# INFORMATION-THEORETIC SECRET SHARING: FUNDAMENTAL LIMITS AND CODING SCHEMES VIA DEEP LEARNING

A Dissertation by

Vidhi Rana

Master of Technology, Govind Ballabh Pant Engineering College, 2016

Bachelor of Technology, College of Engineering Roorkee, 2013

Submitted to the School of Computing
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

December 2023

INFORMATION-THEORETIC SECRET SHARING: FUNDAMENTAL LIMITS AND CODING SCHEMES VIA DEEP LEARNING

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Electrical and Computer Engineering.

Rémi A. Chou, Committee Chair

Edwin Sawan, Committee Member

Abu Asaduzzaman, Committee Member

Alexander Boukhgueim, Committee Member

Ajita Rattani, Committee Member

Accepted for the College of Engineering

Anthony Muscat, Dean

Accepted for the Graduate School

Coleen Pugh, Dean

# ACKNOWLEDGMENTS

I have received a great deal of support and assistance throughout writing this dissertation proposal. Thank you to my advisor, Dr. Remi Chou, for his patience, guidance, and support. I have significantly benefited from his wealth of knowledge and meticulous editing. I am incredibly grateful that he took me on as a student and continued to have faith in me over the years.

Thank you to my committee members, Dr. Edwin Sawan, Dr. Abu Asaduzzaman, Dr. Alexander Boukhgueim, and Dr. Ajita Rattani, for their encouraging words and thoughtful, detailed feedback.

Thank you to my son and my friends, who provided a refreshing environment and happy distractions to rest my mind outside of my research.

Thank you to my husband for the sacrifices he made for me during this journey.

Thank you to my parents for their tremendous support.

# ABSTRACT

This dissertation aims to study and design coding schemes for information-theoretic security. We focus on two models: the secret sharing model and the Gaussian wiretap channel model. The main contribution of this dissertation is to take practical constraints into account. We consider a rate-limited public communication channel to account for bandwidth constraints and finite blocklength for practical applications requiring short packet length or low latency.

## TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

# LIST OF FIGURES

Figure Page

LIST OF FIGURES (continued)

# CHAPTER 1

# INTRODUCTION

Cryptography is the traditional practice and study of schemes for secure communication in the presence of an adversary.

Traditional cryptographic algorithms rely on the assumption of computational complexity, making them hard to break in actual practice by an adversary. In contrast, there exists a class of cryptographic algorithms that rely on an information-theoretic approach that cannot be broken with unlimited computational power. One such example is a one-time pad, which has been proved information-theoretically secure against an unbounded adversary, meaning that an encrypted message provides no information about the confidential message. This notion of security is developed and proved to be true for one-time pad by Claude Shannon.

Later, the information-theoretic approach is extended to wireless communication by leveraging the channel noise. Wyner introduces a wiretap channel, a basic model to account for adversaries in wireless communication. Further, this approach is formulated for the source model, where a sender and legitimate receiver will generate a secret key instead of sending a confidential message (in the channel model).

Furthermore, applying the information-theoretic approach to the problem of secret sharing has been a challenging task. The most famous secret sharing scheme in cryptography is Shamir's secret sharing and is formalized as follows. Consider secret $S$ is divided into $L$ participants such that $S$ can be reconstructed from any $t$ participants, but even complete knowledge of $t-1$ participants reveals nothing about $S$. This scheme enables the construction of robust key management strategies required for encryption. However, in this scheme it is assumed that participants can communicate over an information-theoretically secure channel at no cost. But we are interested in approach that aims at providing a full information-theoretic solution that would not rely on complexity-based cryptography. In other words, we want to avoid the assumption that information-theoretically secure communication channels are available at no cost.

From a practical point of view, taking constraints such as finite blocklength and arbitrarily varying channels into account is a challenging problem. Solving this task directs our interest in a deep learning-based information-theoretic approach.

This dissertation aims to study and design coding schemes under information-theoretic guarantees. We focus on the secret sharing model and the Gaussian wiretap channel model. For the secret-sharing model, we consider two different settings: a) secret sharing with Gaussian sources b) secret sharing over a Gaussian broadcast channel.

The main contribution of this dissertation is to take practical constraints such as rate-limited public communication and finite blocklength.

## 1.1 Outline and Publications

In Chapter 2, we study an information-theoretic secret sharing problem, where a dealer distributes shares of a secret among a set of participants under the following constraints: (i) authorized sets of users can recover the secret by pooling their shares, and (ii) non-authorized sets of colluding users cannot learn any information about the secret. The dealer and the participants observe the realizations of correlated Gaussian random variables and that the dealer can communicate with the participants through a one-way, authenticated, rate-limited, and public channel. Our main result is a closed-form characterization of the fundamental trade-off between secret rate and public communication rate. Chapter 2 is based on the following references [1, 2]:

- V. Rana, R. A. Chou, and H. Kwon, "Information-theoretic secret sharing from correlated Gaussian random variables and public communication" *IEEE Transactions on Information Theory*, vol. 68, no. 1, pp. 549–559, 2021.

