eavesdropper is ensured provided that the wiretap code used for transmission is correctly tailored to the channel. In particular, this requires the Channel State Informations (CSI) about the main channel and the eavesdropper's channel to be known at the emitter. While assuming that the main channel CSI is perfectly known is reasonable since Alice and Bob can always cooperate to characterize their channel, requiring knowledge of the eavesdropper's channel CSI is more questionable; however, in situations where Alice is a wireless base station and the eavesdropper is a user in the network, the CSI is in

[2]Note, however, that allowing the eavesdropper to decode in certain situations does not mean that we impose a decoding strategy.

fact known at the emitter. Moreover, one can replace the exact knowledge of the CSI by a conservative estimation based on geographical information. As an example, one can certainly upper bound the signal-to noise ratio at a receiver if it is known to be located outside a given perimeter.

2. **Authentication.** The wiretap channel mode implicitly assumes that the main channel is authenticated. In principle, this assumption is not restrictive since authentication mechanisms can be implemented in the upper layers of the protocol stack. Note that it is possible to ensure unconditionally secure authentication [10] if a short secret key is available. Typically, the key size required for authentication scales as the logarithm of the message size; therefore, only a small fraction of secrecy capacity needs to be sacrificed to exchange secret keys.

3. **Passive eavesdropping.** The scope of the wiretap channel is restricted to passive eavesdropping strategies where the adversary does not tamper with the main channel or the eavesdropper's channel. Additional techniques are required to cope with jamming.

4. **Availability of random generator.** Unlike traditional encoders, which are deterministic functions, wiretap encoders are stochastic encoders and rely on the availability of perfect random generators. In practice, strong pseudo-random generators could be used, but their initialization mechanism should be carefully considered.

5. **Weak secrecy.** Security is defined in terms of the equivocation rate $\frac{1}{n}H(M_1|\mathbf{Z}^n)$, and a more satisfying criterion would be to use the absolute equivocation $H(M_1|\mathbf{Z}^n)$. The former notion of information-theoretic security is called *weak secrecy*, while the latter is referred to as *strong secrecy*. It is shown in [11] that strong and weak secrecy capacity are equal.

2.3.4 Extensions of the wiretap channel model

There has recently been a renewed interest for the study of wiretap channel models. In this section, we point out several research problems related to the wiretap channel that have been investigated.

- **Secrecy capacity of wireless channels.** Wireless communications are arguably the main application of physical-layer security, and characterizing their fundamental secrecy limits is an important area of research. Barros and Rodrigues [12] provided a detailed characterization of the outage secrecy capacity of slow fading channels and showed that fading alone guarantees that information-theoretic security is achievable, even when the eavesdropper has a better average signal-to-noise ratio than the legitimate receiver. The secrecy capacity of ergodic fading channels was derived independently by Liang *et al.* [13], Li *et al.* [14], and Gopala *et al.* [15], and power and rate allocation schemes for secret communication over fading channels were presented.

- **Multiple-input multiple output wiretap channels.** A natural extension of Theorem 2.2 is the situation where Alice, Bob, and Eve have multiple antennas. A closed-form expression for the secrecy capacity of multiple-input multiple-output Gaussian wiretap channels was derived independently in [16] and [17]. Although the derivation is quite involved, the final result is a convenient and simple generalization of Eq. (2.2).

- **Multi-user wiretap channels.** Numerous multi-user information theory problems have been reconsidered by adding secrecy constraints. Most notably, [18, 19] investigate multiple-access channels with confidential messages, [20, 21] consider secure relaying scenarios, and [22] studies interference channels with confidential messages. In general, it is not possible to obtain a single-letter characterization of the secrecy capacity, and most of the aforementioned works only provide bounds. The problem studied in Chapter 5 of this dissertation falls into the category of multi-user wiretap problems.

- **Wiretap channels with distortion measure.** Rather than enforcing a minimum equivocation rate on the eavesdropper, one could enforce a minimum distortion [23]. This alternative criterion could be useful for securing multimedia content such as video or voice.

2.4 Secret key agreement from common randomness

Contrary to the wiretap channel problem, which considers the communication of secure messages over noisy channels, the focus of secret-key agreement is the distillation of secrecy from common randomness [24, 25, 26]. Specifically, the objective of secret-key agreement is to generate secret keys from common randomness by public discussion over noiseless channels of unlimited capacity, without worrying about the cost of communication.

Two types of models are usually considered for secret-key agreement. The *source-type model* corresponds to a situation where terminals observe the correlated outputs of a source of randomness without having any control on the source. The *channel-type model* considers a scenario where one terminal transmits random symbols to the other terminals over a broadcast channel. The latter situation is similar to the wiretap channel model presented in the previous section, with the notable difference that the memoryless broadcast channel is used only for randomness sharing and that all other communications are performed over a noiseless side channel of unlimited capacity.

2.4.1 Three-terminal secret key agreement

For brevity, we only discuss the source-type model with three terminals, and unless otherwise specified, all results assume a discrete memoryless source. Alice, Bob, and Eve have access to n realizations $\mathbf{X}^n = (X_1, \ldots, X_n)$, $\mathbf{Y}^n = (Y_1, \ldots, Y_n)$, and $\mathbf{Z}^n = (Z_1, \ldots, Z_n)$, of random variables X, Y, and Z, respectively. The correlations between the random variables are governed by the known joint distribution $p_{XYZ}(x, y, z)$.

Definition 2.2. *A permissible secret sharing strategy for the source-type model is an interactive protocol consisting of t rounds, such that at each round k Alice sends a message $\Phi_k(\mathbf{X}^n, \mathbf{\Psi}^{k-1})$ to Bob depending on $\mathbf{X}^n$ and all the messages $\mathbf{\Psi}^{k-1} = (\Psi_1, \ldots, \Psi_{k-1})$ previously sent by Bob, and Bob sends a message $\Psi_k(\mathbf{Y}^n, \mathbf{\Phi}^{k-1})$ to Alice depending on $\mathbf{Y}^n$ and all the messages $\mathbf{\Phi}^{k-1} = (\Phi_1, \ldots, \Phi_{k-1})$ previously sent by Alice. After t rounds, Alice computes her key $K_A = f\left(\mathbf{X}^n, \mathbf{\Phi}^t, \mathbf{\Psi}^t\right)$ and Bob computes his key $K_B = g\left(\mathbf{Y}^n, \mathbf{\Phi}^t, \mathbf{\Psi}^t\right)$ using publicly known functions f and g..*

A rate R_k is an *achievable secret-key rate* of a source-type model if and only if, for any

$\epsilon > 0$, there exists a permissible secret sharing strategy such that

$$\begin{aligned}
\tfrac{1}{n} H(K_A) &> R_k - \epsilon \\
\mathrm{P}[K_A \neq K_B] &< \epsilon \qquad \text{(reliability condition)} \\
\tfrac{1}{n} I\big(K_A; \mathbf{Z}^n, \mathbf{\Phi}^t, \mathbf{\Psi}^t\big) &< \epsilon \qquad \text{(secrecy condition)} \\
\tfrac{1}{n} H(K_A) &> \tfrac{1}{n} \log_2 |\mathcal{K}_A| - \epsilon \quad \text{(uniformity condition)}
\end{aligned}$$

As a counterpart to the secrecy capacity of a wiretap channel, the *secret-key capacity* of X and Y with respect to Z, denoted by $S(X;Y\|Z)$, is defined as the supremum of all achievable secret key rates. In general, one cannot obtain a closed-form expression of the secret key capacity, but the following result can be proven.

Theorem 2.3 ([24], Theorem 2 and 3)**.** *The secret key capacity of a discrete memoryless source-type model with distribution $p_{XYZ}(x,y,z)$ is bounded as follows.*

$$\max\left[I(X;Y) - I(X;Z), I(Y;X) - I(Y;Z)\right] \leq S(X;Y\|Z) \leq \min\left[I(X;Y), I(X;Y|Z)\right].$$

If Eve's observation is a degraded version (*i.e.*, $p_{YZ|X}(y,z|x) = p_{Z|Y}(z|y)\, p_{Y|X}(y|x)$) or a stochastically degraded version (*i.e.*, $p_{YZ|X}(y,z|x) = p_{Z|X}(zxy)\, p_{Y|X}(y|x)$) of Bob's observation the two bounds are equal.

Interestingly, an exact characterization can be obtained for a subclass of protocols in which only a single message from Alice to Bob is allowed. In this case, the supremum of achievable secret key rates is called the *forward secret key capacity.*

Theorem 2.4 ([25], Theorem 2)**.** *The forward secret key capacity of a discrete memoryless source-type model with distribution $p_{XYZ}(x,y,z)$ is given by*

$$\max_{U \to T \to X \to YZ} \left[I(T;Y|U) - I(T;Z|U)\right]$$

When the eavesdropper does not have access to side information ($Z = \emptyset$) and only observes the public discussion, the secret-key capacity is equal to the forward secret-key capacity and is simply the mutual information $I(X;Y)$ [25, Proposition 1].

Among the many salient properties of secret-key agreement highlighted in [25, 24, 26], the most interesting facts are the following.

1. Secret key agreement is tightly related to Slepian-Wolf compression. As we describe in Section 2.5, this innate connection is especially useful for designing practical secret-key agreement schemes.

2. Feedback increases achievable secret-key rates. In particular, certain channels have a zero secrecy capacity without feedback, but secure communications at strictly positive rates are possible with feedback.

3. Interaction is more powerful than one-way communication. In [11], Maurer provides an example where the achievable secret-key rates are strictly larger with interactive protocols than one-way communications; however, it is still not known whether this statement is true in general.

2.4.2 Extensions of secret key agreement results

Several extensions of the aforementioned results have been investigated.

- **Strong secrecy capacity.** As shown in [11, 26], all the above results can be strengthened by replacing the (weak) secrecy and uniformity conditions by

$$I\big(K_A; \mathbf{Z}^n, \mathbf{\Phi}^t, \mathbf{\Psi}^t\big) < \epsilon \qquad \text{(strong secrecy condition)},$$
$$H(K_A) > \log_2 |\mathcal{K}_A| - \epsilon \quad \text{(strong uniformity condition)}.$$

- **Characterization of secret key capacity.** Obtaining a general closed-form expression of the secret-key capacity is still an open problem; however, Maurer showed that an information-theoretic quantity called the *intrinsic conditional information* [27] is useful in many situations for characterizing under what conditions secret key agreement is possible.

- **Multiterminal secret key generation.** Csiszár and Narayan have investigated the problem of key generation among multiple terminals in [26]. In the situation where an eavesdropper only observes the public communication and does not have access to side information, the secret-key capacity has a pleasing and intuitive expression. It is simply the total entropy of the source of common randomness minus the amount of information that must be shared for each terminal to gain complete knowledge of the source.

2.4.3 Beyond classical secret key agreement: quantum cryptography

The main drawbacks of secret-key agreement and wiretap channel models are the implicit assumptions that the probability distribution of the source (or the transition probability of the broadcast channel) is available and that the eavesdropper is purely passive. Interestingly, these issues can be circumvented when the transmission power is so low that quantum effects have to be taken into account. In fact, the fundamental laws of quantum mechanics limit the amount of information simultaneously accessible to Bob and Eve. Examples of such limitations are the no-cloning theorem [28], which forbids the exact duplication of a single quantum, or the Heisenberg uncertainty principle [29], which limits the accuracy of certain joint measurements. Quantum key distribution (also called quantum cryptography) exploits these properties to distribute unconditionally secure keys, even in the presence of an all-powerful eavesdropper only limited by the laws of quantum mechanics [30].

The general principle of a quantum key distribution scheme is the following. Alice and Bob communicate over two distinct channels, a *quantum channel*, which may be under Eve's control, and a *classical channel*, which is assumed to be public, noiseless, and authenticated. Quantum key distribution protocols usually proceed in two steps. First, Alice transmits a sequence of well-chosen random quantum states over the insecure quantum channel. The laws of quantum mechanics guarantee that Eve's operations systematically induce a statistical disturbance in Bob's measurements. Second, Alice and Bob exchange information over the public channel to detect statistical disturbances, indirectly infer the amount of information accessible to Eve, and distill a secret key. The tools used for secret key distillation are not specific to quantum key distribution and are presented in Section 2.5.

Let us re-emphasize that the unconditional authentication and integrity of messages sent over the public channel can be ensured by protocols requiring a short secret key, which implies that Alice and Bob should initially share such a key and should sacrifice a fraction of the subsequently generated keys to authenticate future messages. Hence, quantum key distribution systems are also called *quantum key growing* systems.

Despite being long regarded as an amusement for researchers, quantum cryptography has now been accepted as a viable solution for unconditionally secure communications and can be viewed as the ultimate application of physical-layer security. A few commercial

systems are already available, but their deployment on a large scale is hindered by their cost, the low key generation rates, and the requirement of a dedicated quantum channel.

2.5 Practical information-theoretic tools

As is often the case in information theory, the characterization of secrecy capacity and secret key capacity relies on random coding arguments that are convenient for analysis but fail to translate directly into practical coding schemes. Despite the numerous theoretical contributions, the general problem of wiretap coding has not received much attention. There is still no larger framework to draw on, even with the sustained advances in the area of error control coding.

2.5.1 Codes for the wiretap channel

Even though the random code construction presented in Section 2.3.2 does not provide any practical method for constructing a real code, it is quite natural to attempt to reproduce the binning structure illustrated in Figure 8. In [31], Wei shows how to encode secret information using cosets of certain linear block codes. More recently, this idea has been extended by Thangaraj *et al.* in [32] and Liu *et al.* in [33], where it has been shown how low-density parity-check codes can asymptotically achieve the secrecy capacity of the erasure wiretap channel, and how they can be used to provide secure communications at rates below the secrecy capacity for other channels.

Figure 9 illustrates the coding method of [32] in the case of a wiretap channel with a noiseless main channel and an erasure eavesdropper's channel. In the figure, $\mathbf{M}$ is a k-bit random variable denoting the message to be transmitted. Based on a $(n, n-k)$ code C with parity-check matrix $\mathbf{H}$ and generator matrix $\mathbf{G}$, Alice transmits a codeword $\mathbf{X}$ chosen uniformly at random in the coset of C with syndrome $\mathbf{M}$. Bob retrieves the message by calculating $\mathbf{H}\mathbf{X}^T$.

It is assumed that the eavesdropper observes the codewords through a binary erasure channel with erasure probability $1-\epsilon$. Intuitively, no information is leaked, provided the bits in unerased positions received by the eavesdropper identify codewords in all of the 2^k cosets of C. Actually, this intuition can be formalized [34], and it can be shown that the

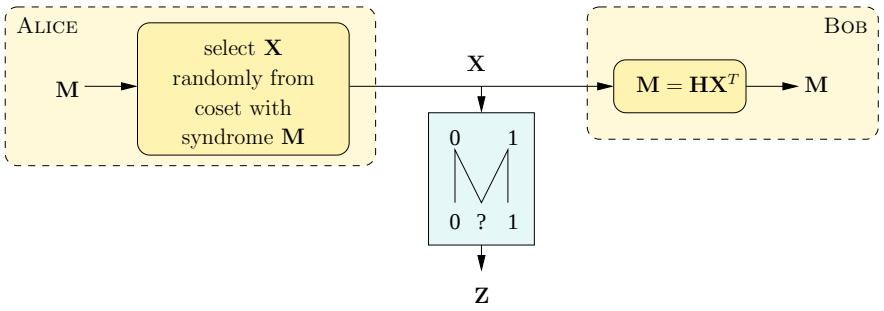


Figure 9. Coding method for erasure wiretap channel.

eavesdropper's equivocation is k bits if any submatrix of the generator matrix $\mathbf{G}$ containing ϵn columns has full rank. It turns out that parity-check matrices of LDPC codes chosen at random from a code ensemble with threshold $\alpha^* > \epsilon$ satisfy this criterion with high probability. Therefore, the dual of such an LDPC code can be used as the code C to secure communications.

2.5.2 Reconciliation and privacy amplification

Although there exist few practical coding schemes for the wiretap channel, there exist practical key distillation methods. Most of these techniques have been proposed in the context on quantum cryptography [35, 36] and are well understood when the source of common randomness is binary. In the rest of this section, it is assumed that all random variables are discrete. The extension of the results to continuous random variables is discussed in Chapter 3.

Following the source-type model for secret-key agreement described earlier, we assume that Alice, Bob, and Eve have access to n i.i.d. realizations of discrete random variables X, Y, and Z, respectively, distributed according to a known distribution p_{XYZ}. Alice, Bob and Eve's sequences are denoted by $\mathbf{X}^n = (X_1, \ldots, X_n)$, $\mathbf{Y}^n = (Y_1, \ldots, Y_n)$, and $\mathbf{Z}^n = (Z_1, \ldots, Z_n)$, respectively. A typical secret-key agreement consists of the two following steps.

Information reconciliation [37] Since Alice and Bob's sequences may not be perfectly correlated, there are discrepancies between Bob's received symbols $\mathbf{Y}^n$ and Alice's symbols $\mathbf{X}^n$; therefore, the first step is for Alice and Bob to correct errors before any further

processing. In the context of secret-key agreement, this operation is called *reconciliation*, and it requires an additional exchange of information between Alice and Bob over the public channel. Note that reconciliation is actually a special case of source coding with side information[3], where Alice compresses her source $\mathbf{X}^n$ and Bob decodes it with the help of correlated side information $\mathbf{Y}^n$. The Slepian-Wolf theorem [38] yields a lower bound on the total number of bits M_{rec} that have to be exchanged:

$$M_{rec} \geq H(\mathbf{X}^n|\mathbf{Y}^n) = nH(X|Y)\,. \tag{2.5}$$

Practical reconciliation algorithms introduce an overhead $\epsilon_{rec} > 0$ and require the transmission of $M_{rec} = nH(X|Y)(1+\epsilon_{rec})$ additional bits. Alternatively, it is also convenient to characterize the reconciliation by its efficiency β, which is defined as

$$\beta(\epsilon_{rec}) = 1 - \epsilon_{rec}\frac{H(X|Y)}{I(X;Y)} \leq 1. \tag{2.6}$$

At the end of the reconciliation step, Alice an Bob share with high probability the common sequence $\mathbf{X}^n$ whose entropy is $n_{rec} = nH(X)$. We assume that $\mathbf{X}^n$ is then compressed into an n_{rec}-bits binary sequence S.

Privacy amplification [39] This second operation allows Alice and Bob to extract a secret key from the binary sequence S. The principle of privacy amplification is to apply a well-chosen compression function $g : \{0,1\}^{n_{rec}} \to \{0,1\}^k$ $(k < n_{rec})$ to the reconciled bit sequence, such that the eavesdropper obtains negligible information about the final k-bit sequence $g(S)$. In practice, this can be achieved by choosing g at random within family of universal hash functions [40, 10], as stated in the following theorem.

Theorem 2.5. *[39, Corollary 4] Let $S \in \{0,1\}^{n_{rec}}$ be the random variable representing the bit sequence shared by Alice and Bob, and let E be the random variable representing the total information available to the eavesdropper. Let e be a particular realization of E. If the Rényi entropy (of order 2) $R(S|E=e)$ is known to be at least c, and Alice and Bob choose $K = G(S)$ as their secret key, where G is a hash function chosen at random from a universal family of hash functions $\mathcal{G} : \{0,1\}^{n_{rec}} \to \{0,1\}^k$, then*

$$H(K|G, E=e) \geq k - \frac{2^{k-c}}{\ln 2}. \tag{2.7}$$

[3]The link between reconciliation and coding with side information is discussed in more detail in Chapter 3.

The total information available to Eve E consists of the sequence $\mathbf{Z}^n$ received from the source of common randomness, as well as the additional bits exchanged during reconciliation, represented by a random variable M. As shown in [41, Theorem 5.2], for any $s > 0$ we have

$$R(S|\mathbf{Z}^n = \mathbf{z}^n, M = m) \geq R(S|\mathbf{Z}^n = \mathbf{z}^n) - \log_2 |M| - 2s - 2 \quad \text{with probability} \quad 1 - 2^{-s}. \tag{2.8}$$

The quantity $\log_2 |M|$ represents the number of bits intercepted by Eve during the reconciliation, which is at most $nH(X|Y)(1 + \epsilon_{rec})$. Evaluating $R(S|\mathbf{Z}^n = \mathbf{z}^n)$ is in general still difficult; however, conditioned on the typicality of the bit sequence, $R(S|\mathbf{Z}^n = \mathbf{z}^n)$ and $H(S|\mathbf{Z}^n = \mathbf{z}^n)$ are equal [11]. Hence, if n is large,

$$nH(X|Z) - nH(X|Y)(1 + \epsilon_{rec}) - 2s - 2$$

is a good lower bound of $R(S|E = e)$, and choosing

$$k = n\beta I(X;Y) - nI(X;Z) - 2s - 2 - r_0, \tag{2.9}$$

with $r_0 > 0$ guarantees that Eve's uncertainty on the key is such that

$$H(K|E) \geq k - 2^{-r_0}/\ln 2 \quad \text{with probability} \quad 1 - 2^{-s}.$$

CHAPTER 3
RECONCILIATION OF CONTINUOUS RANDOM VARIABLES

Reconciliation is an essential ingredient of practical secret-key agreement protocols. The reconciliation of binary random variable has been extensively studied in the context of quantum key distribution, and several efficient methods have been proposed [37, 35]; however, little attention has been devoted to the practical reconciliation of non-binary random variables. In this chapter, we develop an efficient algorithm based on Low-Density Parity-Check (LDPC) codes for the reconciliation of *continuous* random variables. Although our algorithm is generic and could be applied, in principle, to any continuous random variables, we focus most of our discussion on Gaussian random variables.

The scenario considered throughout this chapter is the following. It is assumed that two parties, Alice and Bob, have access to the outcomes of n i.i.d. instances of two distinct correlated continuous random variables $X \in \mathbb{R}$ and $Y \in \mathbb{R}$, with joint probability distribution $p_{XY}(x, y)$. The sequences of realizations obtained by Alice and Bob are denoted by

$$\mathbf{x}^n = (x_1, \ldots, x_n) \in \mathbb{R}^n \quad \text{and} \quad \mathbf{y}^n = (y_1, \ldots, y_n) \in \mathbb{R}^n,$$

respectively. The objective of reconciliation is for Alice and Bob to agree on a common bit sequence based on their observations while minimizing the amount of information exchanged to obtain this sequence.

3.1 Fundamental limit of reconciliation and algorithm design principles

3.1.1 Source coding with side information

Reconciliation can be related to the problem of distributed data compression, which is illustrated in Figure 10. In this situation, the i.i.d realizations of jointly distributed discrete random variables X and Y are to be compressed without loss and reconstructed at a receiver; however, the realizations of the random variable X are available at encoder 1 while the realizations of the random variable Y are available *separately* at encoder 2.

If sources are encoded jointly, the data compression theorem [1, Theorem 2.1] ensures that a total encoding rate $R > H(X, Y)$ should be sufficient to reconstruct the sources perfectly. Surprisingly, the non-intuitive result shown by Slepian and Wolf [38] is that

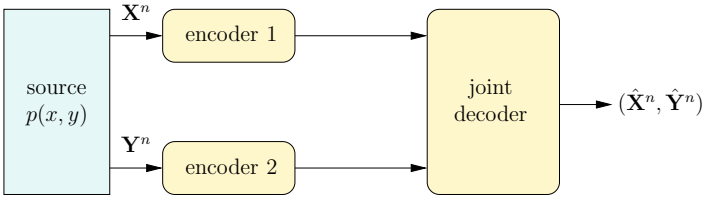


Figure 10. Slepian-Wolf coding of correlated sources.

separate encoding does not incur any rate penalty. Formally, we have the following results.

Definition 3.1. *A* $\left(2^{nR_X}, 2^{nR_Y}, n\right)$ *source code for the joint source* (X, Y) *consists of the following.*

- *Two message sets* $\mathcal{M}_X = \left\{1, 2, \ldots, 2^{nR_X}\right\}$ *and* $\mathcal{M}_Y = \left\{1, 2, \ldots, 2^{nR_Y}\right\}$.
- *Two encoding fucntion* $f_X : \mathcal{X}^n \to \mathcal{M}_X$ *and* $f_Y : \mathcal{Y}^n \to \mathcal{M}_Y$.
- *A decoding fucntion* $g : \mathcal{M}_X \times \mathcal{M}_Y \to \mathcal{X}^n \times \mathcal{Y}^n$.

A rate pair (R_X, R_Y) is achievable for the distributed source if and only if, for any $\epsilon > 0$, there exists a $\left(2^{nR_X}, 2^{nR_Y}, n\right)$ code such that

$$\mathrm{P}[g\left(f_X(\mathbf{X}^n), f_Y(\mathbf{Y}^n)\right) \neq (\mathbf{X}^n, \mathbf{Y}^n)] < \epsilon.$$

Theorem 3.1 (Slepian-Wolf theorem)**.** *The set of achievable compression rates* (R_X, R_Y) *for the distributed source coding problem is given by*

$$\begin{aligned} R_X &\geq H(X|Y)\,, \\ R_Y &\geq H(Y|X)\,, \\ R_X + R_Y &\geq H(X, Y)\,. \end{aligned}$$

The reconciliation of discrete random variables corresponds to the special case where the source Y is directly available at the receiver; this situations is referred to as *source coding with side information*. From Theorem 3.1, it is clear that X can be reconstructed perfectly at the receiver provided that the compression rate R_X is such that

$$R_X \geq H(X|Y)\,.$$

The problem of source coding with side-information can easily be generalized to the situation where X is discrete and Y is continuous. In this case, the metrics $H(X|Y)$ and $I(X;Y)$ are well-defined, and can be calculated as follows.

$$H(X|Y) = \sum_x \int_{\mathbb{R}} p\,(x,y) \log_2 p\,(y|x)\,,$$

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

When X is also a continuous random variable, reconstructing it perfectly would require an infinitely precise quantization; however, the objective of reconciliation is not the lossless compression of the sequence of observations $\mathbf{x}^n$, but the extraction of a common sequence from $\mathbf{x}^n$ and $\mathbf{y}^n$ with little additional communication. For instance, the common sequence can be obtained by first quantizing X into a discrete random variable $\hat{X}$ and then compressing $\hat{X}$ at a rate $R_{rec} \geq H\left(\hat{X}|Y\right)$. As discussed in Section 2.5, the performance of reconciliation is evaluated with a metric called the *reconciliation efficiency* and defined as

$$\beta = \frac{H\left(\hat{X}\right) - R_{rec}}{I(X;Y)} \leq \frac{I\left(\hat{X};Y\right)}{I(X;Y)} \leq 1. \tag{3.1}$$

In principle, $I\left(\hat{X};Y\right)$ can be made as close to $I(X;Y)$ as desired by choosing a fine enough quantizer; however, the main challenge is to design a *practical* scheme such that R_{rec} approaches $H\left(\hat{X}|Y\right)$. This task is non-trivial since the characterization of the ultimate performance of reconciliation algorithms provided by Theorem 3.1 is obtained with non-constructive arguments and provides little practical insight.

One could argue that reconciliation is also related to rate-distortion theory. In fact, we can view the procedure of quantization followed by a lossless compression as a lossy compression subject to a *symbol-wise* distortion constraint of the reconstructed data. However, the results of rate distortion theory consider mainly *average* distortion constraints and do not apply directly to reconciliation.

3.1.2 Reconciliation as coded modulation

In this section, we reformulate the reconciliation problem to provide more insight on the design of practical reconciliation schemes. Let us start by introducing a general description of the quantizer. Let $(I_1, \ldots, I_k)$ be k intervals forming a partition of $\mathbb{R}$. This partition

defines quantization intervals, and without loss of generality, we define the corresponding quantized values $(s_1, \ldots, s_k)$ to be the centers of each interval. We denote the indicator function of each interval I_j by

$$\begin{aligned} \chi_j(x): \quad & \mathbb{R} \longrightarrow \{0,1\} \\ & x \longmapsto 1 \quad \text{if} \quad x \in I_j, \quad 0 \quad \text{otherwise}, \end{aligned} \tag{3.2}$$

and we write the quantizer as

$$\begin{aligned} \mathcal{Q}: \quad & \mathbb{R} \longrightarrow \{\hat{s_j}\}_{j=1\ldots k} \\ & x \longmapsto \textstyle\sum_{j=1}^{k} s_j \chi_j(x). \end{aligned} \tag{3.3}$$

Consequently, the random variable $\hat{X}$ takes the discrete values $(s_1, \ldots, s_k)$ with probability

$$p_j = \Pr\left[\hat{X} = s_j\right] = \int p(x)\chi_j(x)dx, \tag{3.4}$$

respectively.

Notice that Alice's observations $\mathbf{x}^n$ can be viewed as the output of a discrete input/continuous output channel C_Q, characterized by the transition probabilities

$$p_Q(x|s_j) = p(x)\chi_j(x)/p_j, \tag{3.5}$$

when the quantized symbols $\hat{\mathbf{x}}^n = (\mathcal{Q}(x_1), \ldots, \mathcal{Q}(x_n))$ are fed at the input. Likewise, since the joint probability $p(x, y)$ of the continuous random variables X and Y can always be written as the product $p(y|x)p(x)$, the symbols $\mathbf{y}^n$ can also be viewed as the output of a memoryless channel C_S characterized by the transition probability $p_S(y|x) = p(y|x)$, when the i.i.d symbols $\mathbf{x}^n$ are fed at the input.

Combining these two observations, the continuous symbols $\mathbf{y}^n$ could have been generated by sending the discrete symbols $\hat{\mathbf{x}}^n$ through a channel C^*, obtained by concatenating channels C_S and C_Q. Since $\hat{X} \to X \to Y$ is a Markov chain, the transition probabilities of channel C^* are given by

$$p(y|s_j) = \int p_S(y|x)p_Q(x|s_j)dx. \tag{3.6}$$

Finally, each quantized value s_j can be assigned a unique ℓ-bits binary label, with $\ell = \lceil \log_2 k \rceil$. The labeling function that maps an element $x \in \mathbb{R}$ to the mth bit in the label of its quantized value $\mathcal{Q}(x)$ is denoted by

$$\mathcal{L}_m : \mathbb{R} \to \{0,1\} \quad \text{for} \quad m \in \{1, \ldots, \ell\}, \tag{3.7}$$

and the binary sequence obtained by quantizing and labeling $\mathbf{x}^n$ is denoted by

$$(\mathcal{L}(\hat{x}_1), \ldots, \mathcal{L}(\hat{x}_n)) = (v_{11}, \ldots, v_{1\ell}, \ldots, v_{n1}, \ldots, v_{n\ell}) = \mathbf{v}^{n\times\ell}.$$

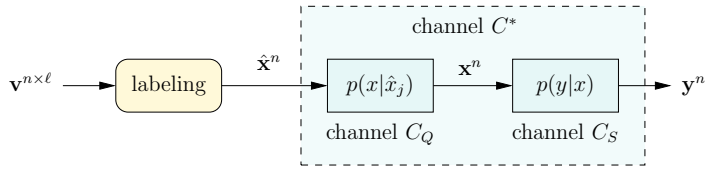


Figure 11. Reconciliation as coded modulation.

The above reformulation of the problem is totally artificial but, as illustrated in Figure 11, it makes apparent the connection with coded modulation techniques. In fact, Figure 11 could represent a coded modulation scheme if the sequence $\mathbf{v}$ was a sequence of codewords obtained *before* transmission over the channel C^*. Coded modulation techniques have been extensively studied in the context of communication over Gaussian channel, and we shall briefly review two powerful techniques called Bit Interleaved Coded Modulation (BICM) [42, 43] and MultiLevel Coding / MultiStage Decoding (MLC/MSD) [44, 45].

Figure 12 illustrates a MLC/MSD coded modulation scheme. Codewords $\mathbf{v}$ are mapped to a sequence of symbols $\mathbf{s}$ before transmission over a channel. The principle of MLC/MSD stems from a simple observation. Let $(V_1, \ldots, V_\ell)$ be the random variables representing the value of each bit in the label of symbol S. The sequence of bits at a given position in the symbol labels is called a *level*. By the chain rule of mutual information, we have

$$I(S;Y) = \sum_{i=1}^{\ell} I(V_i; Y | S_1, \ldots, V_{i-1}) \,, \tag{3.8}$$

which means that we can treat the initial channel as ℓ (dependent) sub-channels, and encoding can be performed on a *level-by-level* basis. Moreover, the above equation shows that, without loss of optimality, the decoder can decode the ℓ bit levels successively, using the result of previously decoded levels.

As illustrated in Figure 13, BICM is a more pragmatic coded modulation scheme where a single code is used. To ensure that the correlations introduced by the mapping and the coding are independent, BICM requires an additional interleaving step. Although BICM

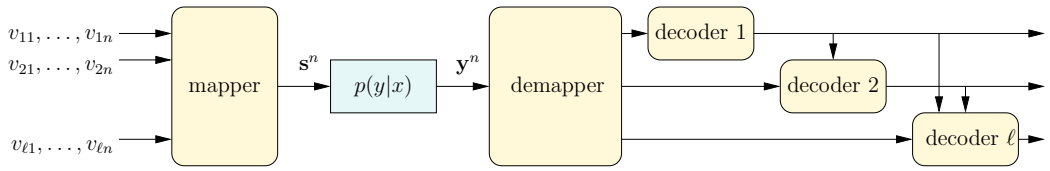


Figure 12. Multilevel coding with multistage decoding.

is a suboptimal scheme, its complexity is lower than MLC/MSD and it achieves relatively good performance when a Gray labeling of the symbols is used.

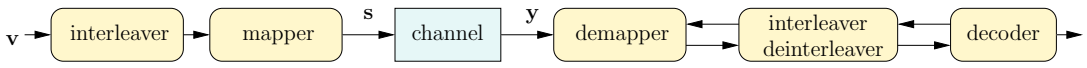


Figure 13. Bit interleaved coded modulation.

Let $\mathbf{H}$ be the parity-check matrix representing the parities satisfied by codewords[1] in a standard coded modulation scheme, such that $\mathbf{H}\mathbf{v}^T = \mathbf{0}$. The only difference between coded modulation and reconciliation is the fact that the binary sequence $\mathbf{v}$ obtained by quantizing and labeling the continuous symbols $\mathbf{x}^n$ is not a codeword in general. Nevertheless, following the idea suggested by Wyner in [46], we can use the syndromes $\mathbf{c} = \mathbf{H}\mathbf{v}^T \neq \mathbf{0}$ as the additional information sent by Alice to allow Bob to decode. For *BICM-like* reconciliation, a single code would be applied to an interleaved version of the whole sequence $\mathbf{v}^{n\times\ell}$, whereas for *MLC/MSD-like reconciliation*, ℓ individual codes would be applied successively to the sequences $(v_{i1}, \ldots, v_{in})$ for $i \in \{1, \ldots, \ell\}$.

3.1.3 Sliced error correction

We point out that the *sliced error correction* algorithm proposed in [36] is just a special case of MLC/MSD-like reconciliation, where interactive binary correction protocols optimized for binary symmetric channels are used as component codes. The algorithm offers a relatively low complexity, but as we show next, this simplification limits its efficiency. Let $(V_1, \ldots, V_\ell)$ be the ℓ random variables corresponding to the ℓ label bits of $\hat{X}$. Using the chain rule of

[1]Here, $\mathbf{H}$ does not necessarily represent a single code. For instance, for a MLC/MSD scheme, $\mathbf{H}$ contains the parity-check matrices of the codes used at each level.

mutual information, we have

$$\begin{aligned} I(V;Y) &= I(V_1,\ldots,V_\ell;Y) \\ &= \sum_{i=1}^{\ell} I(V_i;Y|V_{i-1},\ldots,V_1) \\ &= \sum_{i=1}^{\ell} [H(V_i|V_{i-1},\ldots,V_1) - H(X_i|Y,V_{i-1},\ldots,V_1)]. \end{aligned} \tag{3.9}$$

Sliced error correction assigns label in such a way that $H(V_i|V_{i-1},\ldots,V_1) = 1$, and V_i is estimated from $(Y,V_{i-1},\ldots,V_1)$ using the maximum likelihood estimator ML_i. By Fano's inequality, we have

$$I(V;Y) \geq \sum_{i=1}^{\ell} [1 - h(p_i)], \tag{3.10}$$

with

$$p_i = \mathrm{P}[V_i \neq \mathrm{ML}_i\,(Y,V_{i-1},\ldots,V_1)] \tag{3.11}$$

and h is the binary entropy function. As expected, treating all ℓ levels as binary symmetric channels underestimates $I(\hat{X};Y)$, and $\sum_{i=1}^{\ell}[1-h(p_i)]$ only heads toward $I(X;Y)$ asymptotically as $\ell \to \infty$. For practical values of ℓ (say less than 5) this approximation may not be tight enough to ensure a good reconciliation efficiency, even with perfect codes achieving capacity over binary symmetric channels.

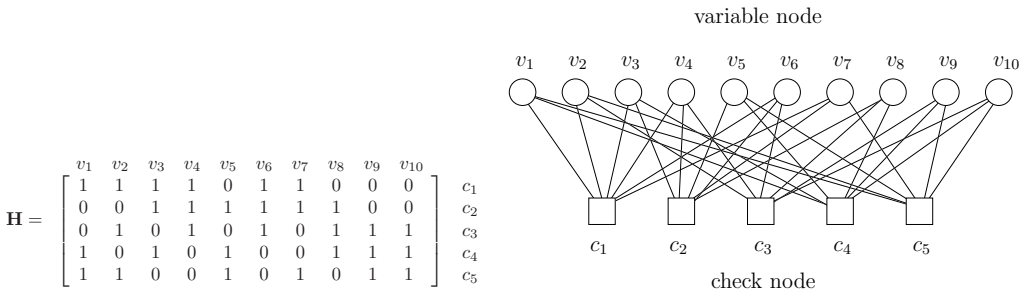
3.2 LDPC-based reconciliation of continuous random variables

Achieving high reconciliation efficiency relies on our ability to design codes and decoders operating at an overall rate close to $I(\hat{X};Y)$. Turbo codes and LDPC codes are promising candidates for this purpose since they have already proven their excellent performance for error correction and side-information coding [47]; however, we limit our investigation to the study of LDPC codes, although we acknowledge that turbo-codes or any other strong channel codes would probably yield similar results.

3.2.1 Review of binary LDPC codes

A binary LDPC code is a binary error correcting code characterized by a *sparse* parity-check matrix $\mathbf{H} = [h_{ij}]_{i=1..n-k}^{j=1..n} \in \{0,1\}^{n-k\times n}$, that is $\mathbf{H}$ contains a small number of ones

compared to the number of zeros. As illustrated in Figure 14, a parity-check matrix can be represented as a bipartite graph.



(a) Parity-check matrix. (b) Tanner graph.

Figure 14. Parity-check matrix and bipartite graph of an LDPC code of blocklength $n = 10$.

This bipartite graph, also called a *Tanner graph*, contains two types of nodes: *variable nodes*, which are used to represent the codewords bits, and *check nodes*, which are used to represent the parity constraints imposed by the parity-check matrix $\mathbf{H}$. A variable node j is connected to a check node i by an *edge* if and only $h_{ij} = 1$ in the parity-check matrix. The *degree* of a node is simply defined as the number of edges connected to it.

A *regular* LDPC code is an LDPC code for which all variables nodes have the same degree d_v and all check nodes have the same degree d_c. When nodes of the same type have different degrees, an LDPC code is called *irregular*. A convenient way of analyzing LDPC codes is to consider ensembles of LDPC codes rather that individual ones. In particular, it is useful to consider ensembles of LDPC codes characterized by the same *edge degree distribution*

$$\lambda(x) = \sum_{i=1}^{d_v^{max}} \lambda_i x^{i-1}, \quad \rho(x) = \sum_{i=1}^{d_c^{max}} \rho_i x^{i-1},$$

where d_v^{max} and d_c^{max} are the maximum degrees of variables nodes and check nodes, respectively, and λ_i and ρ_i are the fraction of edges connected to variables nodes of degree i and check nodes of degree i, respectively. When the block length n is large, all codes within the ensemble tend to have the same properties.

The good error-correcting performance of LDPC codes is mainly due to the fact that there exist low-complexity *soft decoding* algorithms, which exploit the continuous nature of probability distributions and estimate not only the value of each variable node but also the

reliability associated to the estimation. For instance, for a binary variable $v \in \{0,1\}$, the reliability is described by the probabilities $\mathrm{P}[v=0]$ and $\mathrm{P}[v=1]$. Since $\mathrm{P}[v=1]+\mathrm{P}[v=0]=1$, only one parameter is really necessary, and one often chose the log-ratio of probabilities

$$\lambda_v = \log \frac{\mathrm{P}[v=0]}{\mathrm{P}[v=1]}.$$

The sign of λ_v provides the most likely value of the variables, while the magnitude $|\lambda|$ provides a measure of reliability. For instance, when $\lambda = 0$, both values are equally likely, and when $|\lambda| = \infty$ there is no uncertainty of the value of v. The decoding of LDPC codes is performed efficiently over the Tanner graph using a *message-passing* algorithm that we describe in detail in the next section.

Among the many techniques proposed to analyze and design LDPC codes, the most successful ones are probably the methods based on *density evolution* [87] and *EXtrinsic Information Transfer (EXIT) charts* [**?**]. The principle of density evolution is to predict the error-correcting behavior of LDPC codes by tracking the evolution of the distribution of messages exchanged between variables nodes and check nodes at each iteration of the decoding algorithm. EXIT charts have been proposed has a simpler alternative to density evolution, where a single scalar parameter is tracked instead of the full probability distribution of messages. We refer the reader to references [87, **?**] for a detailed presentation of these techniques.

3.2.2 LDPC-based reconciliation

In this section, we derive an efficient decoding algorithm that applies to both MLC/MSD-lie reconciliation and BICM-like reconciliation. This algorithm generalizes the standard decoding algorithm of LDPC codes for Slepian-Wolf compression of non-uniform sources. We recall that the jth label bit obtained by quantizing symbol x_i is $v_{ij} = \mathcal{L}_j\left(\mathcal{Q}(x_i)\right)$. The reconciliation algorithm should estimate the values of variable nodes v_{ij} given the knowledge of observations $\mathbf{y}^n$ and syndromes $\mathbf{c}$. Specifically, we shall compute the *a posteriori* Log-Likelihood Ratios (LLRs)

$$\lambda_{ij} = \log \frac{\mathrm{P}[v_{ij}=0|\mathbf{y}^n,\mathbf{c}]}{\mathrm{P}[v_{ij}=1|\mathbf{y}^n,\mathbf{c}]}, \tag{3.12}$$

for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, \ell\}$. Notice that the numerator and denominator are the marginals of the conditional probability of the whole sequence $(v_{11}, \ldots, v_{n\ell})$ given $\mathbf{y}^n$ and $\mathbf{c}$, that is,

$$\mathrm{P}[v_{ij}|\mathbf{y}^n, \mathbf{c}] = \sum_{v_{kl} \neq v_{ij}} \mathrm{P}[v_{11}, \ldots, v_{n\ell}|\mathbf{y}^n, \mathbf{c}]. \tag{3.13}$$

We shall see that, under the assumption that the Tanner graph of the LDPC code *does not contain cycles*, this marginalization can be computed efficiently. In fact, the above expression factorizes into a product of simpler marginals and can be computed recursively through the graph.

Figure 15 illustrates the Tanner graph from the perspective of a variable node v_{ij}. Notice that the graph is not a bipartite graph. In fact, variable nodes $(v_{i1}, \ldots, v_{i\ell})$ originating from the same symbol x_i are correlated; therefore, a third type of node, called *demapper nodes*, is introduced to account for these correlations.

To analyze the factorization of the marginal probability, we introduce the following notation.

- the subgraph of nodes connected to demapper node y_i through variable node v_{ij} is denoted by $\mathcal{G}_{ij}$;
- the check nodes connected to variable node v_{ij} are denoted by $\left(c_{ij}^{(1)}, \ldots, c_{ij}^{(d)}\right)$;
- the subgraph of nodes connected to variable node node v_{ij} through check node $c_{ij}^{(u)}$ is denoted by $\mathcal{H}_{ij}^{(u)}$;
- $\mathbf{v}\left[\mathcal{G}_{ij}\right]$ denotes the set of variables nodes in subgraph $\mathcal{G}_{ij}$;
- $\mathbf{c}\left[\mathcal{G}_{ij}^{(u)}\right]$ denotes the set of check nodes in subgraph $\mathcal{G}_{ij}^{(u)}$;
- $\mathbf{vy}\left[\mathcal{G}_{ij}^{(u)}\right]$ denotes the set of variable node/observation tuples $(v_{t1}, \ldots, v_{t\ell}, y_t)$ such that $(v_{t1}, \ldots, v_{t\ell}, y_t) \in \mathcal{G}_{ij}^{(u)}$.

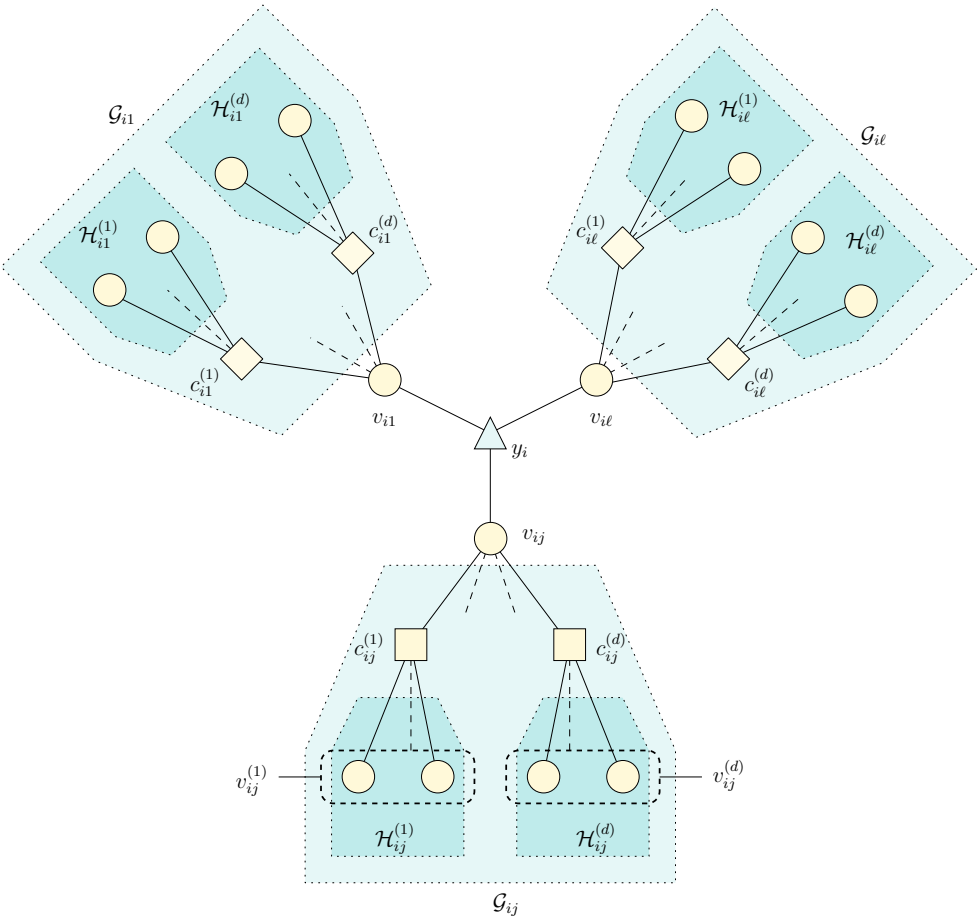


Figure 15. Extended graph of LDPC code from the perspective of variable node v_{ij}.

The total conditional probability in Equation (3.13) can be expanded as follows.

$$
\begin{aligned}
&\mathrm{P}[v_{11},\ldots,v_{n\ell}|\mathbf{y}^n,\mathbf{c}]\\
&\overset{(a)}{\propto} p\left(\mathbf{y}^n|v_{11},\ldots,v_{n\ell}\right)\mathrm{P}[\mathbf{c}|v_{11},\ldots,v_{n\ell}]\mathrm{P}[v_{11},\ldots,v_{n\ell}],\\
&= p\left(\mathbf{y}^n,v_{11},\ldots,v_{n\ell}\right)\mathrm{P}[\mathbf{c}|v_{11},\ldots,v_{n\ell}],\\
&\overset{(b)}{=} p\left(y_i,v_{i1},\ldots,v_{i\ell}\right)\prod_{k=1}^{\ell} p\left(\mathbf{vy}\left[\mathcal{G}_{ik}\right]\right)\mathrm{P}[\mathbf{c}\left[\mathcal{G}_{ik}\right]|v_{ik},\mathbf{x}\left[\mathcal{G}_{ik}\right]],\\
&\overset{(c)}{=} p\left(y_i,v_{i1},\ldots,v_{i\ell}\right)\prod_{k=1}^{\ell} p\left(\mathbf{vy}\left[\mathcal{G}_{ik}\right]|v_{ik}\right)\mathrm{P}[\mathbf{c}\left[\mathcal{G}_{ik}\right]|v_{ik},\mathbf{vy}\left[\mathcal{G}_{ik}\right]],\\
&= p\left(y_i,v_{i1},\ldots,v_{i\ell}\right)\prod_{k=1}^{\ell} p\left(\mathbf{c}\left[\mathcal{G}_{ik}\right],\mathbf{vy}\left[\mathcal{G}_{ik}\right]|v_{ik}\right), \qquad (3.14)
\end{aligned}
$$

where (a) follows from Bayes rule and the fact that $\mathbf{y}$ is independent of $\mathbf{c}$ given $(v_{11},\ldots,v_{n\ell})$, and $(b-c)$ follows from the memoryless property of the source and the fact that for all k,

$\mathbf{c}\left[\mathcal{G}_{ik}\right]$ only depends on $\mathbf{v}\left[\mathcal{G}_{ik}\right]$ given v_{ik}.

Now, using the distributivity of addition and multiplication, the marginal can be written as

$$\sum_{v_{kl}\neq v_{ij}} \mathrm{P}[v_{11},\ldots,v_{n\ell}|\mathbf{y}^n,\mathbf{c}] = \left(\sum_{\mathbf{v}[\mathcal{G}_{ij}]} p\left(\mathbf{c}\left[\mathcal{G}_{ij}\right],\mathbf{vy}\left[\mathcal{G}_{ij}\right]|v_{ij}\right)\right) \left(\sum_{\substack{v_{im},\mathbf{v}[\mathcal{G}_{im}]\\ m\neq j}} p\left(y_i,v_{i1},\ldots,v_{i\ell}\right)\prod_{\substack{k=1\\k\neq j}}^{\ell} p\left(\mathbf{c}\left[\mathcal{G}_{ik}\right],\mathbf{vy}\left[\mathcal{G}_{ik}\right]|v_{ik}\right)\right), \tag{3.15}$$

The first summation, which does not involve the observation y_i corresponding to the bit v_{ij} and only accounts for the correlations specified by the code, is called the *extrinsic* information. The second term, which include the contribution of the observation y_i, is called the *intrinsic* information.

Using the definition of sets $\mathcal{H}_{ij}^{(u)}$, the extrinsic information term can be rewritten as follows.

$$\begin{aligned}
&\sum_{\mathbf{v}[\mathcal{G}_{ij}]} p\left(\mathbf{c}\left[\mathcal{G}_{ij}\right],\mathbf{vy}\left[\mathcal{G}_{ij}\right]|v_{ij}\right)\\
&\quad\overset{(a)}{=} p\left(\mathbf{c}\left[\mathcal{G}_{ij}\right],\mathbf{y}\left[\mathcal{G}_{ij}\right]|v_{ij}\right)\\
&\quad\overset{(b)}{=} \prod_{u=1}^{d} p\left(c_{ij}^{(u)},\mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right],\mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right]|v_{ij}\right)\\
&\quad\overset{(c)}{=} \prod_{u=1}^{d} p\left(c_{ij}^{(u)}|\mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right],\mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right],v_{ij}\right) p\left(\mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right],\mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right]\right),\\
&\quad\propto \prod_{u=1}^{d} p\left(c_{ij}^{(u)}|\mathbf{y}\left[\mathcal{H}_{ij}^{(u)},\right],\mathbf{c}\left[\mathcal{H}_{ij}^{(u)},\right],v_{ij}\right),
\end{aligned} \tag{3.16}$$

where (a) follows from the law of total probability, (b) follows from the independence of $\left(c_{ij}^{(u)},\mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right],\mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right]\right)$ and $\left(c_{ij}^{(v)},\mathbf{c}\left[\mathcal{H}_{ij}^{(v)}\right],\mathbf{y}\left[\mathcal{H}_{ij}^{(v)}\right]\right)$ given v_{ij} for $v\neq u$, and (c) follows from the independence of $\left(\mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right],\mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right]\right)$ and v_{ij}.

Likewise the intrinsic information term can be written as follows.

$$\begin{aligned}
&\sum_{\substack{v_{im},\mathbf{v}[\mathcal{G}_{im}]\\ m\neq j}} p\left(y_i, v_{i1}, \ldots, v_{i\ell}\right) \prod_{\substack{k=1\\ k\neq j}}^{\ell} p\left(\mathbf{c}\left[\mathcal{G}_{ik}\right], \mathbf{vy}\left[\mathcal{G}_{ik}\right] | v_{ik}\right) \\
&\overset{(a)}{=} \sum_{\substack{v_{im}\\ m\neq j}} p\left(y_i, v_{i1}, \ldots, v_{i\ell}\right) \prod_{\substack{k=1\\ k\neq j}}^{\ell} \sum_{\mathbf{v}[\mathcal{G}_{ik}]} p\left(\mathbf{c}\left[\mathcal{G}_{ik}\right], \mathbf{vy}\left[\mathcal{G}_{ik}\right] | v_{ik}\right) \\
&\overset{(b)}{=} \sum_{\substack{v_{im}\\ m\neq j}} p\left(y_i, v_{i1}, \ldots, v_{i\ell}\right) \prod_{\substack{k=1\\ k\neq j}}^{\ell} p\left(\mathbf{c}\left[\mathcal{G}_{ik}\right], \mathbf{y}\left[\mathcal{G}_{ik}\right] | v_{ik}\right) \\
&\overset{(c)}{\propto} \sum_{\substack{v_{im}\\ m\neq j}} p\left(y_i, v_{i1}, \ldots, v_{i\ell}\right) \prod_{\substack{k=1\\ k\neq j}}^{\ell} \prod_{u=1}^{d} p\left(c_{ik}^{(u)} | v_{ik}, \mathbf{y}\left[\mathcal{H}_{ik}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ik}^{(u)}\right]\right), \\
&= \sum_{\hat{x}:\hat{x}_j = v_{ij}} p\left(y_i, \hat{x}\right) \prod_{\substack{k=1\\ k\neq j}}^{\ell} \prod_{u=1}^{d} p\left(c_{ik}^{(u)} | \hat{x}_k, \mathbf{y}\left[\mathcal{H}_{ik}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ik}^{(u)}\right]\right), \qquad (3.17)
\end{aligned}$$

where (a) follows from the distributivity of addition and multiplication, (b) follows from the law of total probability, (c) follows from steps similar to those of Equation (3.16), and (d) follows by introducing a dummy variable $\hat{x}$ spanning all the possible quantized values and letting $\hat{x}_j = \mathcal{L}_j(\hat{x})$ be its jth bit label. By defining

$$\lambda_{ik}^{(u)} = \log \frac{p\left(c_{ik}^{(u)} | \mathbf{y}\left[\mathcal{H}_{ik}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ik}^{(u)}\right], \hat{x}_k = 0\right)}{p\left(c_{ik}^{(u)} | \mathbf{y}\left[\mathcal{H}_{ik}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ik}^{(u)}\right], \hat{x}_k = 1\right)}, \qquad (3.18)$$

and substituting Equations (3.16) and (3.17) in Equation (3.12), we obtain

$$\lambda_{ij} = \underbrace{\log \frac{\sum_{\hat{x}:\hat{x}_j=0} p\left(y_i, \hat{x}\right) \exp\left[\sum_{\substack{k=1\\ k\neq j}}^{\ell} \sum_{u=1}^{d} \left(1-\hat{x}_k\right) \lambda_{ik}^{(u)}\right]}{\sum_{\hat{x}:\hat{x}_j=1} p\left(y_i, \hat{x}\right) \exp\left[\sum_{\substack{k=1\\ k\neq j}}^{\ell} \sum_{u=1}^{d} \left(1-\hat{x}_k\right) \lambda_{ik}^{(u)}\right]}}_{\text{intrinsic LLR}} + \underbrace{\sum_{u=1}^{d} \lambda_{ij}^{(u)}}_{\text{extrinsic LLR}}. \qquad (3.19)$$

Now, let $\left(v_{ij}^{(u)1}, \ldots, v_{ij}^{(u)d_c-1}\right)$ be the variable nodes other than v_{ij} involved in the calculation of parity check $c_{ij}^{(u)}$. Using the fact that

$$c_{ij}^{(u)} = v_{ij} \bigoplus_{k=1}^{d_c-1} v_{ij}^{(u)k},$$

the term $\lambda_{ij}^{(u)}$ can be written as

$$\begin{aligned}
\lambda_{ij}^{(u)} &= \log \frac{p\left(c_{ij}^{(u)} | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right], v_{ij}=0\right)}{p\left(c_{ij}^{(u)} | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right], v_{ij}=1\right)} \\
&= \log \frac{p\left(\bigoplus_{k=1}^{d_c-1} v_{ij}^{(u)k} = c_{ij}^{(u)} \oplus 0 | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right], v_{ij}=0\right)}{p\left(\bigoplus_{k=1}^{d_c-1} v_{ij}^{(u)k} = c_{ij}^{(u)} \oplus 1 | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right], v_{ij}=1\right)} \\
&= \left(1-2c_{ij}^{(u)}\right) \log \frac{p\left(\bigoplus_{k=1}^{d_c-1} v_{ij}^{(u)k} = 0 | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right]\right)}{p\left(\bigoplus_{k=1}^{d_c-1} v_{ij}^{(u)k} = 1 | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right]\right)}
\end{aligned} \tag{3.20}$$

After some simple algebra, it can be shown that

$$\begin{aligned}
&\log \frac{p\left(\bigoplus_{k=1}^{d_c-1} v_{ij}^{(u)k} = 0 | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right]\right)}{p\left(\bigoplus_{k=1}^{d_c-1} v_{ij}^{(u)k} = 1 | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right]\right)} \\
&= -2\tanh^{-1} \prod_{k=1}^{d_c-1} \tanh\left(-\frac{1}{2} \log \frac{p\left(v_{ij}^{(u)k} = 0 | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right]\right)}{p\left(v_{ij}^{(u)k} = 1 | \mathbf{y}\left[\mathcal{H}_{ij}^{(u)}\right], \mathbf{c}\left[\mathcal{H}_{ij}^{(u)}\right]\right)}\right).
\end{aligned} \tag{3.21}$$

Therefore, the terms $\lambda_{ij}^{(u)}$ can be computed as a function of the *a posteriori* LLRs of the variable nodes $v_{ij}^{(u)k}$, with the observations and syndromes in the subgraph $\mathcal{H}_{ij}^{(u)}$. The steps leading to Equations (3.19) and (3.21) can be reapplied to these *a posteriori* LLRs, which allows th computation of λ_{ij} *recursively* through the graph.

In practice, we are interested in obtaining the LLRs of all variable nodes, and instead of applying the above recursion for each LLR, it is possible to use a *message-passing* algorithm that computes all LLRs *simultaneously*. This algorithm, called the *Sum-Product* algorithm, is an iterative algorithm that consists of a set of *local* rules specifying the message computed by each node at each iteration. As illustrated in Figure 16, for each variable node indexed by ij, there are four types of messages to consider: *variable-to-check* messages v_{ijk}, *check-to variable* messages u_{kij}, *demapper-to-variable* messages o_{ij}, and *variable-to-demapper* messages e_{ij}.

The set of check nodes connected to a given variable node ij is denoted by $\mathcal{N}(ij)$, and the set of variable nodes connected to a given check node k is denoted by $\mathcal{M}(k)$. Based on Equations (3.19) and (3.21), the reconciliation message passing algorithm is then the following.

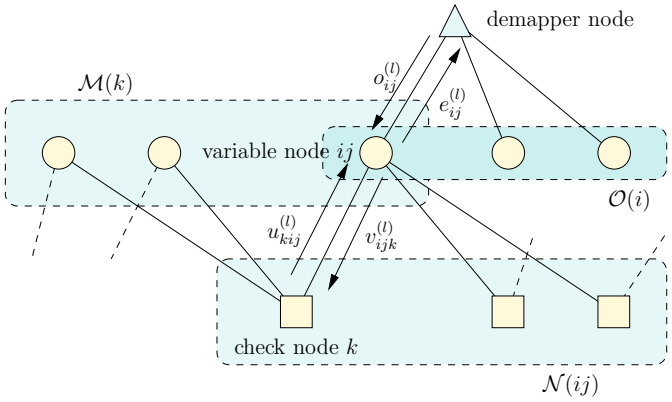


Figure 16. Messages exchanged between nodes.

□ **Initialization.** Initialize all message to zero.

$$\forall i,j,k \quad e_{ij}^{(0)} = o_{ij}^{(0)} = v_{ijk}^{(0)} = u_{kij}^{(0)} = 0. \tag{3.22}$$

□ **Iterations.** For $1 \leq l \leq l_{max}$

1. demapper-to-variable message update

$$o_{ij}^{(l)} = \log \frac{\displaystyle\sum_{\hat{x}:\hat{x}_j=0} p\left(y_i,\hat{x}\right) \exp\left[\sum_{k\neq j}\left(1-\hat{x}_k\right) e_{ik}^{(l-1)}\right]}{\displaystyle\sum_{\hat{x}:\hat{x}_j=1} p\left(y_i,\hat{x}\right) \exp\left[\sum_{k\neq j}\left(1-\hat{x}_k\right) e_{ik}^{(l-1)}\right]}. \tag{3.23}$$

2. variable-to-check message update

$$v_{ijk}^{(l)} = o_{ij}^{(l)} + \sum_{k\in\mathcal{N}(i)\setminus j} u_{kij}^{(l-1)} \tag{3.24}$$

3. check-to variable-message update

$$u_{kij}^{(l)} = 2\tanh^{-1} \prod_{ij\in\mathcal{M}(k)\setminus ij} \tanh\frac{v_{kij}^{(l-1)}}{2} \tag{3.25}$$

4. variable-to-demapper update

$$e_{ij}^{(l)} = \sum_{k\in\mathcal{N}(ij)} u_{kij}^{(l)} \tag{3.26}$$

□ **Hard decoding.** $\forall i \in \{1, \ldots, n\}$ decide

$$v_{ij} = -\frac{1}{2}\left(\text{sign}(e_{ij}^{(l_{max})} + o_{ij}^{(l_{max})}) - 1\right),$$

where

$$\text{sign}(x) = \begin{cases} +1 & \text{if} \quad x \geq 0 \\ -1 & \text{if} \quad x < 0 \end{cases}$$

Note that this algorithm computes the *a posteriori* LLRs exactly only when the graph does not contain cycles; however, most "good" finite-length LDPC codes contain cycles, and the algorithm only provides approximations of the real *a posteriori* LLRs. The LDPC codes used with the above message-passing algorithm naturally admits to a density evolution analysis similar to that of [48], which is sketched in Appendix B.

3.3 Reconciliation of Gaussian random variables

In this section, we investigate the MLC/MSD-like reconciliation and BICM-like reconciliation of Gaussian random variables

$$X \sim \mathcal{N}(0, \Sigma) \quad \text{and} \quad Y = X + N, \quad \text{where} \quad N \sim \mathcal{N}(0, \sigma).$$

We shall see in Chapter 4 that these correlations are useful to distill secret keys over Gaussian channels. In the rest of this section, we adopt the quantization technique proposed in [36]. The set of real numbers is split into an even number k of intervals $\{I_j\}_{1..k}$ symmetric around 0 (this ensures the symmetry of the joint distribution between $\hat{X}$ and Y), and interval bounds are optimized using the simplex method in order to maximize $I(\hat{X}; Y)$.

3.3.1 Choice of codes and rates for MLC/MSD-like reconciliation

Before constructing a set of codes for MLC/MSD-like reconciliation, it is first necessary to identify the rate of each component code. In this section, we compute the *optimal* rates that would be required for *ideal* codes. Although these rates are not achievable with practical codes, we shall see in the next section that they provide a good starting point for optimization. To preserve some symmetry in the probability distribution, it is desirable to consider only labeling strategies satisfying

$$\forall m \qquad p\left(y, \mathcal{L}_m(\mathcal{Q}(x)) = b\right) = p\left(-y, \mathcal{L}_m(\mathcal{Q}(x)) = \overline{b}\right) \tag{3.27}$$

for $b \in \{0,1\}$. In particular, there are two simple labeling strategies fulfilling this requirement: *natural* labeling and *anti-natural* labeling. Both labeling assign to each interval j the ℓ-bit binary representation ($\ell = \lceil \log_2 k \rceil$) of

$$j + (2^n - k)/2,$$

but in the natural labeling case the least significant bit level is decoded first while in the antibinary labeling case the most significant bit level is decoded first.

For a given SNR s, the optimal code rate R^i_{opt} required at each level $i \in \{1, \ldots, \ell\}$ is related to the mutual information

$$I(V_i; Y | V_1 \ldots V_{i-1}, \mathrm{SNR} = s)$$

by

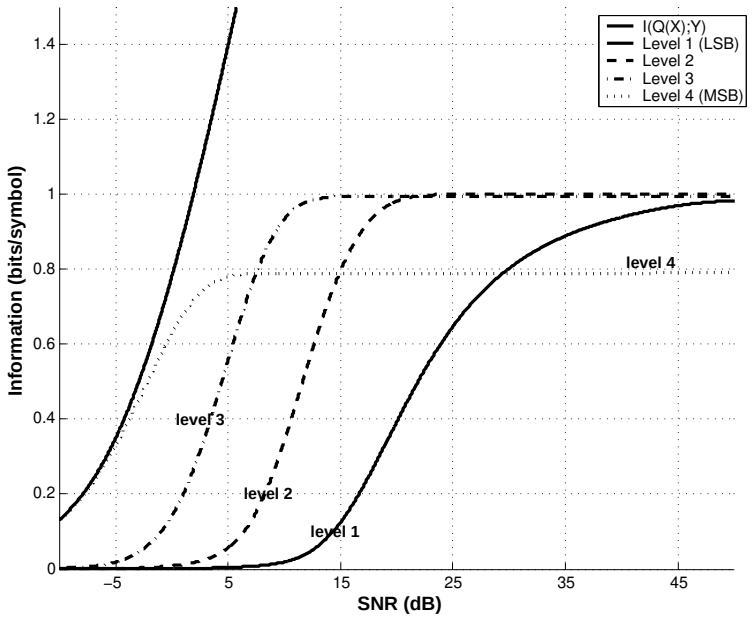
$$R^i_{opt} = 1 - \left(I(V_i; Y | V_1 \ldots V_{i-1}, \mathrm{SNR} = \infty) - I(V_i; Y | V_1 \ldots V_{i-1}, \mathrm{SNR} = s) \right). \qquad (3.28)$$

Figures 17(a) and 17(b) show the mutual information $I(V_i; Y | V_1 \ldots V_{i-1}, \mathrm{SNR} = s)$ as a function of the normalized SNR $10 \log(\Sigma^2/2\sigma^2)$, for 16 quantization intervals ($\ell = 4$ bit labeling). The intervals are optimized for each SNR according to the procedure mentioned earlier.

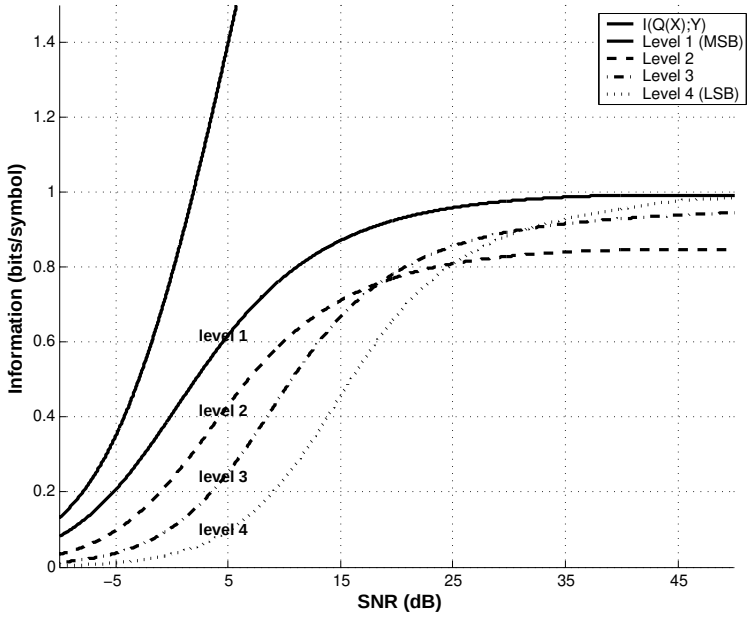
With natural labeling, the mutual information obtained for the first two level is close to zero at low SNR, and consequently, according to Equation (3.28), the code rate required is also extremely small. In this regime, rather than attempting to design efficient low rate codes, it is easier to use a zero rate code that would simply *disclose* the bits of the entire level. This simplification has negligible impact on reconciliation efficiency and reduces the number of codes to be designed. With anti-natural labeling, all levels contribute significantly to the total mutual information and no such simplification is possible; therefore, in all subsequent MLC/MSD simulations, we use natural labeling and carefully choose the number of quantization intervals to operate in a regime where only two codes are really needed.

For instance, at an SNR $\Sigma^2/\sigma^2 = 3$, the rates required with 16 quantization intervals and natural labeling are

$$R^1_{opt} = 0.002, \quad R^2_{opt} = 0.016, \quad R^3_{opt} = 0.259, \quad \text{and} \quad R^4_{opt} = 0.921.$$



(a) Natural labeling.



(b) Anti-natural labeling.

Figure 17. Mutual information by level.

Note that the effect of quantization is negligible since $I(\hat{X};Y)$ differs from $I(X;Y)$ by less than 0.02 bits.

In order to further simplify the code design, we use irregular LDPC codes optimized for a binary-input Gaussian channel as component codes. Good degree distributions with thresholds close to capacity are obtained via density evolution [49]. For our simulations, a large block length of $200,000$ bits is used, and Tanner graphs are generated randomly while avoiding loops of length two and four. In spite of the long block length, the performance of all constructed codes is in general well below that of their capacity achieving counterparts. Consequently, achieving perfect error correction with high probability is only possible at the cost of a code rate reduction. Cutting down the rates of each component code would disclose far too many bits, but, as we discuss in the next section, a careful choice of codes and iterations between levels make it possible to achieve high reconciliation efficiency.

3.3.2 Practical performance of MLC/MSD-like reconciliation

In this section, we present a pragmatic (and heuristic) code choice strategy based on Extrinsic INformation Transfer (EXIT) charts [50], which are a convenient tool to visualize the transfer of mutual information between decoders and demappers involved in MLC/MSD. The behavior of each decoder or demapper is summarized by a curve $I_E = T(I_A)$, characterizing the amount of extrinsic information I_E obtained with a certain *a priori* information I_A at the input.

The demapper transfer curves $I_E = T_d(I_A)$ cannot be computed in closed form but can be obtained via Monte-Carlo simulations using Equation (3.23). Likewise, the transfer curves $I_E = T_c(I_A)$ of the constructed LDPC codes are obtained by Monte-Carlo simulations, assuming with Gaussian *a priori* information. Examples of LDPC transfer curves, obtained after 100 iterations of Sum-Product decoding, are shown in Figure 18. As expected, these curves show that low rate codes gather extrinsic information at a slower pace than high rate codes, which suggests that high reconciliation efficiency can be obtained by compromising only on the rate of high rate codes and by using iterations to compensate for the poor performance of lower rate ones.

Let us now detail how practical codes rates can be found in the case $\Sigma^2/\sigma^2 = 3$ with

16 quantization intervals (4 bits quantization) and a natural labeling. As explained earlier, the first two levels are entirely disclosed, and, in theory, one would need two ideal codes with rates 0.26 and 0.92, respectively, to perform MSD. Instead, following the strategy suggested by the shape of EXIT curves, a rate 0.25 LDPC code is used for the third level, and we select a high rate code that gathers enough extrinsic information to pursue the decoding process and corrects all errors when the *a priori* information is 0.92. The selection is performed heuristically by drawing EXIT charts and ensuring that iterations allow a successful decoding. For instance, a rate 0.86 LDPC code 0.86 offers a good compromise. Figure 18 shows that realistic decoding trajectories are close to the decoding behavior predicted by EXIT charts The practical code rates selected according to the same procedure

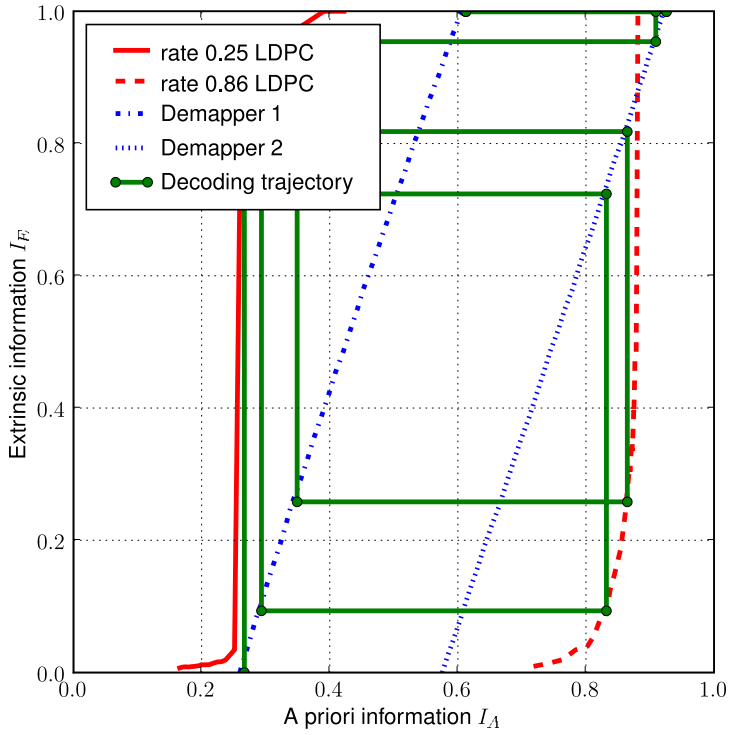


Figure 18. Iterative decoding trajectory when $\Sigma^2/\sigma^2 = 3$ with 16 quantization intervals and binary mapping. Decoding trajectory is averaged over 10 blocks.

for different values of SNR are given in Table 1. Note that when rate 1.0 codes were required, we used algebraic codes with error correcting capability of 1 instead of LDPC codes.

Table 1. Parameters used for MLC/MSD-like reconciliation.

SNR	Intervals	$I(\hat{X};Y)$	$H(\hat{X})$	Optimum rates	Practical rates
1	12 (4 levels)	0.49	3.38	0.001/0.008/0.187/0.915	0/0/0.16/0.86
3	16 (4 levels)	0.98	3.78	0.002/0.016/0.259/0.921	0/0/0.25/0.86
7	22 (5 levels)	1.47	4.23	0.002/0.020/0.295/0.924/1	0/0/0.28/0.86/1
15	32 (5 levels)	1.97	4.68	0.002/0.025/0.332/0.934/1	0/0/0.31/0.86/1

3.3.3 Choice of codes and rates for BICM-like reconciliation

For BICM-like reconciliation, we do not constrain the labeling to satisfy the symmetry condition given in Equation (3.27). In fact, since a single code is applied to an interleaved version of the label bits, a code optimized for a symmetric labeling should perform well, provided the labeling produces a balanced number of zeros and ones. Note that no additional interleaving is needed here since LDPC codes inherently interleave bit nodes to define parity-check equations.

The optimal code rate required at a given SNR is

$$R_{opt} = 1 - \frac{H(\hat{X}) - I(SNR)}{\ell}, \tag{3.29}$$

where $I(s)$ is the maximum BICM-capacity at SNR s. $I(s)$ depends on the mapping and cannot be computed exactly; however, if we let V_m be the binary random variable at level m ($1 \leq m \leq \ell$), we can estimate lower and an upper bounds as follows.

$$H(\hat{X}) - \sum_{m=1}^{\ell} H(V_m|Y) \leq I(s) \leq \min\left\{H(\hat{X}), \sum_{m=1}^{\ell} I(V_m;Y)\right\}, \tag{3.30}$$

Hence the code rate can be bounded as

$$1 - \frac{\sum_{m=1}^{\ell} H(V_m|Y)}{\ell} \leq R_{opt} \leq 1 - \frac{\max\left\{0, H(\hat{X}) - \sum_{m=1}^{\ell} I(V_m;Y)\right\}}{\ell}. \tag{3.31}$$

For instance with 16 quantization intervals, $\Sigma^2/\sigma^2 = 3$ and a gray mapping, we obtain

$$0.26 \leq R_{opt} \leq 0.27.$$

Consequently, the maximum reconciliation efficiency is less than 88%. When choosing a code one has to ensure that the rate is also compatible with the mapping used. Figure 19 shows the transfer curves of a rate 0.16 LDPC code optimized for the Gaussian channel as well as various demapper transfer curves. All transfer curves are obtained via Monte-Carlo

simulations with Gaussian *a priori* information. Perfect decoding is possible if the LDPC code transfer curve remains below the demapper curve. It clearly appears that all mappings cannot be used and that no mapping can gather high extrinsic information for both low and high *a priori* information. Gray mapping gathers the highest extrinsic information without *a priori* information but the slope of its transfer curve is the steepest, which means that it has to be associated with a strong code. The other mappings can be used with weaker but lower rate codes and are therefore not suitable for efficient reconciliation.

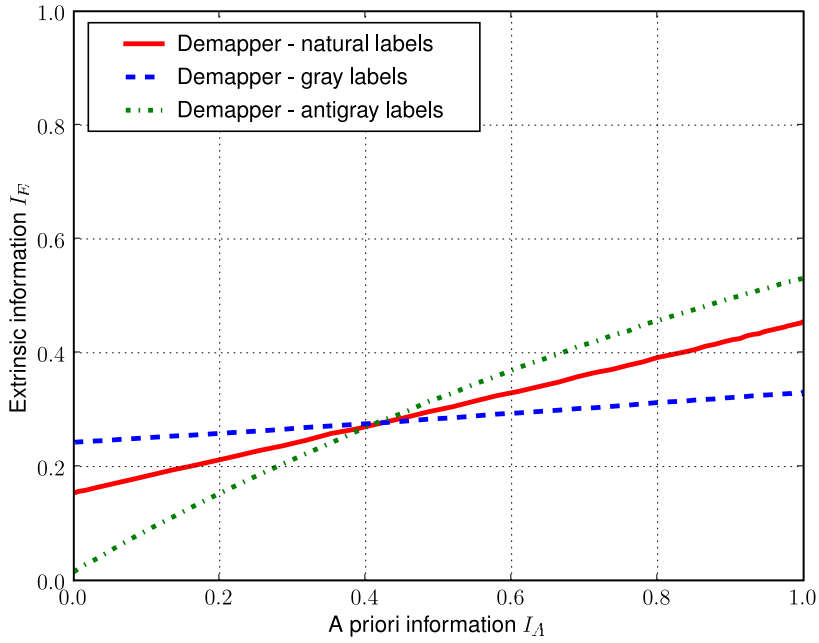


Figure 19. Transfer curves of demapper and code used in BICM-like reconciliation for $\Sigma^2/\sigma^2 = 3$ and 16 quantization intervals.

Unfortunately even with Gray mapping and a strong code we found that the practical code rates were far below the optimal ones. As shown in Figure 19 a rate 0.16 LDPC code is required to ensure full error correction even though the demapper initially feeds the decoder with 0.24 *a priori* information bits.

3.3.4 Simulation results

Table 2 shows the reconciliation efficiency obtained with our MLC/MSD-like procedure for different values of the SNR and compares it with the efficiency of sliced error correction.

Simulations were performed over 50 blocks of size 200,000, and all errors were corrected. When rate-1 codes were required, we used a BCH code with block length 4091 and error correcting capability $t = 1$. This disclosed slightly less than 0.003 additional bits per symbol sent. Since high-rate LDPC codes would sometimes fail to correct a couple of erroneous bits we also applied the same BCH code on top of these LDPC codes.

All sliced error correction results are given for a quantization with 32 intervals. The efficiencies $\eta_{\mathrm{SEC}}^{max}$, η_{SEC}^{1} and η_{SEC}^{2} refer to the efficiency with ideal binary codes, interactive error correction and one-way error correction with Turbo-codes, respectively, as reported in [51]. η_{MLC} is the efficiency obtained with MLS/MSD-like reconciliation using the code rates and quantizers of Table 1, while $\eta_{\mathrm{MLC}}^{max}$ is the maximum efficiency attainable with capacity achieving codes.

SNR	$\eta_{\mathrm{SEC}}^{max}$	η_{SEC}^{1}	η_{SEC}^{2}	$\eta_{\mathrm{MLC}}^{max}$	η_{MLC}
1	75%	60%	<50%	98%	79.4%
3	87%	79%	67%	98%	88.7%
7	90%	84%	76%	98%	90.9%
15	92%	87%	82%	98.5%	92.2%

Table 2. Reconciliation efficiency.

Clearly, the proposed reconciliation procedure achieves close if no better efficiency that sliced error correction with ideal codes.

CHAPTER 4

OPPORTUNISTIC KEY AGREEMENT OVER QUASI-STATIC WIRELESS CHANNELS

This chapter builds upon the results obtained in Chapter 3 and investigates the design of a practical communication scheme providing information-theoretic security for wireless channels. We begin by establishing the fundamental security limits of quasi-static fading channels in terms of average secrecy capacity and probability of outage of secrecy capacity, and subsequently describe a secure communication protocol that exploits the fluctuations of fading coefficients to allow the efficient generation of secret keys. The performance of the protocol is characterized analytically in asymptotic regimes and through Monte-Carlo simulations. We also consider a more realistic situation where the eavesdropper's channel state information is imperfectly known, and we show the effectiveness of the protocol.

4.1 Information-theoretic security over wireless channels

4.1.1 Wireless system setup

Figure 20 illustrates the wireless system setup considered in this chapter. A base station (Alice) wants to send messages, represented by the random variable M to a user in the network (Bob). Messages are encoded into codewords $\mathbf{X}^n$ for transmission over the channel. The main channel is a discrete-time Rayleigh fading channel, and Bob's observations are given by

$$Y_m(i) = H_m(i)X(i) + N_m(i) \quad \forall i \in \{1, \ldots, n\},$$

where $H_m(i)$ is a circularly symmetric complex Gaussian random variable with zero-mean and unit-variance representing the main channel fading coefficient and $N_m(i)$ is a zero-mean circularly symmetric complex Gaussian noise with variance σ_m^2.

Another user in the network (Eve) is also capable of eavesdropping Alice's transmissions. The eavesdropper's channel is an independent discrete-time Rayleigh fading channel, and Eve's obervations are given by

$$Y_w(i) = H_w(i)X(i) + N_w(i) \quad \forall i \in \{1, \ldots, n\},$$

where $H_w(i)$ denotes a circularly symmetric complex Gaussian random variable with zero-mean and unit-variance representing the eavesdropper's channel fading coefficient and $N_w(i)$ denotes a zero-mean circularly symmetric complex Gaussian noise with variance σ_w^2.

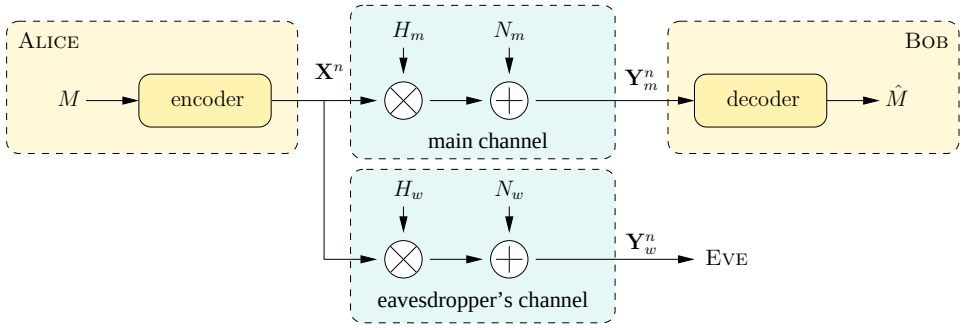


Figure 20. Wireless wiretap channel setup.

It is assumed that the channel input, the channel fading coefficients, and the channel noises are all independent. It is also assumed that both the main and the eavesdropper's channels are quasi-static fading channels, that is, the fading coefficients, albeit random, are constant during the transmission of an entire codeword ($H_m(i) = H_m, \forall i = 1, \ldots, n$ and $H_w(i) = H_w, \forall i = 1, \ldots, n$) and, moreover, independent from codeword to codeword. This corresponds to a situation where the coherence time of the channel is large.

The codewords transmitted by Alice are subject to the average power constraint

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{|X(i)|^2\} \leq P.$$

Consequently, the instantaneous SNR at Bob's receiver is

$$\Gamma_m(i) = P|H_m(i)|^2/\sigma_m^2 = P|H_m|^2/\sigma_m^2 = \Gamma_m$$

and its average value is

$$\overline{\gamma}_m(i) = P\mathbb{E}\{|H_m(i)|^2\}/\sigma_m^2 = P\mathbb{E}\{|H_m|^2\}/\sigma_m^2 = \overline{\gamma}_m.$$

Likewise, the instantaneous SNR at Eve's receiver is

$$\Gamma_w(i) = P|H_w(i)|^2/\sigma_w^2 = P|H_w|^2/\sigma_w^2 = \Gamma_w$$

and its average value is

$$\overline{\gamma}_w(i) = P\mathbb{E}\{|H_w(i)|^2\}/\sigma_w^2 = P\mathbb{E}\{|H_w^2\}/\sigma_w^2 = \overline{\gamma}_w.$$

Since the channel fading coefficients H are zero-mean complex Gaussian random variables and the instantaneous SNR $\Gamma \propto |H|^2$, it follows that Γ is exponentially distributed, specifically

$$p(\gamma_m) = \frac{1}{\overline{\gamma}_m} \exp\left(-\frac{\gamma_m}{\overline{\gamma}_m}\right), \qquad \gamma_m > 0 \tag{4.1}$$

and

$$p(\gamma_w) = \frac{1}{\overline{\gamma}_w} \exp\left(-\frac{\gamma_w}{\overline{\gamma}_w}\right), \qquad \gamma_w > 0. \tag{4.2}$$

The transmission rate between Alice and Bob is defined as

$$R = \frac{H(M)}{n},$$

and secrecy with respect to the eavesdropper is measured in terms of the equivocation rate

$$R_e = \frac{H(M|Y_w^n)}{n}.$$

In the rest of this chapter, we focus our attention on the case of perfectly secure communication ($R_e = R$) and limit our analysis to the study of the secrecy capacity.

4.1.2 Impact of fading on secure communications

In this section, we study the impact of fading on the secrecy capacity of this wireless system by considering two metrics: *average secrecy capacity* and *probability of outage of secrecy capacity*. We assume that Alice and Bob have perfect knowledge of the main channel fading coefficient and that Eve also has perfect knowledge of the eavesdropper's channel fading coefficient. These assumptions are realistic for the slow fading wireless environment under consideration: both receivers can always obtain close to perfect channel estimates, and additionally, the legitimate receiver can also feedback the channel estimates to the legitimate transmitter. Moreover, we assume that Alice and Bob also have partial knowledge of the eavesdropper's channel fading coefficient. Since Eve is assumed to be another active user in the wireless network (*e.g.* in a TDMA environment), Alice can estimate the eavesdropper's channel during Eve's transmissions.

Nevertheless, we shall see that the probability of outage of secrecy capacity allows, in principle, to consider also situations where no Channel State Information (CSI) about the

eavesdropper's channel is available to Alice and Bob. This case corresponds to the situation where Eve is a purely passive and malicious eavesdropper in the wireless network.

We start by deriving the secrecy capacity for one realization of a pair of quasi-static fading channels with complex noise and complex fading coefficients. Based on the result of Theorem 2.2, we obtain the following lemma.

Lemma 4.1 ([12])**.** *The secrecy capacity for one realization* (γ_m, γ_w) *of the quasi-static complex fading wiretap-channel is given by*

$$C_s(\gamma_m, \gamma_w) = \begin{cases} \log(1+\gamma_m) - \log(1+\gamma_w) & \text{if } \gamma_m > \gamma_w \\ 0 & \text{if } \gamma_m \leq \gamma_w. \end{cases} \tag{4.3}$$

Proof. See Section 4.4.1. □

□ **Average secrecy capacity** If perfect CSI of the eavesdropper's channel is available to Alice, the coding scheme achieving the instantaneous secrecy capacity can be adapted to every realization of the fading coefficients. Therefore, in principle, any average secure communication rate below the *average secrecy capacity* of the channel

$$\overline{C}_s = \int_0^\infty \int_0^\infty C_s(\gamma_m, \gamma_w)\, p(\gamma_m) p(\gamma_w) d\gamma_m d\gamma_w$$

is achievable. Note that the average secrecy capacity is easily computable numerically. It can be shown (see the proof of Lemma 4.2 in Section 4.4.2) that

$$\overline{C}_s = F(\overline{\gamma}_m) - F(\frac{\overline{\gamma}_m \overline{\gamma}_w}{\overline{\gamma}_w + \overline{\gamma}_m}), \tag{4.4}$$

where

$$F(x) = \int_0^\infty \log_2(1+u) \frac{1}{x} e^{-\frac{u}{x}} du = e^{\frac{1}{x}} \mathrm{E}_1(x^{-1}) \frac{1}{\log 2} \tag{4.5}$$

and E_1 is the exponential-integral function.

Figure 21 compares the average secrecy capacity of a quasi-static fading channel to the secrecy capacity of a classic wiretap Gaussian channel. Strikingly, the average secrecy rate of the fading channel is higher than or close to the secrecy capacity of the Gaussian wiretap channel. Moreover, contraty to the Gaussian wiretap channel, the average secrecy rate of the fading channel is non-zero even when the average SNR of the main channel is lower than

the average SNR of the eavesdropper's channel. These observations underline the potential of fading channels to secure the transmission of information between two legitimate parties against a possible eavesdropper.

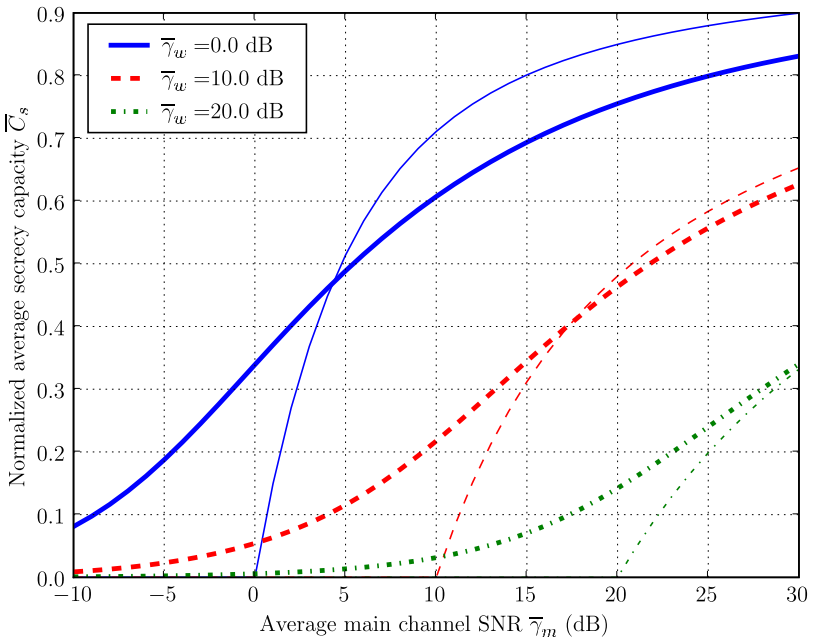


Figure 21. Normalized average secrecy rate versus $\overline{\gamma}_m$, for selected values of $\overline{\gamma}_w$. Thinner lines correspond to the normalized average secrecy rate in the case of Rayleigh fading channels, while thicker lines correspond to the secrecy capacity of the Gaussian wiretap channel. Normalization is effected with respect to the capacity of an AWGN channel with SNR equal to $\overline{\gamma}_m$.

□ **Outage probability of secrecy capacity** The secrecy capacity of a quasi-static Rayleigh fading channel can also be characterized characterized in terms of outage probability.

Proposition 4.1.

$$P[C_s > \tau] = P_0(\tau) \quad = \quad \frac{\bar{\gamma}_m}{\bar{\gamma}_m + \bar{\gamma}_w 2^{\tau}} \exp\left(-\frac{2^{\tau}-1}{\bar{\gamma}_m}\right)$$

Proof.

$$\begin{aligned} \mathrm{P}[C_s > \tau] &= \mathrm{P}\left[\log\frac{1+\Gamma_m}{1+\Gamma_w} > \tau\right] = \mathrm{P}[\Gamma_m > 2^{\tau}(1+\Gamma_w) - 1], \\ &= \int_0^{\infty} p(\gamma_w)\left(\int_{2^{\tau}(1+\gamma_w)-1}^{\infty} p(\gamma_m)d\gamma_m\right)d\gamma_w, \end{aligned}$$

where the last equality exploits the fact that $p(\gamma_m, \gamma_w) = p(\gamma_m)p(\gamma_w)$. The expressions

of $p(\gamma_m)$ and $p(\gamma_w)$ are given by Equation (4.2), and the result follows after some simple algebra. □

Based on this result, it follows immediately that for average signal-to-noise ratios $\overline{\gamma}_m$ and $\overline{\gamma}_w$ on the main channel and the eavesdropper's channel, respectively, the probability of strictly positive secrecy capacity is

$$\mathrm{P}[C_s > 0] \quad = \quad \frac{\overline{\gamma}_m}{\overline{\gamma}_m + \overline{\gamma}_w}. \tag{4.6}$$

It is also useful to express this probability in terms of parameters related to user location. Letting d_m be the distance between Alice and Bob, d_w be the distance between Alice and Eve, and α be the pathloss exponent characterizing the strength of signal attenuation with distance, we have $\overline{\gamma}_m \propto 1/d_m^{\alpha}$ and $\overline{\gamma}_w \propto 1/d_w^{\alpha}$ [52]; therefore, the probability of strictly positive secrecy capacity can be written as

$$\mathrm{P}[C_s > 0] = \frac{1}{1 + (d_m/d_w)^{\alpha}} \tag{4.7}$$

When $\gamma_m \gg \gamma_w$ (or $d_m \ll d_w$) then $\mathrm{P}[C_s > 0] \approx 1$ (or $\mathrm{P}[C_s = 0] \approx 0$). Conversely, when $\gamma_w \gg \gamma_m$ (or $d_w \ll d_m$) then $\mathrm{P}[C_s > 0] \approx 0$ (or $\mathrm{P}[C_s = 0] \approx 1$). This confirms the intuition that greater security is achieved when Eve is further away from Alice than Bob.

We are now ready to characterize the outage probability

$$\mathcal{P}_{\mathrm{out}}(R_s) = \mathrm{P}[C_s < R_s],$$

that is the probability that the instantaneous secrecy capacity is less than a target secrecy rate $R_s > 0$. The operational significance of this definition of outage probability is twofold. First, it provides the fraction of fading realizations for which the wireless channel can support a secure rate of R_s bits/channel use. Second, it provides a security metric for the situation where Alice and Bob have no CSI about the eavesdropper. In fact, in this case, Alice has no choice but to set her secrecy rate to a constant R_s. By doing so, Alice is assuming that the capacity of the wiretap channel is given by $C'_w = C_m - R_s$. As long as $R_s < C_s$, Eve's channel is worse than Alice's estimate, and consequently, $C_w < C'_w$ and the wiretap codes used by Alice ensures perfect secrecy. Otherwise, if $R_s > C_s$, $C_w > C'_w$ and information-theoretic security is compromised.

Proposition 4.2. *The outage probability for a target secrecy rate R_s is given by*

$$\mathcal{P}_{out}(R_s) = P[C_s \leq R_s] = 1 - \frac{\overline{\gamma}_m}{\overline{\gamma}_m + 2^{R_s}\overline{\gamma}_w} \exp\left(-\frac{2^{R_s}-1}{\overline{\gamma}_m}\right). \tag{4.8}$$

It is illustrative to examine the asymptotic behavior of the outage probability for extreme values of the target secrecy rate R_s. From Equation (4.8) it follows that when $R_s \to 0$,

$$\mathcal{P}_{\text{out}} \to \frac{\overline{\gamma}_w}{\overline{\gamma}_m + \overline{\gamma}_w}$$

and when $R_s \to \infty$, we have that $\mathcal{P}_{\text{out}} \to 1$, such that it becomes impossible for Alice and Bob to transmit secret information (at very high rates).

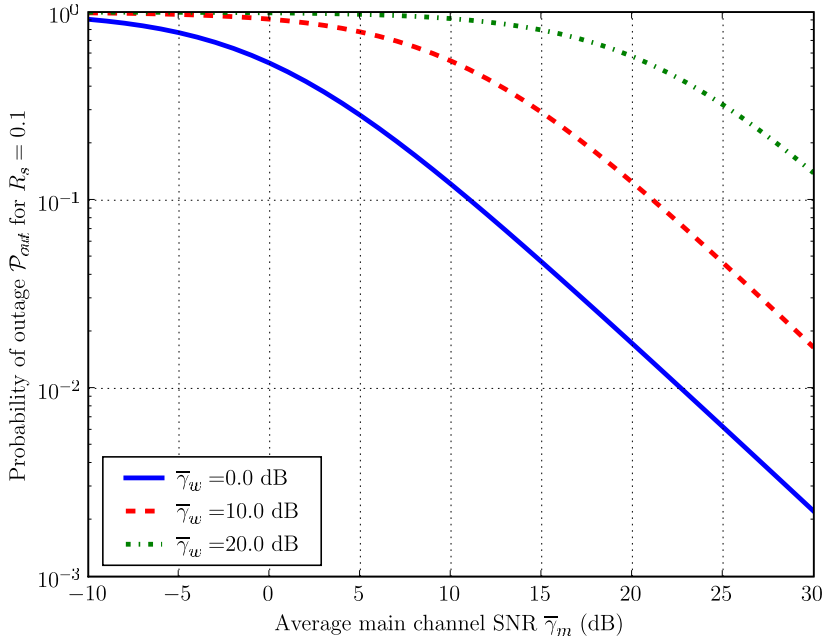


Figure 22. Outage probability versus $\overline{\gamma}_m$, for selected values of $\overline{\gamma}_w$ and for a normalized target secrecy rate equal to 0.1. Normalization is effected with respect to the capacity of an AWGN channel with SNR equal to $\overline{\gamma}_m$.