- V. Rana, R. A. Chou, and H. Kwon, "Secret sharing from correlated Gaussian random variables and public communication." *2020 IEEE Information Theory Workshop (ITW).*

In Chapter 3, we design short blocklength codes for the Gaussian wiretap channel under information-theoretic security guarantees. Our approach consists in decoupling the reliability and secrecy constraints in our code design. Specifically, we handle the reliability constraint via an autoencoder, and handle the secrecy constraint with hash functions. For blocklengths smaller than or equal to $128$, we evaluate through simulations the probability of error at the legitimate receiver and the leakage at the eavesdropper for our code construction. This leakage is defined as the mutual information between the confidential message and the eavesdropper's channel observations, and is empirically measured via a neural network-based mutual information estimator. Our simulation results provide examples of codes with positive secrecy rates that outperform the best known achievable secrecy rates obtained non-constructively for the Gaussian wiretap channel. Additionally, we show that our code design is suitable for the compound and arbitrarily varying Gaussian wiretap channels, for which the channel statistics are not perfectly known but only known to belong to a pre-specified uncertainty set. These models not only capture uncertainty related to channel statistics estimation, but also scenarios where the eavesdropper jams the legitimate transmission or influences its own channel statistics by changing its location. Chapter 3 is based on following references [3, 4]:

- V. Rana and R. A. Chou, "Short blocklength wiretap channel codes via deep learning: Design and performance evaluation," *IEEE Transactions on Communications*, vol. 71, no. 3, pp. 1462–1474, 2023.

- V. Rana and R. A. Chou, "Design of short blocklength wiretap channel codes: Deep learning and cryptography working hand in hand." *2021 IEEE Information Theory Workshop (ITW).*

In Chapter 4, we consider a secret-sharing model where a dealer shares a secret with several participants through a Gaussian broadcast channel such that predefined subsets of participants can reconstruct the secret, and all other subsets of participants cannot learn any information about the secret. Our main contribution is to design a two-layer coding scheme, that rely on two coding layers, namely, a reliability layer and a secrecy layer, where the reliability layer is a channel code for a compound channel without any security constraint at short blocklength. Specifically, we design the reliability layer via an autoencoder and implement the secrecy layer with hash functions. To evaluate the performance of our coding scheme, we evaluate the probability of error and information leakage, which is defined as the mutual information between the secret and the unauthorized sets of users channel outputs. We empirically evaluate this information leakage via a neural network-based mutual information estimator. Our simulation results demonstrate a precise control of the probability of error and leakage thanks to the two-layer coding design. Chapter 4 is based on the following reference.

- R. Sultana, V. Rana and R. A. Chou, ”Secret Sharing Over a Gaussian Broadcast Channel: Optimal Coding Scheme Design and Deep Learning Approach at Short Blocklength,” in *2023 IEEE International Symposium on Information Theory (ISIT)*, Taipei, Taiwan, 2023, pp. 1961-1966.

# CHAPTER 2

# INFORMATION-THEORETIC SECRET SHARING FROM CORRELATED GAUSSIAN RANDOM VARIABLES AND PUBLIC COMMUNICATION

## 2.1 Introduction

Secret sharing has been introduced in [5], [6]. In basic secret-sharing models, a dealer distributes a secret among a set of participants, with the constraint that only pre-defined sets of participants can recover this secret by pooling their shares, while any other set of colluding participants cannot learn any information about the secret.

In most secret-sharing models, including Shamir's scheme [5], it is assumed that the dealer and each participant can communicate over an information-theoretically secure channel at no cost. While complexity-based cryptography techniques, e.g., [7], could be used to implement secure channels without any other resources than a public channel, it would not provide information-theoretically secure channels. In this chapter, we are interested in *another approach that aims at providing a full information-theoretic solution that would not rely on complexity-based cryptography*. In other words, we want to avoid the assumption that information-theoretically secure communication channels are available at no cost. An information-theoretic approach to secret sharing over wireless channels has been introduced in [8] for this purpose. The main idea is to leverage channel noise by remarking that information-theoretic secret sharing over wireless channels is similar to compound wiretap channel models [9]. This information-theoretic approach has also been formulated for source models in [10–12], where participants and dealers share correlated random variables. These models are related to compound secret-key generation, e.g., [13, 14], secure source coding with a multiuser access structure [15], and biometric systems with a multiuser access structure [16], in that multiple reconstruction and security/privacy constraints need to be satisfied simultaneously.