Also of interest is the asymptotic behavior of the outage probability for extreme values of the average SNRs of the main channel and the eavesdropper's channel. When $\overline{\gamma}_m \gg \overline{\gamma}_w$, equation Equation (4.8) yields

$$\mathcal{P}_{\text{out}}(R_s) \approx 1 - \exp\left(-\frac{2^{R_s}-1}{\overline{\gamma}_m}\right),$$

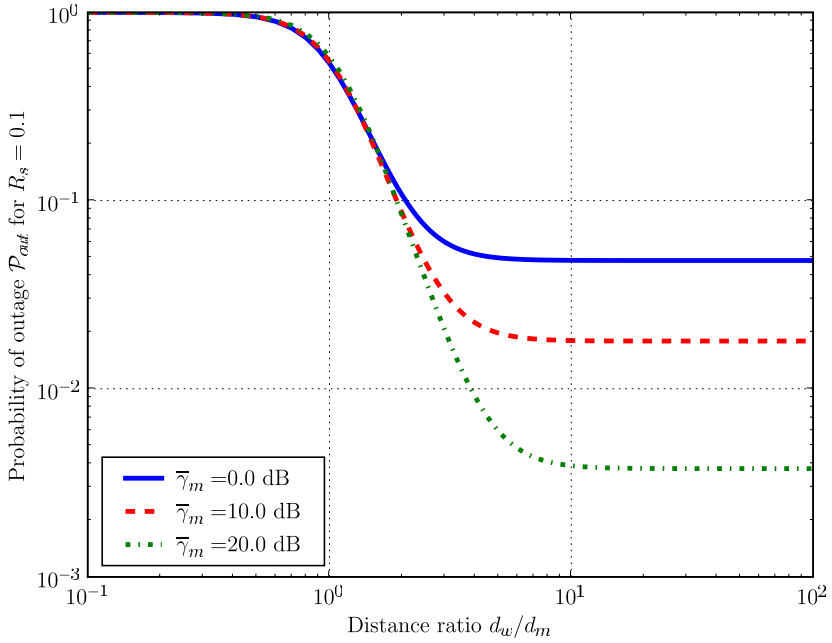


Figure 23. Outage probability versus d_W/d_M, for selected values of $\overline{\gamma}_m$ and for a normalized target secrecy rate equal to 0.1. Normalization is effected with respect to the capacity of an AWGN channel with SNR equal to $\overline{\gamma}_m$.

and in a high SNR regime $\mathcal{P}_{\text{out}} \approx (2^{R_s} - 1)/\overline{\gamma}_m$, i.e. the outage probability decays as $1/\overline{\gamma}_m$. Conversely, when $\overline{\gamma}_w \gg \overline{\gamma}_m$,

$$\mathcal{P}_{\text{out}}(R_s) \approx 1,$$

and confidential communication becomes impossible.

Figure 22 depicts the outage probability versus $\overline{\gamma}_m$, for selected values of $\overline{\gamma}_w$ and for a normalized target secrecy rate equal to 0.1. Observe that the higher $\overline{\gamma}_m$ the lower the outage probability, and the higher $\overline{\gamma}_w$ the higher the probability of an outage. Moreover, if $\overline{\gamma}_m \gg \overline{\gamma}_w$, the outage probability decays as $1/\overline{\gamma}_m$. Conversely, if $\overline{\gamma}_w \gg \overline{\gamma}_m$ the outage probability approaches one. The relationship between outage and distance is highlighted in Figure 23.

The outage probability is also convenient to analyze the situation where Alice might only have imperfect estimates $\hat{H}_m$ and $\hat{H}_w$ of the gains of the main channel and eavesdropper's channels, respectively. We can reasonably assume that Bob cooperates with Alice, which

allows her to obtain a perfect estimate of the main channel fading coefficient. Hence,

$$\hat{H}_m = H_m,$$

where H_m is the true fading coefficient of the main channel. Unfortunately Eve may not be as helpful and Alice's knowledge of the eavesdropper's channel fading is more likely to be noisy. In order to assess the performance of our protocol under more realistic conditions, we model Alice's estimate of Eve's fading coefficient by

$$\hat{H}_w = H_w + Z'_w,$$

where H_w is the true fading coefficient and Z'_w is a zero-mean complex Gaussian noise with known variance σ_e^2 per dimension.

In the absence of additional information allowing Alice to refine her estimation, we have to resort once again to an outage analysis. If Alice communicates by blindly assuming that her estimation is accurate, an outage occurs whenever Alice underestimates the gain of the eavesdropper's channel and attempts to achieve a secure communication rate not supported by the channel.

Proposition 4.3. *The probability of outage is upper bounded by*

$$\mathcal{P}_{out} \leq \frac{1}{2} - \frac{1}{2}\frac{1}{\sqrt{1+2/\sigma_e^2}}. \tag{4.9}$$

Proof. See Section 4.4.3. □

This upper bound on the outage probability is a decreasing function of the variance of the channel estimation error σ_e^2, so that the higher σ_e^2 the lower the outage probability. This counterintuitive result stems from the fact that, at moderate values of the variance of the channel estimation error, Alice tends to consistently underestimate the true wiretap fading coefficient. Consequently, she consistently attempts to communicate at secure rates lower than what the true instantaneous secrecy capacity of the channel would allow.

4.1.3 Opportunistic secret key agreement

In principle, secure communications over wireless quasi-static fading channels can be achieved with codes designed for the Gaussian wiretap channel; however, although the secrecy capacity of the Gaussian wiretap channel has been fully characterized, designing practical coding

schemes is still an open problem. On the other hand, the results on secret key agreement by public discussion and privacy amplification presented in Chapter 2 support the idea that the generation of information theoretically secure keys from common randomness is a somewhat less difficult problem. This naturally suggest a four-step approach to secure communications: *randomness sharing*, *information reconciliation*, *privacy amplification* and *secure communication*.

- **Opportunistic randomness sharing.** To share randomness, Alice transmits discrete random symbols, represented by the random variable X, over the wireless channel. Bob and Eve observe correlated symbols, represented by the random variables Y_m and Y_w, respectively. In theory, as long as Eve and Bob do not share the same information, the amount of secrecy that Alice and Bob can distill from their common randomness is non-zero [24]; however, we are interested in designing a *one-way* secret key agreement protocol, which requires communications from Alice to Bob only. Therefore, the common randomness must be such that $I(X;Y_m) > I(X;Y_w)$. Clearly, this is the case if randomness is shared when the secrecy capacity of the wireless channel is strictly positive. Hence, provided perfect CSI of the eavesdropper's channel is available, Alice and Bob should *opportunistically* exploit the fluctuations of the instantaneous secrecy capacity C_s with time, that is they should attempt to share randomness only when C_s is sufficiently large. Specifically, in the remaining of the paper, we take the set of fading realization (γ_m, γ_w) for which an opportunistic transmission of randomness is performed to be

$$\mathcal{D}(\tau,\kappa) = \{(\gamma_m,\gamma_m) : C_s \geq \tau, C_m > \kappa\}\,. \tag{4.10}$$

  The threshold τ ensures that a minimum amount of secrecy can be distilled from the randomness while the threshold κ ensures that the correlation between Alice and Bob's data is high enough. As discussed in Chapter 3, the latter condition is required for practical algorithms. Finally, let us emphasize that even though we assume perfect CSI of the eavesdropper's channel, the behavior of the protocol is governed by the fading realizations in the set $\mathcal{D}(\tau,\kappa)$ and, therefore, by a probability of outage. This connection will be established explicitly in Section 4.3.

- **Key generation: reconciliation and privacy amplification.** When the estimated fading realizations are such that the secrecy capacity or main channel capacity are too small ($(\gamma_m, \gamma_w) \notin \mathcal{D}(\tau, \kappa)$), Alice and Bob communicate to generate a secure key from the shared randomness previously obtained. As discussed in Chapter 2, key generation is performed in two steps. First, Alice and Bob "reconcile" their randomness, that is they correct the discrepancies in their randomness by exchanging additional error-correction information. Second, Alice and Bob distill secret-bits from their corrected date using privacy amplification.

- **Secure communication.** Alice and Bob can finally use their secret key to transmit messages, using either a one-time pad to ensure perfect secrecy or any symmetric cypher.

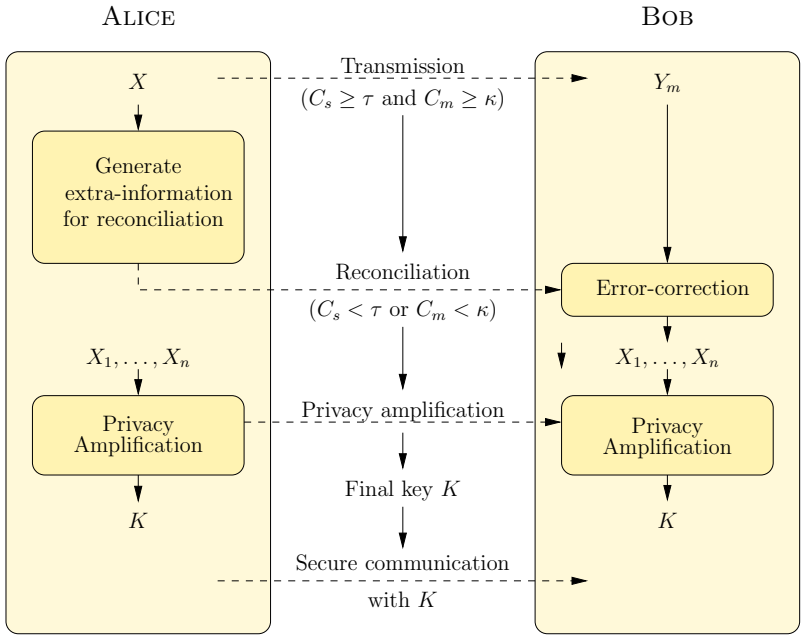


Figure 24. Flowchart of the opportunistic protocol.

The flowchart of the opportunistic protocol is shown in Figure 24. Note that the randomness sharing and privacy amplification steps rely on a perfect estimation of the fading coefficients to calculate the instantaneous secrecy capacity and correctly estimate the amount of secrecy to distill. We shall see in Section 4.3 that this assumption can be somewhat

alleviated to consider a more realistic situation where only imperfect CSI (or a conservative estimate) is available for the eavesdropper's channel.

4.2 Practical algorithms for Secret-key Agreement

In this section, we detail the various steps of the protocol presented in the previous section. To ease the presentation, we present the protocol for a Gaussian wiretap channel, which corresponds to a single realization of the fading coefficients (γ_m, γ_w) in the wireless setup of Section 4.1. Its performance in the quasi-static fading case is evaluated in Section 4.3.

The existence of common information between Alice and Bob is the essential ingredient for secret key agreement. In a wiretap scenario, Alice can generate this shared randomness by transmitting a sequence $X^n = (X_1, \ldots, X_n)$ of n i.i.d. realizations of a discrete random variable X over the main channel, which provides Bob and Eve with sequences of correlated continuous random variables $Y_m^n = (Y_{m,1}, \ldots, Y_{m,n})$ and $Y_w^n = (Y_{w,1}, \ldots, Y_{w,n})$, respectively.

As discussed in Chapter 3, for our application to the Gaussian wiretap channel, we use a MultiLevel Coding (MLC) and MultiStage Decoding (MSD) scheme to reconcile and correct the differences between $\hat{X}$ and X. Since the reconciliation algorithm is not optimal, its efficiency β is stricly less than unity. At the end of the reconciliation step, Alice an Bob share with high probability the common sequence X^n with entropy $n_{rec} = nH(X)$, which is then compressed into a binary sequence S of length n_{rec}.

Finally, the privacy amplification method described in Section 2.5 can be used to distill a secret key of length

$$k = n(\beta I(X;Y_m) - I(X;Y_w)) - 2s - 2 - r_0 \tag{4.11}$$

with $r_0 > 0$,$s > 0$. Letting E be the information accessible to the eavesdropper and G be the random variable denoting a hash function chosen at random from a family of universal hash functions, Eve's uncertainty on the key $K = G(S)$ is guaranteed to satisfy

$$H(K|G,E) \geq k - \frac{2^{-r_0}}{\ln 2} \quad \text{with probability} \quad 1 - 2^{-s}. \tag{4.12}$$

For our protocol, we do not develop anything new and use standard families of hash functions [40, 10].

Finally, the secret key generated $K = G(S)$ can be used to secure Alice's message, using either a one-time pad for perfect secrecy or a standard secret key encryption algorithm. Eve's uncertainty $H(K|G, E)$ about the key is as close to k as we want according to Equation (4.12).

Since the size of the key generated from common randomness is proportional to

$$\beta I(X; Y_m) - I(X; Y_w) \quad \text{bits/symbol},$$

we choose the random variable X such that the mutual information $I(X; Y_m)$ is maximized. Ideally, Alice should choose X achieving the capacity $C_m = 0.5 \log_2(1 + \gamma_m)$ of the main channel, which is possible only with continuous Gaussian random variables; however the discrete support $\mathcal{X}$ and the probability mass function of X can always be optimized so that $I(X; Y_m)$ approaches the channel capacity C_m with arbitrary precision. For instance, for a fixed size $N_c = |\mathcal{X}|$ of the support, this optimization can be performed with the algorithm proposed in [53]. Alternatively, a good approximation of the optimum can be obtained by expanding a uniformly spaced support $\{x_i\}_{i=1\ldots N_c} = \left\{\pm 1, \pm 3, \ldots, \pm \frac{N_c-1}{2}\right\}$ by a factor $\alpha \in \mathbb{R}^+$, and using a Maxwell-Boltzmann probability distribution

$$P(X = x_i) = \frac{\exp\left(-\lambda \alpha^2 |x_i|^2\right)}{\sum_j \exp\left(-\lambda \alpha^2 |x_i|^2\right)}. \tag{4.13}$$

Note that even though $I(X; Y_m)$ is not a convex function of α and λ, non-linear programming seems to be relatively insensitive to the initialization of the optimization. Also, N_c should be large enough so that $I(X; Y_m)$ approaches C_m within the required precision, which is discussed in Section 4.3.

Once again, we point out that the above protocol relies on the results of [38, 39, 41] that were only proven for discrete random variables whereas Y_m and Y_w are continuous random variables; however, it should be noted that these continuous random variables only appear as conditioning random variables in expressions such as $H(X|Y_w)$ where X is discrete. Therefore, the various results are still valid in this case. For instance, Y_m can be quantized into a discrete random variable Y_Δ such that $H(X|Y_\Delta)$ approaches $H(X|Y_m)$ with arbitrary precision as $\Delta \to 0$, and the Slepian-Wolf theorem still holds.

4.3 Performance evaluation

4.3.1 Performance metrics for secure communications

The information-theoretic secure rates of the secret key agreement protocol can be assessed only if the keys are used in conjunction with a one-time pad. However, in principle, the protocol could also be tailored to standard encryption algorithms. Although no information-theoretic security can be guaranteed in this latter case, combining a physical-layer key-generation technique with a symmetric encryption scheme could still be a means of enhancing security. In fact, key-generation rates could be substantially higher than those offered by public-key schemes. Moreover, keys generated from the physical layer are independent from one another, which ensures that the security of the system is re-initialized at each round of key-generation. An attacker getting access to one key would be none the wiser once the key is renewed. Based on these considerations, the performance of the opportunistic protocol is evaluated with the following metric.

Definition 4.1. *The average*[1] *η-secure throughput $\overline{T}_s(\eta)$ of a secret key agreement protocol is the average number of cyphertext bits transmitted per channel use, when the cyphertext is obtained with a symmetric encryption scheme such that the ratio of secret key bits used per cyphertext bits is η.*

In the above definition, generated secret key bits do not contribute to $\overline{T}_s(\eta)$ since keys themselves do not convey any information. The case $\eta = 1$ corresponds to the situation where one bit of secret key is used for each bit of cyphertext. Without loss of generality, we can assume that the encryption scheme is a one-time pad, and therefore, $\overline{T}_s(1)$ measures an average communication rate with perfect security. When $\eta < 1$, $\overline{T}_s(\eta)$ loses all significance in terms of information-theoretically secure communication rate; however, if k_s is the key length required by an encryption scheme, the corresponding key renewal rate is k_s/η channel uses.

Unlike wiretap coding, where messages are transmitted securely directly, secret key agreement requires additional communication to distill a key and send an encrypted message. Here, since we do not assume the existence of another public error-free channel, parts

[1] The average is taken over all channel realizations.

of the available communication rate has to be sacrificed for that purpose. We formalize this constraint by introducing the following metric.

Definition 4.2. *The average η-communication throughput $\overline{T}_c(\eta)$ is the average number of message bits per channel used that can be transmitted in addition to the message required for reconciliation and privacy amplification and to the messages encrypted with the keys.*

Clearly, $\overline{T}_s(\eta)$ and $\overline{T}_c(\eta)$ are not independent and, by definition, take only positive values.

We are now ready to characterize the maximum secure throughput of the protocol. To simplify equations, we use the following notation. For a given parameter $\alpha(\gamma_m, \gamma_w)$ depending on the fading realizations (γ_m, γ_w) and a set $\mathcal{D}$ of fading realizations, we let $\langle \alpha \rangle_{\mathcal{D}}$ denote the average of $X(\gamma_m, \gamma_w)$ over $\mathcal{D}$. We also assume that the coherence time of the channel is large enough, so that the block length n is large and the parameters s, r_0 of privacy amplification can be neglected, and that Alice and Bob can always communicate over the main channel at a rate close to capacity.

Proposition 4.4. *The maximum secure throughput $\overline{T}_s(\eta)$ achievable by the opportunistic secret key agreement protocol is*

$$\max_{\tau \geq 0. \kappa \geq \kappa_{min}} \eta^{-1} \langle \beta I(X;Y_m) - I(X;Y_w) \rangle_{\mathcal{D}(\tau,\kappa)} \tag{4.14}$$

$$\text{subject to} \quad \langle C_m \rangle_{\mathcal{D}^c(\tau,\kappa)} - \langle 2H(X) + \beta(\eta^{-1}-1) I(X;Y_m) - \eta^{-1} I(X;Y_w) \rangle_{\mathcal{D}(\tau,\kappa)} \geq 0, \tag{4.15}$$

where $\mathcal{D}^c(\tau,\kappa)$ denotes the complement of $\mathcal{D}(\tau,\kappa)$ in $\mathbb{R}_+^2$, and κ_{min} is imposed by the reconciliation algorithm.

Proof. When the fading realizations $(\gamma_m, \gamma_w) \in \mathcal{D}(\tau,\kappa)$, an opportunistic transmission is performed. From Equation (4.11), we know that the average number of key bits extractable per channel use is

$$\langle \beta I(X;Y_m) - I(X;Y_w) \rangle_{\mathcal{D}(\tau,\kappa)},$$

and therefore, the average secure throughtput is

$$\overline{T}_s(\eta) = \eta^{-1} \langle \beta I(X;Y_m) - I(X;Y_w) \rangle_{\mathcal{D}(\tau,\kappa)}. \tag{4.16}$$

From Equation (2.6), we also know that the average number of bits per channel use that have to be transmitted for reconciliation is

$$\langle H(X) - \beta I(X;Y_m)\rangle_{\mathcal{D}(\tau,\kappa)}. \tag{4.17}$$

The average number of bits per channel use required by privacy amplification depend on the number of bits required to identify a given universal hash function within its family. The minimum size of a family of universal hash functions $\mathcal{G} : \{0,1\}^{n_{rec}} \to \{0,1\}^{k}$ is known to be at least $2^{n_{rec}-k}$ [54], and identifying a given function therefore requires the transmission of $n_{rec} - k$ bits; however, no hashing scheme is known to achieve this bound for any n_{rec}, therefore we consider the more realistic situation where the identification requires the transmission of n_{rec} bits. For instance, this can be achieved with the following family [10].

$$\mathcal{H}_{\mathrm{GF}(2^{n_{rec}})\to\{0,1\}^{n_{key}}} = \{h_c : c \in \mathrm{GF}(2^{n_{rec}})\}, \tag{4.18}$$

where $h_c(x)$ is defined as n_{key} distinct bits of the product cx in a polynomial representation of $\mathrm{GF}(2^{n_{rec}})$. Consequently the average number of bits per channel use required by privacy amplification is

$$\langle H(X)\rangle_{\mathcal{D}(\tau,\kappa)}. \tag{4.19}$$

Based on our assumption that Alice and Bob can always communicate are a rate equal to the capacity of the main channel, the average number of bits available for communication in addition to the opportunistic transmissions is

$$\langle C_m\rangle_{\mathcal{D}^c(\tau,\kappa)}. \tag{4.20}$$

Therefore, the communication throughput is obtained by subtracting Equations (4.16-4.19) from Equation (4.20) and recalling that $\overline{T}_c(\eta) \geq 0$ yields the desired result. □

4.3.2 Asymptotic performance analysis

Obtaining analytical expression for the optimal performance of the opportunistic communication protocol is non-trivial on several accounts. First, the simplification of the expression in Proposition 4.4 requires the characterization of the trade-off between $H(X)$ and $I(X;Y_m)$ (or $I(X;Y_w)$) for an arbitrary random variable X. For a given $I(X;Y_w)$, we have observed

that the Maxwell-Boltzmann distribution of Equation (4.13) yields a smaller $H(X)$ than most other distributions, but for every pair of fading realizations (γ_m, γ_w) the parameters α and λ have to be optimized, which makes the analytical characterization intractable. Second, the optimal performance depends explicitly on the maximization over the parameters τ and κ.

Therefore, the following analysis considers a (sub-optimal) protocol where the random symbols sent over the channel during the opportunistic transmissions are chosen from a QAM constellation with *uniform* probability. We also assume that reconciliation is performed with efficiency $\beta = 1$ for all SNRs, and $\kappa = 0$.

Proposition 4.5 (adapted from [55])**.** *Let C be the capacity of a complex AWGN channel with input power constraint P, and let $N = \lfloor 2^{C/2+1} \rfloor^2$. If the input symbols X are chosen uniformly at random in the set , where Δ is optimized such that $\mathbb{E}\{X^2\} \leq P$, then the mutual information between the input X and the output Y bounded as*

$$C \geq I(X;Y) \geq C - \xi \quad \textit{with} \quad \xi > 0 \quad \textit{independent of} \quad C,$$

and the entropy of X is bounded as

$$C + 2 \geq H(X) \geq C.$$

Using these inequalities in the equations of Proposition 4.4, we obtain the following bounds

$$F(\overline{\gamma}_m) - \left(2 + \eta^{-1}\right) \langle C_m \rangle_{\mathcal{D}(\tau)} + \eta^{-1} \langle C_w \rangle_{\mathcal{D}(\tau)} - \xi(\eta^{-1} - 1) P_0(\tau) \geq \tag{4.21}$$

$$\langle C_m \rangle_{\mathcal{D}^c(\tau)} - \langle 2H(X) + \beta(\eta^{-1} - 1) I(X;Y_m) - \eta^{-1} I(X;Y_w) \rangle_{\mathcal{D}(\tau)} \tag{4.22}$$

$$\geq F(\overline{\gamma}_m) - \left(2 + \eta^{-1}\right) \langle C_m \rangle_{\mathcal{D}(\tau)} + \eta^{-1} \langle C_w \rangle_{\mathcal{D}(\tau)} - (4 + \xi\eta^{-1}) P_0(\tau) \tag{4.23}$$

and

$$\eta^{-1} \langle C_m \rangle_{\mathcal{D}(\tau)} - \eta^{-1} \langle C_w \rangle_{\mathcal{D}(\tau)} + \xi P_0(\tau) \geq \tag{4.24}$$

$$\eta^{-1} \langle \beta I(X;Y_m) - I(X;Y_w) \rangle_{\mathcal{D}(\tau)} \tag{4.25}$$

$$\geq \eta^{-1} \langle C_m \rangle_{\mathcal{D}(\tau)} - \eta^{-1} \langle C_w \rangle_{\mathcal{D}(\tau)} - \xi P_0(\tau) \tag{4.26}$$

The above bounds allows us to characterize the rates achievable by the protocol in asymptotic regimes.

□ **Secrecy-limited regime.** This regime corresponds to the situation where $\overline{\gamma}_m \to 0$, and therefore, the secrecy capacity over the wireless channel is mainly limited by the capacity of the eavesdropper's channel.

Theorem 4.1. *In the secrecy-limited regime, the secure throughput is bounded from below as*

$$\overline{T}_s(\eta) \geq \eta^{-1}\langle C_m - C_w\rangle_{\mathcal{D}(0)} - \xi P_0(0). \tag{4.27}$$

Proof. By definition of $\langle C_m\rangle_{\mathcal{D}(\tau)}$ and $P_0(\tau)$ we have

$$\forall \tau \geq 0 \quad \lim \langle C_m\rangle_{\mathcal{D}(\tau)} = 0 \quad \text{and} \quad \lim P_0(\tau) = 0 \quad \text{when} \quad \overline{\gamma}_m \to 0. \tag{4.28}$$

Hence, we can take $\tau = 0$ in Equation (4.23), and Equation (4.23) is positive for $\overline{\gamma}_m$ small enough. □

This result is somewhat disappointing since the lower bound can be negative; however, in practice, by using a Maxwell-Boltzmann distribution for the random symbols instead of a uniform distribution, we can expect ξ to be small. Hence, the secure throughput achievable by the protocol in the secrecy limited regime should be close to the average secrecy capacity of the channel.

□ **Communication-limited regime.** By opposition to the secrecy-limited regime, this regime corresponds to the case where $\overline{\gamma}_m \to \infty$, and therefore the secrecy capacity is mainly limited by the capacity of the main channel.

Theorem 4.2. *In the communication limited regimes, the secure throughput achievable by the opportunistic secret key agreement protocol is such that*

$$\overline{T}_s(\eta) = \mathcal{O}\left(\eta^{-1}\log\overline{\gamma}_m\right). \tag{4.29}$$

Moreover, these throughputs are achievable by choosing τ such that $2^\tau = \mathcal{O}\left(\overline{\gamma}_m\right)$ and in this case,

$$\overline{T}_s(\eta) \approx \eta^{-1}\tau P_0(\tau) \quad \textit{when} \quad \overline{\gamma}_m \to \infty. \tag{4.30}$$

Before proving the result, we introduce a proposition that provides bounds for $\langle C_m \rangle_{\mathcal{D}(\tau)}$ and $\langle C_w \rangle_{\mathcal{D}(\tau)}$ depending on $P_0(\tau)$. The proof is provided in Section 4.4.2.

Proposition 4.6. *The average value of the main channel capacity over the set $\mathcal{D}(\tau)$ can be bounded as follows.*

$$P_0(\tau)\left(\tau - \frac{\overline{\gamma}_w^2(\log 2)^{-1}}{\overline{\gamma}_w + \overline{\gamma}_m 2^{-\tau}}\right) \leq \langle C_m \rangle_{\mathcal{D}(\tau)} \leq P_0(\tau)\left(\tau + \overline{\gamma}_w + \overline{\gamma}_m 2^{-\tau}(\log 2)^{-1}\right) \tag{4.31}$$

Likewise, the average value of the wiretap channel capacity over the $\mathcal{D}(\tau)$ can be bounded as follows.

$$0 \leq \langle C_w \rangle_{\mathcal{D}(\tau)} \leq \frac{\overline{\gamma}_w \overline{\gamma}_m}{\overline{\gamma}_m + \overline{\gamma}_w} P_0(\tau)(\log 2)^{-1}, \tag{4.32}$$

Proof of Theorem 4.2. By using the inequalities of Proposition 4.6 in Equation (4.23), we obtain the following lower bound on Equation (4.22).

$$F(\overline{\gamma}_m) - \left(2 + \eta^{-1}\right) P_0(\tau)\left(\tau + \overline{\gamma}_w + \overline{\gamma}_m 2^{-\tau}\right) + \eta^{-1} F(\frac{\overline{\gamma}_w \overline{\gamma}_m}{\overline{\gamma}_m + \overline{\gamma}_w}) P_0(\tau) - (4 + \xi\eta^{-1}) P_0(\tau). \tag{4.33}$$

For any $\overline{\gamma}_m$, to satisfy the constraint in the maximization of Proposition 4.4, it suffices to take τ such that

$$\frac{F(\overline{\gamma}_m)}{\left(2 + \eta^{-1}\right)\left(\tau + \overline{\gamma}_w + \overline{\gamma}_m 2^{-\tau}(\log 2)^{-1}\right) - \eta^{-1} F(\frac{\overline{\gamma}_w \overline{\gamma}_m}{\overline{\gamma}_m + \overline{\gamma}_w}) + (4 + \xi\eta^{-1})} \geq P_0(\tau). \tag{4.34}$$

For any $c_0 > 0$, we can choose τ such that $\overline{\gamma}_m 2^{-\tau} = c_0$, and $\tau = \log_2 \overline{\gamma}_m - \log_2 c_0 > 0$ for $\overline{\gamma}_m$ large enough. Since

$$\lim \frac{F(\overline{\gamma}_m)}{\log \overline{\gamma}_w} = 1 \quad \text{when} \quad \overline{\gamma}_m \to \infty, \tag{4.35}$$

the left-hand side of Equation (4.34) converges to

$$\frac{\log 2}{2 + \eta^{-1}} \quad \text{when} \quad \overline{\gamma}_m \to \infty, \tag{4.36}$$

From Proposition 4.1, the right-hand side of Equation (4.34) is equal to

$$\frac{c_0}{c_0 + \overline{\gamma}_w} \exp\left(-\frac{1 - 2^{-\tau}}{c_0}\right); \tag{4.37}$$

therefore, we can always choose c_0 (independent of η) such that Equation (4.34) is satisfied when $\overline{\gamma}_m \to \infty$. Substituting such a τ in Equation (4.24) and (4.26), we have.

$$\langle C_m \rangle_{\mathcal{D}(\tau)} = \mathcal{O}(\tau P_0(\tau)) \quad \text{and} \quad \langle C_m \rangle_{\mathcal{D}(\tau)} = \mathcal{O}(P_0(\tau)), \tag{4.38}$$

when $\overline{\gamma}_m \to \infty$. Using these scaling laws in Equations (2.3) and (4.26), we obtain the second part of the theorem,

$$\eta^{-1}\langle I(X;Y_m) - I(X;Y_w)\rangle_{\mathcal{D}(\tau)} \approx \eta^{-1}\tau P_0(\tau) \quad \text{when} \quad \overline{\gamma}_m \to \infty. \tag{4.39}$$

The first part of the theorem follows by recalling that $P_0(\tau) = \mathcal{O}(1)$ and $\tau = \mathcal{O}(\log \overline{\gamma}_m)$ when $\overline{\gamma}_m \to \infty$. □

For $\eta = 1$, the result of Theorem 4.2 states that, in the communication-limited regime, the information-theoretic secure rates achievable by the protocol scale as $\mathcal{O}\left(\log \overline{\gamma}_m\right)$, and therefore as $\mathcal{O}\left(\overline{C}_s\right)$. Hence, even if secret key agreement incurs a rate penalty compared to the direct use of wiretap codes, this penalty is a constant fraction of the the average secrecy capacity.

4.3.3 Simulation results

In this section, we characterize the secure throughput achievable by the protocol through Monte-Carlo simulations. Simulations are performed using a Maxwell-Boltzmann distribution of the random symbols, for which

$$I(X;Y_m) \approx C_m, \quad I(X;Y_w) \approx C_w, \quad \text{and} \quad H(X) \leq C_m + 2. \tag{4.40}$$

The maximum average secure throughput for $\eta = 1$ achievable by the opportunistic protocol is shown Figure 25.

As expected the protocol is in general sub-optimal since most of the main channel capacity has to be sacrificed for key agreement. Interestingly when the wiretap channel average SNR $\overline{\gamma}_w$ is well above the main channel average SNR $\overline{\gamma}_m$, all the additional communication required for reconciliation and privacy amplification as well as the communication secured by a one-time pad, can be performed when the secrecy capacity is zero. In this case, the protocol incurs little loss of secure communication rate.

Figure 26 shows the secure throughputs obtained for different values of η. Strictly speaking, the protocol does not provide any information theoretic security in this regime, since the keys generated are used to encode several bits; nevertheless, this result shows that the protocol provides an efficient and potentially fast way of exchanging information-theoretically secure keys. In this mode of operation, it could be tailored with standard

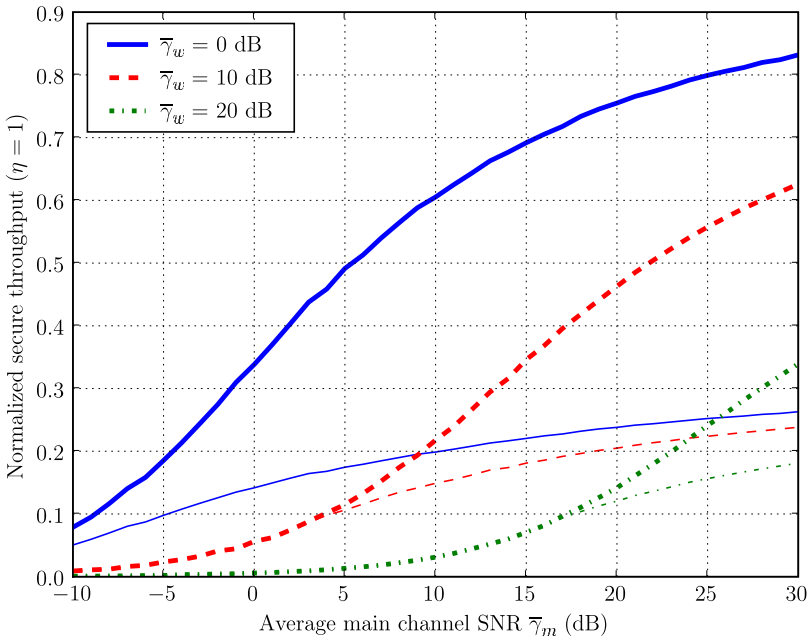


Figure 25. Average secure throughput (thin lines) and average secrecy capacity (thick lines). All throughputs are normalized to the channel capacity of a Gaussian channel with same average SNR $\overline{\gamma}_m$.

secure encryption algorithms (such as AES with 192 bits) to strengthen the current level of security of wireless communications.

4.3.4 Mitigating the effects of imperfect CSI

In this last section, we consider the situation discussed in Section 4.1, where Alice has perfect CSI about the main channel fading coefficient, but only partial CSI about the eavesdropper's channel fading coefficient. As mentioned in Section before, Alice has little choice but to apply the opportunistic protocol blindly, and the keys generated have length

$$\hat{k} = n\left(\beta I(X;Y_m) - \hat{I}\,(X;Y_w)\right) - 2s - 2 - r_0. \tag{4.41}$$

Unfortunately, the lower bound on Eve's R$\acute{\text{e}}$nyi entropy is in reality

$$n\left(\beta I(X;Y_m) - I(X;Y_w)\right) - 2s - 2. \tag{4.42}$$

Therefore, from Theorem 2.5, Eve's uncertainty on the final key is

$$H(K|E) \geq \hat{k} - \frac{2^{n\left(I(X;Y_w) - \hat{I}(X;Y_w)\right) - r_0}}{\ln 2} \tag{4.43}$$

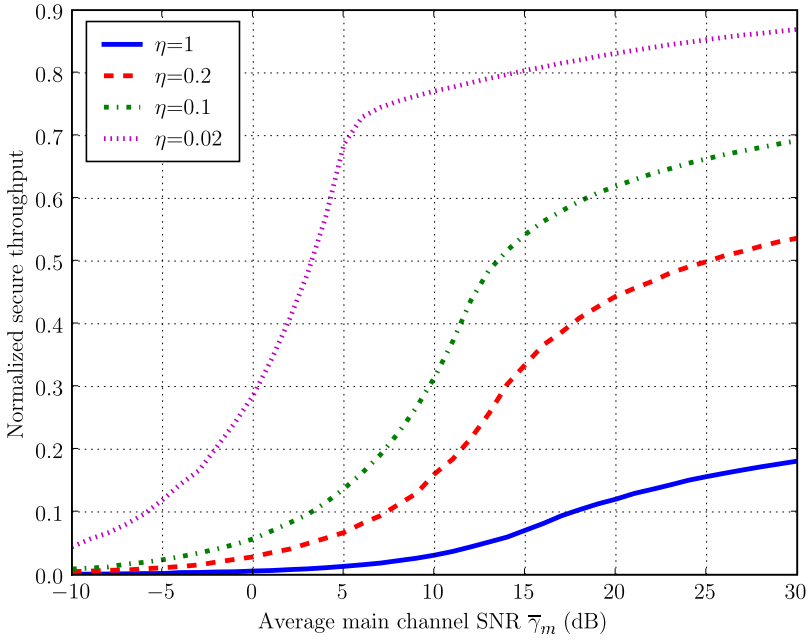


Figure 26. Secure throughput for various values of η.

Clearly, when $I(X;Y_w) < \hat{I}(X;Y_w)$, Alice unnecessarily reduces her secure throughput, but this does not compromise the secrecy of the key; however, when $\hat{I}(X;Y_w) > I(X;Y_w)$, Alice understimates the information leaked to the eavesdropper and subsequently generate keys whose entropy is not maximum.

Until now, we have assumed that the parameter r_0 was chosen such that $r_0 \ll n$. To mitigate the effect of imperfect CSI, let us now consider the situation where $r_0 \propto n$ and let use define

$$\alpha = \frac{r_0}{n}$$

From Equation (4.43), we see that as long as $\hat{I}(X;Y_w) - I(X;Y_w) < \alpha$, the lower bound on $H(\hat{K}|G, E=e)$ approaches $\hat{k}$ exponentially as $n \to \infty$.

The introduction of imperfect CSI and the use of the parameter α slightly modify the expression of communication throughput given in Proposition 4.4. $\overline{T}_s(\eta)$ is now given by

$$\max_{\tau} \eta^{-1} \langle \beta I(X;Y_m) - I(X;Y_w) - \alpha \rangle_{\mathcal{D}(\tau,\kappa)} \tag{4.44}$$

$$\text{subject to} \quad \langle C_m \rangle_{\mathcal{D}^c(\tau,\kappa)} - \langle 2H(X) + \beta(\eta^{-1}-1) I(X;Y_m) - \eta^{-1} I(X;Y_w) - \eta^{-1}\alpha \rangle_{\mathcal{D}(\tau,\kappa)} \geq 0 \tag{4.45}$$

Contrary to the situation where perfect CSI is available, the average secure throughput defined above is not sufficient to characterize the security of the system. In fact it only represents Alice's *targeted* secure communication rate, which might be different from the true secure communication rate. Hence, we need to introduce the true average secure throughput $\overline{\mathcal{R}}_s$ and the average leaked throughput $\overline{\mathcal{R}}_l$ defined as:

$$\overline{\mathcal{R}}_s = \eta^{-1}\langle \beta I(X;Y_m) - I(X;Y_w) - \alpha \rangle_{\mathcal{D}_s}, \tag{4.46}$$

$$\overline{\mathcal{R}}_c = \eta^{-1}\langle \beta I(X;Y_m) - I(X;Y_w) - \alpha \rangle_{\mathcal{D}_l}, \tag{4.47}$$

where

$$\mathcal{D}_s = \left\{ (\hat{\gamma_m}, \gamma_w) : \hat{C}_s \geq \tau, C_m \geq \kappa, I(X;Y_w) - \hat{I}\left(X;Y_w\right) < \alpha \right\} \tag{4.48}$$

$$\mathcal{D}_l = \left\{ (\hat{\gamma_m}, \gamma_w) : \hat{C}_s \geq \tau, C_m \geq \kappa, I(X;Y_w) - \hat{I}\left(X;Y_w\right) \geq \alpha \right\} \tag{4.49}$$

These expressions cannot be computed in close form but can be obtained with Monte-Carlo simulations. Figure 27 shows the results obtained for an estimation noise variance of $\sigma^2 = 10$ and $\sigma^2 = 0.0001$ when $\eta = 1$ and $\alpha = 0$ (i.e. the safety parameter $r_0 \ll n$).

Interestingly, as already pointed out in Section 4.1.2, when Alice has a bad estimation of the eavesdropper's channel fading coefficient, and if the main channel SNR is large, most of the keys generated are still secure. This unexpected behavior is created by the asymmetry of the distribution $p(\hat{\gamma}_w|\gamma_w)$, which forces Alice to underestimate the eavesdropper fading coefficient most of the time. On the other hand, when the estimation of the wiretap CSI improves, the impact of imperfect CSI is somewhat mitigated by increasing the parameter α, which simply plays the role of a safety margin and reduces the length of the generated keys. By increasing α, the average leaked throughput can be made arbitrarily small, at the cost of a decreased secure throughput. Figure 28 shows the results obtained for $\alpha = 0.1$.

When $\sigma^2 = 0.0001$, the secure throughput loss is negligible, however this slight increase in α suffices to ensure the secrecy of the keys generated. The mitigation is less effective when $\sigma^2 = 10$, and a further increase of α would be necessary to reduce the leaked throughput.

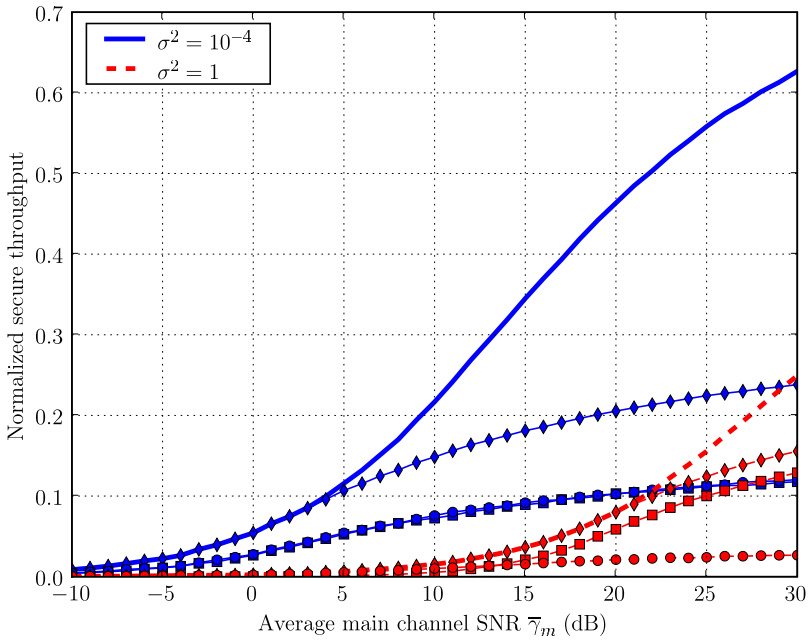


Figure 27. Impact of imperfect CSI. Thicker lines represent the estimated average secrecy capacity. The diamond lines (⋄) represent Alice's targeted average secure throughput with her imperfect CSI, the square lines (⊡) and circle lines (∘) respectively represent the true average secure throughput and average leaked throughput. All throughputs are normalized to the channel capacity of a Gaussian channel with same average SNR $\overline{\gamma}_m$.

4.4 Proofs for Chapter 4

4.4.1 Proof of Lemma 4.1

Suppose that both the main and the wiretap channel are complex AWGN channels, i.e. transmit and receive symbols are complex and both additive noise processes are zero mean circularly symmetric complex Gaussian. The power of the complex input X is constrained according to $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{|X(i)|^2\} \leq P$. Since each use of the complex AWGN channel can be viewed as two uses of a real-valued AWGN channel [56, Appendix B], the secrecy capacity of the complex wiretap channel follows from Theorem 2.2 as

$$C_s = \log\left(1+\frac{P}{N_m}\right) - \log\left(1+\frac{P}{N_w}\right),$$

per complex dimension[2].

To complete the proof, we introduce complex fading coefficients for both the main

[2]Alternatively, this result can be proven by repeating step by step the proofs of [7] using complex-valued random variables instead of real-valued ones.

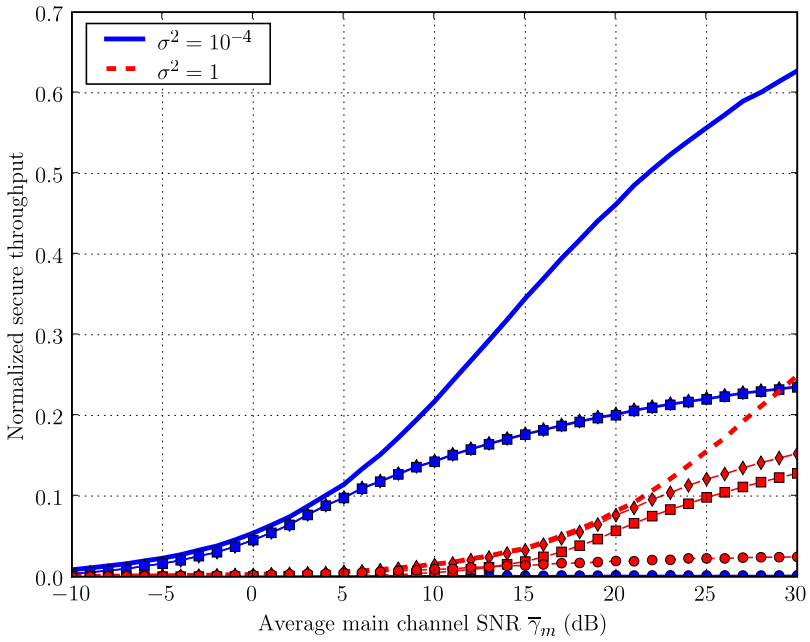


Figure 28. Mitigation of imperfect CSI. Thicker lines represent the estimated average secrecy capacity. The diamond lines (⋄) represent Alice's targeted average secure throughput with her imperfect CSI, the square lines (⊡) and circle lines (∘) respectively represent the true average secure throughput and average leaked throughput. All throughputs are normalized to the channel capacity of a Gaussian channel with same average SNR $\overline{\gamma}_m$.

channel and the eavesdropper's channel, as detailed in Section 4.1.1. Since in the quasi-static case H_m and H_w are random but remain constant for all time, it is perfectly reasonable to view the main channel (with fading) as a complex AWGN channel [56, Chapter 5] with SNR $\gamma_m = P|h_m|^2/N_m$ and capacity

$$C_m = \log\left(1 + |h_m|^2 \frac{P}{N_m}\right).$$

Similarly, the capacity of the eavesdropper's channel is given by

$$C_w = \log\left(1 + |h_w|^2 \frac{P}{N_w}\right),$$

with SNR $\gamma_w = P|h_w|^2/N_w$. Thus, once again based on Theorem 2.2 and the nonnegativity of channel capacity, we may write the secrecy capacity for one realization of the quasi-static fading scenario as Equation (4.3).

4.4.2 Proof of Proposition 4.6

The main channel capacity averaged over the realization in $\mathcal{D}(\tau)$ can be expanded as follows.

$$\begin{aligned}
\langle C_m \rangle_{\mathcal{D}(\tau)} &= \int_{\mathcal{D}(\tau)} \log_2(1+\gamma_m)\, p(\gamma_m) p(\gamma_w) d\gamma_m d\gamma_w \\
&= \int_{2^\tau-1}^{\infty} \log_2(1+\gamma_m)\, p(\gamma_m) \left(\int_0^{2^{-\tau}(\gamma_m+1)-1} p(\gamma_w) d\gamma_w \right) d\gamma_m \\
&= \int_{2^\tau-1}^{\infty} \log_2(1+\gamma_m)\, \frac{1}{\overline{\gamma}_m} e^{-\gamma_m/\overline{\gamma}_m} \left(1 - e^{\frac{2^{-\tau}(\gamma_m+1)-1}{\overline{\gamma}_m}} \right) d\gamma_m \\
&= \int_{2^\tau-1}^{\infty} \log_2(1+\gamma_m)\, \lambda_1 e^{-\gamma_m \lambda_1} d\gamma_m \\
&\qquad - \frac{\overline{\gamma}_w}{\overline{\gamma}_w + 2^{-\tau}\overline{\gamma}_m} e^{-\frac{2^{-\tau}-1}{\overline{\gamma}_w}} \int_{2^\tau-1}^{\infty} \log_2(1+\gamma_m)\, \lambda_2 e^{-\lambda_2 \gamma_m} d\gamma_m, \quad (4.50)
\end{aligned}$$

where

$$\lambda_1 = \frac{1}{\overline{\gamma}_m} \quad \text{and} \quad \lambda_2 = \frac{\overline{\gamma}_w + \overline{\gamma}_m 2^{-\tau}}{\overline{\gamma}_w \overline{\gamma}_m}.$$

To obtain simple bounds of this expression, we introduce a simple lemma.

Lemma 4.2.

$$e^{-\lambda x} \log(1+x) \le \int_x^{\infty} \log(1+z)\, \lambda e^{-\lambda z} dz \le e^{-\lambda x} \log(1+x) + \frac{e^{-\lambda x}}{\lambda(x+1)} \quad (4.51)$$

Proof. The upper bound in the lemma follows by integrating the left-hand side by parts as

$$\int_x^{\infty} \log(1+z)\, \lambda e^{-\lambda z} dz = e^{-\lambda x} \log(1+x) + e^{\lambda} \mathrm{E}_1\left((x+1)\lambda\right), \quad (4.52)$$

where $\mathrm{E}_1(x)$ is the exponential-integral function. The result follows by bounding the exponential-integral function as $\mathrm{E}_1(x) \le e^{-x}/x$. The lower bound follows by noting that $\log(1+z) \ge \log(1+x)$ for $z \ge x$, therefore

$$\int_x^{\infty} \log(1+z)\, \lambda e^{-\lambda z} dz \ge \log(1+x) \int_x^{\infty} \lambda e^{-\lambda z} dz = e^{-\lambda x} \log(1+x) \quad (4.53)$$

□

By applying the lemma on each of the two terms of the right-hand-side, we obtain

$$\begin{aligned}
\langle C_m \rangle_{\mathcal{D}(\tau)} &\le \tau e^{-\frac{2^\tau-1}{\overline{\gamma}_m}} + \frac{e^{-\frac{2^\tau-1}{\overline{\gamma}_m}}}{2^\tau \log 2} \overline{\gamma}_m - \frac{\overline{\gamma}_w}{\overline{\gamma}_w + 2^{-\tau}\overline{\gamma}_m} e^{-\frac{2^{-\tau}-1}{\overline{\gamma}_w}} \tau e^{-\frac{\overline{\gamma}_w + \overline{\gamma}_m 2^{-\tau}}{\overline{\gamma}_m \overline{\gamma}_m}(2^\tau-1)} \\
&= e^{-\frac{2^\tau-1}{\overline{\gamma}_m}} \left(\tau + \frac{\overline{\gamma}_m}{2^\tau \log 2} - \tau \frac{\overline{\gamma}_w}{\overline{\gamma}_w + 2^{-\tau}\overline{\gamma}_m} \right) \\
&= P_0(\tau) \left(\tau + \overline{\gamma}_w + \overline{\gamma}_m 2^{-\tau} (\log 2)^{-1} \right). \quad (4.54)
\end{aligned}$$

Likewise, by reversing the bounds we obtain

$$\begin{aligned}\langle C_m\rangle_{\mathcal{D}(\tau)} &\geq \tau e^{-\frac{2^\tau-1}{\overline{\gamma}_m}} - \frac{\overline{\gamma}_w}{\overline{\gamma}_w+2^{-\tau}\overline{\gamma}_m}e^{-\frac{2^{-\tau}-1}{\overline{\gamma}_w}}\left(\tau+\frac{1}{2^\tau\log 2}\frac{\overline{\gamma}_w\overline{\gamma}_m}{\overline{\gamma}_w+\overline{\gamma}_m 2^{-\tau}}\right)e^{-\frac{\overline{\gamma}_w+\overline{\gamma}_m 2^{-\tau}}{\overline{\gamma}_m\overline{\gamma}_m}(2^\tau-1)} \\ &= e^{-\frac{2^\tau-1}{\overline{\gamma}_m}}\left(\tau-\tau\frac{\overline{\gamma}_w}{\overline{\gamma}_w+2^{-\tau}\overline{\gamma}_m}-\frac{\overline{\gamma}_w}{\overline{\gamma}_w+2^{-\tau}\overline{\gamma}_m}\frac{1}{2^\tau\log 2}\frac{\overline{\gamma}_w\overline{\gamma}_m}{\overline{\gamma}_w+\overline{\gamma}_m 2^{-\tau}}\right) \\ &= P_0(\tau)\left(\tau-\frac{1}{\log 2}\frac{\overline{\gamma}_w^2}{\overline{\gamma}_w+\overline{\gamma}_m 2^{-\tau}}\right) \qquad (4.55)\end{aligned}$$

To bound the wiretap channel capacity averaged over the realizations in $\mathcal{D}(\tau)$ we write

$$\begin{aligned}\langle C_w\rangle_{\mathcal{D}(\tau)} &= \int_{\mathcal{D}(\tau)}\log_2(1+\gamma_w)\,p(\gamma_m)p(\gamma_w)d\gamma_m d\gamma_w \\ &= \int_0^\infty \log_2(1+\gamma_w)p(\gamma_w)\left(\int_{2^\tau(1+\gamma_w)-1}^\infty p(\gamma_m)d\gamma_m\right)d\gamma_w \\ &= \int_0^\infty \log_2(1+\gamma_w)p(\gamma_w)\left(e^{-\frac{2^\tau(1+\gamma_w)-1}{\overline{\gamma}_m}}\right)d\gamma_w \\ &= P_0(\tau)\int_0^\infty \log_2(1+\gamma_w)\frac{\overline{\gamma}_m+\overline{\gamma}_w 2^\tau}{\overline{\gamma}_w\overline{\gamma}_m}e^{-\frac{\overline{\gamma}_m+\overline{\gamma}_w 2^\tau}{\overline{\gamma}_w\overline{\gamma}_m}\gamma_w}d\gamma_w \qquad (4.56)\end{aligned}$$

The result follows by noting that for any $\tau \geq 0$

$$\begin{aligned}0 &\leq \int_0^\infty \log_2(1+\gamma_w)\frac{\overline{\gamma}_m+\overline{\gamma}_w 2^\tau}{\overline{\gamma}_w\overline{\gamma}_m}e^{-\frac{\overline{\gamma}_m+\overline{\gamma}_w 2^\tau}{\overline{\gamma}_w\overline{\gamma}_m}\gamma_w}d\gamma_w \\ &\leq \frac{1}{\log 2}\int_0^\infty \gamma_w\frac{\overline{\gamma}_m+\overline{\gamma}_w 2^\tau}{\overline{\gamma}_w\overline{\gamma}_m}e^{-\frac{\overline{\gamma}_m+\overline{\gamma}_w 2^\tau}{\overline{\gamma}_w\overline{\gamma}_m}\gamma_w}d\gamma_w \\ &= \frac{1}{\log 2}\frac{\overline{\gamma}_w\overline{\gamma}_m}{\overline{\gamma}_m+\overline{\gamma}_w 2^\tau} \leq \frac{1}{\log 2}\frac{\overline{\gamma}_w\overline{\gamma}_m}{\overline{\gamma}_m+\overline{\gamma}_w} \qquad (4.57)\end{aligned}$$

4.4.3 Proof of Proposition 4.3

An outage event occurs whenever Alice overestimates the amount of secrecy she can distill from an opportunistic transmission. Therefore,

$$\mathcal{P}_{\text{out}} = \mathrm{P}\left[\hat{C}_s > C_s, C_s \geq \tau_s, C_m > \tau_m\right] \leq \mathrm{P}\left[\hat{C}_s > C_s\right] = \mathrm{P}\left[\hat{\Gamma}_w < \Gamma_w\right]$$

Now, $\mathrm{P}\left[\hat{\Gamma}_w < \Gamma_w\right]$ can be written as follows

$$\begin{aligned}\mathrm{P}\left[\hat{\Gamma}_w < \Gamma_w\right] &= \int_0^\infty \mathrm{P}\left[\hat{\Gamma}_w < \Gamma_w | \Gamma_w = \gamma_w\right]p(\gamma_w)d\gamma_w \\ &= \int_0^\infty\left(\int_0^{\gamma_w} p(\hat{\gamma}_w|\gamma_w)d\hat{\gamma}_w\right)p(\gamma_w)d\gamma_w\end{aligned}$$

where $p(\gamma_w)$ is the probability density function of Γ_w (see Equation 4.2) and $p(\hat{\gamma}_w|\gamma_w)$ is the probability density function of $\hat{\Gamma}_w$ conditioned on Γ_w. This probability density function

is non-central χ^2 with two degrees of freedom, *i.e.*

$$p(\hat{\gamma}_w|\gamma_w) = \frac{1}{2\bar{\gamma}_w\sigma_e^2} e^{-\frac{(\gamma_w+\hat{\gamma}_w)}{2\bar{\gamma}_w\sigma_e^2}} I_0\left(\frac{\sqrt{\gamma_w\hat{\gamma}_w}}{\bar{\gamma}_w\sigma_e^2}\right), \hat{\gamma}_w > 0$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind [57]. Thus, the probability $\mathrm{P}[\hat{\gamma}_w < \gamma_w|\gamma_w]$ reduces to

$$\mathrm{P}\Big[\hat{\Gamma}_w < \Gamma_w|\Gamma_{w=}\gamma_w\Big] = 1 - Q_1\big(\sqrt{\gamma_w/(\bar{\gamma}_w\sigma_e^2)}, \sqrt{\gamma_w/(\bar{\gamma}_w\sigma_e^2)}\big)$$

where $Q_1(\cdot,\cdot)$ is the generalized Marcum Q function [57]. Using standard results for integrals involving the generalized Marcum Q function [58], the upper bound to the outage probability reduces to

$$\mathcal{P}_{\text{out}} \leq \frac{1}{2} - \frac{1}{2}\frac{1}{\sqrt{1+2/\sigma_e^2}}. \tag{4.58}$$

CHAPTER 5
COOPERATION VS. SECRECY TRADE-OFFS

In Chapters 3 and 4, we studied the design of *practical* schemes providing some level of information-theoretic security. In this chapter, we tackle a more theoretical problem and attempt to grasp the fundamental trade-offs between cooperation and information-theoretic security in multi-user scenarios. This investigation is motivated by the fact that the positive impact of cooperation among multiple users on transmission reliability is now well understood [59, 60], but it is not clear whether cooperation is still beneficial when secrecy constraints are added to the problems. While the fundamental secure communication rates over wireless channels have been specifically investigated in [12, 8, 15] for three-terminal networks, little attention has been devoted to the study of multi-terminal situations where several players with different (and possibly conflicting) security requirements interact at the same time.

We shed light on this fundamental problem by considering a three-terminal communication scenario, where the intended receiver of a private message also acts as a relay for another node. We acknowledge that similar situations have already been investigated. For instance, the situation where a relay helps a destination node while being kept ignorant of some private message is studied in [21], a four-player scenario where a trusted relay helps a transmitter and a legitimate receiver to conceal messages from an eavesdropper is considered in [20], and finally, coordination among users is shown to be helpful for increasing secrecy rates in [19, 22]. However, unlike all aforementioned contributions, our approach attempts to analyze the *fundamental* security compromise that a user must accept if he is willing to cooperate with other nodes. For discrete memoryless channels, we provide a single-letter characterization of the exact region of achievable rates, which generalizes the results of Csiszár and Körner on the broadcast channel with confidential messages [6]. For the Gaussian channel, we analyze the performance of several strategies and obtain simply computable bounds on the secrecy capacity region. Our results are still mainly of theoretical interest, as the design of practical codes seems to be beyond the reach of today's capabilities; nevertheless, the relaying strategies that we analyze provide some insight on

how to design secure relaying schemes.

5.1 Channel model

Definition 5.1 ([60])**.** *A partially cooperative relay discrete broadcast channel is a channel with a discrete source input alphabet $\mathcal{X}$, a discrete relay input alphabet $\mathcal{X}_1$, two discrete channel output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, and a transition probability function $p(y, z|x, x_1)$. Such a channel is denoted by the set $(\mathcal{X}, \mathcal{X}_1, p(y, z|x, x_1), \mathcal{Y}, \mathcal{Z})$.*

We shall restrict our attention to memoryless channels, for which the transition probability of a sequence of n symbols is given by

$$p(\mathbf{y}^n, \mathbf{z}^n|\mathbf{x}^n, \mathbf{x}_1^n) = \prod_{i=1}^{n} p(y_i, z_i|x_i, x_{1,i}).$$

Definition 5.2. *A Gaussian memoryless partially cooperative relay broadcast channel is a partially cooperative relay broadcast channel such that*

$$\begin{aligned} Y &= X + Z_1, \\ Z &= X + X_1 + Z_2, \end{aligned}$$

where Z_1 and Z_2 and independent zero-mean Gaussian random variables with variances N_1 and N_2, respectively. The channel input sequences are subject to average power constraints

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{X_i^2\} \leq P \quad \textit{and} \quad \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{X_{1,i}^2\} \leq P_1.$$

Notice that our definition of a Gaussian partially cooperative relay broadcast channel assumes that the receiver observing output Y is able to cancel X_1 completely. In practice, there might still be a residual contribution of X_1 to Y, and a more accurate model would be

$$Y = X + \eta X_1 + Z_1,$$

where $\eta < 1$ models the efficiency of the cancellation; however, for simplicity, we do not consider this situation here.

Definition 5.3. *A $\left(2^{nR_0}, 2^{nR_1}, n\right)$ code for the partially cooperative relay broadcast channel consists of the following.*

- *Two message sets* $\mathcal{W}_0 = \left\{1, 2, \ldots, 2^{nR_0}\right\}$ *and* $\mathcal{W}_1 = \left\{1, 2, \ldots, 2^{nR_1}\right\}$*;*

- *An encoding function (possibly stochastic)* $f_n : \mathcal{W}_0 \times \mathcal{W}_1 \to \mathcal{X}^n$*, which maps each message pair* $(w_0, w_1) \in \mathcal{W}_0 \times \mathcal{W}_1$ *to a codeword* $\mathbf{x}^n \in \mathcal{X}^n$*;*

- *A set of relay functions* $\{g_i\}_{i=1}^n$ *such that*

$$x_{1,i} = g_i\left(y_1, \ldots, y_{i-1}\right), \quad \text{for } 1 \leq i \leq n;$$

- *Two decoding functions* $h_y : \mathcal{Y}^n \to \mathcal{W}_0 \times \mathcal{W}_1$ *and* $h_z : \mathcal{Z}^n \to \mathcal{W}_0$ *mapping the observation* $\mathbf{y}^n$ *to the message pair* $(\hat{w}_0, \hat{w}_1)$ *and the observation* $\mathbf{z}^n$ *to the message* $\hat{\hat{w}}_0$*, respectively.*

Definition 5.4. *The average probability of error* $P_e^{(n)}$ *is defined as the probability that the decoded messages at the intended receivers are different for the transmitted ones, that is*

$$P_e^{(n)} \quad = \quad P[h_y(\mathbf{Y}^n) \neq (W_0, W_1) \ \text{ or } h_z(\mathbf{Z}^n) \neq W_0].$$

Definition 5.5. *A rate tuple* (R_0, R_1, R_e) *is said to be achievable for a partially cooperative relay broadcast channel if* $\forall \epsilon_0 > 0$*, there exists a* $\left(2^{nR_0}, 2^{nR_1}, n\right)$ *code such that*

$$P_e^{(n)} \leq \epsilon_0,$$
$$\frac{1}{n} H(W_1|Z^n) \geq R_e - \epsilon_0.$$

Figure 29 illustrates a partially cooperative broadcast channel with confidential messages. We emphasize that the secrecy level of message W_1 is measured in terms of the equivocation rate $\frac{1}{n} H(W_1|\mathbf{Z}^n)$, which corresponds to the notion of *weak* secrecy. A stronger measure (and better suited for cryptographic purposes) would be the absolute equivocation $H(W_1\mathbf{Z}^n)$ [11]; however, we do not consider it here.

Let us also point out that the model shown in Figure 29 differs from the one studied in [21] in that the secrecy constraint is placed on the relay node and not on the destination node. Although this might appear to be a minor modification, the problem becomes

significantly different on several accounts. First, the results of [21] are only valid provided the adversary relay follows a fixed protocol. This requirement is somewhat relaxed in our setting since the relay is the receiver of the secret messages. Second, the results presented in Section 5.2 do not follow directly from those of [21], and we shall see that the efficient relaying strategies for the Gaussian case are quite different.

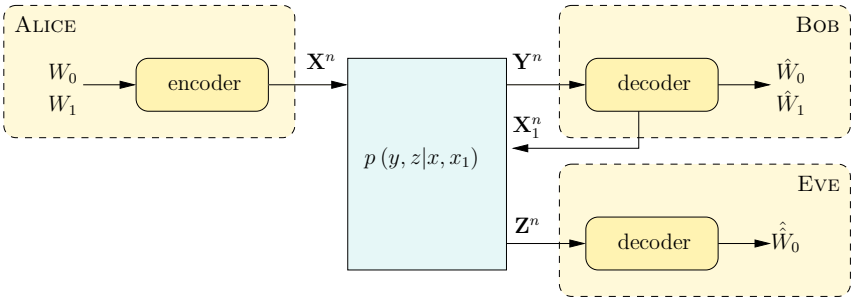


Figure 29. Channel model 1: partially cooperative relay broadcast channel with confidential messages.

Definition 5.6. *The secrecy capacity region is defined as the set of rates* (R_0, R_1) *such that the message* W_1 *can be communicated secretly, i.e.*

$$\mathcal{C}_s = \{(R_0, R_1) : (R_0, R_1, R_1) \text{ is achievable}\}\,.$$

5.2 Equivocation-rate region of discrete memoryless channels

Theorem 5.1. *The region of achievable rates* (R_0, R_1, R_e) *for the partially cooperative broadcast channel with confidential messages is given by*

$$\mathcal{C}_0 = \bigcup_{p(u,x_1,v)p(x|v)p(y,z|x,x_1)} \left\{ \begin{array}{l} 0 \le R_e \le R_1 \\ R_e \le I(V;Y|U,X_1) - I(V;Z|U,X_1) \\ R_1 + R_0 \le I(V;Y|U) + \min\left(I(U;Y|X_1)\,, I(U,X_1;Z)\right) \\ R_0 \le \min\left(I(U;Y|X_1)\,, I(U,X_1;Z)\right) \end{array} \right\}$$

Proof. See Section 5.4. □

By setting $X_1 = \emptyset$ in the above, we obtain the region of achievable rates for the broadcast channel with confidential messages. It is actually not surprising that the result of [6] generalizes to account for partial cooperation of the receiver with a private message. In fact,

by assuming that the relay decodes the common message, we ensure that there is no loss of optimality by considering a relaying relaying scheme where the relay decodes his message before forwading information.

Corollary 5.1. *The secrecy capacity region of the partially cooperative broadcast channel with confidential messages is given by*

$$\mathcal{C}_s = \bigcup_{p(u,x_1,v)p(x|v)p(y,z|x,x_1)} \left\{ \begin{array}{l} R_1 \leq I(V;Y|U,X_1) - I(V;Z|U,X_1) \\ R_0 \leq \min\left(I(U;Y|X_1)\,, I(U,X_1;Z)\right) \end{array} \right\}$$

As is often the case in information-theory, the single-letter characterizations of the equivocation-rate regions $\mathcal{C}_0$ and $\mathcal{C}_s$ obtained above are considered an acceptable solution to the problem since they are computable with numerical algorithms. Clearly, it would be preferable to obtain a general closed-form expression of the regions, but such a characterization is not possible in general. Therefore, in the next section, we specialize these results to the case of Gaussian channels, for which simpler expressions can be obtained.

5.3 Achievable Equivocation-rate region for Gaussian channels

The derivation of achievable rates in Theorem 5.1 relies on the notion of typical set decoding, which can be readily extended to continuous random variables. In particular, we can obtain *achievable* equivocation-rate regions for the Gaussian partially cooperative broadcast channel by substituting well-chosen random variables in Theorem 5.1. Although this does not characterize, the exact equivocation-rate region, the bounds that we obtain offer some interesting insight on the design of practical relaying schemes.

In the following, we simplify notation by introducing the function $C(x) = \frac{1}{2}\log_2(1+x)$, where $C(\text{SNR})$ represents the capacity of a point-to-point Gaussian channel with signal-to-noise ratio SNR. For any constant c such that $0 \leq c \leq 1$, we also define $\bar{c} = 1 - c$.

Proposition 5.1 (Decode-and-Forward)**.** *A region of achievable rates with a "Decode-and-Forward" relaying strategy is given by*

$$\mathcal{C}_{DF} = \bigcup_{0\leq\alpha\leq 1} \left\{ \begin{array}{l} R_0 \leq \max\limits_{0\leq\beta\leq 1} \min \left(\begin{array}{l} C\left(\frac{P_1+\bar{\alpha}P+2\sqrt{\beta\bar{\alpha}PP_1}}{\alpha P+N_2}\right), \\ C(\frac{\beta\bar{\alpha}P}{\alpha P+N_1}) \end{array} \right) \\ R_1 \leq C\left(\frac{\alpha P}{N_1}\right) \\ R_e \leq C\left(\frac{\alpha P}{N_1}\right) - C\left(\frac{\alpha P}{N_2}\right) \end{array} \right\}$$

Proof. The region is obtained by computing the bound of Theorem 5.1 with the following random variables.

$$\begin{aligned} X_1 &\sim \mathcal{N}(0,P_1) \quad U' \sim \mathcal{N}(0,\beta\bar{\alpha}P) \quad X' \sim \mathcal{N}(0,\alpha P) \\ U &= \sqrt{\frac{\bar{\beta}\bar{\alpha}P}{P_1}}X_1 + U' \quad X = U + X' \quad V = U \end{aligned}$$

□

Figure 30 illustrates the rates achievable using the Decode-and-Forward strategy for parameters $P = 1$, $P_1 = 1$, $N_1 = 0.01$ and $N_2 = 0.05$. As expected, since the relay messages are based on the (correctly) decoded common message, no additional information is leaked to the eavesdropper.

Instead of cooperating with the destination node, the relay might decide to confuse the destination node by jamming the channel with white Gaussian noise. It is arguable whether this can be considered as a meaningful relaying strategy, but it is definitely allowed in our channel model. We note that this idea has already appeared in the literature as "cooperative jamming" [19], "noise forwarding" [20], or "artificial noise" [22].

Proposition 5.2 (Jamming)**.** *A region of achievable rates obtained with a "Jamming" strategy is given by*

$$\mathcal{C}_J = \bigcup_{0\leq\alpha\leq 1} \left\{ \begin{array}{l} R_0 \leq \min\left(C\left(\frac{\bar{\alpha}P}{\alpha P+N_2+P_1}\right), C(\frac{\bar{\alpha}P}{\alpha P+N_1})\right) \\ R_1 \leq C\left(\frac{\alpha P}{N_1}\right) \\ R_e \leq C\left(\frac{\alpha P}{N_1}\right) - C\left(\frac{\alpha P}{N_2+P_1}\right) \end{array} \right\}$$

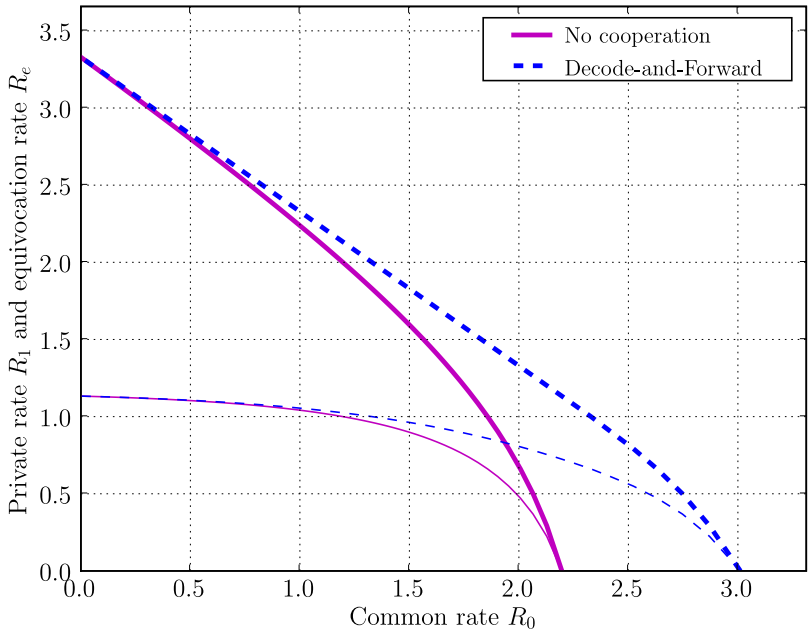


Figure 30. Achievable rate regions with a Decode-and-Forward relaying strategy. The private message rate R_1 is represented by thick lines while the equivocation rate R_e is represented by thin lines.

Proof. If the relay emits white Gaussian noise with power P_1, the resulting channel is equivalent to a Gaussian broadcast channel (without relaying) characterized by

$$\begin{aligned} Y &= X + Z_1, \\ Z &= X + Z_2' \quad \text{with} \quad Z_2' \sim \mathcal{N}(0, N_2 + P_1). \end{aligned}$$

The results follows directly from [6, Theorem 1] by using the following random variables.

$$U \sim \mathcal{N}(0, \bar{\alpha} P) \quad X' \sim \mathcal{N}(0, \alpha P) \quad X = U + X'.$$

□

The boundaries of the region achieved with the Jamming relaying strategy are shown in Figure 31 for parameters $P = 1$, $P_1 = 1$, $N_1 = 0.01$ and $N_2 = 0.05$. Jamming obviously increases the equivocation of the private message but the cost of a tremendous reduction of common rate.

In principle, time-sharing between the two strategies allows to achieve all the equivocation rates in the convex hull of $\mathcal{C}_{DF} \cup \mathcal{C}_J$; however, this is not acceptable from a security

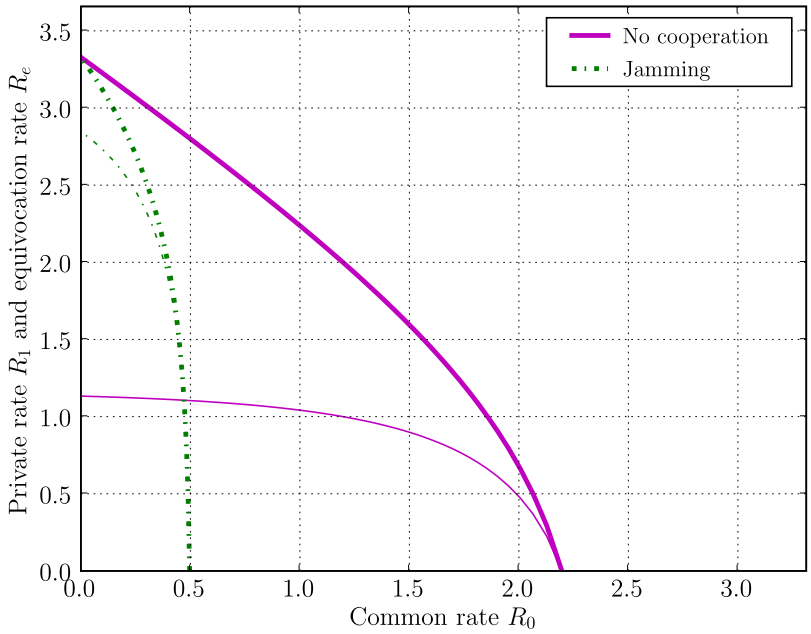


Figure 31. Achievable rate regions with a Jamming relaying strategy. The private message rate R_1 is represented by thick lines, while the equivocation rate R_e is represented by thin lines.

perspective since the same level of security should be ensured for all transmitted messages. Intermediate points can still be achieved by a mixed strategy, as shown by the following proposition.

Proposition 5.3 (Opportunistic Jamming)**.** *A region of achievable rates with an "Opportunistic Jamming" relaying strategy is given by*

$$\mathcal{C}_{OJ} = \bigcup_{0\leq\alpha,\beta,\gamma\leq 1} \left\{ \begin{array}{l} R_0 \leq \min \left(\begin{array}{l} C\left(\frac{\gamma P_1+\bar{\alpha}P+2\sqrt{\bar{\beta}\bar{\alpha}P\gamma P_1}}{\alpha P+\bar{\gamma}P_1+N_2}\right), \\ C(\frac{\beta\bar{\alpha}P}{\alpha P+N_1}) \end{array} \right) \\ R_1 \leq C\left(\frac{\alpha P}{N_1}\right) \\ R_e \leq C\left(\frac{\alpha P}{N_1}\right) - C\left(\frac{\alpha P}{N_2+\bar{\gamma}P_1}\right) \end{array} \right\}$$

Proof. If the relay uses a fraction $\bar{\gamma}$ of its total power P_1 to confuse the eavesdropper and the remaining fraction γ to perform Decode-and-Forward, then the equivalent channel is given by

$$\begin{aligned} Y &= X + Z_1, \\ Z &= X + X_1' + Z_2' \quad \text{with} \quad Z_2' \sim \mathcal{N}(0, N_2 + \bar{\gamma}P_1), \end{aligned}$$

and X_1' subject to the average power constraint $\mathbb{E}\{X_1\} \leq \gamma P_1$. The result follows by computing the bounds in Theorem 5.1 using the following random variables.

$$\begin{aligned} X_1 &\sim \mathcal{N}(0, \gamma P_1), \quad U' \sim \mathcal{N}(0, \beta\bar{\alpha}P), \quad X' \sim \mathcal{N}(0, \alpha P), \\ U &= \sqrt{\frac{\bar{\beta}\bar{\alpha}P}{\gamma P_1}} X_1 + U', \quad X = U + X' \quad V = U \end{aligned}$$

□

In certain regimes where the common rate is mostly limited by the relay node alone, the Opportunistic Jamming strategy can significantly improve the equivocation rates without reducing the message rates achievable with Decode-and-Forward. More precisely, we have the following lemma.

Lemma 5.1. *If $N_2 < N_1$, define*

$$\alpha^* = \frac{\frac{P}{N_1} - \frac{P_1}{N_2 - N_1}}{\frac{P}{N_1} + \frac{P}{N_1}\frac{P_1}{N_2 - N_1}}.$$

then $\forall \alpha \in [\alpha^*, 1]$, the common and private rates in Proposition 5.1 are bounded as

$$R_0 \leq C\left(\frac{\bar{\alpha}P}{\alpha P + N_1}\right), \quad R_1 \leq C\left(\frac{\alpha P}{N_1}\right),$$

and the relay node can jam without reducing the common message rate as long as the fraction of power γ satisfies

$$\gamma \geq \gamma^* = \bar{\alpha}P\frac{P_1 + N_2 - N_1}{P_1 P + N_1 P_1}.$$

Proof. If $N_2 > N_1$, it can be easily verified that

$$C\left(\frac{P_1 + \bar{\alpha}P}{\alpha P + N_2}\right) \geq C(\frac{\bar{\alpha}P}{\alpha P + N_1}) \Longleftrightarrow \alpha \geq \alpha^*,$$

which implies that $\beta = 1$ is optimal for $\alpha \geq \alpha^*$. Now, one can check that

$$C\left(\frac{\gamma P_1 + \bar{\alpha}P}{\alpha P + \overline{\gamma} P_1 + N_2}\right) \geq C(\frac{\bar{\alpha}P}{\alpha P + N_1}) \Longleftrightarrow \gamma \geq \gamma^*.$$

□

The boundary of the maximum equivocation-rate region achievable by Opportunistic Jamming without sacrificing cooperation is shown in Figure 32 for $P = 1$, $P_1 = 1$, $N_1 = 0.01$ and $N_2 = 0.05$.

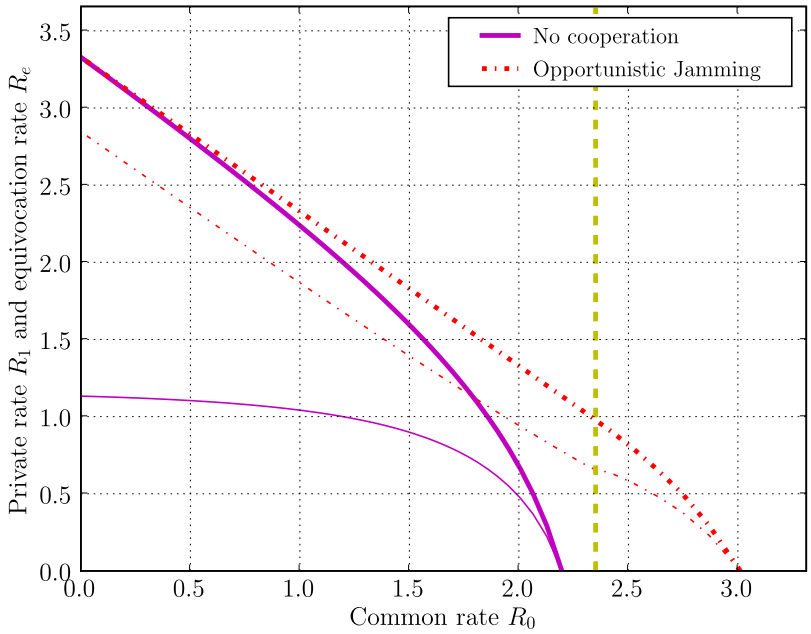


Figure 32. Achievable rate regions with Opportunistic Jamming. The private message rate R_1 is represented by thick lines while the equivocation rate R_e is represented by thin lines. The dashed vertical line corresponds to the common rate when $\alpha = \alpha^*$.

The opportunistic jamming strategy is interesting, since one would intuitively expect relaying to be harmful for secrecy and beneficial for reliability; however, our analysis shows that carefully designing relaying strategies can lead to better performance than no relaying at all, both in term of secrecy and achievable rates.

5.4 Proof Theorem 5.1

5.4.1 Achievability part

The direct part of the proof follows from a random coding argument combining Csiszár and Körner's wiretap coding [6] with Willem's backward decoding strategy for the relay channel [61]. Since the wiretap coding scheme is slightly involved, we refer the reader unfamiliar with wiretap code construction to Appendix **A**, where a detailed proof of [6] is provided. An outline of the proof is as follows.

- **Step 1.** We first show the existence of an *inner code* with a specific structure allowing the reliable transmission of three messages (V_0, V_1, V_2) at certain rates (R'_0, R'_1, R'_2) over the partially cooperative broadcast channel;

- **Step 2.** We construct a simple *outer code* that exploits the aforementioned inner code to guarantee the reliable transmission of a common message W_0 and of a private message W_1, with certain rates R_0 and R_1, respectively;
- **Step 3.** We compute the equivocation rate R_e guaranteed by the coding scheme;
- **Step 4.** We derive a convex region of achievable rates (R_0, R_1, R_e).

Fix $\frac{1}{2} > \epsilon_0 > 0$, $\epsilon > 0$, and $\delta > 0$ such that

$$h(\epsilon) < \frac{\epsilon_0}{5}, \qquad 0 < \delta < \min\left[\frac{\epsilon}{16}, \frac{\epsilon_0}{5\log_2|\mathcal{X}|}\right]. \tag{5.1}$$

Let U, X_1, X, Y, and Z be random variable with joint probability distribution

$$p(u, x_1, x, y, z) = p(y, z|x, x_1)\, p(x|x_1, u)\, p(u|x_1)\, p(x_1),$$

and such that $I(X;Y|U,X_1) \geq I(X;Z|U,X_1)$.

Step 1: Inner code construction.

Define the following rates.

$$\begin{cases} R_0' = \min\left[I(U;Y|X_1), I(U,X_1;Z)\right] - \epsilon \\ R_1' = I(X;Y|U,X_1) - I(X;Z|U,X_1) - \frac{\epsilon}{2} \\ R_2' = I(X;Z|U,X_1) - \frac{\epsilon}{2}. \end{cases}$$

□ *Random code generation.*

1. Generate $2^{nR_0'}$ independent sequences of length n at random in $\mathcal{X}_1^n$ according to the distribution $p(\mathbf{x}_1^n) = \prod_{i=1}^n p_{X_1}(x_{1,i})$. Label[1] the sequences $\mathbf{X}_1^n(i)$ with $i \in \left\{1, 2, \ldots, 2^{nR_0'}\right\}$.
2. For each $i \in \left\{1, 2, \ldots, 2^{nR_0'}\right\}$, generate $2^{nR_0'}$ independent sequences of length n at random in $\mathcal{U}^n$ according to the distribution $p(\mathbf{u}^n|\mathbf{x}_1^n) = \prod_{i=1}^n p_{U|X_1}(u_i|x_{1,i})$. Label the sequence $\mathbf{U}^n(j|i)$ with $j \in \left\{1, 2, \ldots, 2^{nR_0'}\right\}$ and $i \in \left\{1, 2, \ldots, 2^{nR_0'}\right\}$.

[1]We assume that all rate R are such that 2^{nR} is an integer. This approximation greatly improves the clarity of the proof, and we refer the reader to Appendix A for an example of a more careful derivation.

3. For each $(j,i) \in \left\{1,2,\ldots,2^{nR_0'}\right\} \times \left\{1,2,\ldots,2^{nR_0'}\right\}$, generate $2^{n(R_1'+R_2')}$ independent sequences of length n at random according to the distribution $p\left(\mathbf{x}^n|\mathbf{u}^n,\mathbf{x}_1^n\right) = \prod_{i=1}^n p_{X|U,X_1}\left(x_i|u_i,x_{1,i}\right)$ Label the sequences $\mathbf{X}^n(k,l|i,j)$ with $k \in \left\{1,2,\ldots,2^{nR_1'}\right\}$ and $l \in \left\{1,2,\ldots,2^{nR_2'}\right\}$

□ *Alice's encoding procedure.* Encoding is performed using a block-Markov scheme over B consecutive blocks of length n. The same codebooks are used in each of the B blocks. In block $b \in \{1,\ldots,B\}$, to transmit messages V_0^b, V_1^b, and V_2^b, the encoder simply sends the codeword $\mathbf{X}^n(V_1^b,V_2^b|V_0^b,V_0^{b-1})$ over the channel, with the convention that $V_0^0 = V_0^B = 1$.

□ *Bob's encoding procedure.* Let $\hat{V}_0^{b-1}$ be Bob's estimates of V_0^{b-1}. Then in block $b \in \{1,\ldots,B\}$, Bob transmits $\mathbf{X}_1^n(\hat{V}_0^{b-1})$, with the convention that $\hat{V}_0^0 = 1$.

□ *Bob's decoding procedure.* Assuming that Bob has estimated $\hat{V}_0^{b-1}$, he determines the unique message $\hat{V}_0^b$ such that

$$\left(\mathbf{X}_1^n(\hat{V}_0^{b-1}),\mathbf{U}^n(\hat{V}_0^b|\hat{V}_0^{b-1}),\mathbf{Y}^n(b)\right) \in A_\delta^{(n)}. \tag{5.2}$$

Bob then determines the unique message pair $\left(V_1^b,V_2^b\right)$ such that

$$\left(\mathbf{X}^n(\hat{V}_1^b,\hat{V}_2^b|\hat{V}_0^b,\hat{V}_0^{b-1}),\mathbf{X}_1^n(\hat{V}_0^{b-1}),\mathbf{U}^n(\hat{V}_0^b|\hat{V}_0^{b-1}),\mathbf{Y}^n(b)\right) \in A_\delta^{(n)}. \tag{5.3}$$

□ *Eve's decoding procedure.* Eve's decoding starts once she has received all channel observations $\mathbf{Y}^n(b)$ with $b \in \{1,\ldots,B\}$. She then proceeds to decode codewords V_0^b in *backward order* as follows.

1. Eve knows that in the Bth block, the message $V_0^B = 1$ was sent; therefore, she determines the unique $\hat{\hat{V}}_0^{B-1}$ such that

$$\left(\mathbf{X}_1^n(\hat{\hat{V}}_0^{B-1}),\mathbf{U}^n(1|\hat{\hat{V}}_0^{B-1}),\mathbf{Z}^n(B)\right) \in A_\delta^{(n)}.$$

2. In the bth block with $b \in \{1,\ldots,B-2\}$, Eve determines the unique message $\hat{\hat{V}}_0^b$ such that

$$\left(\mathbf{X}_1^n(\hat{\hat{V}}_0^b),\mathbf{U}^n(\hat{\hat{V}}_0^{b+1}|\hat{\hat{V}}_0^b),\mathbf{Z}^n(b)\right) \in A_\delta^{(n)}. \tag{5.4}$$

□ *Charlie's decoding procedure.* To impose a structure on the code, we introduce a *virtual* decoder, Charlie, who is assumed to have access to V_0^b and V_1^b for $b \in \{1, \ldots, B\}$ and to decode message V_2^b; therefore, in each block b, Charlie determines the unique $\hat{\hat{V}}_2^b$ such that

$$\left(\mathbf{X}^n(V_1^b, \hat{\hat{V}}_2^b | V_0^b, V_0^{b-1}), \mathbf{X}_1^n(V_0^{b-1}), \mathbf{U}^n(V_0^b | V_0^{b-1}), \mathbf{Z}^n(b)\right) \in A_\delta^{(n)}. \tag{5.5}$$

□ *Analysis of probability of error.* Let $\mathbf{V}_0 = \left(V_0^1, \ldots, V_0^B\right)$ denote the sequence of messages V_0^b sent in the B blocks, and define similarly $\mathbf{V}_1$, $\mathbf{V}_2$, $\hat{\mathbf{V}}_0$, $\hat{\mathbf{V}}_1$, $\hat{\mathbf{V}}_2$, $\hat{\hat{\mathbf{V}}}_0$, $\hat{\hat{\mathbf{V}}}_2$. We shall start by simplifying the calculation of the average probability of error of the sequence of B blocks by relating it to the average probability or error of *individual* blocks. Define the following sequence-error events.

$$\mathcal{E}_B = \left\{(\mathbf{V}_0, \mathbf{V}_1, \mathbf{V}_2) \neq \left(\hat{\mathbf{V}}_0, \hat{\mathbf{V}}_1, \hat{\mathbf{V}}_2\right)\right\}, \mathcal{E}_{B,0} = \left\{\mathbf{V}_0 \neq \hat{\mathbf{V}}_0\right\},$$
$$\mathcal{E}_{B,1} = \left\{(\mathbf{V}_1, \mathbf{V}_2) \neq \left(\hat{\mathbf{V}}_1, \hat{\mathbf{V}}_2\right)\right\}, \mathcal{E}_E = \left\{\mathbf{V}_0 \neq \hat{\hat{\mathbf{V}}}_0\right\}, \mathcal{E}_C = \left\{\mathbf{V}_2 \neq \hat{\hat{\mathbf{V}}}_2\right\}.$$

The total probability of error can be written as

$$\begin{aligned}
\overline{P}_{err}^{(n)} &= \mathrm{P}[\mathcal{E}_B \cup \mathcal{E}_E \cup \mathcal{E}_C] \\
&\overset{(a)}{\leq} \mathrm{P}[\mathcal{E}_B] + \mathrm{P}[\mathcal{E}_E] + \mathrm{P}[\mathcal{E}_C] \\
&\overset{(b)}{\leq} \mathrm{P}[\mathcal{E}_B] + \mathrm{P}[\mathcal{E}_E|\mathcal{E}_B]\mathrm{P}[\mathcal{E}_B] + \mathrm{P}\big[\mathcal{E}_E|\overline{\mathcal{E}}_B\big]\mathrm{P}\big[\overline{\mathcal{E}}_B\big] + \mathrm{P}[\mathcal{E}_C|\mathcal{E}_B]\mathrm{P}[\mathcal{E}_B] + \mathrm{P}\big[\mathcal{E}_C|\overline{\mathcal{E}}_B\big]\mathrm{P}\big[\overline{\mathcal{E}}_B\big] \\
&\overset{(c)}{\leq} 3\mathrm{P}[\mathcal{E}_B] + \mathrm{P}\big[\mathcal{E}_E|\overline{\mathcal{E}}_B\big] + \mathrm{P}\big[\mathcal{E}_C|\overline{\mathcal{E}}_B\big],
\end{aligned} \tag{5.6}$$

where (a) follows from the union bound, (b) follows from the law of total probability, and (c) follows from the fact that probabilities are upper bounded by 1. Equation (5.6) greatly simplifies the calculations since it implies that we can bound the average probability of error at Eve and Charlie's side by assuming that Bob's relaying signal is correct.

Now, define the following block-error events

$$\mathcal{E}_{B,0}^b = \left\{V_0^b \neq \hat{V}_0^b\right\}, \quad \mathcal{E}_{B,1}^b = \left\{\left(V_1^b, V_2^b\right) \neq \left(\hat{V}_1^b, \hat{V}_2^b\right)\right\},$$
$$\mathcal{E}_E^b = \left\{V_0^b \neq \hat{\hat{V}}_0^b\right\}, \quad \mathcal{E}_C^b = \left\{V_2^b \neq \hat{\hat{V}}_2^b\right\}.$$

By applying the union bound and the law of total probability, we obtain

$$\mathrm{P}[\mathcal{E}_B] = \mathrm{P}[\mathcal{E}_{B,0} \cup \mathcal{E}_{B,1}] \leq \mathrm{P}[\mathcal{E}_{B,0}] + \mathrm{P}[\mathcal{E}_{B,1}] \leq 2\mathrm{P}[\mathcal{E}_{B,0}] + \mathrm{P}\big[\mathcal{E}_{B,1} \cap \overline{\mathcal{E}}_{B,0}\big]. \tag{5.7}$$

Using basic properties of sets and probabilities, we can write

$$\begin{aligned}
\mathrm{P}[\mathcal{E}_{B,0}] &= \mathrm{P}\left[\bigcup_{b=1}^{B} \mathcal{E}_{B,0}^{b} \setminus \left[\bigcup_{b'=1}^{b-1} \mathcal{E}_{B,0}^{b'}\right]\right], \\
&= \mathrm{P}\left[\bigcup_{b=1}^{B} \mathcal{E}_{B,0}^{b} \bigcap_{b'=1}^{b-1} \overline{\mathcal{E}}_{B,0}^{b'}\right], \\
&= \sum_{b=1}^{B} \mathrm{P}\left[\mathcal{E}_{B,0}^{b} \bigcap_{b'=1}^{b-1} \overline{\mathcal{E}}_{B,0}^{b'}\right], \\
&\leq \sum_{b=1}^{B} \mathrm{P}\Big[\mathcal{E}_{B,0}^{b} \cap \overline{\mathcal{E}}_{B,0}^{b-1}\Big], \\
&\leq \sum_{b=1}^{B} \mathrm{P}\Big[\mathcal{E}_{B,0}^{b} | \overline{\mathcal{E}}_{B,0}^{b-1}\Big].
\end{aligned} \tag{5.8}$$

Likewise,

$$\begin{aligned}
\mathrm{P}\big[\mathcal{E}_{B,1} \cap \overline{\mathcal{E}}_{B,0}\big] &= \mathrm{P}\left[\left(\bigcup_{b=1}^{B} \mathcal{E}_{B,1}^{b}\right) \cap \overline{\mathcal{E}}_{B,0}\right], \\
&\leq \sum_{b=1}^{B} \mathrm{P}\Big[\mathcal{E}_{B,1}^{b} \cap \overline{\mathcal{E}}_{B,0}\Big], \\
&\leq \sum_{b=1}^{B} \mathrm{P}\Big[\mathcal{E}_{B,1}^{b} \cap \overline{\mathcal{E}}_{B,0}^{b} \cap \overline{\mathcal{E}}_{B,0}^{b-1}\Big], \\
&\leq \sum_{b=1}^{B} \mathrm{P}\Big[\mathcal{E}_{B,1}^{b} | \overline{\mathcal{E}}_{B,0}^{b} \cap \overline{\mathcal{E}}_{B,0}^{b-1}\Big].
\end{aligned} \tag{5.9}$$

The same operations can be performed to bound $\mathrm{P}\big[\mathcal{E}_E|\overline{\mathcal{E}}_B\big]$, and we obtain

$$\begin{aligned}
\mathrm{P}\big[\mathcal{E}_E|\overline{\mathcal{E}}_B\big] &\leq \mathrm{P}\Bigg[\left(\bigcup_{b=1}^{B}\mathcal{E}_E^b\right)|\overline{\mathcal{E}}_B\Bigg], \\
&\leq \mathrm{P}\Bigg[\left(\bigcup_{b=1}^{B}\mathcal{E}_E^b \setminus \left[\bigcup_{b'=b+1}^{B}\mathcal{E}_E^{b'}\right]\right)|\overline{\mathcal{E}}_B\Bigg], \\
&= \mathrm{P}\Bigg[\left(\bigcup_{b=1}^{B}\mathcal{E}_E^b \bigcap_{b'=b+1}^{B}\overline{\mathcal{E}}_E^{b'}\right)|\overline{\mathcal{E}}_B\Bigg], \\
&= \sum_{b=1}^{B}\mathrm{P}\Bigg[\mathcal{E}_E^b \bigcap_{b'=b+1}^{B}\overline{\mathcal{E}}_E^{b'}|\overline{\mathcal{E}}_B\Bigg], \\
&\leq \sum_{b=1}^{B}\mathrm{P}\Big[\mathcal{E}_E^b \cap \overline{\mathcal{E}}_E^{b+1}|\overline{\mathcal{E}}_B\Big], \\
&\leq \sum_{b=1}^{B}\mathrm{P}\Big[\mathcal{E}_E^b|\overline{\mathcal{E}}_E^{b+1} \cap \overline{\mathcal{E}}_B\Big]. \qquad (5.10)
\end{aligned}$$

Finally,

$$\mathrm{P}\big[\mathcal{E}_C|\overline{\mathcal{E}}_B\big] = \mathrm{P}\Bigg[\bigcup_{b=1}^{B}\mathcal{E}_C^b|\overline{\mathcal{E}}_B\Bigg] \leq \sum_{b=1}^{B}\mathrm{P}\Big[\mathcal{E}_C^b|\overline{\mathcal{E}}_B\Big]. \qquad (5.11)$$

We shall now proceed to bound the individual probabilities of block error events obtained in Equation (5.8-5.11). Because of the symmetry of the random coding argument, we can assume that the message triple $(1,1,1)$ is sent in each of the B blocks; therefore, we have

$$\begin{aligned}
&\mathrm{P}\Big[\mathcal{E}_{B,0}^b|\overline{\mathcal{E}}_{B,0}^{b-1}\Big] \\
&= \mathrm{P}\Big[\mathcal{E}_{B,0}^b|\overline{\mathcal{E}}_{B,0}^{b-1}, V_0^b = V_1^b = V_2^b = 1\Big], \\
&\overset{(a)}{=} \mathrm{P}\Big[\exists i \neq 1 \text{ such that } (\mathbf{X}_1^n(1), \mathbf{U}^n(i|1), \mathbf{Y}^n(b)) \in A_\delta^{(n)}|\overline{\mathcal{E}}_{B,0}^{b-1}, V_0^b = V_1^b = V_2^b = 1\Big], \\
&\overset{(b)}{\leq} \sum_{i\neq 1}\mathrm{P}\Big[(\mathbf{X}_1^n(1), \mathbf{U}^n(i|1), \mathbf{Y}^n(b)) \in A_\delta^{(n)}|\overline{\mathcal{E}}_{B,0}^{b-1}, V_0^b = V_1^b = V_2^b = 1\Big], \\
&\overset{(c)}{=} \sum_{i\neq 1}\sum_{(\mathbf{x}_1^n,\mathbf{u}^n,\mathbf{y}^n)\in A_\delta^{(n)}} p\left(\mathbf{y}^n|\mathbf{x}_1^n\right) p\left(\mathbf{u}^n|\mathbf{x}_1^n\right) p\left(\mathbf{x}_1^n\right), \qquad (5.12)
\end{aligned}$$

where (a) follows from the definition of decoding in Equation (5.2), (b) follows from the union bound, and (c) follows the code construction that ensures that $\mathbf{U}^n(i|1)$ is independent of

$\mathbf{Y}^n$ given $\mathbf{X}^n(1)$. Now, the AEP ensure that there $\exists n_0 \in \mathbb{N}$, such that $\forall n \geq n_0$

$$\forall (\mathbf{u}^n, \mathbf{x}_1^n, \mathbf{y}^n) \in A_\delta^{(n)} \qquad \begin{cases} p\left(\mathbf{x}_1^n\right) \leq 2^{-n(H(X_1)-\delta)}, \\ p\left(\mathbf{u}^n|\mathbf{x}_1^n\right) \leq 2^{-n(H(U|X_1)-2\delta)}, \\ p\left(\mathbf{y}^n|\mathbf{x}_1^n\right) \leq 2^{-n(H(Y|X_1)-2\delta)}, \end{cases}$$

$$\text{and} \qquad |A_\delta^{(n)}(U, X_1, Y)| \leq 2^{n(H(U,X_1,Y)+\delta)}.$$

Substituting these bounds in Equation (5.12), we obtain

$$\begin{aligned} \mathrm{P}\Big[\mathcal{E}_{B,0}^b | \overline{\mathcal{E}}_{B,0}^{b-1}\Big] &\leq \sum_{i \neq 1} 2^{-n(I(U;Y|X)-6\delta)}, \\ &\leq 2^{n(R_0'-I(U;Y|X)+6\delta)} = 2^{n(6\delta-\epsilon)}. \end{aligned} \tag{5.13}$$

Since $6\delta < \epsilon$, $\exists n_1 \geq n_0$ such that

$$\forall n \geq n_1 \qquad \mathrm{P}\Big[\mathcal{E}_{B,0}^b | \overline{\mathcal{E}}_{B,0}^{b-1}\Big] \leq \frac{\epsilon}{36B} \quad \text{and} \quad \mathrm{P}[\mathcal{E}_{B,0}] \leq \frac{\epsilon}{36}. \tag{5.14}$$

Ve can bound $\mathrm{P}\Big[\mathcal{E}_{B,1}^b | \overline{\mathcal{E}}_{B,0}^b \cap \overline{\mathcal{E}}_{B,0}^{b-1}\Big]$ in a similar way.

$$\begin{aligned} &\mathrm{P}\Big[\mathcal{E}_{B,1}^b | \overline{\mathcal{E}}_{B,0}^b \cap \overline{\mathcal{E}}_{B,0}^{b-1}\Big] \\ &= \mathrm{P}\Big[\mathcal{E}_{B,1}^b | \overline{\mathcal{E}}_{B,0}^b \cap \overline{\mathcal{E}}_{B,0}^{b-1}, V_0^b = V_1^b = V_2^b = 1\Big] \\ &\overset{(a)}{=} \mathrm{P}\Big[\exists (j,k) \neq (1,1) : \mathbf{X}^n(j,k|1,1), \mathbf{X}_1^n(1), \mathbf{U}^n(1|1), \mathbf{Y}^n(b) | \overline{\mathcal{E}}_{B,0}^b \cap \overline{\mathcal{E}}_{B,0}^{b-1}, V_0^b = V_1^b = V_2^b = 1\Big] \\ &\overset{(b)}{\leq} \mathrm{P}\Big[\mathbf{X}^n(j,k|1,1), \mathbf{X}_1^n(1), \mathbf{U}^n(1|1), \mathbf{Y}^n(b) | \overline{\mathcal{E}}_{B,0}^b \cap \overline{\mathcal{E}}_{B,0}^{b-1}, V_0^b = V_1^b = V_2^b = 1\Big] \\ &\overset{(c)}{\leq} \sum_{(j,k)\neq(1,1)} \sum_{(\mathbf{x}^n, \mathbf{x}_1^n, \mathbf{u}^n, \mathbf{y}^n) \in A_\delta^{(n)}} p\left(\mathbf{y}^n|\mathbf{x}_1^n, \mathbf{u}^n\right) p\left(\mathbf{x}^n|\mathbf{u}^n, \mathbf{x}_1^n\right) p\left(\mathbf{u}^n|\mathbf{x}_1^n\right) p\left(\mathbf{x}_1^n\right), \\ &\overset{(d)}{\leq} \sum_{(j,k)\neq(1,1)} 2^{n(H(X,X_1,U,Y)+\delta)} 2^{-n(H(Y|U,X_1)-2\delta)} 2^{-n(H(X|U,X_1)-2\delta)} 2^{-n(H(U|X_1)-2\delta)} 2^{-n(H(X_1)-\delta)} \\ &\leq 2^{n(R_1'+R_2'-I(X;Y|U,X_1)+8\delta)} = 2^{n(8\delta-\epsilon)}, \end{aligned} \tag{5.15}$$

where (a) follows from the definition of decoding in Equation (5.3), (b) follows from the union bound, (c) follows from the code construction that ensures that $\mathbf{X}^n(j,k|1,1)$ is independent of $\mathbf{Y}^n$ given $\mathbf{X}^n(1)$ and $\mathbf{U}^n(1|1)$, and (d) follows from the AEP. Since $8\delta < \epsilon$, $\exists n_2 \geq n_1$ such that

$$\forall n \geq n_2 \qquad \mathrm{P}\Big[\mathcal{E}_{B,1}^b | \overline{\mathcal{E}}_{B,0}^b \cap \overline{\mathcal{E}}_{B,0}^{b-1}\Big] \leq \frac{\epsilon}{18B} \quad \text{and} \quad \mathrm{P}\big[\mathcal{E}_{B,1} \cap \overline{\mathcal{E}}_{B,0}\big] \leq \frac{\epsilon}{18}. \tag{5.16}$$

Using the definition of decoding in Equation (5.4) and (5.5) and the AEP, it can be shown that

$$\mathrm{P}\Big[\mathcal{E}_E^b|\overline{\mathcal{E}}_E^{b+1}\cap\overline{\mathcal{E}}_B\Big]\leq 2^{n(R_0'-I(U;Y|X_1)+6\delta)}=2^{n(6\delta-\epsilon)}, \tag{5.17}$$

$$\mathrm{P}\Big[\mathcal{E}_C^b|\overline{\mathcal{E}}_B\Big]\leq 2^{n(R_2'-I(X;Z|U,X_1)+8\delta)}=2^{n(8\delta-\frac{\epsilon}{2})}. \tag{5.18}$$

Since $\epsilon>16\delta>6\delta$, $\exists n_3\geq n_2$ such that

$$\forall n\geq n_3\quad\begin{cases}\mathrm{P}\Big[\mathcal{E}_E^b|\overline{\mathcal{E}}_E^{b+1}\cap\overline{\mathcal{E}}_B\Big]\leq\dfrac{\epsilon}{3B}\\ \mathrm{P}\Big[\mathcal{E}_C^b|\overline{\mathcal{E}}_B\Big]\leq\dfrac{\epsilon}{3B}\end{cases}\text{ and }\quad\begin{cases}\mathrm{P}\big[\mathcal{E}_E|\overline{\mathcal{E}}_B\big]\leq\dfrac{\epsilon}{3}\\ \mathrm{P}\big[\mathcal{E}_C|\overline{\mathcal{E}}_B\big]\leq\dfrac{\epsilon}{3}\end{cases} \tag{5.19}$$

Finally, substituting Equations (5.14), (5.16), and (5.19) in Equations (5.7) and then (5.6), we obtain

$$\overline{P}_{err}^{(n)}\leq\epsilon.\leq\epsilon_0 \tag{5.20}$$

Therefore, $\forall n\geq n_3$, one can argue that there exists a specific code $\mathcal{C}_{inner}^*$ with average probability of error less than ϵ_0.