In this chapter, we consider the information-theoretic secret sharing model in [11] with Gaussian sources. Specifically, the dealer and the participants observe realizations of correlated Gaussian random variables, and the dealer can communicate with the participants over an authenticated, one-way, rate-limited, and public communication channel. In wireless networks, independently and identically distributed realizations of correlated random variables can, for instance, be obtained from channel gain measurements after appropriate manipulations [17, 18]. Our approach for the achievability part consists in handling the reliability and security requirements separately. Specifically, reliability is obtained via a coding scheme akin to a compound version of Wyner-Ziv coding [19], and security relies on universal hashing via extractors [20]. Interestingly, the converse shows that there is no loss of optimality in decoupling the reliability and security requirements. The achievability is first obtained for discrete random variables and then extended to continuous random variables via fine quantization. In principle, one cannot assume a specific quantization strategy to ensure the security requirement in an information-theoretic manner; hence, the key step in this extension is to show that information-theoretic security holds, provided that the quantization is sufficiently fine. For the converse part, we can partly rely on techniques developed in [21], [22]. However, unlike in [21], [22], our setting involves multiple security constraints that need to be satisfied simultaneously; hence, the main task in the converse is to prove a saddle point property without any degradation assumption on the source model.

The main differences between our work and [11,13,14,16] are that [11,13,14,16] consider discrete memoryless sources, whereas we consider Gaussian sources. As described above, handling Gaussian random variables calls for different proof techniques and considerations. Additionally, unlike [11, 13, 14, 16], it also allows us to derive capacity results without assuming any source degradation properties. We also highlight that unlike [11, 14], we consider rate-limited public communication, and unlike [11,16], we handle arbitrary access structures.

The main features of our work can be summarized as follows: (i) Our model relies on correlated Gaussian random variables and, similar to [11] but unlike traditional secret-sharing schemes [5], does not rely on the assumption that information-theoretically secure channels between the dealer and the participants are available. (ii) Similar to the model in [11] but unlike traditional secret-sharing models, we consider a model that requires information-theoretic security for the secret with respect to unauthorized sets of participants during the distribution phase, i.e., when the dealer distributes shares of the secret to participants. (iii) We establish a closed-form expression that characterizes the optimal trade-off between secret rate and public communication rate. (iv) The size of the shares in our coding scheme scales linearly with the size of the secret for any access structure similar to the model in [11]. Indeed, a share comprises the public communication from the dealer and $n$ quantized realizations of a Gaussian random variable, which can be shown to both linearly scale with $n$. The size of the shares does depend on the specific access structure considered but not on the number of participants. Specifically, the public communication must ensure that the set of authorized users with the least amount of information about the secret is able to reconstruct the secret. By contrast, the best-known traditional secret-sharing schemes may require a share size that grows exponentially with the number of the participants for some access structures [23] – note, however, that it is unknown whether or not there exist traditional secret-sharing schemes that require a smaller share size. (v) For threshold access structures, i.e., when a fixed number of participants $t$ is needed to reconstruct the secret (independently from the specific identities of those participants), we establish that the size of the secret that can be exchanged is, in general, *not* a monotonic function of the threshold $t$.

The remainder of the chapter is organized as follows. We set the notation in Section 2.2 and formally introduce the problem statement in Section 2.3. We present our main results in Section 2.4, and proofs in Sections 2.5 and 2.6. Finally, we provide concluding remarks in Section 2.7.

### 2.2 Notation

For any $a, b \in \mathbb{R}$, define $[\![a, b]\!] \triangleq [\lfloor a \rfloor, \lceil b \rceil] \cap \mathbb{N}$. For $x \in \mathbb{R}$, define $[x]^+ \triangleq \max(0, x)$. For a set $\mathcal{S}$, let $2^{\mathcal{S}}$ denote the power set of $\mathcal{S}$. All logarithms are taken in base 2 throughout the chapter. Let $I_m$ denote the identity matrix of dimension $m \in \mathbb{N}$. Let $\det(W)$ denote the determinant of a matrix $W$ and $|\mathcal{S}|$ denote the cardinality of a set $\mathcal{S}$. For two random variables $X$ and $V$, $\sigma_X^2$ and $\sigma_{XV}$ denote $\mathbb{E}[(X - \mathbb{E}[X])^2]$ and $\mathbb{E}[(X - \mathbb{E}[X])(V - \mathbb{E}[V])]$, respectively. $N \sim \mathcal{N}(0, \Sigma)$ indicates that $N$ is a zero-mean Gaussian random vector with covariance matrix $\Sigma$. The indicator function is denoted by $\mathbb{1}\{\omega\}$, which is equal to $1$ if the predicate $\omega$ is true and $0$ otherwise. Let $H(X)$ (respectively, $h(X)$) denote the Shannon entropy (respectively, the differential entropy) of a discrete (respectively continuous), random variable $X$. Also, let $I(X; Y)$ denote the mutual information between $X$ and $Y$, which are either continuous or discrete random variables.