Step 2: Outer code construction

In this section, we simply specify the mapping from message pair $\big(W_0^b,W_1^b\big)$ to a message triple $\big(V_0^b,V_1^b,V_2^b\big)$ performed by the outer code. To achieve rates

$$\begin{cases}0\leq R_0\leq\min\left[I(U;Y|X_1),I(U,X_1;Z)\right]-\epsilon\\ I(X;Y|U,X_1)-I(X;Z|U,X_1)\leq R_1\leq I(X;Y|U,X_1)-\epsilon\end{cases}$$

To achieve rates

$$\begin{cases}0\leq R_0\leq\min\left[I(U;Y|X_1),I(U,X_1;Z)\right]-\epsilon\\ 0\leq R_1+R_0\leq I(X;Y|U,X_1)+\min\left[I(U;Y|X_1),I(U,X_1;Z)\right]-\epsilon\\ I(X;Y|U,X_1)\leq R_1\end{cases}$$

Step 3: Equivocation calculation

We are now ready to bound Eve's equivocation $H\big(W_1^b|\mathbf{Z}^n(b)\big)$. To simplify notation, we shall write $\mathbf{X}^n$, $\mathbf{X}_1^n$, $\mathbf{U}^n$, and $\mathbf{Z}^n$ instead of $\mathbf{X}^n(V_1^b,V_2^b|V_0^b,V_0^{b-1})$, $\mathbf{X}_1^n(V_0^{b-1})$, $\mathbf{U}^n(V_0^b|V_0^{b-1})$

and $\mathbf{Z}^n(b)$.

$$\begin{aligned}
H\Big(W_1^b|\mathbf{Z}^n\Big) &\geq H\Big(W_1^b|\mathbf{Z}^n, V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) \\
&= H\Big(W_1^b, \mathbf{Z}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) - H\Big(\mathbf{Z}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) \\
&= H\Big(W_1^b, \mathbf{Z}^n, \mathbf{X}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) - H\Big(\mathbf{X}^n|W_1^b, V_0^b, V_0^{b-1}, \mathbf{X}_1^n, \mathbf{Z}^n\Big) \\
&\qquad - H\Big(\mathbf{Z}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) \\
&= H\Big(\mathbf{X}^n, W_1^b|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) + H\Big(\mathbf{Z}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n, W_1^b, \mathbf{X}^n\Big) \\
&\qquad - H\Big(\mathbf{X}^n|W_1^b, V_0^b, V_0^{b-1}, \mathbf{X}_1^n, \mathbf{Z}^n\Big) - H\Big(\mathbf{Z}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) \\
&\geq H\Big(\mathbf{X}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) - H\Big(\mathbf{X}^n|W_1^b, V_0^b, V_0^{b-1}, \mathbf{Z}^n\Big) - I\Big(\mathbf{X}^n; \mathbf{Z}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big),
\end{aligned} \tag{5.21}$$

By construction of the inner code, we have

$$\frac{1}{n}H\Big(\mathbf{X}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) = \frac{1}{n}H\Big(V_1^b, V_2^b\Big) = n\left(R_1' + R_2'\right) \geq I(X;Y|U,X_1) - \epsilon. \tag{5.22}$$

Now, the two outer codes described in Step 2 have been constructed such that W_1^b uniquely identifies V_1^b; therefore,

$$\begin{aligned}
\frac{1}{n}H\Big(\mathbf{X}^n|W_1^b, V_0^b, V_0^{b-1}, \mathbf{Z}^n\Big) &\leq \frac{1}{n}H\Big(\mathbf{X}^n|V_1^b, V_0^b, V_0^{b-1}, \mathbf{Z}^n\Big) && (5.23)\\
&= \frac{1}{n}H\Big(V_2^b|V_1^b, V_0^b, V_0^{b-1}, \mathbf{Z}^n\Big), && (5.24)\\
&\overset{(a)}{\leq} \frac{1}{n} + h\left(\mathrm{P}\Big[V_2^b \neq g(V_0^b, V_0^{b-1}, V_1^b, \mathbf{Z}^n)\Big]\right), && (5.25)
\end{aligned}$$

where (a) follows from Fano's inequality and g is any function of V_0^b, V_0^{b-1}, V_1^b, and $\mathbf{Z}^n$. In particular, g could be the decoding function used by Charlie. Since Charlie's probability of error is at at most ϵ, we obtain

$$\frac{1}{n}H\Big(\mathbf{X}^n|W_1^b, V_0^b, V_0^{b-1}, \mathbf{Z}^n\Big) \leq \frac{1}{n} + h(\epsilon). \tag{5.26}$$

Recall that by construction of the inner code, $\Big(V_0^b, V_0^{b-1}\Big)$ uniquely determines $\mathbf{U}^n$ and vice-versa; therefore, $I\Big(\mathbf{X}^n; \mathbf{Z}^n|V_0^b, V_0^{b-1}, \mathbf{X}_1^n\Big) = I(\mathbf{X}^n; \mathbf{Z}^n|\mathbf{U}^n, \mathbf{X}_1^n)$. Now, let J be an indicator function such that

$$J = \begin{cases} 1 & \text{if} \quad (\mathbf{u}^n, \mathbf{x}^n, \mathbf{x}_1^n, \mathbf{z}^n) \in A_\delta^{(n)}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.27}$$

By the AEP, $\exists n_4 \geq n_3$ such that $\forall n \geq n_4$

$$\forall(\mathbf{u}^n, \mathbf{x}^n, \mathbf{x}_1^n, \mathbf{z}^n) \in A_\delta^{(n)} \qquad \begin{cases} p\left(\mathbf{x}, \mathbf{z}^n | \mathbf{u}^n, \mathbf{x}_1^n\right) \leq 2^{-n(H(X,Z|U,X_1)-2\delta)}, \\ p\left(\mathbf{x}^n | \mathbf{u}^n, \mathbf{x}_1^n\right) \geq 2^{-n(H(X|U,X_1)+2\delta)}, \\ p\left(\mathbf{z}^n | \mathbf{u}^n, \mathbf{x}_1^n\right) \geq 2^{-n(H(Z|U,X_1)+2\delta)}, \end{cases}$$

$$\text{and} \qquad \mathrm{P}[J=1] = \mathrm{P}\Big[(\mathbf{u}^n, \mathbf{x}^n, \mathbf{x}_1^n, \mathbf{z}^n) \in A_\delta^{(n)}\Big] \geq 1-\delta.$$

Therefore,

$$\frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, \mathbf{X}_1^n, J=1) \leq I(X; Z|U, X_1) + 6\delta. \tag{5.28}$$

Now,

$$\begin{aligned} \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, \mathbf{X}_1^n) &\leq \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n, J | \mathbf{U}^n, \mathbf{X}_1^n) && (5.29) \\ &= \frac{1}{n} I(\mathbf{X}^n; J | \mathbf{U}^n, \mathbf{X}_1^n) + \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, \mathbf{X}_1^n, J), && (5.30) \\ &\leq \frac{1}{n} I(\mathbf{X}^n; J | \mathbf{U}^n, \mathbf{X}_1^n) + \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, \mathbf{X}_1^n, J=1)\, \mathrm{P}[J=1] \\ &\qquad + \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, \mathbf{X}_1^n, J=0)\, \mathrm{P}[J=0], && (5.31) \\ &\overset{(a)}{\leq} \frac{1}{n} + I(X; Z|U, X_1) + 6\delta + \delta \log_2 |\mathcal{X}|, && (5.32) \end{aligned}$$

where (a) follows from the fact that

$$I(\mathbf{X}^n; J | \mathbf{U}^n, \mathbf{X}_1^n) \leq H(J) \leq 1,$$

$$I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, \mathbf{X}_1^n, J=0) \leq H(\mathbf{X}^n) \leq n \log_2 |\mathcal{X}|,$$

$$\mathrm{P}[J=0] \leq \delta \qquad \text{and} \qquad \mathrm{P}[J=1] \leq 1.$$

Substituting the bounds obtained in Equations (5.22), (5.26), and (5.32) in Equation (5.21), we obtain that $\forall n \geq n_4$

$$\frac{1}{n} H(W_1 | \mathbf{Z}^n) \geq I(X; Y | U.X_1) - I(X; Z|U, X_1) - \epsilon - \frac{2}{n} - h(\epsilon) - 6\delta - \delta \log_2 |\mathcal{X}|. \tag{5.33}$$

Using the bounds in Equation (5.1), it is clear that $\exists n_5 \geq n_4$ such that $\forall n \geq n_5$

$$\frac{1}{n} H\left(W_1^b | \mathbf{Z}^n\right) \geq I(X; Y|U, X_1) - I(X; Z|U, X_1) - \epsilon_0, \tag{5.34}$$

and the equivocation rate $R_e \geq I(X; Y|U, X_1) - I(X; Z|U, X_1)$ is achievable. We summarize the results obtained thus far with the following lemma.

Lemma 5.2. *If X, U, X_1, Y, and Z are random variables governed by a joint probability distribution*

$$p\left(u, x_1, x, y, z\right) = p\left(y, z|x, x_1\right) p\left(x|x_1, u\right) p\left(u|x_1\right) p\left(x_1\right),$$

and such that $I(X;Y|U,X_1) \geq I(X;Z|U,X_1)$, then the rates (R_0, R_1, R_e) satisfying the conditions below are achievable.

$$\begin{cases} 0 \leq R_0 \leq \min\left[I(U;Y|X_1), I(U,X_1;Z)\right] \\ 0 \leq R_1 + R_0 \leq I(X;Y|U,X_1) + \min\left[I(U;Y|X_1), I(U,X_1;Z)\right] \\ R_e = I(X;Y|U,X_1) - I(X;Z|U,X_1) \leq R_1 \end{cases}$$

Step 4: Achievability of full region

In this last step, we characterize a larger convex region of achievable rates.

Lemma 5.3. *If X, U, V, X_1, Y, and Z are random variables governed by a joint probability distribution*

$$p\left(u, x_1, x, y, z\right) = p\left(y, z|x, x_1\right) p\left(v|x_1, u\right) p\left(x|v\right) p\left(u|x_1\right) p\left(x_1\right),$$

and such that $I(V;Y|U,X_1) \geq I(V;Z|U,X_1)$, then the rates (R_0, R_1, R_e) satisfying the conditions below are achievable.

$$\begin{cases} 0 \leq R_0 \leq \min\left[I(U;Y|X_1), I(U,X_1;Z)\right] \\ 0 \leq R_1 + R_0 \leq I(V;Y|U,X_1) + \min\left[I(U;Y|X_1), I(U,X_1;Z)\right] \\ 0 \leq R_e = I(V;Y|U,X_1) - I(V;Z|U,X_1) \leq R_1 \end{cases}$$

Proof. The result follows by prefixing the partially cooperative broadcast channel with a discrete memoryless channel with transition probability $p(x|v)$. □

Lemma 5.4. *The region $\mathcal{R}$ defined below is convex.*

$$\mathcal{R} = \bigcup_{X_1 U \to V \to X \to YZ} \left\{ \begin{array}{l} 0 \leq R_0 \leq \min\left[I(U;Y|X_1), I(U,X_1;Z)\right] \\ 0 \leq R_1 + R_0 \leq I(V;Y|U,X_1) + \min\left[I(U;Y|X_1), I(U,X_1;Z)\right] \\ 0 \leq R_e \leq I(V;Y|U,X_1) - I(V;Z|U,X_1) \leq R_1 \end{array} \right\} \tag{5.35}$$

Proof. The result can be proved by showing that the rates achieved with time-shared random variables are also in $\mathcal{R}$. See the proof of Lemma A.3 in Appendix A for details. □

Lemma 5.5. *The region $\mathcal{R}$ is equal to the region $\mathcal{C}$ defined below.*

$$\mathcal{C} = \bigcup_{X_1 U \to V \to X \to YZ} \left\{ \begin{array}{l} 0 \leq R_e \leq R_1, \\ 0 \leq R_e \leq I(V;Y|U,X_1) - I(V;Z|U,X_1) \\ 0 \leq R_0 \leq \min\left[I(U;Y|X_1), I(U,X_1;Z)\right] \\ 0 \leq R_1 + R_0 \leq I(V;Y|U,X_1) + \min\left[I(U;Y|X_1), I(U,X_1;Z)\right] \end{array} \right\} \tag{5.36}$$

Proof. This lemma is consequence of the convexity of $\mathcal{R}$. See the proof of of Lemma A.4 in Appendix A for details. □

5.4.2 Converse part

The converse part of the proof combines the techniques used in [59] to prove the upper bound on the capacity region of the relay channel, and in [6] to prove the upper bound on the achievable region of the broadcast channel with confidential messages. We also make extensive use of the fact that the relay signal $X_{1,i}$ is a deterministic function of the observations $\mathbf{Y}^{i-1} = (Y_1, \ldots, Y_{i-1})$ only.

We consider a sequence of $\left(2^{nR_0}, 2^{nR_1}, n\right)$ codes for the partially cooperative broadcast channel with $P_e^{(n)} \to 0$. From Fano's inequality [1] we obtain

$$H(W_1, W_0|\mathbf{Y}^n) \leq 1 + n\left(R_0 + R_1\right) P_e^{(n)} \stackrel{\Delta}{=} n\epsilon_1, \tag{5.37}$$

$$H(W_0|\mathbf{Z}^n) \leq 1 + nR_0 P_e^{(n)} \stackrel{\Delta}{=} n\epsilon_2, \tag{5.38}$$

where $\epsilon_1, \epsilon_2 \to 0$ as $n \to \infty$. By the chain rule of entropy, Equation (5.37) also implies that

$$H(W_1|W_0, \mathbf{Y}^n) \leq n\epsilon_1 \quad \text{and} \quad H(W_0|\mathbf{Y}^n) \leq n\epsilon_1. \tag{5.39}$$

Let us introduce the shorthand notation $\tilde{\mathbf{Z}}^{i+1} = (Z_{i+1}, \ldots, Z_n)$. Now, consider the

common message rate nR_0.

$$\begin{aligned}
nR_0 &= H(W_0) = I(W_0;\mathbf{Y}^n) + H(W_0|\mathbf{Y}^n) \\
&\leq I(W_0;\mathbf{Y}^n) + n\epsilon_1 = \sum_{i=1}^{n} I\big(W_0;Y_i|\mathbf{Y}^{i-1}\big) + n\epsilon_1 \\
&= \sum_{i=1}^{n}\Big\{I\Big(W_0,\tilde{\mathbf{Z}}^{i+1};Y_i|\mathbf{Y}^{i-1}\Big) - I\Big(\tilde{\mathbf{Z}}^{i+1};Y_i|W_0,\mathbf{Y}^{i-1}\Big)\Big\} + n\epsilon_1 \\
&= \sum_{i=1}^{n}\Big\{I\Big(W_0,\tilde{\mathbf{Z}}^{i+1};Y_i|\mathbf{Y}^{i-1}X_{1,i}\Big) - I\Big(\tilde{\mathbf{Z}}^{i+1};Y_i|W_0,\mathbf{Y}^{i-1}\Big)\Big\} + n\epsilon_1 \\
&\leq \sum_{i=1}^{n}\Big\{I\Big(W_0,\tilde{\mathbf{Z}}^{i+1},\mathbf{Y}^{i-1};Y_i|X_1,i\Big) - I\Big(\tilde{\mathbf{Z}}^{i+1};Y_i|W_0,\mathbf{Y}^{i-1}\Big)\Big\} + n\epsilon_1 \qquad (5.40)
\end{aligned}$$

Likewise, using $\mathbf{Z}^n$ instead of $\mathbf{Y}^n$, on can show that

$$nR_0 \leq \sum_{i=1}^{n}\Big\{I\Big(W_0,\tilde{\mathbf{Z}}^{i+1},\mathbf{Y}^{i-1},X_{1,i};Z_i\Big) - I\Big(\mathbf{Y}^{i-1};Z_i|W_0,\tilde{\mathbf{Z}}^{i+1}\Big)\Big\} + n\epsilon_2 \qquad (5.41)$$

Notice that Equations (5.40) and (5.41) also imply

$$nR_0 \leq \sum_{i=1}^{n} I\Big(W_0,\tilde{\mathbf{Z}}^{i+1},\mathbf{Y}^{i-1};Y_i|X_1,i\Big) + n\epsilon_1, \qquad (5.42)$$

$$nR_0 \leq \sum_{i=1}^{n} I\Big(W_0,\tilde{\mathbf{Z}}^{i+1},\mathbf{Y}^{i-1},X_{1,i};Z_i\Big) + n\epsilon_2. \qquad (5.43)$$

Now, let us consider $H(W_1|W_0)$.

$$\begin{aligned}
H(W_1|W_0) &= I(W_1;\mathbf{Y}^n|W_0) + H(W_1|W_0,\mathbf{Y}^n) \\
&\leq I(W_1;\mathbf{Y}^n|W_0) + n\epsilon_1 \\
&= \sum_{i=1}^{n} I\big(W_1;Y_i|W_0,\mathbf{Y}^{i-1}\big) + n\epsilon_1 \\
&= \sum_{i=1}^{n}\Big\{I\Big(W_1,\tilde{\mathbf{Z}}^{i+1};Y_i|W_0,\mathbf{Y}^{i-1}\Big) - I\Big(\tilde{\mathbf{Z}}^{i+1};Y_i|W_0,\mathbf{Y}^{i-1},W_1\Big)\Big\} + n\epsilon_1 \\
&\leq \sum_{i=1}^{n}\Big\{I\Big(W_1,\tilde{\mathbf{Z}}^{i+1};Y_i|W_0,\mathbf{Y}^{i-1}\Big)\Big\} + n\epsilon_1 \\
&= \sum_{i=1}^{n}\Big\{I\Big(\tilde{\mathbf{Z}}^{i+1};Y_i|W_0,\mathbf{Y}^{i-1}\Big) + I\Big(W_1;Y_i|W_0,\tilde{\mathbf{Z}}^{i+1},\mathbf{Y}^{i-1}\Big)\Big\} + n\epsilon_1 \qquad (5.44)
\end{aligned}$$

Combining Equation (5.40) and Equation (5.44) we obtain

$$\begin{aligned}
n\,(R_0+R_1) &= H(W_0,W_1) = H(W_1|W_0) + H(W_0) \\
&\leq \sum_{i=1}^{n}\Big\{I\Big(W_0,\tilde{\mathbf{Z}}^{i+1},\mathbf{Y}^{i-1};Y_i|X_1,i\Big) + I\Big(W_1;Y_i|W_0,\tilde{\mathbf{Z}}^{i+1},\mathbf{Y}^{i-1}\Big)\Big\} + 2n\epsilon_1 \qquad (5.45)
\end{aligned}$$

We shall now introduce the following lemma.

Lemma 5.6 (adapted from [6], Lemma 7)**.**

$$\sum_{i=1}^{n} I\left(\tilde{\mathbf{Z}}^{i+1}; Y_i | W_0, \mathbf{Y}^{i-1}\right) = \sum_{i=1}^{n} I\left(Y^{i-1}; Z_i | W_0, \tilde{\mathbf{Z}}^{i+1}\right), \tag{5.46}$$

$$\sum_{i=1}^{n} I\left(\tilde{\mathbf{Z}}^{i+1}; Y_i | W_0, W_1, \mathbf{Y}^{i-1}\right) = \sum_{i=1}^{n} I\left(Y^{i-1}; Z_i | W_0, W_1, \tilde{\mathbf{Z}}^{i+1}\right). \tag{5.47}$$

Utilizing Lemma 5.6 in Equation (5.44) and combining the result with Equation (5.41) we obtain

$$\begin{aligned} n\left(R_0 + R_1\right) &= H(W_0, W_1) = H(W_1|W_0) + H(W_0) \\ &\leq \sum_{i=1}^{n} \left\{ I\left(W_0, \tilde{\mathbf{Z}}^{i+1}, \mathbf{Y}^{i-1}, X_{1,i}; Z_i\right) + I\left(W_1; Y_i | W_0, \tilde{\mathbf{Z}}^{i+1}, \mathbf{Y}^{i-1}\right) \right\} + n\left(\epsilon_1 + \epsilon_2\right) \end{aligned} \tag{5.48}$$

Finally, by using a method similar to those in [6] and Lemma 5.6, we can bound the equivocation rate as

$$\begin{aligned} nR_e &\leq H(W_1|\mathbf{Z}^n) = H(W_1|W_0, \mathbf{Z}^n) + I(W_1; W_0|\mathbf{Z}^n) \\ &= I(W_1; \mathbf{Y}^n|W_0) - I(W_1; \mathbf{Z}^n|W_0) + H(W_1|W_0, \mathbf{Y}^n) + I(W_1; W_0|\mathbf{Z}^n) \\ &\leq I(W_1; \mathbf{Y}^n|W_0) - I(W_1; \mathbf{Z}^n|W_0) + H(W_1|W_0, \mathbf{Y}^n) + H(W_0|\mathbf{Z}^n) \\ &\leq I(W_1; \mathbf{Y}^n|W_0) - I(W_1; \mathbf{Z}^n|W_0) + n\left(\epsilon_1 + \epsilon_2\right) \\ &= \sum_{i=1}^{n} \left\{ I\left(W_1; Y_i | W_0, \mathbf{Y}^{i-1}\right) - I\left(W_1; Z_i | W_0, \tilde{\mathbf{Z}}^{i+1}\right) \right\} + n\left(\epsilon_1 + \epsilon_2\right) \\ &= \sum_{i=1}^{n} \left\{ I\left(W_1, \tilde{\mathbf{Z}}^{i+1}; Y_i | W_0, \mathbf{Y}^{i-1}\right) - I\left(\tilde{\mathbf{Z}}^{i+1}; Y_i | W_0, W_1, \mathbf{Y}^{i-1}\right) \right. \\ &\qquad \left. - I\left(W_1, \mathbf{Y}^{i-1}; Z_i | W_0, \tilde{\mathbf{Z}}^{i+1}\right) + I\left(\mathbf{Y}^{i-1}; Z_i | W_0, W_1, \tilde{\mathbf{Z}}^{i+1}\right) \right\} + n\left(\epsilon_1 + \epsilon_2\right) \\ &= \sum_{i=1}^{n} \left\{ I\left(W_1, \tilde{\mathbf{Z}}^{i+1}; Y_i | W_0, \mathbf{Y}^{i-1}\right) - I\left(W_1, \mathbf{Y}^{i-1}; Z_i | W_0, \tilde{\mathbf{Z}}^{i+1}\right) \right\} + n\left(\epsilon_1 + \epsilon_2\right) \\ &= \sum_{i=1}^{n} \left\{ I\left(\tilde{\mathbf{Z}}^{i+1}; Y_i | W_0, \mathbf{Y}^{i-1}\right) + I\left(W_1; Y_i | W_0, \tilde{\mathbf{Z}}^{i+1}, \mathbf{Y}^{i-1}\right) \right. \\ &\qquad \left. - I\left(\mathbf{Y}^{i-1}; Z_i | W_0, \tilde{\mathbf{Z}}^{i+1}\right) - I\left(W_1; Z_i | W_0, \tilde{\mathbf{Z}}^{i+1}, \mathbf{Y}^{i-1}\right) \right\} + n\left(\epsilon_1 + \epsilon_2\right) \\ &= \sum_{i=1}^{n} \left\{ I\left(W_1; Y_i | W_0, \tilde{\mathbf{Z}}^{i+1}, \mathbf{Y}^{i-1}\right) - I\left(W_1; Z_i | W_0, \tilde{\mathbf{Z}}^{i+1}, \mathbf{Y}^{i-1}\right) \right\} + n\left(\epsilon_1 + \epsilon_2\right) \\ &= \sum_{i=1}^{n} \left\{ I\left(W_1; Y_i | W_0, \tilde{\mathbf{Z}}^{i+1}, \mathbf{Y}^{i-1}, X_{1,i}\right) - I\left(W_1; Z_i | W_0, \tilde{\mathbf{Z}}^{i+1}, \mathbf{Y}^{i-1}, X_{1,i}\right) \right\} + n\left(\epsilon_1 + \epsilon_2\right). \end{aligned}$$

Let us now introduce the random variables $U_i = \left(W_0, \tilde{\mathbf{Z}}^{i+1}, \mathbf{Y}^{i-1}\right)$ and $V_i = (W_1, U_i)$. One can check that random variables U_i, V_i, X_i, $X_{1,i}$, Y_i, and Z_i are such that $U_i X_{1,i} \to V_i \to$

$X_i \to Y_i Z_i$. We can rewrite Equations (5.42), (5.43), (5.45), (5.48), and (5.49) as

$$\begin{aligned}
nR_0 &\leq \sum_{i=1}^{n} I(U_i; Y_i|X_1, i) + n\epsilon_1, \\
nR_0 &\leq \sum_{i=1}^{n} I(U_i, X_{1,i}; Z_i) + n\epsilon_2. \\
n\left(R_0 + R_1\right) &\leq \sum_{i=1}^{n} \left\{I(U_i; Y_i|X_1, i) + I(V_i; Y_i|U_i)\right\} + 2n\epsilon_1, \\
n\left(R_0 + R_1\right) &\leq \sum_{i=1}^{n} \left\{I(U_i, X_{1,i}; Z_i) + I(V_i; Y_i|U_i)\right\} + n\left(\epsilon_1 + \epsilon_2\right), \\
nR_e &\leq \sum_{i=1}^{n} \left\{I(V_i; Y_i|U_i, X_{1,i}) - I(V_i; Z_i|U_i, X_{1,i})\right\} + n\left(\epsilon_1 + \epsilon_2\right).
\end{aligned}$$

We now introduce a random Q independent of $(W_0, W_1, X^n, \mathbf{Y}^n, \mathbf{Z}^n, X_1^n)$ and uniformly distributed over the set $\{1, \ldots, n\}$, and define $V = V_Q$, $Y = Y_Q$, $Z = Z_Q$ and $X_1 = X_{1_Q}$. Therefore, letting

$$\epsilon_3 = \max\left[2\epsilon_1, \epsilon_1 + \epsilon_2\right]$$

we obtain the desired result

$$\begin{aligned}
R_0 &\leq \min\left[I(U; Y|X_1), I(U, X_1; Z)\right] + \epsilon_3, \\
\left(R_0 + R_1\right) &\leq \min\left[I(U; Y|X_1), I(U, X_1; Z)\right] + I(V; Y|U) + \epsilon_3, \\
R_e &\leq I(V; Y|U, X_1) - I(V; Z|U, X_1) + \epsilon_3,
\end{aligned}$$

where $\epsilon_3 \to 0$ as $n \to \infty$.

CHAPTER 6
INFORMATION-THEORETIC COMMITMENT

In Chapters 3-5, we discussed the beneficial role of noisy channels for secure communication; however, the usefulness of noisy channels, and more generally of sources of correlated randomness, for cryptographic purposes goes well beyond the scope of secure transmissions. For instance, it is known that information theoretically secure cryptographic primitives such as bit commitment, oblivious transfer, and coin tossing can be built upon non-trivial noisy correlations.

The aforementioned cryptographic primitives have been extensively studied in the cryptography community from a computational security perspective. Following the work of Crépeau [62], the information theory has gained interest for these problems by reconsidering security from an information-theoretic perspective. Recently, information-theoretically secure protocols for oblivious transfer and bit commitment have been obtained for non-trivial discrete channels, and the highest rate at which a channel can be used for bit commitment have been characterized in [63, 64]. Although there have been a few contributions dealing with the construction of practical bit commitment schemes [62, 65], the design of such schemes in a more general setting, and especially for the Gaussian channel, remains an open problem.

In this chapter, we provide a partial solution to this problem by presenting a new achievability proof of bit commitment rates over discrete memoryless and Gaussian channels, which highlights the innate connection with secret-key agreement. In the case of discrete memoryless channels, where contrary to the Gaussian case [64] the commitment capacity is not infinite, our proof also shows the achievability of the bit commitment capacity. Although non-constructive, our proof yields useful insight on how to design *practical* bit commitment schemes over general noisy channels.

6.1 Principle of bit commitment

In this section, we briefly review the bit commitment models proposed in [63, 12]. As shown in Figure 33, we assume that there exists a unidirectional memoryless channel $(\mathcal{X}, p(z|x), \mathcal{Z})$

connecting the players Alice and Bob. This channel can be either discrete of Gaussian with a input power constraint. Furthermore, we assume that a bidirectional noiseless channel is also available.

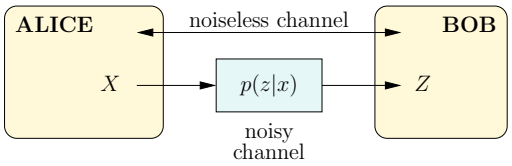


Figure 33. Bit commitment setup.

A commitment scheme consists of two phases, a *commit* phase and a *reveal* phase. During the commit phase, Alice commits to a message a belonging to a set $\mathcal{A}$ and sends Bob an evidence of her commitment. The set of messages received by Bob and his private randomness are denoted by the random variable V_b. During the reveal phase, Alice discloses a and any other private randomness R she may have used in the commit phase. Bob then checks if Alice's data is consistent with the evidence received during the commit phase. He finally accepts a if it is consistent, and rejects it otherwise. Bob's test is denoted by the function $\beta(V_b, a, R) \in \{\text{ACC}, \text{REJ}\}$.

A bit commitment protocols should satisfy the following properties:

- **Concealing.** A bit commitment protocol is ϵ-concealing if after receiving Alice's commitment V_b, Bob is unable to learn more than ϵ bits of information on A:

$$I(A; V_b) \leq \epsilon. \tag{6.1}$$

- **Binding.** A protocol is δ-binding if Alice cannot cheat once she has committed to a value:

$$\forall a, a' \neq a \in \mathcal{A} \qquad \mathrm{P}\big[\beta(V_b, a, R) = \text{ACC}, \beta(V_b, a', R) = \text{ACC}\big] \leq \delta. \tag{6.2}$$

- **Correct.** A protocol is η-correct if, when Alice and Bob behave according to the protocol, the probability of Bob accepting Alice's commitment is higher than $1 - \eta$:

$$\forall a \in \mathcal{A} \quad \mathrm{P}[\beta(V_b, a, R) = ACC] \geq 1 - \eta. \tag{6.3}$$

Definition 6.1. *A bit commitment protocol is secure if for any $\epsilon > 0$, $\delta > 0$, and $\eta > 0$ there exists a commitment that is ϵ-concealing, δ-binding, and η-correct. The commitment rate of a secure protocol is defined as*

$$R = \frac{\log_2 |\mathcal{A}|}{n}, \tag{6.4}$$

and the commitment capacity is the maximum rate achievable by a secure protocol.

In the following, we implicitly assume that the discrete memoryless channels considered a *non-redundant*, meaning that none of their output distributions is a convex combination of their other output distributions:

$$\forall x \in \mathcal{X}, \forall p \text{ s.t. } p(x) \neq 0,\ p(z|x) \neq \sum_{x'} p(x')p(z|x'). \tag{6.5}$$

As discussed in [63], this assumption does not limit the scope of the results, and one can note that most channels of interest (binary symmetric, Gaussian, etc.) satisfy this property. Also, we point out that we do not allow Alice or Bob to modify the channel.

6.2 Bit commitment from secret key agreement

The earlier achievability proofs of bit commitment rates over discrete memoryless and Gaussian channels [63, 64] exploit the connections with the achievability proofs of secrecy capacity over the wiretap channel [5, 7]. In fact in both cases, the proofs rely on the existence of "wiretap codes" capable of ensuring simultaneously reliability and security. Unfortunately these proofs provide little insight on how to design bit commitment protocols, since the design of practical wiretap codes is still a challenging problem [32]. However, the link between bit commitment and wiretap codes turns out to be a mathematical convenience rather than a fundamental connection. More specifically, as we will see shortly, the concealing and binding properties of a bit commitment scheme need not be enforced by a single code, revealing the innate connection with secret key agreement and Slepian-Wolf coding [38] for which practical constructions exist.

We now introduce a simple bit commitment protocol over discrete memoryless channels.

PROTOCOL 1 (discrete memoryless channels)

Commit phase

1. Alice generates a vector of n i.i.d symbols $\mathbf{x}^n \in \mathcal{X}^n$ according to a uniform probability distribution and sends it through the noisy channel; Bob receives a correlated vector $\mathbf{z}^n$;

2. Alice sends a vector $\mathbf{s} = s(\mathbf{x}^n)$ of $\alpha n + \log_2 n + \xi$ bits to Bob, calculated according to a predefined Slepian-Wolf compression code (the exact choice of s, $\alpha > 0$ and $\xi > 0$ will be discussed later);

3. Alice chooses a function g uniformly at random among a universal$_2$ family of hash functions $\mathcal{G} : \mathcal{X}^n \to \{0,1\}^k$, where $k = nH(X|Z) - \alpha n - \log_2 n - \xi - 2r_0 - 2 - r_1$ ($r_0, r_1 > 0$), and reveals g to Bob;

4. Alice distills a key $\mathbf{k} = g(\mathbf{x}^n)$ and computes $\mathbf{c} = \mathbf{k} \oplus \pi(\mathbf{a})$, where $\mathbf{a}$ is the message she wants to commit to and $\pi : \mathcal{A} \to \{0,1\}^k$ is a one to one mapping;

5. Alice reveals $\mathbf{c}$ to Bob;

Reveal phase

1. Alice reveals $\mathbf{x}^n$ and $\mathbf{m}$ to Bob;

2. Bob accepts $\mathbf{m}$ if and only if

   (a) $\mathbf{z}^n \in A_\epsilon^{*(n)}(Z|\mathbf{x}^n)$, where $A_\epsilon^{*(n)}(Z|\mathbf{x}^n)$ is the set of sequences $\mathbf{z}^n$ strongly conditionally typical with $\mathbf{x}^n$,

   (b) $s(\mathbf{x}^n) = \mathbf{s}$,

   (c) $g(\mathbf{x}^n) \oplus \mathbf{c} = \mathbf{m}$.

Lemma 6.1. *Protocol 1 is ϵ-concealing.*

Proof. The concealing property depends only on the secrecy of the key $\mathbf{k}$ distilled by Alice, hence the result follows directly the properties of reconciliation and privacy amplification.

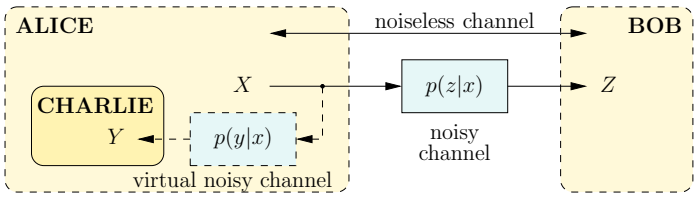


Figure 34. Bit commitment setup with third party.

More specifically, as shown in Figure 34, let us introduce a third party (Charlie) observing the vector $\mathbf{x}^n$ through a virtual discrete memoryless channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ such that $nH(X|Y) = \alpha n + \log_2 n + \xi$. After the first step of the protocol, Alice, Charlie and Bob have access respectively to the n i.i.d. realizations of the correlated random variable X, Y and Z, distributed according to $p(x, y, z) = p(z|x)p(y|x)p(x)$. Let G be the random variable denoting the random choice of the hash function. From [41, Theorem 5.2] and [39, Corollary 4], we know that Alice and Charlie can distill a secret key K of length

$$k = nH(X|Z) - \alpha n - \log_2 n - \xi - 2r_0 - 2 - r_1,$$

such that

$$H(K|\mathbf{z}^n, \mathbf{s}, G) \geq k - 2^{-r_1}/\ln 2 \tag{6.6}$$

with probability $1 - 2^{-r_0}$. It is important to note that this result does not depend on the actual function s, but only on the number of bits disclosed. □

We now introduce a simple lemma that we shall use to study the binding property of the protocol.

Lemma 6.2. *The number of sequences of length $n \geq 1$ in $H_\sigma(\mathbf{x})$ is upper bounded by $\kappa\sqrt{n}2^{(h(\sigma)+\sigma\log_2|\mathcal{X}|)n}$ for some $\kappa > 0$.*

Proof. Without loss of generality we can assume that σn ($\sigma < 1/2$) is an integer. Therefore,

$$|H^*_\sigma(\mathbf{x})| = \sum_{i=1}^{\sigma n} \binom{n}{i} (|\mathcal{X}| - 1)^i \leq \sigma n \binom{n}{\sigma n} |\mathcal{X}|^{\sigma n}, \tag{6.7}$$

and using Stirling's approximation, we have that

$$\binom{n}{\sigma n} \leq \frac{1}{\sqrt{2\pi n\sigma(1-\sigma)}} 2^{h(\sigma)n} \exp\frac{1}{12n}. \tag{6.8}$$

For $n \geq 1$ $\exp(1/12n) \leq \sqrt{2\pi}$, hence setting $\kappa = \sqrt{\sigma/(1-\sigma)}$ we get the desired result. □

Lemma 6.3. *There exists a Slepian-Wolf compression scheme such that Protocol 1 is δ-binding.*

Proof. Let $\delta > 0$ and $\alpha > 0$. Let $\sigma > 0$ be such that

$$h(\sigma) + \sigma \log_2 |\mathcal{X}| < \alpha. \tag{6.9}$$

We denote the Hamming distance between between two discrete vectors $\mathbf{x}^n$ and $\mathbf{x}_0^n$ by $d_H(\mathbf{x}^n, \mathbf{x}_0^n)$. Now, for any $\mathbf{x}^n \in \mathcal{X}^n$ define

$$H_\sigma(\mathbf{x}^n) = \{\mathbf{x}_0^n \in \mathcal{X}^n : d_H(\mathbf{x}_0^n, \mathbf{x}^n) < \sigma n\} \tag{6.10}$$

$$\text{and} \quad H_\sigma^*(\mathbf{x}^n) = H_\sigma(\mathbf{x}^n) \setminus \{\mathbf{x}^n\}. \tag{6.11}$$

For any $\mathbf{x}^n, \mathbf{x}_0^n \in \mathcal{X}^n$, if $\mathbf{x}_0^n \notin H_\sigma(\mathbf{x}^n)$ we know from [63, Lemma 14] that $\exists n_1 \in \mathbb{N}$ such that

$$\forall n \geq n_1 \quad \mathrm{P}\Big[\mathbf{Z}^n \in A_\epsilon^{(n)}(Z|\mathbf{x}_0^n)\Big] < \delta \tag{6.12}$$

Hence, Alice cannot pass Bob's first test but with arbitrarily low probability. We shall now show that there exists a function s such that if $\mathbf{x}_0^n \in H_\sigma^*(\mathbf{x}^n)$, then Alice cannot pass Bob's second test but with arbitrarily low probability. To do so, we first use a random binning argument to show that for n large enough,

$$\mathrm{P}\Big[\exists \mathbf{x}_0^n \in H_\sigma^*(\mathbf{X}^n) : \mathbf{Z}^n \in A_\epsilon^{(n)}(Z|\mathbf{X}^n), S(\mathbf{x}_0^n) = S(\mathbf{X}^n)\Big] \leq \frac{\delta}{2},$$

where S is a random variable representing a Slepian-Wolf compression code s chosen uniformly at random among all codes disclosing α_n bits. In fact,

$$\begin{aligned}
&\mathrm{P}\Big[\exists \mathbf{x}_0^n \in H_\sigma^*(\mathbf{X}^n) : \mathbf{Z}^n \in A_\epsilon^{(n)}(Z|\mathbf{X}^n), S(\mathbf{x}_0^n) = S(\mathbf{X}^n)\Big] \\
&\overset{(a)}{\leq} \sum_{\mathbf{x}^n, \mathbf{z}^n} \sum_{\mathbf{x}_0^n \in H_\sigma^*(\mathbf{x}^n)} p(\mathbf{x}^n, \mathbf{z}^n) \mathrm{P}[S(\mathbf{x}_0^n) = S(\mathbf{x}^n)], \\
&\overset{(b)}{\leq} \sum_{\mathbf{x}^n, \mathbf{z}^n} p(\mathbf{x}^n, \mathbf{z}^n) \kappa \sqrt{n} 2^{(h(\sigma) + \sigma \log_2 |\mathcal{X}|)n} 2^{-\alpha n}, \\
&\leq \kappa \sqrt{n} 2^{(h(\sigma) + \sigma \log_2 |\mathcal{X}| - \alpha)n},
\end{aligned} \tag{6.13}$$

where (a) follows from the union bound and

$$\mathrm{P}\Big[\mathbf{Z}^n \in A_\epsilon^{(n)}(Z|\mathbf{X}^n), S(\mathbf{x}_0^n) = S(\mathbf{x}^n)\Big] \leq \mathrm{P}[S(\mathbf{x}_0^n) = S(\mathbf{x}^n)], \tag{6.14}$$

and (b) follows from the bound on $|H^*_\sigma(\mathbf{x})|$ derived in Lemma 6.2 and

$$\mathrm{P}\big[S(\mathbf{x}') = S(\mathbf{x})\big] = 2^{-\alpha n}. \tag{6.15}$$

Since $h(\sigma) + \sigma \log_2 |\mathcal{X}| < \alpha$, $\exists n_2 \geq n_1$ such that

$$\forall n \geq n_2 \qquad \kappa\sqrt{n} 2^{(h(\sigma)+\sigma \log_2 |\mathcal{X}|-\alpha)n} \leq \frac{\delta}{2}. \tag{6.16}$$

It is now standard procedure [1] to argue that there exists at least one function s_1 such that

$$\forall n \geq n_2 \quad \mathrm{P}\Big[\exists \mathbf{x}_0^n \in H^*_\sigma(\mathbf{X}^n) : \mathbf{Z}^n \in A^{(n)}_\epsilon(Z|\mathbf{X}^n), s_1(\mathbf{x}_0^n) = s_1(\mathbf{X}^n)\Big] \leq \frac{\delta}{2}. \tag{6.17}$$

Since the random variable $\mathbf{X}^n$ is uniformly distributed on $\mathcal{X}^n$, notice that Equation (6.17) only implies that

$$\forall n \geq n_2 \quad \mathrm{P}\Big[\exists \mathbf{x}_0^n \in H^*_\sigma(\mathbf{x}^n) : \mathbf{Z}^n \in A^{(n)}_\epsilon(Z|\mathbf{x}^n), s_1(\mathbf{x}_0^n) = s_1(\mathbf{x}^n)\Big] \leq \delta \tag{6.18}$$

for half of the sequence $\mathbf{x}^n \in \mathcal{X}^n$. Let $\mathcal{X}_1$ be the set containing such sequences, and note that $|\mathcal{X}_1| = \lceil |\mathcal{X}^n|/2 \rceil$. By applying the previous random coding argument to the set $\mathcal{X}^n \setminus \mathcal{X}_1$, one can show the existence of a function s_2 and a set $\mathcal{X}_2 \subset \mathcal{X}^n \setminus \mathcal{X}_1$ such that $|\mathcal{X}_2| = \lceil |\mathcal{X}^n|/4 \rceil$ and the sequences $\mathbf{x} \in \mathcal{X}_2$ satisfy Equation (6.18) with s_2. By induction, one can then construct a sequences of disjoint sets $\mathcal{X}_i$ of decreasing size and a sequence of functions s_i such that all the sequences $\mathbf{x} \in \mathcal{X}_i$ satisfy Equation (6.18) with s_i. It is easy to check that $n \log_2 |\mathcal{X}|$ sets are sufficient to ensure that any

$$\forall \mathbf{x}^n \in \mathcal{X}^n \quad \exists i \in \{1, \ldots, n \log_2 |\mathcal{X}|\} \quad \text{such that} \quad \mathbf{x}^n \in \mathcal{X}_i.$$

Let s be a Slepian-Wolf compression code defined as follows:

$$\forall \mathbf{x}^n \in \mathcal{X}^n \; s(\mathbf{x}^n) = (i, s_i(\mathbf{x}^n)) \text{ if } \mathbf{x}^n \in \mathcal{X}_i. \tag{6.19}$$

By construction $(i, s_i(\mathbf{x}^n))$ can be described with only $\alpha n + \log_2 n + \log_2 \log_2 |\mathcal{X}|$ bits. Hence, we have shown the existence of a Slepian-Wolf code s such that

$$\forall \mathbf{x}^n \in \mathcal{X}^n, \qquad \mathrm{P}\Big[\exists \mathbf{x}_0^n \in H^*_\sigma(\mathbf{x}^n) : \mathbf{Z}^n \in A^{(n)}_\epsilon(Z|\mathbf{x}^n), s(\mathbf{x}_0^n) = s(\mathbf{x}^n)\Big] \leq \delta. \tag{6.20}$$

Setting $\xi = \log_2 \log_2 |\mathcal{X}|$ concludes the proof. □

Lemma 6.4. *Protocol 1 is η-correct.*

Proof. Let $\eta > 0$. If Alice and Bob behave according to the protocol, clearly the only test which might fail is Bob's first typicality test. However the Asymptotic Equipartition Principle guarantees that the probability of the set of non jointly-typical sequences gets negligible as $n \to \infty$ [1]. □

Based on the previous lemmas, we can now prove the following theorem.

Theorem 6.1. *Protocol 1 achieves the commitment capacity of a discrete memoryless channel.*

Proof. Since the parameters ξ, r_0 and r_1 are independent of n, Alice can commit to messages at a rate $k/n \approx H(X|Z) - \alpha$ in the limit of large n . Furthermore, α can be chosen as small as desired, hence protocol 1 achieves the bit commitment rate $H(X|Z)$ as $n \to \infty$. As it is, our protocol does not achieve the bit commitment capacity since we have assumed that the sequences $\mathbf{x}^n$ were chosen uniformly at random; however, this assumption can be removed since it is not required in the derivation of Lemmas 6.1 and 6.4, and the proof of Lemma 6.3 can be modified to operate on the set of typical sequences $\mathbf{x}^n$ sequences rather than on the whole set $\mathcal{X}^n$. □

We note that Imai *et al* have proposed a similar scheme in [65], however their protocol differs from Protocol 1 in that the additional bits computed in step 2 are chosen randomly by Bob for each committed message. Their protocol has the advantage of being constructive, however its extension to the Gaussian case is not straightforward because of the required interactivity. On the other hand, Protocol 1 can be easily modified to operate over the Gaussian channel as described below.

Protocol 2 (Gaussian channel)

Commit phase

1. Alice generates a vector of n i.i.d continuous symbols $\mathbf{x}^n = \{x_i\}_{i=1..n}$, where the x_i are chosen in a finite set $\mathcal{X} \in \mathbb{R}$ according to a *discrete* probability distribution $p(x)$ such

that $\mathbb{E}\{X^2\} \leq P$; Alice then sends $\mathbf{x}^n$ through the Gaussian channel; Bob receives a correlated vector of symbols $\mathbf{z}^n$;

2. Alice chooses a function g uniformly at random among a universal$_2$ family of hash functions $\mathcal{G} : \mathcal{X}^n \to \{0,1\}^k$, where $k = nH(X|Z) - 2r_0 - 2 - r_1$ $(s, r > 0)$, and reveals g to Bob;

3. Alice distills a key $\mathbf{k} = g(\mathbf{x})$ and computes $\mathbf{c} = \mathbf{k} \oplus \pi(\mathbf{a})$, where $\mathbf{a}$ is the message she wants to commit to and $\pi : \mathcal{A} \to \{0,1\}^k$ is a one to one mapping;

4. Alice reveals $\mathbf{c}$ to Bob;

Reveal phase

1. Alice reveals $\mathbf{x}$ and $\mathbf{m}$ to Bob;

2. Bob accepts $\mathbf{m}$ if

   (a) $\mathbf{z} \in A_\epsilon^{(n)}(Z|\mathbf{x})$, where $A_\epsilon^{(n)}(Z|\mathbf{x}^n)$ is the set of sequences $\mathbf{z}^n$ weakly conditionally typical with $\mathbf{x}^n$,

   (b) $g(\mathbf{x}) \oplus \mathbf{c} = \mathbf{m}$.

Lemma 6.5. *Protocol 2 is ϵ-concealing.*

Proof. The proof is identical to the proof of Lemma 6.1 by introducing a noiseless channel between Alice and Charlie, and will be omitted. □

Lemma 6.6. *Protocol 2 is δ-binding.*

Proof. The proof is simpler than the proof of Lemma 6.3, since the Hamming distance is replaced by the Euclidean distance and since no binning procedure is required. In the following, we denote the Euclidean distance between two continuous vectors $\mathbf{x}^n$ and $\mathbf{x}_0^n$ by $\|\mathbf{x}^n - \mathbf{x}_0^n\|$. Let $\delta > 0$ and let σ_Q be the minimum Euclidean distance between two scalar

symbols x and x_0 such that $p(x) \neq 0$ and $p(x_0) \neq 0$. Let $\sigma_Q > \mu > 0$. For any $\mathbf{x}^n \in \mathbb{R}^n$ define

$$E_\mu(\mathbf{x}^n) = \left\{\mathbf{x}_0^n : \|\mathbf{x}_n - \mathbf{x}_0^n\| < \mu\sqrt{n}\right\} \tag{6.21}$$

$$\text{and} \qquad E_\mu^*(\mathbf{x}^n) = E_\mu(\mathbf{x}^n) \setminus \{\mathbf{x}^n\}. \tag{6.22}$$

For any $\mathbf{x}^n, \mathbf{x}_0^n \in \mathbb{R}^n$, if $\mathbf{x}_0^n \notin E_\mu(\mathbf{x}^n)$, we know from [64, Proposition 3] that $\exists n_1 \in \mathbb{N}$ such that

$$\forall n \geq n_1 \quad \mathrm{P}\Big[\mathbf{Z}^n \in A_\epsilon^{(n)}(Z|\mathbf{x}_0^n)\Big] < \delta. \tag{6.23}$$

By choosing $\mu < \sigma_Q$, is it clear that Alice cannot choose a sequence $\mathbf{x}_0^n \in E_\mu^*(\mathbf{x}^n)$, which concludes the proof. □

It is interesting to highlight that the binding property is solely enforced by the use of a finite set of symbols $\mathcal{X} \subset \mathbb{R}$, and one can legitimately wonder if Protocol 1 could not be simplified to avoid the requirement of a binning scheme. In fact, instead of sending n i.i.d realization of the same random variable through the discrete memoryless channel, Alice could send a sequence $\mathbf{x}^n$ chosen at random among a set of sequences with known minimum distance; however, achieving the bit commitment capacity would require imposing a structure on the set of sequences, which is likely to be as difficult in practice as the construction of wiretap codes.

Lemma 6.7. *Protocol 2 is η-correct.*

Proof. The proof is identical to the proof of Lemma 6.4 and will be omitted. □

For a fixed choice of $\mathcal{X}$ and $p(x)$, the commitment rate achievable with Protocol 2 is $k/n \approx H(X|Z)$ in the limit of large n. This rate can be made as large as desired by increasing the size of the set $\mathcal{X}$, which is in accordance with the fact that the commitment capacity of a Gaussian channel is infinite, as shown in [64].

From a practical perspective, it is convenient that bit commitment does not require the use of codes ensuring *simultaneously* the concealing and binding properties. By highlighting

the connection between bit commitment and secret key agreement, we have provided non-constructive protocols for both non-redundant discrete memoryless channels and Gaussian channels based on privacy amplification and Slepian-Wolf coding. Although these protocols are in essence non-constructive, they provide useful guidelines for developing practical bit commitment schemes, as we discuss in the next section.

6.3 Practical commitment schemes

6.3.1 Bit commitment over binary memoryless channels

As it is often the case in information theory, the random binning argument used in the previous section is convenient for analysis purposes but it fails to translate directly into practical coding schemes. Nevertheless, it should be noted that the only non-constructive part in our protocols concerns the existence of "good" binning codes, since the existence of universal$_2$ families of hash functions for privacy amplification is a widely accepted cryptographic premise [10]. Furthermore, the construction of binning codes is arguably better understood than the construction of wiretap codes required by the previous works [63, 64], and it is known that there exists "good" structured *algebraic* binning schemes.

The desired property of structured binning schemes is illustrated in Figure 35, where the smaller balls represent sequences $\mathbf{x}$ and the different colors represent different bins. The bins should be assigned in such a way that no two identical bins can be found in a Hamming ball of radius σ.

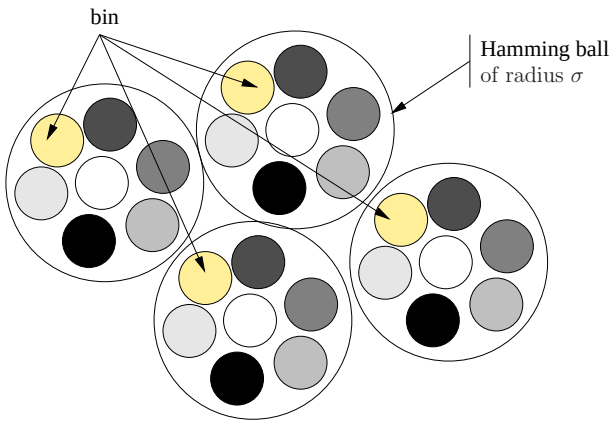


Figure 35. Binning through coset codes.

As shown in the following example, this property can be obtained by binning according

to the cosets of some parity-check codes. This procedure was first suggested by Wyner in [46]. Given a parity-check code with $(n-k) \times n$ parity check matrix $\mathbf{H}$, a coset code $C_{\mathbf{s}}$ is defined as $\left\{\mathbf{x} : \mathbf{x}\mathbf{H}^T = \mathbf{s}\right\}$. There are 2^{n-k} cosets and each of these is a shifted version of the original code, hence conserving his distance properties. One should notice that the coset coding scheme imposes a tradeoff between the practical value n and the achievable commitment rate. In fact, achieving a high commitment rate requires the use of a high rate code, while operating with a reasonable value of n requires a code with a linearly growing minimum distance δn with δ close to 1.

Example 6.1 (Binary symmetric channel.)**.** *Let us consider a binary symmetric channel with cross-over probability $p = 0.4$, whose bit commitment capacity is $h(p) \approx 0.97$. Gallager showed in [66] that for large k and large n the ensemble of (n, j, k) regular Low-Density Parity-Check (LDPC) codes has a typical minimum distance ratio of δ_{jk} such that $1 - R \approx h(\delta_{jk})$, and that most codes in the ensemble have a minimum distance close to or greater than $\delta_{jk} n$. For instance, any $(n, 5, 6)$ LDPC code is likely to have a minimum distance greater that $0.25n$ and has a rate $R \approx 0.2$. One can therefore hope to achieve a bit commitment rate of $h(0.4) - 0.8 \approx 0.17$.*

6.3.2 Bit commitment over Gaussian channels

It is clear from the properties of Protocol 2 that designing a high rate practical scheme only relies on the construction of a good quantizer. Yet, even a particularly simple quantization allows to achieve reasonable rates.

Example 6.2 (Gaussian channel.)**.** *Let us consider a Gaussian channel with noise variance 1 and power constraint $P = 1$, and a 8-PAM constellation of uniformly spaced symbols*

$$\mathcal{X} = \{-7d, -5d, -3d, -d, d, 3d, 5d, 7d\} ,$$

where the inter-symbol distance $2d$ is optimized to satisfy the power constraint P. If Alice transmits n symbols chosen uniformly at random in $\mathcal{X}$ during the commit phase of Protocol 2, one can easily check that $I(X;Z) \approx 0.5$ bits, and therefore we can hope to achieve a commitment rate of approximately 2.5 bits.

CHAPTER 7
SERVER-CLIENT ARCHITECTURES BASED ON WIRETAP CODES

Based on the previous chapters, it may appear that physical-layer security is indeed a specificity of the physical layer; however, as we illustrate in this final chapter, some of the tools developed for the physical layer can find applications in upper layers of a communication protocol stacks. In particular, we show that the wiretap channel model naturally appears in certain client-server network communications subject to *active* attacks. In this context, the secrecy capacity admits to a different meaning, as it can be used as an optimization metric rather than simply identify maximum secure communication rates.

7.1 Client-server networks under attack

Among the many attacks which infect the Internet and aim at disrupting packet traffic, the most disruptive ones are host compromise and Denial of Service (DoS) attacks. In a host compromise scenario the attacker exploits weaknesses of a node to gain control of the host. Once the host is compromised attacks are launched on the neighboring nodes. On the other hand, in a DoS attack an attacker tries to direct a large amount of bogus traffic to a susceptible node, with the intention of consuming a large amount of bandwidth and rendering the node unable to service legitimate traffic. In a more harmful manner, host compromise and DoS attacks can be combined to cause a Distributed DoS attack (DDoS). The frequency and magnitude of DoS attacks have been steadily increasing for the last couple of years [67]. For instance, there has been a significant number of DoS attacks on popular e-commerce websites and governmental websites in 2000 and 2001, and more recently these attacks have targeted the root Domain Name Servers (DNSs) and the DNS backbone network. DoS attacks can target either Transmission Control Protocol/Internet Protocol (TCP/IP) [68], overlay network [69] or application layers [70].

Both host compromise and DoS attacks are often caused by *worms*, which are self-replicating computer programs specifically designed to damage the network. Worms may have various types of spreading behavior over the network that are intended to compromise a host. *Random* scanning simply selects a victim node at random. In contrast, *topological*

scanning [71] uses address information contained in the victims for propagation. With the advent of IPv6 whose address space is much larger than that of IPv4, topological scanning or other more sophisticated types of scanning methods may supersede random scanning as a method of infection. Specifically, topological worms exploit "susceptible neighboring" nodes, and then infect those neighbors. Vital information such as routing tables, email addresses, a list of peers, and Uniform Resource Locators (URLs) about other hosts, can be exploited to identify new potential victims.

Most of the previous research in countering DoS attacks requires significant changes to the existing network infrastructure (see [72, 73] and references therein) or collaboration across ISPs [74], and hence may be farfetched or impractical. Several DoS attacks detection schemes and mitigating mechanisms have been proposed, using filtering [75], Honeypots [76], or specialized overlay nodes that have capabilities to resist and survive attacks [70, 77, 78]. Among the recent works based on overlay network for protection against DoS attacks are a survivable DoS-resistant overlay network called "rewire" [78] and a generic DDoS protection service called "OverDoSe" [79].

Despite all these protective measures, the Internet is intrinsically an open network and will never be completely protected from these attacks. Therefore designing network architectures capable of mitigating the impact of these unavoidable attacks has become a crucial issue. Here, we specifically consider the effect of these attacks upon the design of resilient and secure *client-server* architectures. The typical client-server architecture of interest is shown in Figure 36. In order to communicate with a distant *server*, a *client* sends packets to *access points*, which then route these packets through an *interconnection network* and to a set of *targets* directly connected to the server.

The clients have a secure protocol to access the content or service from the servers, and they have knowledge of the location of several access points. Therefore, only legitimate clients can talk to the access points. Each access point knows the path to certain targets and forwards the traffic from clients to servers and *vice-versa* once the authentication protocol is established. Access points forward traffic to multiple targets to forward the traffic, and targets can serve multiple access points for service responses. Typically, the ratio of access points per target is roughly about 10 and access points may choose to forward traffic only to

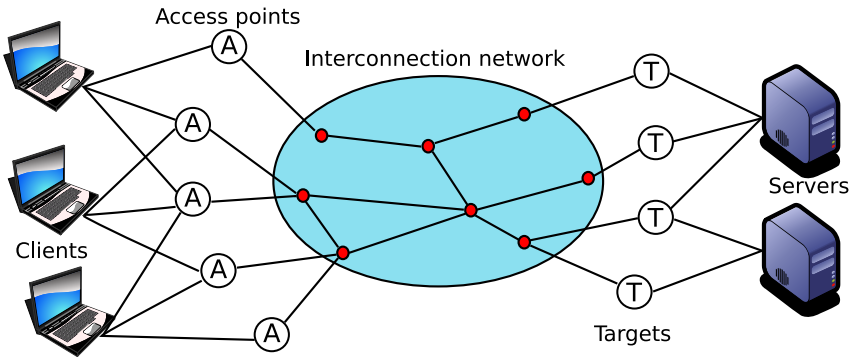


Figure 36. Client-server architecture

a subset of targets. The interconnection network is assumed to employ the Internet Protocol (IP) network. Routers in the IP network forward the packets and do not distinguish between normal and secure traffic from the access points. The server is usually a single node (*e.g.* a cluster of several computers) and the service offered cannot be replicated for security purpose (dynamically changing content). In some cases, servers with the same content may also be distributed geographically to prevent service denials in the case of failure/attack of one of them. Servers use packet filters to ensure that only traffic from the clients is received.

In [80], Bu *et al.* proposed to design attack-resistant client-server architectures that are both *resilient*, *i.e.* able to provide an alternative communication path should one of them be disrupted, and *secure*, *i.e.* able to reduce or prevent the number of compromised nodes. Clearly, these two requirements are contradictory since resiliency calls for an increased connectivity between targets and access point, while security calls for a limited connectivity. In order to measure quantitatively the resistance of a given assignment of access points to targets, Bu *et al.* introduced a metric called the *blocking probability* which is defined as the percentage of client requests that cannot be relayed to the servers in case of attacks.

In this chapter, we consider a similar approach but we include an additional *secrecy* constraint. Specifically, we want to ensure that a malicious attacker hacking the packet information at compromised nodes is unable to retrieve information about the content being exchanged between client and server. Rather than using traditional cryptographic tools to encrypt information contained in the packet, we also want to exploit the fact that the attacker only gets parts of the packets sent by the client.

7.2 Notation and attacker model

In the following, we use notations similar to those of [80]. The set of access points and the set of targets are denoted by $\mathcal{A}$ and $\mathcal{T}$, respectively. Furthermore, we let $n_a = |\mathcal{A}|$ and $n_t = |\mathcal{T}|$. The assignment relationship between the access points and the targets is specified by an $n_a \times n_t$ matrix $\mathbf{M} = \{m_{ij}\}$ called the *assignment matrix*, such that $m_{ij} = 1$ if and only if the access point i is connected to the target j. The set of access points connected to a target j is denoted by $\mathcal{A}(j)$ and the set of targets connected to an access point i is denoted by $\mathcal{T}(i)$. The assignment matrix can be conveniently represented as a bipartite graph where each node is either an access-point or a target, as shown in Figure 37. In what follows, we adopt this representation and refer to the links between access points and targets as "edges".

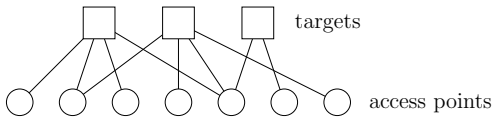


Figure 37. Bipartite graph representation of an assignment matrix.

We assume that attacking the interconnection network would not be fruitful from the attacker point of view. Namely, if an attacker compromises a random host or launches a DoS attack against a random node in a sufficiently large interconnection network, the probability that this attack actually disrupts the traffic between access points and targets is very low. We note that if an overlay network is used for client-server services [69, 79, 78] this probability might not be negligible anymore. We do not consider this scenario here, and leave its analysis for future work. We focus only on DoS and host compromise attacks on the access points and targets. For $k \in \{\mathcal{A}, \mathcal{T}\}$, we denote the event of a node being compromised by C_k, and by D_k if it is subject to a DoS attack.

We also make the following additional assumptions:

- The tools used to launch a compromise attack or a DoS attack on an access point are usually different, hence the corresponding events are assumed to be independent:

$$\forall\, i \in \mathcal{A} \quad \mathrm{P}[C_i, D_i] = \mathrm{P}[C_i]\mathrm{P}[D_i].$$

- To improve the resiliency and to reduce the probability of correlated failures, access points are often placed in different domains. In general, the access points do not have information about each other and as a result we assume that the probabilities that access points are compromised or under DoS attack are independent:

$$\forall\,(i,j) \in \mathcal{A}^2 \quad i \neq j \Rightarrow \begin{cases} \mathrm{P}[C_i, C_j] = \mathrm{P}[C_i]\mathrm{P}[C_j], \\ \mathrm{P}[C_i, D_j] = \mathrm{P}[C_i]\mathrm{P}[D_j], \\ \mathrm{P}[D_i, D_j] = \mathrm{P}[D_i]\mathrm{P}[D_j]. \end{cases}$$

- We assume that the traffic from access points to targets is secure, but the publicly known assignment is a vulnerability that can be exploited for topological scanning. Since the location of access points is publicly known, the probability of launching an attack directly on a target is substantially smaller than the probability of first compromising an access point. Therefore, we make the simplifying assumption that targets are never directly attacked:

$$\forall\, j \in \mathcal{T} \quad \mathrm{P}\big[C_j | \forall\, i \in \mathcal{A}(j)\ \bar{C}_i\big] = \mathrm{P}\big[D_j | \forall\, i \in \mathcal{A}(j)\ \bar{C}_i\big] = 0.$$

  Moreover, we assume that an attack on a target depends only on the attacks on its connected access points:

$$\forall\,(i,j) \in \mathcal{A} \times \mathcal{T} \quad m_{i,j} = 0 \Rightarrow \begin{cases} \mathrm{P}[C_i, C_j] = \mathrm{P}[C_i]\mathrm{P}[C_j], \\ \mathrm{P}[C_i, D_j] = \mathrm{P}[C_i]\mathrm{P}[D_j], \\ \mathrm{P}[D_i, C_j] = \mathrm{P}[D_i]\mathrm{P}[C_j], \\ \mathrm{P}[D_i, D_j] = \mathrm{P}[D_i]\mathrm{P}[D_j], \end{cases}$$

- Finally, we simplify the problem by adopting an Internet Service Provider (ISP) centric approach, where all the nodes are assumed to have the same probability of attack.

$$\begin{aligned} \forall\,(i,j) \in \mathcal{A} \times \mathcal{T}\ \text{s.t.}\ m_{ij} = 1 \quad \mathrm{P}[C_j|C_i] &\triangleq \eta_c, \\ \forall\,(i,j) \in \mathcal{A} \times \mathcal{T}\ \text{s.t.}\ m_{ij} = 1 \quad \mathrm{P}[D_j|C_i] &\triangleq \eta_d, \\ \forall\, i \in \mathcal{A} \quad \mathrm{P}[C_i] &\triangleq \rho_c, \\ \forall\, i \in \mathcal{A} \quad \mathrm{P}[D_i] &\triangleq \rho_d. \end{aligned}$$

Another approach would be to consider nodes in very different domains and potentially with different levels of protection. This latter situation is more difficult to investigate analytically, and is only addressed in Section 7.5.2.

The consequences of the attacks are twofold. First, a packet is *blocked* whenever it reaches a node that is compromised or under DoS. Second, a packet is *read* whenever it is routed through a compromised node. We assume that attackers can launch distributed attacks on a large number of nodes and that they have the ability to collude, *i.e.* they can coordinate their attacks and share their information about the packets intercepted. For instance, if a packet reaches a compromised node that is also under DoS, then the packet is still read and all read packets can be sent to a centralized attacker.

Given an assignment matrix $\mathbf{M}$, let $p_b(\mathbf{M})$ and $p_r(\mathbf{M})$ denote the probability of a packet being blocked and read, respectively, and let us assume that each packet is sent uniformly at random to one of the access points. A packet is blocked either if the access point is under attack or if all neighboring targets are attacked from other access points. Likewise, a packet is read if the access point is compromised or if one of the connected targets is compromised from another access point. Therefore, we obtain:

$$\begin{aligned} p_b(\mathbf{M}) = \frac{1}{n_a} \sum_{i \in \mathcal{A}} & \left[\mathrm{P}[C_i] + \mathrm{P}[D_i] - \mathrm{P}[C_i]\mathrm{P}[D_i] \right. \\ & \left. + (1 - \mathrm{P}[D_i])\,(1 - \mathrm{P}[C_i])\,\left(\mathrm{P}\left[\forall k \in \mathcal{A}(i)\, D_k | \bar{C}_i \bar{D}_i\right]\right)\right], \\ p_r(\mathbf{M}) = \frac{1}{n_a} \sum_{i \in \mathcal{A}} & \left[\mathrm{P}[C_i] + (1 - \mathrm{P}[D_i])\,(1 - \mathrm{P}[C_i])\,\left(\mathrm{P}\left[\exists k \in \mathcal{A}(i) \,\text{s.t.}\, D_k | \bar{C}_i \bar{D}_i\right]\right)\right]. \end{aligned}$$

In general, it is not possible to derive closed-form expressions of $p_r(\mathbf{M})$ and $p_b(\mathbf{M})$ without further assumptions on the structure of the assignment. Moreover, computing the exact values numerically is computationally intensive. In practice, one has to resort to approximation algorithms similar to those of [80] to obtain an estimation of these parameters. In this chapter, numerical values of $p_r(\mathbf{M})$ and $p_b(\mathbf{M})$ are obtained through Monte Carlo simulations.

7.3 Client-server communication over a wiretap channel

7.3.1 Packet coding with scrambled codewords

The notions of blocking probability and reading probably given in the previous sections are naturally defined packet-wise. However, to achieve the promise of physical-layer security, and eventually exploit wiretap codes, we need a packet coding scheme such that these notions are translated to the bit level for each codeword. Such a scheme is illustrated in Figure 38.

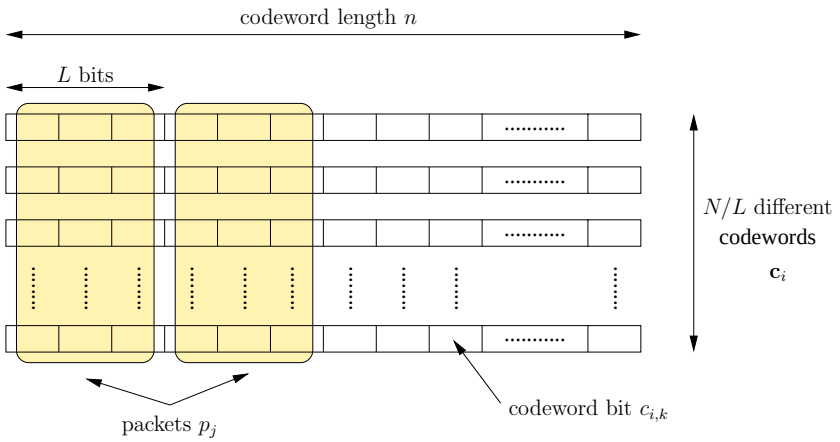


Figure 38. Packet encoding scheme with scrambled codewords.

Let us assume that each packet carries N message bits, and let $L \leq N$ be an integer that divides N. To generate packet bits, the client first generates N/L independent messages of length k bits, $\mathbf{v}_i = (v_{i1}, \ldots, v_{ik})$ for $i \in \{0, \ldots, N/L - 1\}$. He then encodes each of them into a codeword of length n bits, $\mathbf{c}_i = (c_{i1}, \ldots, c_{in})$ for $i \in \{0, \ldots, N/L - 1\}$, with an error correcting code to be specified later. The coded bits of each codeword $\mathbf{c}_i$ are finally grouped into subblocks $\mathbf{b}_{ik} = \{c_{ij}, j \in \{kL, \ldots, (k+1)(L-1)\}\}$ of length L, and each subblock is assigned to a different packet. Hence, the jth packet ($j \in \{0, \ldots, \lceil n/L - 1 \rceil\}$) contains the subblocks $\mathbf{p}_j = \{\mathbf{b}_{ij} : i \in \{0, \ldots, N/L - 1\}\}$. Each packet j is finally routed uniformly at random to an access point.

The overall effect of this encoding is simply to scramble the bits within the same codeword among different packets. Therefore, from the perspective of a codeword $\mathbf{c}_i$, the probability of each codeword subblock $\mathbf{b}_{ij}$ being blocked or read is

$$\begin{aligned} \mathrm{P}[\mathbf{b}_{ij}\,\text{blocked}] &= \mathrm{P}[\mathbf{p}_j\,\text{blocked}] = p_b, \\ \mathrm{P}[\mathbf{b}_{ij}\,\text{read}] &= \mathrm{P}[\mathbf{p}_j\,\text{read}] = p_r. \end{aligned}$$

The delay introduced by this coding scheme is in the order of n/L packets since the server needs to buffer n/L packets before being able to decode N/L messages. Consequently, this approach may not perform well in its current form for delay-sensitive services, however the coding scheme can be applied to situations where high-throughput dedicated links are available between access points and targets. Coding the messages across packets also introduces some overhead, which might decrease the communication throughput. The characterization of overhead is further discussed in Section 7.4. In the remaining of this chapter, we assume that $L = 1$ unless otherwise specified.

7.3.2 Equivalent wiretap channel model

As we describe next, the coding scheme of Section 7.3.1 allows us to model the communications between the client and the server as a particular instance of communications over a wiretap channel. In fact, the bits of each codeword, from the server perspective, are obtained as if they were transmitted through a binary erasure channel with erasure probability $\epsilon_1 = p_b(\mathbf{M})$. Likewise, from the attacker perspective, the bits are read as if they were transmitted through a binary erasure channel with probability $\epsilon_2 = 1 - p_r(\mathbf{M})$. This holds even though our model assumes the presence of an active attacker launching various attacks on the access points and targets. An equivalent formulation of the problem is shown in Figure 39.

It should be noted that the assumptions of Section 7.2 do not guarantee that these two erasure channels are independent in general. Therefore, computing the exact secrecy capacity of this general broadcast channel is a non-trivial problem. Nevertheless, we still have the following proposition:

Proposition 7.1. *Given a network assignment matrix* $\mathbf{M}$*, if* $\epsilon_2 \geq \epsilon_1$*, or equivalently*

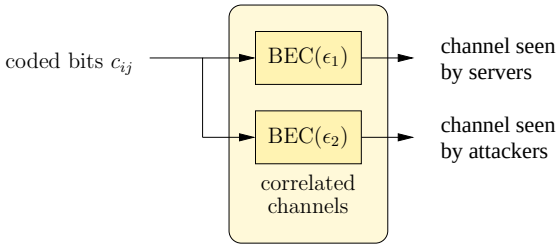


Figure 39. Equivalent wiretap channel model.

$p_r(\mathbf{M}) + p_b(\mathbf{M}) \leq 1$, *then the secrecy capacity of the client-server architecture is given by*

$$C_s(\mathbf{M}) = \epsilon_2 - \epsilon_1 = 1 - p_r(\mathbf{M}) - p_b(\mathbf{M}). \tag{7.1}$$

Proof. The main channel BEC(ϵ_1) between the client and the server, and the wiretap channel BEC(ϵ_2) between the client and the attacker are binary erasure channels; therefore, for any input probability distribution ($\mathrm{P}[X = 0] = p, \mathrm{P}[X = 1] = 1 - p$) of the binary variable X, we have [1]

$$\begin{aligned} I(X;Y) - I(X;Z) &= (1-\epsilon_1)h(p) - (1-\epsilon_2)h(p), \\ &= (\epsilon_2 - \epsilon_1)h(p). \end{aligned}$$

Since $h(p)$ is a convex-$\cap$ function of p, $I(X;Y) - I(X;Z)$ is also a convex-$\cap$ function of p if and only if $\epsilon_2 \geq \epsilon_1$. If this is the case, then by [81, Theorem 2] the main channel BEC(ϵ_1) is less noisy than the wiretap channel BEC(ϵ_2). The result finally follows by [81, Theorem 3]. □

This proposition provides only a partial characterization of the secrecy capacity, since we do not know the value of the secrecy capacity when $\epsilon_2 < \epsilon_1$. However, as we illustrate numerically in Section 5.2, this characterization is sufficient to analyze most situations of interest.

Note that these results can be generalized to the case where codewords are scrambled in blocks of length $L > 1$. The equivalent wiretap channel is then a 2^L-ary erasure wiretap channel, and if $\epsilon_2 \geq \epsilon_1$ the secrecy capacity is given by

$$C_s(\mathbf{M}) = (\epsilon_2 - \epsilon_1)\, L = (1 - p_r(\mathbf{M}) - p_b(\mathbf{M}))\, L. \tag{7.2}$$

7.4 Assignment optimization based on secrecy capacity

Unlike a traditional wiretap model, where the secrecy capacity depends on the fixed *a priori* properties of the underlying channel, the secrecy capacity of the equivalent wiretap channel $C_s(\mathbf{M})$ depends on the chosen assignment matrix $\mathbf{M}$ through the probabilities $p_r(\mathbf{M})$ and $p_b(\mathbf{M})$. Therefore, rather than being viewed only as the fundamental limit on secure communication rates, the secrecy capacity should also be understood as a metric quantifying the quality of a specific assignment matrix. Hence, the design of the assignment matrix $\mathbf{M}$ can be optimized with the secrecy capacity as the objective function. From an operational perspective, maximizing the secrecy capacity of an assignment is equivalent to minimizing the rate penalty inflicted by the packet coding strategy.

As we pointed out earlier, there exists no closed-form expression for the secrecy capacity $C_s(\mathbf{M})$ in general. Consequently, it is difficult to characterize the optimal assignment with respect to the secrecy capacity without resorting to numerical approximations of $C_s(\mathbf{M})$ and specific optimization algorithms. Nevertheless, in the remainder of this section, we derive upper and lower bounds of the secrecy capacity. The derivations of these limits provide useful insight on the practical construction of assignment matrices.

7.4.1 Bounds on secrecy capacity

Let us introduce the probability of an access point being under attack $\rho_a = \rho_c + \rho_d - \rho_c\rho_d$ and the probability of a target being successfully attacked from an access point $\eta_a = \eta_c + \eta_d - \eta_c\eta_d$. An upper bound of the secrecy capacity $C_s^{(u)}$ is given by the following proposition.

Proposition 7.2. *The secrecy capacity of the optimal assignment is bounded from above as follows:*

$$\max_{\mathbf{M}} C_s(\mathbf{M}) \leq C_s^{(u)} = 1 - \rho_a - \rho_c. \tag{7.3}$$

Proof. Clearly, ignoring the effect of the attacks on the targets upon $p_b(\mathbf{M})$ and $p_r(\mathbf{M})$ provides valid lower bounds of these probabilities. By definition,

$$\forall\, \mathbf{M} \qquad p_b(\mathbf{M}) \geq \rho_a \qquad \text{and} \qquad p_r(\mathbf{M}) \geq \rho_c,$$

and the results follows directly from the expression for the secrecy capacity. □

Notice that our upper bound is currently independent of any specific assignment and therefore, it might not be tight. In future work, we will address this issue and develop an upper bound taking into account network parameters.

To obtain a lower bound $C_s^{(l)}$ of the secrecy capacity, we can compute the secrecy capacity of any specific assignment matrix $\mathbf{M}$. By construction, this secrecy capacity is achievable and therefore, it provides a valid lower bound. We introduce the following definitions to make the characterization of the lower bound tractable analytically.

Definition 7.1. *A* cluster *is a set of access points connected to a single target, see Figure 40. A cluster containing exactly k access points is called a k-cluster. A* clustered *assignment is an assignment containing only clusters.*

Definition 7.2. *An assignment is* balanced *if the number of access points connected to any two different targets differs at most by 1. Formally,*

$$\forall\, i, j \in \mathcal{T} \quad ||\mathcal{A}(i)| - |\mathcal{A}(j)|| \leq 1.$$

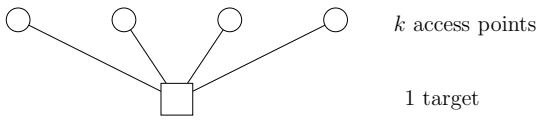


Figure 40. A k-cluster.

The exact blocking and reading probabilities of a k-cluster are then an immediate consequence of the previous definition.

Proposition 7.3. *For a k-cluster, the blocking probability $p_b^{(k)}$ and the reading probability $p_r^{(k)}$ are given by*

$$p_b^{(k)} = \rho_a + (1-\rho_a)\Big[1-(1-\rho_c\eta_a)^{k-1}\Big], \tag{7.4}$$

$$p_r^{(k)} = \rho_c + (1-\rho_a)\Big[1-(1-\rho_c\eta_c)^{k-1}\Big]. \tag{7.5}$$

Consequently, $p_b^{(k)}$ and $p_r^{(k)}$ are increasing function of k and $\lim_{k\to\infty} p_b^{(k)} = \lim_{k\to\infty} p_r^{(k)} = 1$.

The secrecy capacity of a balanced clustered assignment follows readily and provides the desired lower bound.

Proposition 7.4. *Let* $k = \lfloor \frac{n_a}{n_t} \rfloor$ *be the smallest cluster size of the balanced clustered assignment; then*

$$\max_{\mathbf{M}} C_s(\mathbf{M}) \geq C_s^{(l)} \quad = \quad \frac{kn_t}{n_a}\Big(1 - p_b^{(k)} - p_r^{(k)}\Big) + \frac{n_a - kn_t}{n_a}\Big(1 - p_b^{(k+1)} - p_r^{(k+1)}\Big) \quad (7.6)$$

When all clusters are of equal size (*i.e.* n_t/n_a is an integer), then this expression simplifies to

$$\begin{aligned} C_s^{(l)} &= 1 - \rho_a - (1-\rho_a)\Big[1 - (1 - \rho_c\eta_a)^{k-1}\Big] - \rho_c - (1-\rho_a)\Big[1 - (1-\rho_c\eta_c)^{k-1}\Big], \\ &= C_s^{(u)} - (1-\rho_a)\Big[2 - (1-\rho_c\eta_a)^{k-1} - (1-\rho_c\eta_c)^{k-1}\Big]. \end{aligned}$$

The latter equation clearly shows the secrecy rate penalty inflicted by the attacks on targets.

Figure 41 shows these two bounds for several sets of parameters $(\rho_c, \rho_d, \eta_c, \eta_d)$ as a function of the ratio n_t/n_a. As the values of $(\rho_c, \rho_d, \eta_c, \eta_d)$ increase, assignments become more vulnerable and therefore, their secrecy capacity decreases. Also, as expected, as the ratio n_t/n_a decreases, the size of the clusters in a balanced clustered assignment increases and therefore, a target is more likely to be blocked.

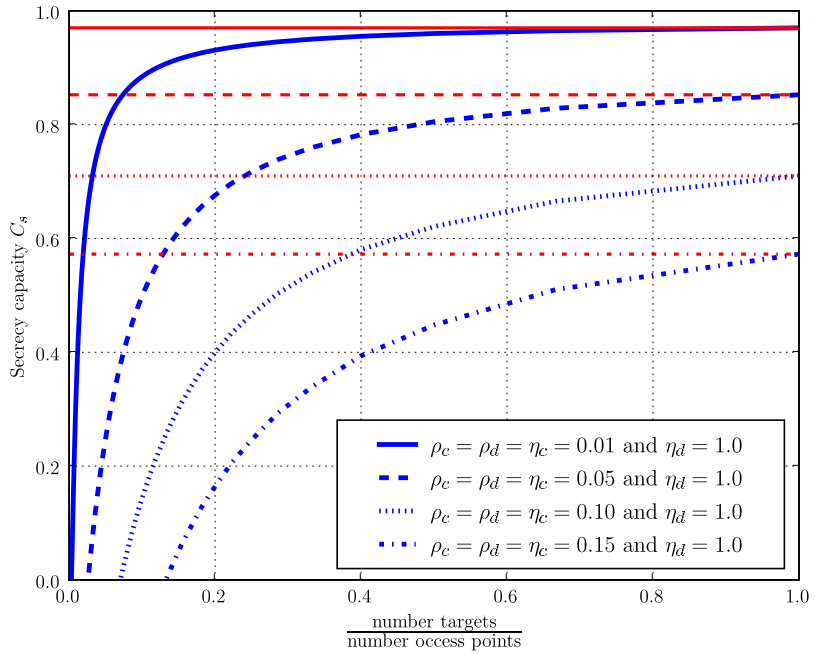


Figure 41. Lower and upper bounds of secrecy capacity. The thick lines correspond to lower bounds, while the thin ones correspond to upper bounds.

Notice that the gap between the upper and lower bounds becomes more significant as the ratio n_t/n_a decreases; however, in practice, it might be difficult to find an assignment whose secrecy capacity is significantly greater than our lower bound. In fact, although a balanced clustered assignment is one of the simplest assignment that one can implement, the following theorem shows that it is "nearly optimal" in some situations.

Theorem 7.1. *If the size k of the clusters in a balanced clustered assignment satisfies the following (sufficient) conditions*

$$1 \leq (1-\rho_c\eta_a)^{k-1}(1-(k+1)\rho_c\eta_a) + (1-\rho_c\eta_c)^{k-1}(1-(k+1)\rho_c\eta_c), \quad (7.7)$$

$$k+1 \leq \min\left(\frac{2-\rho_c\eta_a}{\rho_c\eta_a}, \frac{2-\rho_c\eta_c}{\rho_c\eta_c}\right), \quad (7.8)$$

then the assignment is "nearly optimal", in the sense that no greedy algorithm operating edge-by-edge can improve its secrecy capacity.

Proof. (sketch) A greedy algorithm operating edge-by-edge attempts to improve the secrecy capacity of an assignment by modifying one edge at a time. Therefore, we need only to consider three operations on edges: *removing* an edge, *adding* and edge and *rewiring* an edge. The theorem is then a consequence of the following propositions, whose proofs are relegated to the appendix.

Proposition 7.5. *If condition (7.7) is fulfilled, then removing an edge cannot improve the secrecy capacity of a balanced clustered assignment.*

Proposition 7.6. *If condition (7.8) holds, then rewiring a link cannot improve the secrecy capacity of a balanced clustered assignment.*

Proposition 7.7. *Adding an extra link cannot improve the secrecy capacity of a balanced clustered assignment.*

□

To model practical situations, we can assume that $\rho_c \ll 1$ and $\rho_d \ll 1$, and conditions (7.7-7.8) reduce to an upper bound of the size of the clusters k.

In [80], the metric used for assignment optimization is the blocking probability alone. Our analysis still holds in this case and yields bounds of the minimum blocking probability blocking probability.

$$\rho_a \leq \min_{\mathbf{M}} p_b(\mathbf{M}) \leq 1 - (1-\rho_a)(1-\rho_c\eta_a)^{k-1}\left(\frac{kn_t}{n_a} - \frac{n_a - kn_t}{n_a}(1-\rho_c\eta_a)\right), \tag{7.9}$$

A "near optimality" result, similar to that of Theorem 7.1, also holds.

Theorem 7.2. *If the size k of the clusters in a balanced clustered assignment satisfies the following conditions*

$$1 \leq (1-\rho_c\eta_a)^{k-2}(1-k\rho_c\eta_a) + (1-\rho_c\eta_c)^{k-2}(1-k\rho_c\eta_c), \tag{7.10}$$

$$k \leq \frac{1}{\rho_c\eta_a}, \tag{7.11}$$

then the assignment is "nearly optimal", in the sense that no greedy algorithm operating edge-by-edge can decrease its blocking probability.

Proof. See appendix. □

7.4.2 Overhead of packet coding scheme

The communication scheme described in Section 7.3.1 requires the introduction of coding across packets and the overhead introduced by the codes reduces the communication throughput between access points and targets. Yet, a fair estimation of this overhead should be made with respect to the capacity of the equivalent main erasure channel BEC(ϵ_1). In fact, during the communication some information redundancy is required to compensate for the lost packets, regardless of the error-control system employed (coding, re-transmit, etc.).

The capacity of a BEC(ϵ_1) is $C_1 = 1 - p_b(\mathbf{M})$, therefore the overhead is at least $\Delta_{min} = p_b(\mathbf{M})$ bits per packet bit. The proposed coding scheme is based on codes with rates arbitrarily close to the secrecy capacity $C_s = 1 - p_b(\mathbf{M}) - p_r(\mathbf{M})$, hence the overhead is at least

$$\Delta = p_b(\mathbf{M}) + p_r(\mathbf{M}) \quad \text{bits per packet bits.}$$

Figure 42 shows the increase in overhead ($p_r(\mathbf{M})/p_b(\mathbf{M})$ in percent) inflicted by the use of wiretap codes when using a balanced clustered assignment in different scenarios.

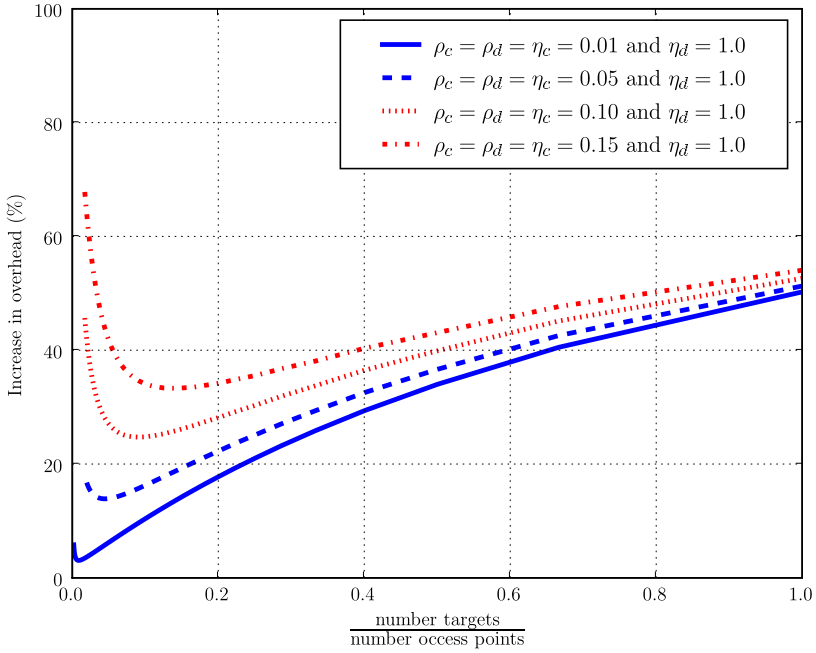


Figure 42. Increase in overhead inflicted by wiretap codes.

Table 3. Set of parameters used in simulations.

	Vulnerable architecture	Robust architecture
ρ_c	0.04	0.02
ρ_d	0.04	0.02
η_c	0.04	0.02
η_d	1.0	0.2
Conditions (7.7-7.8)	$k \leq 21$	$k \leq 169$

As expected, more vulnerable architectures require more overhead since their secrecy capacity is lower. The overhead tends to decrease with the ratio n_t/n_a as the blocking probability increases at a faster pace than the reading probability. It is also interesting to notice that all curves exhibit a minimum, in the low n_t/n_a regime. In all cases, Figure 42 clearly shows that information-theoretic secrecy has a significant cost in terms of overhead.

7.5 Simulation results

7.5.1 Near-optimality of balanced clustered assignments

We start this section by numerically analyzing the lower and upper bounds derived in the previous section. In our simulations, we consider two situations whose parameters are described in Table 3.

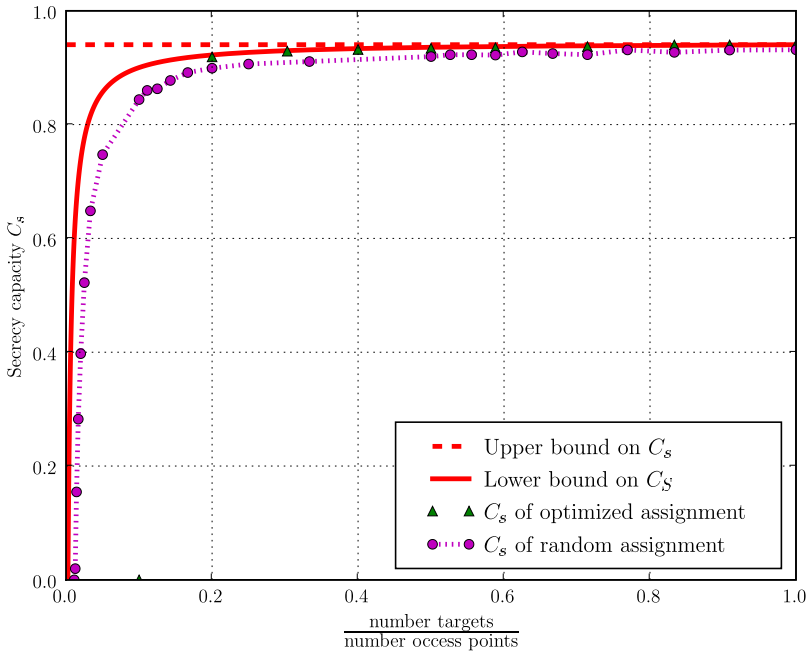


Figure 43. Secrecy capacities of several assignments for a robust client-server architecture.

The first set of parameters represents a vulnerable architecture where access points are quite likely to be attacked and where a compromised access point is used to launch a DoS attack with probability 1 on the connected targets. On the other hand, the second set of parameters describes a more robust system, where a compromised access point does not have as much impact on the neighboring targets. The secrecy capacities of several network assignments in these two scenarios are shown in Figures 43 and 44 as a function of the ratio n_t/n_a. All results are obtained assuming $n_t = 10$.

In addition to the bounds derived in Section 7.4, we also plot the secrecy capacity of a random assignment, where each edge between an access point and a target is created according to a Bernoulli 1/2 probability distribution[1], and the secrecy capacity of an optimized assignment, where the positions of the edges are the result of a greedy optimization (see [80] for details).

For a robust client-server architecture, the choice of assignment has little effect and a random assignment performs almost as well as an optimized one. On the other hand,

[1]We also ensure that each access point is connected to at least one target.

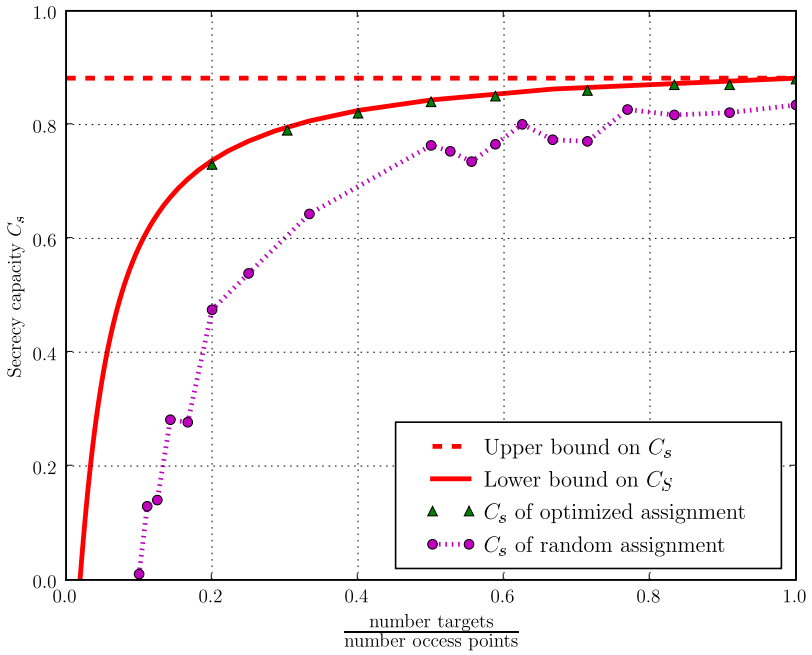


Figure 44. Secrecy capacities of several assignments for a vulnerable client-server architecture.

as the architecture becomes more vulnerable, the gap between the secrecy capacity of a random assignment and of optimized one becomes more significant. Notice that none of our optimizations yielded an assignment with a higher secrecy capacity than a balanced clustered assignment. This supports the argument that our lower bound might be tight.

7.5.2 Non-ISP centric situations

The ISP centric approach investigated thus far, where all access points or targets are equally likely to be attacked, may not always be valid in practice. In fact, in a non-ISP centric approach, access points are in different domains and therefore, their robustness to an attack is likely to differ. This scenario is harder to analyze, nevertheless a balanced clustered assignment still provides a valid sub-optimal assignment that can be used as the starting point for optimization algorithms.

In the following simulation we choose the DoS attack probability ρ_d and the compromise probability ρ_c of each access point according to a uniform distribution on $[0, 0.08]$. We also set the probability of each target being under DoS attack from an access point to $\eta_d = 1.0$, and a choose the compromise probability η_d according to a uniform distribution on $[0, 0.08]$.

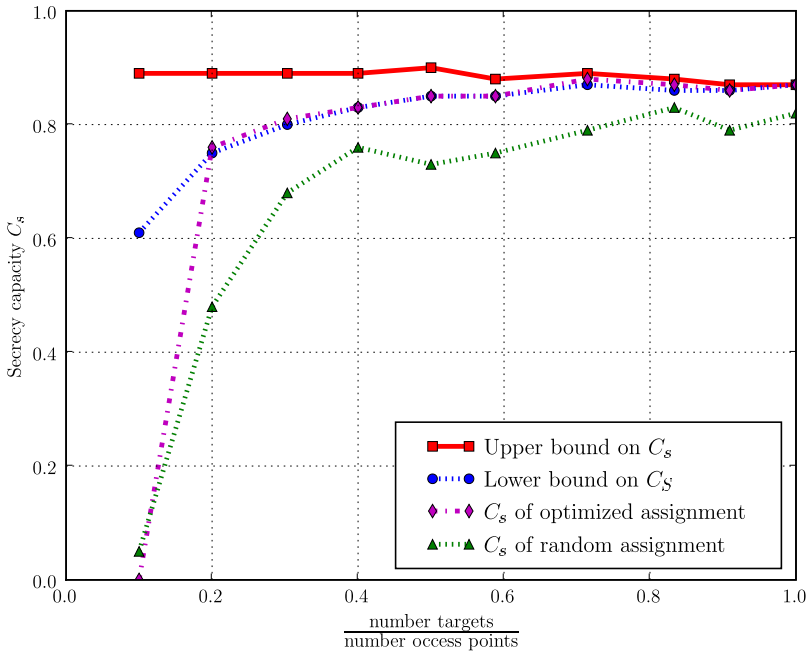


Figure 45. Secrecy capacities of several assignments for an architecture with different vulnerabilities.

Even though we do not have results characterizing the goodness of our lower bound of the secrecy capacity in this case, notice that none of the assignments obtained after optimization exhibit a secrecy capacity significantly higher than that of a balanced cluster assignment.

7.6 Proof of Theorems 7.1 and 7.2

7.6.1 Effect of edge removal

Let us consider the effect of removing an edge from a $k+1$ cluster. As shown in Figure 46, this operation leaves an access point unconnected. The new reading probability is therefore

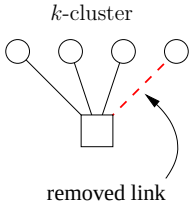


Figure 46. Removal of a link in a cluster.

given by

$$p_b^{rem} = \frac{k}{k+1} p_b^{(k)} + \frac{1}{k+1}. \tag{7.12}$$

Likewise, the new reading probability is

$$p_r^{rem} = \frac{k}{k+1} p_r^{(k)} + \frac{\rho_c}{k+1}. \tag{7.13}$$

After some calculation, one can verify that $p_b^{rem} - p_b^{(k+1)}$ is positive if and only if $k \leq \frac{1}{\rho_c \eta_a}$, and that $p_r^{rem} - p_r^{(k+1)}$ is always negative. Still, $p_b^{rem} - p_b^{(k+1)} + p_r^{rem} - p_r^{(k+1)}$ is positive as long as condition (7.7) is satisfied:

$$1 \leq (1-\rho_c\eta_a)^{k-1}(1-(k+1)\rho_c\eta_a) + (1-\rho_c\eta_c)^{k-1}(1-(k+1)\rho_c\eta_c). \tag{7.14}$$

7.6.2 Effect of edge addition

By adding an edge to a balanced clustered assignment, one necessarily connects two clusters together. For simplicity we focus on the case with two k-clusters, but the connection of two $k+1$-clusters, or a k-cluster with a $k+1$-cluster leads to the same conclusions. As

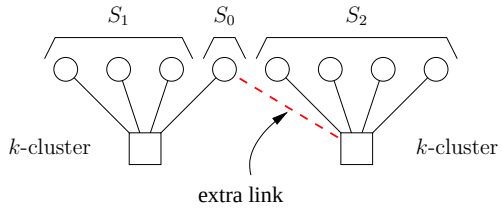


Figure 47. Addition of a cross-link between clusters.

shown in Figure 47, a packet can be routed in one of three access point sets labeled S_0, S_1 and S_2. A packet routed to an access point in S_1 (S_2) is blocked either if the node in S_0 is compromised and attacks the targets, or if the sub-cluster made of the access point in S_1 (S_2) and of the connected target is itself blocked. Therefore,

$$\begin{aligned} \mathrm{P}[\text{block}|\text{packet routed to } S_1] &= \rho_c\eta_a + (1-\rho_c\eta_a)p_b^{(k-1)}, & (7.15) \\ \mathrm{P}[\text{block}|\text{packet routed to } S_2] &= \rho_c\eta_a + (1-\rho_c\eta_a)p_b^{(k)}. & (7.16) \end{aligned}$$

Likewise, a packet routed to the node S_0 is blocked if the node S_0 is under attack or if *both* targets are attacked by other nodes. Hence,

$$\text{P[block|packet routed to } S_0] = \rho_a + (1-\rho_a)\left(1-(1-\rho_c\eta_a)^{k-1}\right)\left(1-(1-\rho_c\eta_a)^{k}\right). \quad (7.17)$$

After some calculations, one can show that the change in blocking probability after adding the edge is positive if

$$k\rho_c\eta_a \geq \left(1-(1-\rho_c\eta_a)^{k-1}\right)(1-\rho_c\eta_a). \quad (7.18)$$

Note that $\rho_c\eta_a \leq 1$, therefore $(1-\rho_c\eta_a)^{k-1} \geq 1-(k-1)\rho_c\eta_a$ and the condition is always satisfied for $k \geq 1$.

A similar analysis shows that adding an edge also increases the reading probability.

7.6.3 Effect of edge rewiring

Finally, consider the effect of rewiring an edge between 2 clusters as illustrated in Figure 48. We examine the case of rewiring between two $k+1$-clusters. The rewiring of two k-clusters or a $k+1$-cluster with a k-cluster can be treated in a similar way, but leads to less stringent conditions. Essentially, the rewiring introduces an unbalance by creating a $k+2$-cluster and a k-cluster. The new blocking probability p_b^{rw} and reading probability p_r^{rw} are now given respectively by:

$$p_b^{rw} = \frac{k}{2k+2}p_b^{(k)} + \frac{k+2}{2k+2}p_b^{(k+2)}, \quad (7.19)$$

$$p_r^{rw} = \frac{k}{2k}p_r^{(k)} + \frac{k+2}{2k+2}p_r^{(k+2)}. \quad (7.20)$$

After some lengthy but straightforward calculations one can check that $p_b^{rw} - p_b^{(k+1)} \geq 0$ if $k+1 \leq \frac{2-\rho_c\eta_a}{\rho_x\eta_a}$. Likewise, $p_r^{rw} - p_r^{(k+1)} \geq 0$ if $k+1 \leq \frac{2-\rho_c\eta_c}{\rho_x\eta_c}$. The sufficient condition (7.8) follows directly.

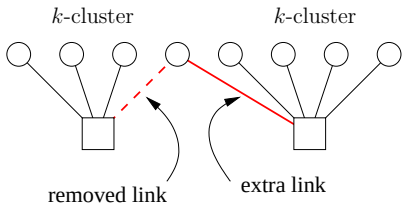


Figure 48. Rewiring an edge between clusters

CHAPTER 8
CONCLUSION

This final chapter summarizes our contributions and points out several areas for future research.

8.1 Contributions

The notion of *physical-layer security*, which is based on the idea that noise and losses are *resources* for information-theoretic security, advocates a paradigm shift in cryptography and calls for a *cross-layer* design of security schemes. Although physical-layer security has the potential of significantly strengthening the security level of current systems, by introducing some level of information-theoretic security instead of computational security, it is fair to acknowledge that its practicality is sometimes questionable. In this dissertation, we investigated several aspects of physical-layer security, with an emphasis on the design of *practical* schemes exploiting the imperfections of communication channels. In particular, the core of this dissertation focuses on wireless channels, whose specific nature is ideally suited for the design of effective physical-layer security schemes.

In Chapters 3 and 4, we developed and analyzed a practical secure communication protocol for quasi-static wireless channels, which is based on secret-key agreement from common randomness and *opportunistically* exploits the fluctuations of fading to allow the distillation of information theoretically secure keys. The protocol relies heavily on a specifically designed message-passing algorithm for the reconciliation of continuous and non-binary random variables, which exploits the ingredients of powerful error-control coding techniques, such as soft and iterative decoding, to achieve performance close to the fundamental limits imposed by information theory. Our analysis of the protocol showed that, in some instances, a pragmatic yet practical secret-key agreement approach does not incur any penalty in terms of secure rate.

Chapter 5 examined the interplay between cooperation requirements and security constraints, which is a particularly relevant issue in wireless communications. Cooperation has

already been proven to be beneficial for reliability, but intuitively, the redundancy introduced by relaying and cooperation schemes is likely to impair secrecy. To shed light on this problem, we analyzed a tripartite relaying situation with secrecy constraints, and we characterized the fundamental security compromise that a relay must accept if he desires to help an adversarial node. Strikingly, our analysis showed that the notion of cooperation and secrecy are not necessarily antagonistic. In particular, for Gaussian channels, we exhibited relaying strategies that provide both increased reliability and improved security.

Chapter 6 built upon the insight on secret-key agreement schemes obtained from Chapter 2, 3, and 4, to design an information-theoretically secure commitment primitive. Interestingly, although bit commitment might seem totally unrelated to physical-layer security, the tools developed for secure communication are equally useful for information-theoretic commitment. We highlighted the innate connection between secret-key agreement and bit commitment, which shed light on the design of practical schemes. Most notably, we showed that pragmatic commitment protocols based on secret-key agreement incur no loss of optimality and achieve the commitment capacity of discrete memoryless channels. Moreover, such protocols easily generalize to operate over Gaussian channels.

Finally, our investigation of secure client-server-architectures in Chapter 7 led to several interesting observations regarding the scope of applications of physical-layer security. Most importantly, we showed that the transmission of packets at the *network layer* sometimes admits to a natural erasure wiretap channel representation, which is in sharp contrast with the analysis of secure communications at the physical layer, where the wiretap channel model is usually introduced as an *a priori* model. This equivalent representation advocates the use of wiretap codes at the upper layers of protocol stacks, to prevent attackers from retrieving information based on a fraction of intercepted packets. Although the introduction of coding across different packets requires a redefinition of the role of the network layer, this idea is not dissimilar to the increasingly popular notion of network coding, which calls for the replacement of routing mechanisms by coding techniques. Furthermore, we emphasize that the use of wiretap channel models for packet-based communications seems less restricting than for physical-layer communications over noisy channels. For instance,

the attacker model adopted for our analysis of client-server architectures explicitly considers *active attacks* on the network. Moreover, the parameters of the wiretap channel model used to represent the network depend on the network characteristics; therefore, these parameters, such as the secrecy capacity, can be used as *optimization metrics* to improve the robustness and security of the network architecture.

8.2 Future Research

The work presented in this dissertation could be extended in many interesting directions.

- **Design of practical wiretap codes.** In this dissertation, we often circumvented the problem of designing wiretap codes by using a pragmatic yet more practical secret-key agreement approach; however, we point out that this simplification always came at the cost of increased communication and protocol complexity. Therefore, finding practical wiretap code constructions is still extremely relevant for practical purposes. Research on the construction of practical wiretap codes has been mostly driven by the insights provided by the early work of Wyner and Csiszár and Körner, but few practical solutions have been proposed so far, and the issue of coding for the wiretap channel should probably be addressed from a different perspective. For instance, rather than trying to design codes achieving a certain level of information-theoretic security, it would be useful to develop tools for analyzing the equivocation of a given code over a fixed channel. The insights brought by this analysis should shed light on the construction of finite-length codes with desirable security properties.

- **Multi-user information-theoretic security.** There are a number of challenging problems in the area of multi-user information-theoretic security. Most of the current work on information-theoretically secure communications is related to the wiretap channel model, and little attention has been devoted to information-theoretic security for networks. Generalizing the results obtained for secure point-to-point communications to secure network communications is all the more challenging as there is currently no framework to draw on. The notions of feedback, cooperation, and trust, are of paramount importance in multi-user scenarios and are not yet well understood.

- **Cryptographic primitives based on noisy channels.** There are many theoretical and practical issues that can be further explored regarding information-theoretically secure primitives. First, the identification of the fundamental security limits of noisy channels is still an largely open problem. Most of the work thus far has focused on discrete channels and, in general, the methods developed do not extend to continuous channels. Second, the design of efficient practical schemes has seldom been addressed, although this is of paramount importance to demonstrate the usefulness of information-theoretic security. This problem is non-trivial since most theoretical analysis rely on mathematical arguments that do not translate directly into practical schemes. As discussed in Chapter 6, secret-key agreement techniques might be useful for that purpose. Finally, it is essential to find convincing engineering applications for these schemes. As an example, some of the aforementioned cryptographic primitives could be tailored to cognitive radio protocols. In fact, in cognitive radio scenarios, devices compete to use a shared medium and maximize bandwidth usage, and ensuring that all players behave honestly and observe a fair behavior is a crucial issue. Bandwidth allocation is often performed with bidding schemes, and bit commitment protocols would be ideally suited to prevent devices from cheating and monopolizing bandwidth.

- **Cross-layer protocols for physical-layer security.** Physical-layer security has the potential of strengthening the security of communications by relaxing the assumptions on the computational power of the eavesdropper; however, it relies on other assumptions about the communication channels which may not be extremely accurate in practice. In light of these considerations, it is likely that the implementation of physical-layer security in a real system will be part of a layered-approach, and the design of protocols that combine traditional cryptographic techniques with physical-layer techniques is an interesting research direction. A key part of this research is the definition of relevant metrics that would make it possible to assess the performance of these hybrid schemes.

- **Experimental validation.** A crucial step towards establishing the validity of physical-layer security is the demonstration of secure protocols in real-life situations; however, with the notable exception of [82], experimental information-theoretic security has been left relatively unexplored. A more exhaustive study of the hardware requirements for physical-layer security is required to assess the weaknesses of realistic systems. As is often the case in cryptography, practical devices inevitably present security vulnerabilities that are not taken into account by theoretical models but are equally important.

APPENDIX A

ACHIEVABLE RATE-EQUIVOCATION REGION OF BROADCAST CHANNEL WITH CONFIDENTIAL MESSAGES

In this appendix, we provide and alternative proof for the result obtained in [6] by Csiszár and Körner. Our proof is based on the notion of *typical set decoding* popularized by Cover and Thomas. This approach has the advantage of being conceptually simpler than the maximal code construction used by Csiszár and Körner and allows us to combine wiretap codes with many other coding schemes, such as multiple-access schemes, relaying schemes, etc.

We first introduce modified definitions of codes and achievable rates for the broadcast channel with confidential messages, which ensure that message sets always contain an integer number of messages.

Definition A.1. *An (M_0, M_1, n) code for the broadcast channel with confidential messages consists of the following.*

- *Two message sets $\mathcal{M}_0 = \{1, 2, \ldots, M_0\}$ and $\mathcal{M}_1 = \{1, 2, \ldots, M_1\}$.*
- *An encoding function (possibly stochastic) $f_n : \mathcal{M}_0 \times \mathcal{M}_1 \to \mathcal{X}^n$, which maps each message pair $(m_0, m_1) \in \mathcal{M}_0 \times \mathcal{M}_1$ to a codeword $\mathbf{x}^n \in \mathcal{X}^n$.*
- *Two decoding functions $g_n : \mathcal{Y}^n \to \mathcal{M}_0 \times \mathcal{M}_1$ and $h_n : \mathcal{Z}^n \to \mathcal{M}_0$, which map an observation $\mathbf{y}^n$ to a message pair $(\hat{m}_0, \hat{m}_1)$ and an observation $\mathbf{z}^n$ to a message $\tilde{m}_0$.*

In the rest of this appendix, we denote the two messages sent by the transmitter by random variables W_0 and W_1, the messages estimated by decoder g_n by $\hat{W}_0$ and $\hat{W}_1$, and the message estimated by decoder h_n by $\tilde{W}_0$.

Definition A.2. *A rate triple $(R_0.R_1, R_e)$ is said to be achievable if, $\forall \epsilon > 0$, there exists*

an (M_0, M_1, n) *code such that*

$$R_0 - \epsilon \leq \frac{1}{n} \log_2 M_0 \leq R_0,$$
$$R_1 - \epsilon \leq \frac{1}{n} \log_2 M_1 \leq R_1,$$
$$\frac{1}{n} H(W_1 | \mathbf{Z}^n) - \epsilon \leq R_e,$$
$$P[(W_0, W_1) \neq g_n(\mathbf{Y}^n) \quad or \quad W_0 \neq f_n(\mathbf{Z}^n)] \leq \epsilon.$$

The main technical difficulty in obtaining the region of achievable rates is the proof of the following lemma.

Lemma A.1 ([6], Lemma 3)**.** *Let* U*,* X*,* Y*, and* Z *be random variables such that* $U \to X \to YZ$ *and* $I(X;Y|U) \geq I(X;Z|U)$*. The following rate tuples* (R_0, R_1, R_e) *are achievable.*

$$\begin{cases} R_0 \leq \min\left[I(U;Y), I(U;Z)\right], \\ R_1 + R_0 \leq I(X;Y|U) + \min\left[I(U;Y), I(U;Z)\right], \\ R_e = I(X;Y|U) - I(X;Z|U) \leq R_1. \end{cases}$$

To show the achievability of rates in Lemma A.1, we shall consider the coding scheme illustrated in Figure 49.

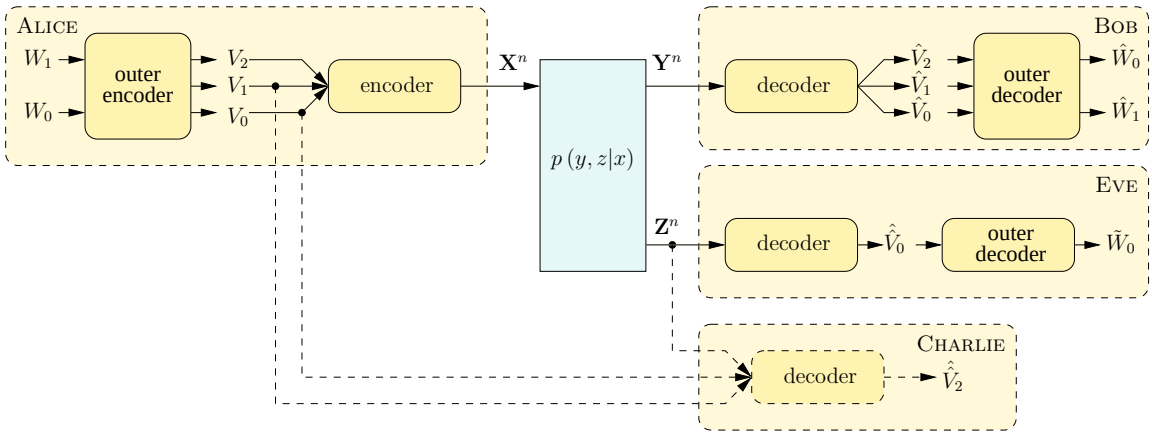


Figure 49. Coding scheme for wiretap code construction.

The encoder consists of the concatenation of an *outer code*, which maps the message pair $(W_0, W_1) \in \{1, \ldots, M_0\} \times \{1, \ldots, M_1\}$ to a message triple $(V_0, V_1, V_2) \in \{1, \ldots, K_0\} \times \{1, \ldots, K_1\} \times \{1, \ldots, K_2\}$, and an *inner code*, which maps the message triple (V_0, V_1, V_2)

into a codeword $\mathbf{X}^n$ for transmission over the broadcast channel. We shall design the inner code to have the following properties.

1. Bob, observing the output $\mathbf{Y}^n$ of the broadcast channel, recovers messages V_0, V_1, and V_2 with arbitrarily small probability of error;

2. Eve, observing the output $\mathbf{Z}^n$ of the broadcast channel recovers message V_0 with arbitrarily small probability of error;

3. Charlie, a *virtual* receiver observing $\mathbf{Z}^n$ and having access to V_0 and V_1, recovers V_2 with arbitrarily small probability of error.

We point out that the third receiver Charlie is simply introduced to set a constraint on the structure of the inner code. We shall see that this specific structure is essential to compute the equivocation of Eve.

Fix $\frac{1}{2} > \epsilon > 0$ and fix probability distributions $p_U(u)$ and $p_{X|U}(x|u)$. Let κ, δ and ϵ' be such that

$$0 < \kappa < \frac{\epsilon}{20}, \quad 0 < \delta < \min\left[\frac{\kappa}{6}, \frac{\epsilon}{5\log|\mathcal{X}|}\right], \quad 0 < \epsilon' < \epsilon \quad \text{with} \quad h(\epsilon') < \frac{\epsilon}{5}. \tag{A.1}$$

$\exists n_0 \in \mathbb{N}$ such that $\forall n \geq n_0$, one can find $K_0, K_1, K_2 \in \mathbb{N}$ satisfying

$$\begin{aligned}
\min\left(I(U;Y), I(U;Z)\right) - 2\kappa &\leq \frac{1}{n}\log_2 K_0 \leq \min\left(I(U;Y), I(U;Z)\right) - \kappa, & \text{(A.2)}\\
I(X;Y|U) - I(X;Z|U) - 2\kappa &\leq \frac{1}{n}\log_2 K_1 \leq I(X;Y|U) - I(X;Z|U) - \kappa, & \text{(A.3)}\\
I(X;Z|U) - 2\kappa &\leq \frac{1}{n}\log_2 K_2 \leq I(X;Z|U) - \kappa. & \text{(A.4)}
\end{aligned}$$

The above equations also imply

$$I(X;Y|U) - 4\kappa \leq \frac{1}{n}\log_2 K_1K_2 \leq I(X;Y|U) - 2\kappa. \tag{A.5}$$

1. Random inner code construction

Generate K_0 sequences of length n independently at random in $\mathcal{U}^n$ according to the distribution $p(\mathbf{u}^n) = \prod_{i=1}^n p(u_i)$. Label the sequences $\mathbf{U}^n(i)$ with $i \in \{1, \ldots, K_0\}$. For each sequence $\mathbf{U}^n(i)$, generate K_1K_2 sequences of length n independently at random in $\mathcal{X}^n$ according to the distribution $p(\mathbf{x}^n|\mathbf{u}^n) = \prod_{i=1}^n p_{X|U}(x_i|u_i)$. Label the sequences $\mathbf{X}^n(i,j,k)$ with $i \in \{1, \ldots, K_0\}$, $j \in \{1, \ldots, K_1\}$, and $k \in \{1, \ldots, K_2\}$.

□ *Encoding procedure.* It is assumed that messages V_0, V_1, and V_2 are chosen at random according to a uniform distribution in $\{1,\ldots,K_0\}$, $\{1,\ldots,K_1\}$, and $\{1,\ldots,K_2\}$, respectively. To transmit a message triple (V_0,V_1,V_2), the encoder simply transmits the codeword $\mathbf{X}^n(V_0,V_1,V_2)$.

□ *Bob's decoding procedure.* Bob determines the unique message tuple $\left(\hat{V}_0,\hat{V}_1,\hat{V}_2\right)$ such that

$$\left(\mathbf{U}^n(\hat{V}_0),\mathbf{X}^n\left(\hat{V}_0,\hat{V}_1,\hat{V}_2\right),\mathbf{Y}^n\right)\in A_\delta^{(n)}.$$

If there are none such tuple or more than one, an error is declared.

□ *Eve's decoding procedure.* Eve determines the unique message $\tilde{V}_0$ such that

$$\left(\mathbf{U}^n(\tilde{V}_0),\mathbf{Z}^n\right)\in A_\delta^{(n)}.$$

If there is none such codeword or more than one, an error is declared.

□ *Charlie's decoding procedure.* Charlie determines the unique message $\tilde{V}_2$ such that

$$\left(\mathbf{U}^n(V_0),\mathbf{X}^n\left(V_0,V_1,\tilde{V}_2\right),\mathbf{Z}^n\right)\in A_\delta^{(n)}.$$

If there si none such message or more than one an error is declared.

2. Analysis of probability of error

Define the following error event.

$$E(i)=\left\{(\mathbf{U}^n(i),\mathbf{Y}^n)\in A_\delta^{(n)}\right\} \tag{A.6}$$

$$\tilde{E}(i,j,k)=\left\{(\mathbf{U}^n(i),\mathbf{X}^n(i,j,k),\mathbf{Y}^n)\in A_\delta^{(n)}\right\} \tag{A.7}$$

$$F(i,j,k)=\left\{(\mathbf{U}^n(i),\mathbf{X}^n(i,j,k),\mathbf{Z}^n)\in A_\delta^{(n)}\right\} \tag{A.8}$$

$$G(i)=\left\{(\mathbf{U}^n(i),\mathbf{Z}^n)\in A_\delta^{(n)}\right\} \tag{A.9}$$

By symmetry of the random code construction, the average probability of error[1] is independent of the codeword send; therefore, we can assume without loss of generality that the

[1]The probability of error is averaged over all codewords and all codebooks generated according to $p_U(u)$ and $p_{X|U}(x|u)$

codeword $\mathbf{X}^n(1,1,1)$ is sent and write the total probability of error as follows.

$$\begin{aligned}\overline{P}_e^{(n)} &= \mathrm{P}[\text{error}|\mathbf{X}^n(1,1,1)\text{ sent}] \\ &= \mathrm{P}\left[E^c(1)\bigcup_{i\neq 1}E(i)\bigcup\tilde{E}^c(1,1,1)\bigcup_{(j,k)\neq(1,1)}\tilde{E}(1,j,k)\bigcup F^c(1,1,1)\right. \\ &\qquad \left.\bigcup_{k\neq 1}F(1,1,k)\bigcup G^c(1)\bigcup_{i\neq 1}G(i)|\mathbf{X}^n(1,1,1)\text{ sent}\right] \qquad (A.10)\end{aligned}$$

By applying the union bound on the expression above, the total probability of error can be bounded as follows.

$$\begin{aligned}\overline{P}_e^{(n)} &\leq \mathrm{P}[E^c(1)|\mathbf{X}^n(1,1,1)\text{ sent}] + \sum_{i\neq 1}\mathrm{P}[E(i)|\mathbf{X}^n(1,1,1)\text{ sent}] \\ &\quad + \mathrm{P}\Big[\tilde{E}^c(1,1,1)|\mathbf{X}^n(1,1,1)\text{ sent}\Big] + \sum_{(j,k)\neq(1,1)}\mathrm{P}\Big[\tilde{E}(1,j,k)|\mathbf{X}^n(1,1,1)\text{ sent}\Big] \\ &\quad + \mathrm{P}[F^c(1,1,1)|\mathbf{X}^n(1,1,1)\text{ sent}] + \sum_{(j,k)\neq(1,1)}\mathrm{P}[F(1,1,k)|\mathbf{X}^n(1,1,1)\text{ sent}] \\ &\quad + \mathrm{P}[G^c(1)|\mathbf{X}^n(1,1,1)\text{ sent}] + \sum_{i\neq 1}\mathrm{P}[G(i)|\mathbf{X}^n(1,1,1)\text{ sent}]. \qquad (A.11)\end{aligned}$$

By the joint AEP, $\exists n_1 \geq n_0$ such that $\forall n \geq n_1$

$$\mathrm{P}[E^c(1)|\mathbf{X}^n(1,1,1)\text{ sent}] \leq \frac{\epsilon'}{8}, \qquad (A.12)$$

$$\mathrm{P}\Big[\tilde{E}^c(1,1,1)|\mathbf{X}^n(1,1,1)\text{ sent}\Big] \leq \frac{\epsilon'}{8}, \qquad (A.13)$$

$$\mathrm{P}[F^c(1,1,1)|\mathbf{X}^n(1,1,1)\text{ sent}] \leq \frac{\epsilon'}{8}, \qquad (A.14)$$

$$\mathrm{P}[G^c(1)|\mathbf{X}^n(1,1,1)\text{ sent}] \leq \frac{\epsilon'}{8}. \qquad (A.15)$$

Now, for $(i,j,k) \neq (1,1,1)$, the code generation process ensures that $\mathbf{X}^n(i,j,k)$ is independent of $\mathbf{X}^n(1,1,1)$ and $\mathbf{U}^n(1)$. Since $\mathbf{Y}^n$ and $\mathbf{Z}^n$ are obtained by transmitting $\mathbf{X}^n(1,1,1)$ over the broadcast channel, they are also independent of $\mathbf{X}^n(i,j,k)$ and $\mathbf{U}^n(i)$ for $(i,j,k) \neq (1,1,1)$. Consequently,

$$\begin{aligned}\forall i \neq 1 \qquad \mathrm{P}[E(i)|\mathbf{X}^n(1,1,1)\text{ sent}] &= \mathrm{P}\Big[(\mathbf{U}^n(i),\mathbf{Y}^n) \in A_\delta^{(n)}|\mathbf{X}^n(1,1,1)\text{ sent}\Big], \\ &= \sum_{(\mathbf{u}^n,\mathbf{y}^n)\in A_\delta^{(n)}} p\left(\mathbf{u}^n\right)p\left(\mathbf{y}^n\right). \qquad (A.16)\end{aligned}$$

Again, by the joint AEP, $\exists n_2 \geq n_1$ such that $\forall n \geq n_2$

$$\begin{aligned}
&\forall(\mathbf{u}^n, \mathbf{y}^n) \in A_\delta^{(n)} \qquad p\left(\mathbf{u}^n\right) \leq 2^{-n(H(U)-\delta)}, \quad p\left(\mathbf{y}^n\right) \leq 2^{-n(H(Y)-\delta)}, \\
&\text{and} \qquad |A_\delta^{(n)}(U,Y)| \leq 2^{n(H(U,Y)+\delta)},
\end{aligned}$$

Therefore, substituting the above inequalities in Equation (A.16), we obtain

$$\begin{aligned}
\forall i \neq 1 \qquad \mathrm{P}[E(i)|\mathbf{X}^n(1,1,1) \text{ sent}] &\leq 2^{n(H(U,Y)-H(U)-H(Y)+4\delta)}, \\
&= 2^{-n(I(U;Y)-3\delta)}. \qquad \text{(A.17)}
\end{aligned}$$

Likewise, $\forall(j,k) \neq (1,1)$,

$$\begin{aligned}
\mathrm{P}\Big[\tilde{E}(1,j,k)|\mathbf{X}^n(1,1,1) \text{ sent}\Big] &= \mathrm{P}\Big[(\mathbf{U}^n(1), \mathbf{X}^n(1,j,k), \mathbf{Y}^n) \in A_\delta^{(n)}|\mathbf{X}^n(1,1,1) \text{ sent}\Big], \\
&= \sum_{(\mathbf{u}^n, \mathbf{x}^n, \mathbf{y}^n) \in A_\delta^{(n)}} p\left(\mathbf{u}^n\right) p\left(\mathbf{x}^n|\mathbf{u}^n\right) p\left(\mathbf{y}^n|\mathbf{u}^n\right). \qquad \text{(A.18)}
\end{aligned}$$

Applying the AEP, we know that $\exists n_3 \geq n_2$ such that $\forall n \geq n_3$

$$\begin{aligned}
&\forall(\mathbf{u}^n, \mathbf{x}^n, \mathbf{y}^n) \in A_\delta^{(n)} \qquad \begin{cases} p\left(\mathbf{u}^n\right) \leq 2^{-n(H(U)-\delta)}, \\ p\left(\mathbf{x}^n|\mathbf{u}^n\right) \leq 2^{-n(H(X|U)-2\delta)}, \\ p\left(\mathbf{y}^n|\mathbf{u}^n\right) \leq 2^{-n(H(Y|U)-2\delta)}, \end{cases} \\
&\text{and} \qquad |A_\delta^{(n)}(U,X,Y)| \leq 2^{n(H(U,X,Y)+\delta)},
\end{aligned}$$

Again, substituting these inequalities in Equation (A.18), we obtain

$$\begin{aligned}
\forall(j,k) \neq (1,1) \quad \mathrm{P}\Big[\tilde{E}(1,j,k)|\mathbf{X}^n(1,1,1) \text{ sent}\Big] &\leq 2^{n(H(U,X,Y)-H(U)-H(X|U)-H(Y|U)+6\delta)} \\
&= 2^{-n(I(X;Y|U)-6\delta)}. \qquad \text{(A.19)}
\end{aligned}$$

Applying the same technique to the events $F(1,1,k)$ and $G(i)$, we can show the existence of $n_4 \geq n_3$ such that $\forall n \geq n_4$

$$\forall(j,k) \neq (1,1) \qquad \mathrm{P}[F(1,1,k)|\mathbf{X}^n(1,1,1) \text{ sent}] \leq 2^{-n(I(X;Z|U)-6\delta)}, \qquad \text{(A.20)}$$

$$\forall i \neq 1 \qquad \mathrm{P}[G(i)|\mathbf{X}^n(1,1,1) \text{ sent}] \leq 2^{-n(I(U;Z)-4\delta)}. \qquad \text{(A.21)}$$

Finally, substituting Equations (A.12-A.15), (A.17), (A.19), (A.20) and (A.21) in the right-hand side of Equation (A.11), we have

$$\begin{aligned}
\overline{P}_e^{(n)} &\leq \frac{\epsilon'}{2} + K_0 2^{-n(I(U;Y)-4\delta)} + K_1 K_2 2^{-n(I(X;Y|U)-6\delta)} + K_2 2^{-n(I(X;Z|U)-6\delta)} + K_0 2^{-n(I(U;Z)-4\delta)}, \\
&\leq \frac{\epsilon}{2} + 2^{-n(\kappa-4\delta)} + 2^{-n(2\kappa-6\delta)} + 2^{-n(\kappa-6\delta)} + 2^{-n(\kappa-4\delta)}, \qquad \text{(A.22)}
\end{aligned}$$

where the second inequality follows from Equations (A.2-A.4). Since $\kappa > 6\delta$, $\exists n_5 \geq n4$ such that $\forall n \geq n_5$

$$2^{-n(2\kappa-6\delta)} \leq \frac{\epsilon'}{8}, \quad 2^{-n(\kappa-6\delta)} \leq \frac{\epsilon'}{8}, \quad \text{and} \quad 2^{-n(\kappa-4\delta)} \leq \frac{\epsilon'}{8},$$

and consequently,

$$\overline{P}_e^{(n)} \leq \epsilon'.$$

It is now standard procedure to argue that $\forall n \geq n_5$ there exists at least one specific code $\mathcal{C}^*_{inner}$ of length n such that the probability of error averaged over all codewords satisfies $P_e^{(n)} \leq \epsilon'$.

3. Outer code construction

We are now ready to design the outer code. To ensure that the outer code does not increase the probability of error, the following properties should be satisfied.

1. The mapping of (W_0, W_1) to (V_0, V_1, V_2) is an injective function such that V_0, V_1, and V_2 are uniformly distributed in $\{1, \ldots, K_0\}$, $\{1, \ldots, K_1\}$, and $\{1, \ldots, K_2\}$, respectively.

2. The outer decoders at Bob and Eve's side are surjective functions.

We shall now consider the construction of the outer code in detail. Recall that the transmission rate of messages W_0 and W_1 are denoted by R_0 and R_1, respectively.

□ Case 1. Transmission at rates (R_0, R_1) such that

$$\begin{cases} 0 \leq R_0 \leq \min\left[I(U;Y), I(U;Z)\right], \\ I(X;Y|U) - I(X;Z|U) \leq R_1 \leq I(X;Y|U). \end{cases}$$

Let $0 \leq \alpha \leq \min\left[I(U;Y), I(U;Z)\right] - 4\kappa$ and $0 \leq \beta \leq I(X;Z|U) - 4\kappa$. $\exists n_6 \geq n_5$ such that $\forall n \geq n_6$ one can find $L_0', L_0'', L_2', L_2'' \in \mathbb{N}$ satisfying

$$\begin{aligned} & L_0' L_0'' = K_0, && \alpha + \kappa \leq \frac{1}{n} \log_2 L_0'' \leq \alpha + 2\kappa, \\ \text{and} \quad & L_2' L_2'' = K_2, && \beta + \kappa \leq \frac{1}{n} \log_2 L_2'' \leq \beta + 2\kappa. \end{aligned}$$

To transmit messages w_0 and $w_1 = (w_1', w_1'')$ chosen uniformly at random in sets $\{1, \ldots, L_0'\}$ and $\{1, \ldots, K_1\} \times \{1, \ldots, L_2'\}$, respectively, Alice computes the indices

$$\begin{aligned} v_0 &= (w_0 - 1)L_0'' + r_0, \\ v_1 &= w_1', \\ v_2 &= (w_1'' - 1)L_2'' + r_2, \end{aligned}$$

where r_0, and r_2 are chosen uniformly at random in sets $\{1, \ldots, L_0''\}$ and $\{1, \ldots, L_2''\}$, respectively. Then, she uses the code $\mathcal{C}_{inner}^*$ identified earlier and transmits the codeword

$$\mathbf{x}^n(v_0, v_1, v_2).$$

One can easily check that v_0 and v_2 are uniformly distributed in $\{1, \ldots, K_0\}$ and $\{1, \ldots, K_2\}$, respectively, and it is obvious how to define surjective decoding functions. The concatenation of this outer code with $\mathcal{C}_{inner}^*$ achieves transmission rates

$$\min\left[I(U;Y), I(U;Z)\right] - \alpha - 4\kappa \leq R_0 = \frac{1}{n}\log_2 L_0' \leq \min\left[I(U;Y), I(U;Z)\right] - \alpha - 2\kappa, \tag{A.23}$$

$$I(X;Y|U) - \beta - 4\kappa \leq R_1 = \frac{1}{n}\log_2 K_1 L_2' \leq I(X;Y|U) - \beta - 2\kappa. \tag{A.24}$$

□ Case 2. Transmission at rates (R_0, R_1) such that

$$\begin{cases} 0 \leq R_0 + R_1 \leq I(X;Y|U) + \min\left[I(U;Y), I(U;Z)\right] \\ \text{with} \quad R_1 \geq I(X;Y|U). \end{cases}$$

Let $0 \leq \alpha \leq \min\left[I(U;Y), I(U;Z)\right] - 4\kappa$ and $0 \leq \beta \leq \min\left[I(U;Y), I(U;Z)\right] - \alpha - 6\kappa$. $\exists n_7 \geq n_6$ such that $\forall n \geq n_7$ one can find $L_0', L_0'', L_0''', L_0'''' \in \mathbb{N}$ satisfying

$$\begin{aligned} L_0' L_0'' &= K_0, \qquad & \alpha + \kappa \leq \frac{1}{n}\log_2 L_0'' \leq \alpha + 2\kappa, \\ L_0''' L_0'''' &= L_0'', \qquad & \beta + \kappa \leq \frac{1}{n}\log_2 L_0'''' \leq \beta + 2\kappa. \end{aligned}$$

To transmit messages w_0 and $w_1 = (w_1', w_1'', w_1''')$ chosen uniformly at random in sets $\{0, \ldots, L_0'''\}$ and $\{0, \ldots, L_0'\} \times \{0, \ldots, K_1\} \times \{0, \ldots, K_2\}$, respectively, Alice computes the

indices

$$v_0 = (w_1' - 1)L_0'' + (w_0 - 1)L_0'''' + r_0,$$
$$v_1 = w_1'',$$
$$v_2 = w_1''',$$

where r_0, is chosen uniformly at random in $\{1, \ldots, L_0''''\}$. She then uses the code $\mathcal{C}_{inner}^*$ identified earlier and transmits the codeword

$$\mathbf{x}^n(v_0, v_1, v_2),$$

Once again, one can check that v_0 is uniformly distributed in $\{1, \ldots, K_0\}$, and it is obvious how to define surjective decoding functions. The concatenation of this outer code with $\mathcal{C}_{inner}^*$ achieves transmission rates

$$\min\left[I(U;Y), I(U;Z)\right] - \alpha - \beta - 6\kappa \leq R_0 = \frac{1}{n}\log_2 L_0''' \leq \min\left[I(U;Y), I(U;Z)\right] - \alpha - \beta - 3\kappa, \tag{A.25}$$

$$I(X;Y|U) + \alpha - 4\kappa \leq R_1 = \frac{1}{n}\log_2 L_0' K_1 K_2 \leq I(X;Y|U) + \alpha - 2\kappa. \tag{A.26}$$

Combining Equations (A.23-A.24) and (A.25-A.26), it is clear that transmission rates (R_0, R_1) satisfying the following inequalities are achievable.

$$\begin{cases} 0 \leq R_0 + R_1 \leq I(X;Y|U) + \min\left[I(U;Y), I(U;Z)\right], \\ 0 \leq R_0 \leq \min\left[I(U;Y), I(U;Z)\right], \\ R_1 \geq I(X;Y|U) - I(X;Z|U). \end{cases}$$

4. Analysis of equivocation

We shall now bound the equivocation $H(W_1|\mathbf{Z}^n)$ obtained with the concatenated coding schemes described above. Using basic properties of the entropy, the equivocation can be

bounded as follows.

$$\begin{aligned} H(W_1|\mathbf{Z}^n) &\geq H(W_1|\mathbf{Z}^n, V_0) \\ &= H(W_1, \mathbf{Z}^n|V_0) - H(\mathbf{Z}^n|V_0) \\ &= H(W_1, \mathbf{Z}^n, \mathbf{X}^n|V_0) - H(\mathbf{X}^n|W_1, V_0, \mathbf{Z}^n) - H(\mathbf{Z}^n|V_0) \\ &= H(\mathbf{X}^n, W_1|V_0) + H(\mathbf{Z}^n|V_0, W_1, \mathbf{X}^n) - H(\mathbf{X}^n|W_1, V_0, \mathbf{Z}^n) - H(\mathbf{Z}^n|V_0) \\ &\geq H(\mathbf{X}^n|V_0) - H(\mathbf{X}^n|W_1, V_0, \mathbf{Z}^n) - I(\mathbf{X}^n; \mathbf{Z}^n|V_0)\,, \end{aligned} \tag{A.27}$$

where the last inequality follow from the fact that $(V_0, W_1) \to \mathbf{X}^n \to \mathbf{Z}^n$.

By construction of the inner code, we have

$$\frac{1}{n}H(\mathbf{X}^n|V_0) = \frac{1}{n}H(V_1, V_2) = \frac{1}{n}\log_2 K_1 K_2 \geq I(X;Y|U) - 4\kappa. \tag{A.28}$$

Now, the two outer codes described above have been constructed such that W_1 is uniquely identified by (V_0, V_1, V_2), but not by (V_0, V_1) alone. Therefore,

$$\begin{aligned} \frac{1}{n}H(\mathbf{X}^n|W_1, V_0, \mathbf{Z}^n) &\leq \frac{1}{n}H(\mathbf{X}^n|V_1, V_0, \mathbf{Z}^n) && \text{(A.29)} \\ &= \frac{1}{n}H(V_2|V_1, V_0, \mathbf{Z}^n)\,, && \text{(A.30)} \\ &\overset{(a)}{\leq} \frac{1}{n} + h\left(\mathrm{P}[V_2 \neq g(V_0, V_1, \mathbf{Z}^n)]\right), && \text{(A.31)} \end{aligned}$$

where (a) follows from Fano's inequality and g is any function of V_0, V_1, and $\mathbf{Z}^n$. In particular, g could be the decoding function used by Charlie. Since Charlie's probability of error is at at most ϵ', we obtain

$$\frac{1}{n}H(\mathbf{X}^n|V_1, V_0, \mathbf{Z}^n) \leq \frac{1}{n} + h(\epsilon'). \tag{A.32}$$

Recall that by construction of the inner code, V_0 uniquely determines $\mathbf{U}^n$ and vice-versa; therefore, $I(\mathbf{X}^n; \mathbf{Z}^n|V_0) = I(\mathbf{X}^n; \mathbf{Z}^n|\mathbf{U}^n)$. Now, let J be an indicator function such that

$$J = \begin{cases} 1 & \text{if} \quad (\mathbf{u}^n, \mathbf{x}^n, \mathbf{z}^n) \in A_\delta^{(n)}, \\ 0 & \text{otherwise.} \end{cases} \tag{A.33}$$

By the AEP, $\exists n_8 \geq n_7$ such that $\forall n \geq n_8$

$$\forall (\mathbf{u}^n, \mathbf{x}^n, \mathbf{z}^n) \in A_\delta^{(n)} \qquad \begin{cases} p\left(\mathbf{x}, \mathbf{z}^n | \mathbf{u}^n\right) \leq 2^{-n(H(X,Z|U)-2\delta)}, \\ p\left(\mathbf{x}^n | \mathbf{u}^n\right) \geq 2^{-n(H(X|U)+2\delta)}, \\ p\left(\mathbf{z}^n | \mathbf{u}^n\right) \geq 2^{-n(H(Z|U)+2\delta)}, \end{cases}$$

$$\text{and} \qquad \mathrm{P}[J=1] = \mathrm{P}\Big[(\mathbf{u}^n, \mathbf{x}^n, \mathbf{z}^n) \in A_\delta^{(n)}\Big] \geq 1-\delta.$$

Therefore,

$$\frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, J=1) \leq I(X; Z|U) + 6\delta. \tag{A.34}$$

Now,

$$\begin{aligned} \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n) &\leq \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n, J | \mathbf{U}^n) && \text{(A.35)} \\ &= \frac{1}{n} I(\mathbf{X}^n; J | \mathbf{U}^n) + \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, J), && \text{(A.36)} \\ &\leq \frac{1}{n} I(\mathbf{X}^n; J | \mathbf{U}^n) + \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, J=1)\, \mathrm{P}[J=1] \\ &\qquad + \frac{1}{n} I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, J=0)\, \mathrm{P}[J=0], && \text{(A.37)} \\ &\overset{(a)}{\leq} \frac{1}{n} + I(X; Z|U) + 6\delta + \delta \log_2 |\mathcal{X}|, && \text{(A.38)} \end{aligned}$$

where (a) follows from the fact that

$$I(\mathbf{X}^n; J | \mathbf{U}^n) \leq H(J) \leq 1,$$

$$I(\mathbf{X}^n; \mathbf{Z}^n | \mathbf{U}^n, J=0) \leq H(\mathbf{X}^n) \leq n \log_2 |\mathcal{X}|,$$

$$\mathrm{P}[J=0] \leq \delta \qquad \text{and} \qquad \mathrm{P}[J=1] \leq 1.$$

Substituting the bounds obtained in Equations (A.28), (A.32), and (A.38) in Equation (A.27), we obtain that $\forall n \geq n_8$

$$\frac{1}{n} H(W_1 | \mathbf{Z}^n) \geq I(X; Y|U) - I(X; Z|U) - 4\kappa - \frac{2}{n} - h(\epsilon') - 6\delta - \delta \log_2 |\mathcal{X}|. \tag{A.39}$$

Using the bounds in Equation (A.1), it is clear that

$$\frac{1}{n} H(W_1 | \mathbf{Z}^n) \geq I(X; Y|U) - I(X; Z|U) - \epsilon. \tag{A.40}$$

The equivocation rate $R_e \geq I(X; Y|U) - I(X; Z|U)$ is achievable, which concludes the proof of Lemma A.1.

For completeness, we provide the remaining of the proof leading to the characterization of the full achievable region, but these last steps are identical to those provided in [6].

Lemma A.2 ([6], Lemma 4)**.** *If U, V, X, Y, and Z are random variables such that $U \to V \to X \to YZ$ and $I(V;Y|U) \geq I(V;Z|U)$, the following rate tuple (R_0, R_1, R_e) are achievable.*

$$\begin{cases} R_0 \leq \min\left[I(U;Y), I(U;Z)\right], \\ R_1 + R_0 \leq I(V;Y|U) + \min\left[I(V;Y), I(V;Z)\right], \\ R_e = I(V;Y|U) - I(V;Z|U) \leq R_1. \end{cases}$$

Proof. Consider the discrete memoryless channel with transition probability $p_{YZ|V}(y,z|v)$, obtained by concatenating a channel with transition probability $p_{X|V}(x|v)$ before the original broadcast channel with transition probability $p_{YZ|X}(y,z|x)$. The result follows by applying Lemma A.1 to the new channel. □

Lemma A.3 ([6], Lemma 5)**.** *The region $\mathcal{R}$ defined below is convex.*

$$\mathcal{R} = \bigcup_{U \to V \to X \to YZ} \left\{ \begin{array}{l} R_0 \leq \min\left[I(U;Y), I(U;Z)\right], \\ R_1 + R_0 \leq I(V;Y|U) + \min\left[I(U;Y), I(U;Z)\right], \\ R_e \leq I(V;Y|U) - I(V;Z|U) \leq R_1. \end{array} \right\} \tag{A.41}$$

Proof. First, since $I(V;Y|U) - I(V;Z|U)$ is an achievable equivocation rate, the fact that all equivocation rates $R_e \leq I(V;Y|U) - I(V;Z|U)$ follows from the definition of achievability.

Now let $\left(R_0^{(1)}, R_1^{(1)}, R_e^{(1)}\right)$ and $\left(R_0^{(2)}, R_1^{(2)}, R_e^{(2)}\right)$ be rate tuples in $\mathcal{R}$, achievable with random variables $U_1 \to V_1 \to X_1 \to Y_1Z_1$ and $U_2 \to V_2 \to X_2 \to Y_2Z_2$, respectively. Let J be a random variable independent of all others taking values 1 and 2 with probability p and $1-p$, respectively, and define the time-shared random variables

$$U = U_J, \quad V = V_J, \quad X = X_J, \quad Y = Y_J, \quad Z = Z_J. \tag{A.42}$$

Then, by definition of conditional mutual information, we have

$$I(V;Y|U) = pI(V_1;Y_1|U_1) + (1-p)I(V_2;Y_2|U_2) \tag{A.43}$$

$$I(V;Z|U) = pI(V_1;Z_1|U_1) + (1-p)I(V_2;Z_2|U_2) \tag{A.44}$$

$$I(U;Y) \geq I(U;Z|J) = pI(U_1;Z_1) + (1-p)I(U_2;Z_2)\,, \tag{A.45}$$

$$I(U;Y) \geq I(U;Z|J) = pI(U_1;Z_1) + (1-p)I(U_2;Z_2)\,. \tag{A.46}$$

Therefore, using the definition of the achievable rates in $\mathcal{R}$, we obtain

$$pR_0^{(1)} + (1-p)R_0^{(2)} \leq pI(U_1;Y_1) + (1-p)I(U_2;Y_2) \leq I(U;Y)\,, \tag{A.47}$$

$$pR_0^{(1)} + (1-p)R_0^{(2)} \leq pI(U_1;Z_1) + (1-p)I(U_2;Z_2) \leq I(U;Z)\,. \tag{A.48}$$

Likewise, one can show that

$$p(R_0^{(1)} + R_1^{(1)}) + (1-p)(R_0^{(2)} + R_1^{(1)}) \leq I(U;Y) + I(V;Y|U)\,, \tag{A.49}$$

$$p(R_0^{(1)} + R_1^{(1)}) + (1-p)(R_0^{(2)} + R_1^{(1)}) \leq I(U;Z) + I(V;Y|U)\,, \tag{A.50}$$

$$pR_e^{(1)} + (1-p)R_e^{(2)} \leq I(V;Y|U) - I(V;Z|U) \tag{A.51}$$

Therefore, the time-shared random variables are also in $\mathcal{R}$ which concludes the proof. □

Lemma A.4 ([6], Lemma 5)**.** *The region $\mathcal{R}$ is equal to the region $\mathcal{C}$ defined below.*

$$\mathcal{C} = \bigcup_{U\to V\to X\to YZ} \left\{ \begin{array}{l} 0 \leq R_e \leq R_1, \\ R_e \leq I(V;Y|U) - I(V;Z|U)\,, \\ R_1 + R_0 \leq I(V;Y|U) + \min\left[I(U;Y)\,, I(U;Z)\right], \\ R_0 \leq \min\left[I(U;Y)\,, I(U;Z)\right]. \end{array} \right\} \tag{A.52}$$

Proof. Clearly, $\mathcal{R} \subset \mathcal{C}$ and the only difference between regions $\mathcal{R}$ and $\mathcal{C}$ are the rates R_1 such that

$$0 \leq R_1 \leq I(V;Y|U) - I(V;Z|U)\,.$$

Now let $(R_0, R_1, R_e) \in \mathcal{C}$ and define

$$R_1^* = I(V;Y|U) + \min\left[I(U;Y)\,, I(U;Z)\right] - R_0,$$

$$R_e^* = I(V;Y|U) - I(V;Z|U)\,.$$

Since $R_1^* \geq R_e^*$, the rate tuple $(R_0, R_1^*, R_e^*) \in \mathcal{R}$. By definition of $\mathcal{R}$, the triples (R_0, R_e^*, R_e^*), $(R_0, R_1^*, 0)$, and $(R_0, 0, 0)$ also belong to $\mathcal{R}$. Recalling that $R_1 \leq R_1^*$ and $R_e \leq R_e^*$, the triple (R_0, R_1, R_e) is in the convex hull defined by the four triple above. Therefore $\mathcal{R} \subset \mathcal{C}$ and the proof is complete. □

APPENDIX B

THRESHOLDS OF ITERATIVELY DEMODULATED CODED-MODULATION SCHEMES

In this appendix, we discuss the density evolution analysis of the algorithm described in Chapter 3. For simplicity, we consider the application of the algorithm to coded modulation schemes rater than reconciliation. We restrict our investigation to Bit-Interleaved Coded Modulation (BICM) [42], which is commonly used to achieve good error-rate performance and bandwidth efficiency over Gaussian channels. We recall that the good performance of this techniques relies mainly on the use of powerful error-correcting component codes, and LDPC codes have been proven to be especially useful for this purpose; however, to achieve performance close to capacity, the edge degree distributions of the bipartite graphs associated with the constituent code has to be carefully optimized.

In the particular case where the receiver does not iterate between the demapper and the error correcting code, which is usually the case for BICM with gray mapping, the optimization of asymptotically long codes can be performed exactly using density evolution [83]; however, in other situations, such as BICM with non-Gray mapping, iterations result in significant improvement. Several optimization techniques based on EXtrinsic Information Transfer (EXIT) charts have been proposed for those iteratively demodulated schemes [84, 85, 86]. The major advantage of the EXIT chart approach is the reduction of the initial code design problem to a less complex curve-fitting problem; however, despite their effectiveness, these methods rely on the assumption that curve-fitting yields capacity-approaching codes, which has only be proven for the binary erasure channel. Moreover, EXIT chart analysis approximates the densities of messages propagated during the decoding algorithm by symmetric Gaussian densities.

B.1 Coded Modulation with LDPC Codes

B.1.1 System Model

The density evolution algorithm presented here applies to any concatenation of demapper and decoders (e.g. iterative MultiStage Decoding); however, for simplicity, we focus our discussion on BICM schemes, see Figure 50. At the transceiver, k-bit messages $\mathbf{m}$ are

encoded into n-bit codewords $\mathbf{c} = (c_1, \ldots, c_n)$ by a single LDPC encoder. The bits of each codeword are then grouped into sub-blocks of ℓ bits. Each sub-block is mapped to one of 2^ℓ complex-valued symbols $\mathbf{x}$ according to a predefined *mapping*. The set $\mathcal{X}$ of 2^ℓ symbols is called a *constellation* in the context of coded modulation. The symbol sequence $\mathbf{x}$ is corrupted by additive white gaussian noise $\mathbf{n}$ with variance $\sigma^2 = N_0/2$ during the transmission, and the receiver observes the sequence $\mathbf{y} = \mathbf{x} + \mathbf{n}$. The demapper plays the role of a metric computer, and outputs the *intrinsic* Log-Likelihood Ratios (LLRs) of each bit c_i. These LLRs depend not only on the observations $\mathbf{y}$ coming from the channel but also on the *extrinsic* LLRs that may be available from a previous decoding attempt.

Figure 50. Iteratively demodulated BICM scheme.

B.1.2 Message Passing Algorithm

The iterative demodulating and decoding algorithm considered here is a message passing algorithm over an extended Tanner graph, which computes the LLRs

$$\log \frac{\mathrm{P}[c_i = 0|\mathbf{y}]}{\mathrm{P}[c_i = 1|\mathbf{y}]} \quad \forall i \in \{1, \ldots, n\}.$$

The graph is obtained by adding *demapper nodes* to the standard bipartite Tanner graph of an LDPC code, see Figure 51. Contrary to the standard Sum-Product decoding algorithm [87], where the intrinsic information available at each variable node is fixed during the decoding process, the introduction of demapper nodes allows this information to be updated at a later stage.

We use the following notations to describe the message passing algorithm.

- $\mathcal{N}(i)$ ($\mathcal{M}(j)$) denotes the set of indices of check (variable) nodes connected to a variable node i (check node j),
- $\mathcal{O}(i)$ denotes the set of indices of variable nodes connected to the same demapper node as the variable node i,
- y_i represents the channel observation associated to a variable node i,

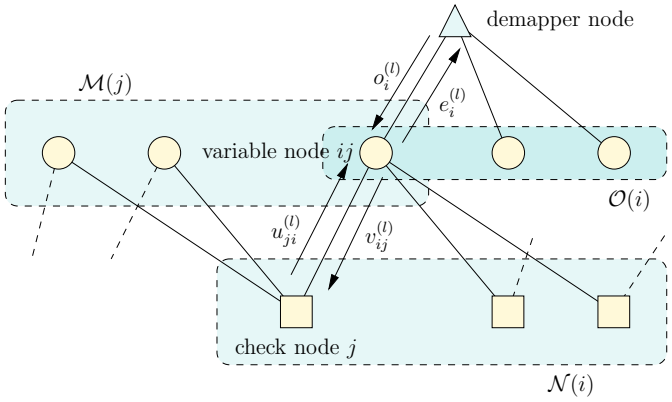


Figure 51. Example of extended Tanner Graph including demapper nodes.

- if x is a symbol in the constellation $\mathcal{X}$, then x_k denotes its kth label bit,
- messages are defined as shown in Figure 51.

The message passing algorithm presented in Chapter 3 can then be rewritten as follows.

□ **Initialization.** Initialize all message to zero.

$$\forall i, j \quad e_i^{(0)} = o_i^{(0)} = v_{ij}^{(0)} = u_{ji}^{(0)} = 0. \tag{B.1}$$

□ **Iterations.** For $1 \leq l \leq l_{max}$

1. demapper-to-variable message update

$$o_i^{(l)} = \log \frac{\displaystyle\sum_{x \in \mathcal{X}: x_j = 0} p\left(y_i | x\right) \exp\left[\sum_{k \neq j} \left(1 - x_k\right) e_{ik}^{(l-1)}\right]}{\displaystyle\sum_{x \in \mathcal{X}: x_j = 1} p\left(y_i | x\right) \exp\left[\sum_{k \neq j} \left(1 - x_k\right) e_{ik}^{(l-1)}\right]}. \tag{B.2}$$

2. variable-to-check message update

$$v_{ij}^{(l)} = o_i^{(l)} + \sum_{k \in \mathcal{N}(i) \setminus j} u_{ki}^{(l-1)} \tag{B.3}$$

3. check-to variable-message update

$$u_{ji}^{(l)} = 2 \tanh^{-1} \prod_{k \in \mathcal{M}(j) \setminus ij} \tanh \frac{v_{ik}^{(l-1)}}{2} \tag{B.4}$$

4. variable-to-demapper update

$$e_i^{(l)} = \sum_{k \in \mathcal{N}(i)} u_{ki}^{(l)} \tag{B.5}$$

□ **Hard decoding.** $\forall i \in \{1, \ldots, n\}$ decide

$$v_i = -\frac{1}{2}\left(\text{sign}(e_i^{(l_{max})} + o_i^{(l_{max})}) - 1\right),$$

where

$$\text{sign}(x) = \begin{cases} +1 & \text{if} \quad x \geq 0 \\ -1 & \text{if} \quad x < 0 \end{cases}$$

Notice that the scheduling of the message-passing algorithm assumes that the messages $o_i^{(\ell)}$ are calculated at every iteration. In practice, for large constellations, the computation of $o^{(\ell)}$ according to Equation (B.2) quickly becomes prohibitive, and it is computationally more efficient to perform this update less often.

B.2 Density Evolution and Threshold Computation

The principle of *density evolution* is to analyze the performance of the message-passing algorithm by tracking the probability densities of the messages computed at each iteration. When the underlying channel is symmetric, the probability densities of messages are codeword-independent [48], and it is sufficient to analyze the transmission of the all-zero codeword. However, as we illustrate in Figure 52, the underlying channel in a BICM scheme is clearly asymmetric when there are more than two symbols in a constellation, and the densities depend on the codeword sent. Notice that the shape of the curves shown above also questions the Gaussian assumption used in EXIT chart approaches. To circumvent this problem, one can either track the messages densities averaged over the LDPC code and its cosets [88], or averaged over all valid codewords [89]. The former approach is easier to analyze, but, in practice, we are interested in the behavior of linear codes only and there is no guarantee that the results averaged over all coset codes are still relevant for linear codes, in general; therefore, we consider the latter approach and average the densities over all possible codewords.

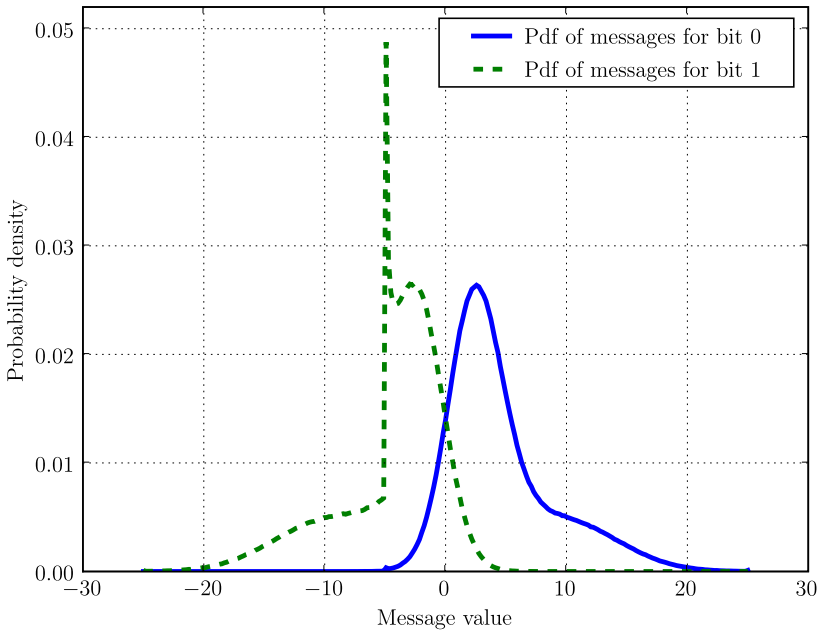


Figure 52. Probability densities of messages at the output of the demapper, for a 4-PAM constellation with Gray mapping and noise variance $\sigma^2 = 0.016$.

B.2.1 Density Evolution

We recall that LDPC codes are characterized by their edge degree distributions

$$\lambda(x) = \sum_i \lambda_i x^{i-1} \quad \text{and} \quad \rho(x) = \sum_i \rho_i x^{i-1}, \tag{B.6}$$

where λ_i and ρ_i represent the fraction of edges connected to degree-i variable and check nodes, respectively.

The probability of decoding error during the ℓth iteration is analyzed by considering the message $v^{(\ell+1)}$ flowing from a variable node to a check node during the ℓth iteration. The ensemble of node and edges contributing to $v^{(\ell+1)}$ forms a *message flow neighborhood* of depth ℓ. Figure 53 represents a message flow of depth 1, and a neighborhood of depth ℓ can be constructed by branching several of these elementary neighborhoods. It should be noted that a neighborhood of depth ℓ depends on the input sequence, since from Equation (B.2) the message sent by a demapper node depends on the value of its connected bits. Hence, the decoder behavior varies from one input sequence to another.

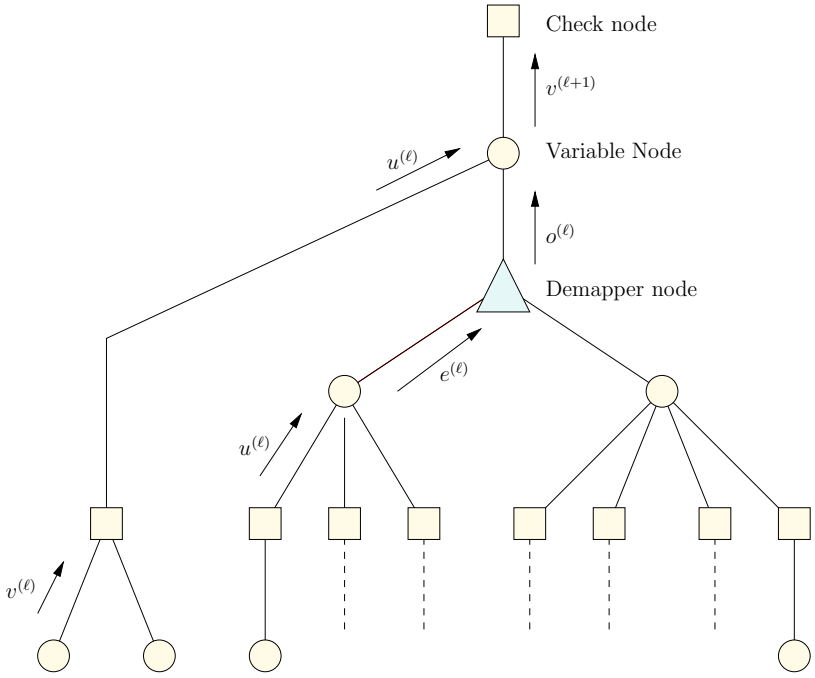


Figure 53. Demapping neighborhood of depth 1.

Assuming that the value of the variable node at the root of the message flow neighborhood of depth ℓ is $x \in \{0,1\}$, we let $f_V^{(\ell)}(x)$, $f_U^{(\ell)}(x)$, $f_O^{(\ell)}(x)$, and $f_E^{(\ell)}(x)$ be the respective densities of the messages $v^{(\ell)}$, $u^{(\ell)}$, $o^{(\ell)}$ and $e^{(\ell)}$, averaged over all valid codewords such that the value of the variable node at the root of the neighborhood is x, .

Under the assumption that the message flow neighborhood is *perfectly projected* (see [89] for a formal definition), it is possible to show that the density $f_V^{(\ell+1)}(x)$ is obtained by evolving $f_V^{(\ell)}(x)$ as follows. At the root variable node,

$$\begin{aligned} f_V^{(\ell+1)}(x) &= f_{\mathcal{O}}^{(\ell+1)}(x) \otimes \left[\sum_i \lambda_i \left(\bigotimes_{k=1}^{i-1} f_U^{(\ell)}(x) \right) \right] \\ &= f_{\mathcal{O}}^{(\ell+1)}(x) \otimes \lambda\Big(f_U^{(\ell)}(x) \Big), \end{aligned} \tag{B.7}$$

where $\otimes$ denotes the convolution operator on probability density functions. Note that this operation can be computed efficiently using Fourier transforms. The closed-form expression of the evolution through a check node requires a change of measure (see [89][Equation 15]), and, for simplicity, we denote the operation at a check node of degree i simply by $\mathcal{E}_c^{i-1}$.

Hence, we write

$$\begin{aligned} f_U^{(\ell)}(x) &= \sum_i \rho_i \mathcal{E}_c^{i-1}\left(f_V^{(\ell)}(0), f_V^{(\ell)}(1)\right) \\ &= \rho\Big[\mathcal{E}_c\left(f_V^{(\ell)}(0), f_V^{(\ell)}(1)\right)\Big]. \end{aligned} \tag{B.8}$$

This evolution can be calculated with a look-up table, as discussed in Section B.4. The evolution of the density through the demapper can likewise be written as

$$f_{\mathcal{O}}^{(\ell)}(x) = \mathcal{E}_d\Big(f_E^{(\ell)}(0), f_E^{(\ell)}(1)\Big). \tag{B.9}$$

In general, there exists no closed-form expression for $\mathcal{E}_d$, and the numerical evaluation must be done via Monte-Carlo simulations of the demapping function given by Equation (B.2). More precisely, the ℓ-th density evolution through the demapper is performed as follows. We first generate n_b bits c_i uniformly at random, which are then mapped to n_s symbols and corrupted by channel noise. We then generate n_b independent realizations $e_i^{(\ell)}$ according to $f_E^{(\ell)}(c_i)$. Using these values, we compute the n_b outputs $o_i^{(\ell)}$ of the demapper for each bit c_i. The output density $f_{\mathcal{O}}^{(\ell)}(x)$ ($x \in \{0,1\}$) is finally obtained by reconstructing the histogram of the values $o_i^{(\ell)}$ for which $c_i = x$. The number of bits n_b used in the simulation has to be carefully chosen, since a bad estimation of the density $f_{\mathcal{O}}^{(\ell)}(x)$ may lead to an overestimation of the true threshold. These numerical issues are discussed in Section B.3.1.

Finally $f_E^{(\ell)}(x)$ is obtained from $f_U^{(\ell)}(x)$ by

$$f_E^{(\ell)}(x) = \sum_i \frac{\lambda_i}{i\int_0^1 \lambda(x)dx}\left(\bigotimes_{k=1}^{i} f_U^{(\ell)}(x)\right). \tag{B.10}$$

Notice that, in the equation above, the averaging is performed from a node perspective and not from an edge perspective.

B.2.2 Concentration and Threshold

The key result justifying the validity of the density evolution algorithm is the fact that, if the message bits $\mathbf{m}$ are independent and uniformly distributed, then, for almost all graphs with given edge degree distributions $\lambda(x)$ and $\rho(x)$, the decoder behaves close to its expected behavior. The derivation of this results follows directly from the proof of [89][Section IV], and will be omitted.

Table 4. Mappings of 4-PAM constellation

Symbol	-1.34	-0.45	0.45	1.34
Natural	0	0	1	1
mapping	0	1	0	1
Gray	0	0	1	1
mapping	0	1	1	0
Antigray	0	1	1	0
mapping	0	1	0	1

It can be shown that the probability of decoding error concentrates around the value $p_e^{(\ell)}$, defined as

$$p_e^{(\ell)} = \frac{1}{2}\left(\int_{m=0}^{\infty} f_V^{(\ell)}(0) + \int_{m=-\infty}^{0} f_V^{(\ell)}(1)\right).$$

The threshold σ^* of the iteratively demodulated coded modulation scheme is defined as the supremum of all noise standard deviations such that

$$\lim_{\ell\to\infty} p_e^{(\ell)} = 0. \tag{B.11}$$

As a consequence of the concentration theorem, if we can find σ^* then almost all input sequence will be decoded reliably if the noise is such that $\sigma < \sigma^*$

B.3 Simulation Results

B.3.1 Iterative and Non-Iterative Thresholds

In this section, we apply density evolution to analyze BICM coded modulation with 4-PAM constellations and rate-0.5 codes. We consider two different LDPC codes; the first code is a regular LDPC code with degree distributions

$$\lambda(x) = x^2 \quad \text{and} \quad \rho(x) = x^5,$$

whereas the second code is an irregular LDPC codes with degree distributions

$$\lambda(x) = 0.251828x + 0.211152x^2 + 0.537020x^9,$$
$$\text{and} \quad \rho(x) = x^7.$$

The 4-PAM constellation contains 4 symbols with amplitudes $\{-1.34, -0.45, 0.45, 1.34\}$, and the various mappings used in our simulations are shown in Table 4.

All simulations are performed using a 8-bit discretized density evolution, with $2^8 - 1$ quantization bins spanning the range $[-25; 25]$ uniformly. We allow at most 1000 decoding

Table 5. Thresholds of 4-PAM iterative BICM scheme with regular code.

Mapping	Non-Iter. threshold	Iter. threshold
Natural	4.78 dB	3.41 dB
Gray	3.41 dB	3.34 dB
Anti-Gray	6.08 dB	4.39 dB

Table 6. Thresholds of 4-PAM iterative BICM scheme with irregular code.

Mapping	Non-Iter. threshold	Iter. threshold
Natural	4.25 dB	3.69 dB
Gray	2.80 dB	2.77 dB
Anti-Gray	5.62 dB	4.95 dB

iterations, and densities are evolved until the probability of error falls below 10^{-6} or until a fixed point is reached. The number n_b of realizations used in Monte-Carlo simulation for the evolution of densities through demappers is $n_s \times 5\,10^6$, which turns out to be sufficient to reconstruct the histograms of $f_{\mathcal{O}}^{(\ell)}(0)$ and $f_{\mathcal{O}}^{(\ell)}(1)$ accurately. The thresholds obtained with our algorithm are presented in Table 5 and 6.

As expected, Gray mapping yields the best thresholds and, in this case, negligible gain is obtained by performing iterations between the demapper and decoder. Notice that the best threshold of 2.77 dB obtained with the irregular code is less than 0.5 dB away from the parallel independent decoding capacity (2.27 dB) of the channel, which is the ultimate performance of the BICM scheme. For all other mappings, iterations lead to more significant improvements of the thresholds, although better results could be expected by using codes specifically optimized for each scheme. This will be briefly discussed in section B.3.2.

In order to validate the threshold calculations, we simulated the 4-PAM iterative BICM scheme for LDPC codes of size 500,000. The codes were randomly generated by avoiding cycles of length 2 and 4. In order to speed up the decoding, we limited the number of iterations of the LDPC code to 100, and the number of iterations between the decoder and the demapper to 10. As shown in Figure 54 and Figure 55, the thresholds previously obtained accurately predict the performance of a BICM scheme with long LDPC codes.

B.3.2 Optimization Results

As mentioned earlier, given a fixed modulation scheme, one can optimize the degree distribution of an LDPC code in order to obtain the best possible threshold. However, the

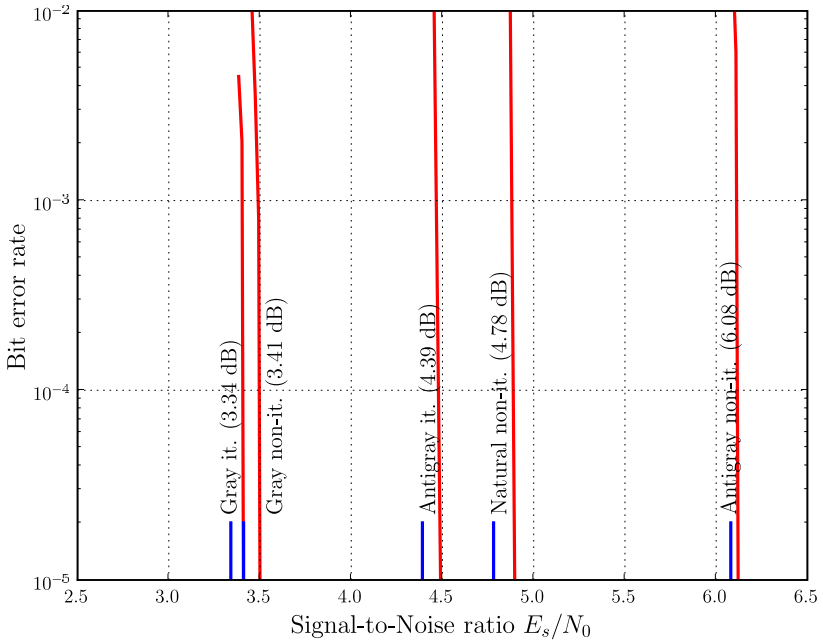


Figure 54. Simulation of 4-PAM BICM scheme with regular code.

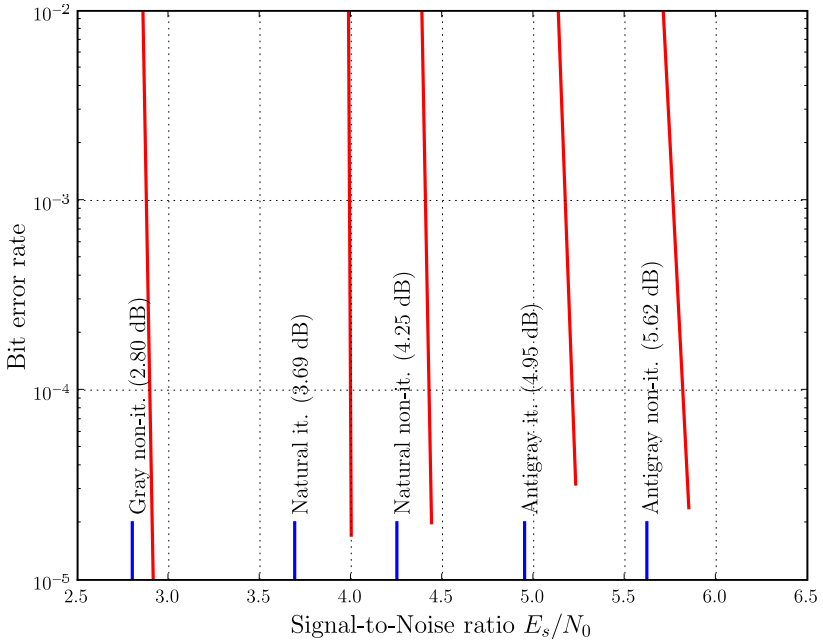


Figure 55. Simulation of 4-PAM BICM scheme with irregular code.

threshold is not a linear function of the edge degree distributions, which makes it impossible to apply well-know fast optimization techniques. Following [87], we used the Differential Evolution (DE) optimization method. DE is a mixture of a genetic algorithm and a hill-climbing optimization, and has been proven to be quite effective for solving real-valued non-linear problems.

We applied DE to maximize the threshold of the anti-Gray 4-PAM BICM scheme at a spectral efficiency of 1 bit/symbol. In order to limit the search space, we limited our optimization to variable node degree distributions with maximum degree 10, and three non-zero elements (degrees 2,3 and 10). We also restricted the check node degree distribution to be concentrated. One good degree distribution found was:

$$\begin{aligned} \lambda(x) &= 0.307274x + 0.400869x^2 + 0.291857x^9 \\ \rho(x) &= 0.645361x^5 + 0.354639x^6. \end{aligned}$$

The thresholds obtained are given in Table 7 and compared against the results of the irregular code of the previous section.

Table 7. Thresholds of 4-PAM anti-Gray BICM.

	Optimized code	BPSK-AWGN code
Iterative threshold	4.05 dB	4.95 dB
Non iter. threshold	5.77 dB	5.62 dB

Clearly, the optimized code outperforms the previous code, which confirms the importance of specific code optimization for iterative receivers.

B.4 Discretized density evolution

In this section, we briefly describe the discretized density evolution algorithm used to compute thresholds. Densities are evolved through variable nodes by using discrete Fourier transforms [90]. However, the evolution through the check nodes (according to [89][Equation 15]) requires an averaging over all possible values of the bits involved in the parity-check equation. This can be computed efficiently by performing a change of measure, however, we use a different technique in our simulations.

Following [90], we define the operator $\mathcal{R}$ as

$$\mathcal{R}(a,b) = \mathcal{Q}\left(2\tanh^{-1}\left(\tanh\frac{a}{2}\tanh\frac{b}{2}\right)\right),$$

where $\mathcal{Q}$ is the quantizer used for the discretized density evolution. If a and b have densities p_a and p_b, respectively, we abuse notations and write $\mathcal{R}(p_a, p_b)$ to denote the density of $\mathcal{R}(a,b)$. Also, the kth fold calculation of $\mathcal{R}$ $\mathcal{R}(p_a, \ldots, \mathcal{R}(p_a, \mathcal{R}(p_a, p_a)))$ is denoted by $\mathcal{R}^k(p_a)$. $\mathcal{R}$ is efficiently calculated with a look-up table.

The evolution of the input densities $f_V^{(\ell)}(0)$ and $f_V^{(\ell)}(1)$ through the check node of degree d_c is then performed as follows:

$$f_U^{(\ell)}(1) = \frac{1}{2^{d_c-2}} \sum_{\substack{v=1 \\ v \text{ odd}}}^{d_{c-1}} \binom{d_c-1}{v} \mathcal{R}\left(\mathcal{R}^{d_c-1-v}\left(f_V^{(\ell)}(0)\right), \mathcal{R}^{v}\left(f_V^{(\ell)}(1)\right)\right), \tag{B.12}$$

$$f_U^{(\ell)}(0) = \frac{1}{2^{d_c-2}} \sum_{\substack{v=1 \\ v \text{ even}}}^{d_{c-1}} \binom{d_c-1}{v} \mathcal{R}\left(\mathcal{R}^{d_c-1-v}\left(f_V^{(\ell)}(0)\right), \mathcal{R}^{v}\left(f_V^{(\ell)}(1)\right)\right) \tag{B.13}$$

Clearly, this operation becomes fairly complex as the maximum degree of check nodes increases, but is perfectly tractable for degrees less than 20.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2nd ed., 2006.

[2] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. CRC Press, Inc., 5th ed., October 1996.

[3] A. Hodjat and I. Verbauwhede, "Area-throughput trade-offs for fully pipelined 30 to 70 gbits/s aes processors," *IEEE Transactions on Computers*, vol. 55, pp. 366–372, April 2006.

[4] C. E. Shannon, "Communication theory of secrecy systems," *Bell Systems Technical Journal*, vol. 28, pp. 656–715, 1948.

[5] A. D. Wyner, "The Wire-Tap Channel," *The Bell System Technical Journal*, vol. 54, pp. 1355–1367, October 1975.

[6] I. Csiszár and J. Körner, "Broadcast Channels with Confidential Messages," *IEEE Trans. Inf. Theory*, vol. 24, pp. 339–348, May 1978.

[7] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian Wire-Tap Channel," *IEEE Trans. Inf. Theory*, vol. 24, pp. 451–456, July 1978.

[8] Y. Liang and V. H. Poor, "Secure communication over fading channels.," in *Proc. of 44th Allerton Conference on Communication, Control and Computing*, (Urbana, IL, USA), September 2006.

[9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akademiai Kiado, December 1997.

[10] M. N. Wegman and J. Carter, "New hash functions and their use in authentication and set equality," *Journal of Computer Sciences and Systems*, vol. 22, pp. 265–279, 1981.

[11] U. M. Maurer and S. Wolf, "Information-Theoretic Key Agreement: From Weak to Strong Secrecy for Free," in *Advances in Cryptology - Eurocrypt 2000*, p. 351, Lecture Notes in Computer Science, B. Preneel, 2000.

[12] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *Proc. IEEE International Symposium on Information Theory*, (Seattle, USA), pp. 356–360, 2006.

[13] Y. Liang, H. V. Poor, and S. Shamai, "Secrecy capacity region of fading broadcast channels," in *Proc. IEEE International Symposium on Information Theory*, (Nice, France), June 2007.

[14] Z. Li, R. Yates, and W. Trappe, "Secrecy capacity of independent parallel channels," in *Proc. of 44th Annual Allerton Conference*, (Monticello, IL, USA), September 2006.

[15] P. K. Gopala, L. Lai, and H. El-Gamal, "On the secrecy capacity of fading channels," in *Proc. of IEEE International Symposium on Information Theory*, (Nice, France), June 2007.

[16] F. Oggier and B. Hassibi, "The secrecy capacity of the mimo wiretap channel," in *Proc. 2007 Allerton Conference on Communication, Control and Computing*, (Allerton, IL, USA), 2007.

[17] A. Khisti and G. Wornell, "The mimome channel," in *Proc of the 45th Allerton Conference on Communication, Control and Computing*, 2007.

[18] Y. Liang and V. H. Poor, "Generalized multiple access channels with confidential messages," in *Proc. IEEE International Symposium on Information Theory*, pp. 952–956, July 2006.

[19] E. Tekin and A. Yener, "The multiple access wire-tap channel: Wireless secrecy and cooperative jamming," in *Information Theory and Applications Workshop*, (San Diego, CA, USA), 2007.

[20] L. Lai and H. El-Gamal, "Cooperative secrecy: The relay-eavesdropper channel," in *Proc. of IEEE International Symposium on Information Theory*, (Nice, France), July 2007.

[21] Y. Oohama, "Coding for relay channels with confidential messages," in *Proc. IEEE Information Theory Workshop*, (Cairns, Australia), pp. 87–89, September 2001.

[22] R. Liu, I. Marić, P. Spasojević, and R. D. Yates, "Discrete memoryless interference and broadcast channels with confidential messages," in *Proc. of 44th Allerton Conference on Communication, Control and Computing*, (Urbana, IL, USA), September 2006. see also arXiv:cs/0702099.

[23] H. Yamamoto, "Rate-distortion theory for the shannon cipher system," *IEEE Trans. Info. Theory*, vol. 43, no. 3, pp. 827–835, 1997.

[24] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, pp. 733–742, May 1993.

[25] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. i. secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, pp. 1121–1132, July 1993.

[26] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3047–3061, Dec. 2004.

[27] U. M. Maurer and S. Wolf, "Unconditionally Secure Key Agreement and Intrinsic Conditional Information," *IEEE Trans. Inf. Theory*, vol. 45, pp. 499–514, March 1999.

[28] W. K. Wooters and W. H. Zurek, "A single quantum cannot be cloned," *Nature*, vol. 299, pp. 802–803, October 1982.

[29] J. J. Sakurai, *Modern Quantum Mechanics*. Addison Wesley Longman, revised edition ed., 1994.

[30] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum Cryptography," *Rev. Mod. Phys.*, vol. 74, pp. 145–195, January 2002.

[31] V. Wei, "Generalized hamming weights for linear codes," *IEEE Transactions on Information Theory*, vol. 37, pp. 1412–1418, Sept. 1991.

[32] A. Thangaraj, S. Dihidar, A. R. Calderbank, S. W. McLaughlin, and J.-M. Merolla, "Applications of ldpc codes to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2933–2945, Aug. 2007.

[33] R. Liu, Y. Liand, H. V. Poor, and P. Spasojević, "Secure nested codes for type ii wiretap channels," in *Proc. 2007 IEEE Information Theory Workshop*, (Lake Tahoe, California, USA), September 2007.

[34] L. H. Ozarow and A. D. Wyner, "Wire Tap Channel II," *AT&T Bell Laboratories Technical Journal*, vol. 63, pp. 2135–2157, December 1984.

[35] W. T. Buttler, S. K. Lamoreaux, J. R. Torgerson, G. H. Nickel, C. H. Donahue, and C. G. Peterson, "Fast, efficient error reconciliation for quantum cryptography," *Phys. Rev. A*, vol. 67, pp. 052303/1–8, May 2003.

[36] G. Van Assche, J. Cardinal, and N. J. Cerf, "Reconciliation of a Quantum-Distributed Gaussian Key," *IEEE Trans. Inf. Theory*, vol. 50, pp. 394–400, February 2004.

[37] G. Brassard and L. Salvail, "Secret-key Reconciliation by Public Discussion," in *Advances in Cryptology-Eurocrypt'93* (T. Helleseth, ed.), pp. 411–423, Springer-Verlag, 1993.

[38] D. Slepian and J. K. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Trans. Inf. Theory*, vol. 19, pp. 471–480, July 1973.

[39] C. H. Bennett, G. Brassard, C. Crépeau, and U. Maurer, "Generalized Privacy Amplification," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1915–1923, 1995.

[40] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *Journal of Computer and System Sciences*, vol. 18, pp. 143–154, 1979.

[41] C. Cachin and U. M. Maurer, "Linking Information Reconciliation and Privacy Amplification," *Journal of Cryptology*, vol. 10, pp. 97–110, March 1997.

[42] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved Coded Modulation," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927–946, May 1998.

[43] S. ten Brink, "Iterative demapping and decoding for multilevel modulation," in *Proc. IEEE Globecom Conference*, vol. 1, (Sydney), pp. 579–584, November 1998.

[44] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel Codes: Theoretical Concepts and Practical Design Rules," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1361–1391, July 1999.

[45] T. Wörz and J. Hagenauer, "Iterative decoding for multilevel codes using Reliability information," in *Proc. IEEE Globecom Conference*, (Orlando), December 1992.

[46] A. Wyner, "Recent results in the shannon theory," *IEEE Trans. Inf. Theory*, vol. 20, pp. 2–10, Jan 1974.

[47] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of Binary Sources With Side Information at the Decoder Using LDPC Codes," *IEEE Comm. Lett.*, vol. 6, pp. 440–442, October 2002.

[48] T. J. Richardson and R. L. Urbanke, "The Capacity of Low-Density Parity-Check Codes Under Message-Passing Decoding," *IEEE Trans. Inf. Theory*, vol. 47, pp. 599–618, February 2001.

[49] LTHC, Communications Theory Lab.

[50] S. ten Brink, "Convergence Behavior of Iteratively Decoded Parallel Concatenated Codes," *IEEE Trans. Comm.*, vol. 49, pp. 1727–1737, October 2001.

[51] K.-C. Nguyen, G. Van Assche, and N. J. Cerf, "Side-Information Coding with Turbo Codes and its Application to Quantum Key Distribution," in *Proc. International Symposium on Information Theory and its Applications*, 2004.

[52] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 2nd ed., December 2001.

[53] N. Varnica, X. Ma, and A. Kavcic, "Capacity of power constrained memoryless awgn channels with fixed input constellations," in *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, vol. 2, pp. 1339–1343vol.2, 17-21 Nov. 2002.

[54] D. R. Stinson, "Universal hashing and authentication codes," *Lecture Notes in Computer Science*, vol. 576, pp. 74–85, 1991.

[55] L. Ozarow and A. Wyner, "On the capacity of the gaussian channel with a finite number of input levels," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1426–1428, 1990.

[56] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[57] J. G. Proakis, *Digital Communications*. McGraw-Hill, 4th ed., 2001.

[58] M. Simon and M.-S. Alouini, "Some new results for integrals involving the generalized marcum q function and their application to performance evaluation over fading channels," *IEEE Journal of Wireless Communications*, vol. 2, no. 4, pp. 611–615, 2003.

[59] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, 1979.

[60] Y. Liang and V. V. Veeravalli, "Cooperative relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, pp. 900–928, March 2007.

[61] F. M. Willems, *Information theoretical results for the discrete memoryless multiple access channel*. PhD thesis, Katholieke Universiteit Leuven, 1982. available online http://www.sps.ele.tue.nl/members/F.M.J.Willems/.

[62] C. Crépeau, "Efficient cryptographic protocols based on noisy channels," in *Proc. of EUROCRYPT 1997* (Springer, ed.), pp. 306–317, 1997.

[63] A. Winter, A. C. A. Nascimento, and H. Imai, "Commitment capacity of discrete memoryless channels," in *Proc. of 9th IMA international conference*, (Cirencester, UK), pp. 33–51, 2003.

[64] J. Barros, H. Imai, A. C. A. Nascimento, and S. Skludarek, "Bit commitment over Gaussian channels," in *Proc. of 2006 IEEE International Symposium on Information Theory*, (Seattle, USA), pp. 1437–1441, July 2006.

[65] H. Imai, K. Morozov, A. C. A. Nascimento, and A. Winter, "Efficient protocols achieving the commitment capacity of noisy correlations," in *Proc. of 2006 IEEE International Symposium on Information Theory*, (Seattle, USA), pp. 1432–1436, July 2006.

[66] R. G. Gallager, *Low Density Parity Check Codes*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1963.

[67] D. Moore, G. Voelker, and S. Savage, "Inferring internet denial-of-service activity," *USENIX Security Symposium*, pp. 9–22, 2001.

[68] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Practical network support for ip traceback," *In Proceedings of the 2000 ACM SIGCOMM Conference*, 2000.

[69] A. D. Keromytis, V. Misra, and D. Rubenstein, "Sos: Secure overlay services," *Proceedings of ACM SIGCOMM*, 2002.

[70] S. Kandula, D. Katabi, M. Jacob, and A. Berger, "Botz-4-sale: Surviving organized ddos attacks that mimic flash crowds," *2nd Symposium on Networked Systems Design and Implementation (NSDI)*, May 2005.

[71] R. Narasimha, Z. Chen, and C. Ji, "Topological malware propagation on networks: Spatial dependence and its significance," *submitted*, 2007.

[72] J. Ioannidis and S. M. Bellovin, "Implementing pushback: Router-based defense against ddos attacks," *Proceedings of Network and Distributed System Security Symposium, San Diego, California*, 2002.

[73] A. C. Snoeren, C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio, S. T. Kent, and W. T. Strayer, "Hash-based ip traceback," *Proceeding of ACM/SIGCOMM*, 2001.

[74] Y. Chen, K. Hwang, and W.-S. Ku, "Collaborative detection of ddos attacks over multiple network domains," *IEEE Transactions on Parallel and Distributed Systems*, to appear, 2007.

[75] P. Ferguson and D. Senie, "Network ingress filtering:defeating denial of service attacks which employ ip source address spoofing," *RFC 2827*, May 2000.

[76] N. Weiler, "Honeypots for distributed denial of service attacks," *Proceedings of the 11th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pp. 109 –114, 2002.

[77] R. Stone, "Centertrack: An ip overlay network for tracking dos floods," *USENIX Security Symposium*, 2000.

[78] T. Bu, S. Norden, and T. Woo, "A survivable dos-resistant overlay network," *Computer Networks*, vol. 50, no. 9, pp. 1281–1301, June 2006.

[79] E. Shi, I. Stoica, D. Anderson, and A. Perrig, "Overdose: A generic ddos protection service using an overlay network," *CMU-CS-06-114*, Feb 2006.

[80] T. Bu, S. Norden, and T. Woo, "Trading resiliency for security: Model and algorithms," in *Proc. 12th IEEE International Conference on Network Protocols*, pp. 218–227, 2004.

[81] M. van Dijk, "On a special class of broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 43, pp. 712–714, March 1997.

[82] H. Imai, K. Kobara, and K. Morozov, "On the possibility of key agreement using variable directional antenna," in *in Proc. of 1st Joint Workshop on Information Security*, (Korea), 2006.

[83] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2141–2155, September 2003.

[84] F. Schreckenbach and G. Bauch, "Exit charts for iteratively decoded multilevel modulation," in *Proc. of XII European Signal Processing Conference*, (Vienna, Austria), September 2004.

[85] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of Low-Density Parity-Check Codes for Modulation and Detection," *IEEE Trans. Comm.*, vol. 52, pp. 670–678, April 2004.

[86] G. Lechner, J. Sayir, and I. Land, "Optimization of ldpc codes for receiver frontends," in *Information Theory, 2006 IEEE International Symposium on*, pp. 2388–2392, July 2006.

[87] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes," *IEEE Trans. Inf. Theory*, vol. 47, pp. 619–637, February 2001.

[88] A. Kavčič, X. Ma, and M. Mitzenmacher, "Binary Intersymbol Interference Channels: Gallager Codes, Density Evolution, and Code Performance Bounds," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1636–1651, July 2003.

[89] C.-C. Wang, S. R. Kulkarni, and H. V. Poor, "Density Evolution for Asymmetric Memoryless Channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4216–4236, December 2005.

[90] S.-Y. Chung, J. G. David Forney, T. J. Richardson, and R. Urbanke, "On the Design of Low-Density Parity-Check Codes within 0.0045 dB of the Shannon Limit," *IEEE Comm. Lett.*, vol. 5, pp. 58–60, February 2001.

VITA

Matthieu Bloch received the Engineering degree from Supélec (France) and the M.S. in Electrical Engineering from the Georgia Institute of Technology in 2003. In December 2006, he received the Ph.D. in Engineering Science from the Université de Franche-Comté, where his doctoral research explored the design of robust and efficient quantum key distribution systems. In 2008, he received the Ph.D. degree in Electrical and Computer Engineering from the Georgia Institute of Technology. His research interests lie in the area of quantum cryptography and information theory, with an emphasis on the design of practical coding schemes ensuring information-theoretic security.