### 2.3 Problem Statement

Consider a dealer and $L$ participants. Define $\mathcal{L} \triangleq [\![1, L]\!]$, $\mathcal{X} \triangleq \mathbb{R}$, and $\mathcal{Y} \triangleq \mathbb{R}$. Consider a Gaussian memoryless source model $(\mathcal{X} \times \mathcal{Y}_{\mathcal{L}}, p_{XY_{\mathcal{L}}})$, where $Y_{\mathcal{L}} \triangleq (Y_l)_{l \in \mathcal{L}}$ and $(X, Y_{\mathcal{L}})$ are jointly Gaussian random variables with a non-singular covariance matrix. Let $\mathbb{A}$ be a set of subsets of $\mathcal{L}$ such that for any $\mathcal{T} \subseteq \mathcal{L}$, if $\mathcal{T}$ contains a set that belongs to $\mathbb{A}$, then $\mathcal{T}$ also belongs to $\mathbb{A}$, i.e., $\mathbb{A}$ is a monotone access structure [24]. We also define $\mathbb{U} \triangleq 2^{\mathcal{L}} \backslash \mathbb{A}$ as the set of all colluding subsets of users who must not learn any information about the secret. In the following, for any $\mathcal{A} \in \mathbb{A}$ and for any $\mathcal{U} \in \mathbb{U}$, we use the notation $Y_{\mathcal{A}}^n \triangleq (Y_l^n)_{l \in \mathcal{A}}$ and $Y_{\mathcal{U}}^n \triangleq (Y_l^n)_{l \in \mathcal{U}}$. Moreover, we assume that the dealer can communicate with the participants over an authenticated, one-way, rate-limited, noiseless, and public communication channel.

**Definition 2.1.** *A $(2^{nR_s}, R_p, \mathbb{A}, n)$ secret-sharing strategy is defined as follows:*

- *The dealer observes $X^n$ and Participant $l \in \mathcal{L}$ observes $Y_l^n$.*

- *The dealer sends over the public channel the message $M$ to the participants with the band-*

*width constraint $H(M) \leq nR_p$.*

- *The dealer computes a secret $S \in \mathcal{S} \triangleq [\![1, 2^{nR_s}]\!]$ from $X^n$.*
- *Any subset of participants $\mathcal{A} \in \mathbb{A}$ can compute an estimate $\widehat{S}(\mathcal{A})$ of $S$ from their observations $(Y_l^n)_{l \in \mathcal{A}}$ and $M$.*

**Definition 2.2.** *A rate pair $(R_p, R_s)$ is achievable if there exists a sequence of $(2^{nR_s}, R_p, \mathbb{A}, n)$ secret-sharing strategies such that*

$$\begin{aligned} \lim_{n \to \infty} \max_{\mathcal{A} \in \mathbb{A}} \mathbb{P}[\widehat{S}(\mathcal{A}) \neq S] &= 0, & (2.1) \\ \lim_{n \to \infty} \max_{\mathcal{U} \in \mathbb{U}} I(S; M, Y_{\mathcal{U}}^n) &= 0, & (2.2) \\ \lim_{n \to \infty} \log |\mathcal{S}| - H(S) &= 0. & (2.3) \end{aligned}$$

(2.1) means that any subset of participants in $\mathbb{A}$ is able to recover the secret, (2.2) means that any subset of participants in $\mathbb{U}$ cannot obtain information about the secret, while (2.3) means that the secret is nearly uniform and that its entropy is nearly equal to its length.

**Remark 2.1.** *The uniformity condition (2.3) ensures that a secret-sharing strategy that maximizes the length of the secret, will also maximize the entropy of the secret. Without this condition, maximizing the length of the secret would not be meaningful as one could always increase the length of the secret by adding redundancy to it. This is the same reason why in secret-key generation, one requires uniformity of the secret key [25, 26].*

The secret capacity region is defined as

$$\mathcal{R}(p_{XY_{\mathcal{L}}}, \mathbb{A}) \triangleq \{(R_p, R_s) : (R_p, R_s) \text{ is achievable}\}.$$

Moreover, for a fixed $R_p$, the supremum of secret rates $R_s$ such that $(R_p, R_s) \in \mathcal{R}(p_{XY_{\mathcal{L}}}, \mathbb{A})$ is called the secret capacity and is denoted by $C_s(\mathbb{A}, R_p)$.

Additionally, one can write for any $\mathcal{A} \in \mathbb{A}$ and for any $\mathcal{U} \in \mathbb{U}$ (see appendix A for the derivation)

$$Y_{\mathcal{A}} = H_{\mathcal{A}} X + W_{Y_{\mathcal{A}}}, \tag{2.4}$$

$$Y_{\mathcal{U}} = H_{\mathcal{U}} X + W_{Y_{\mathcal{U}}}, \tag{2.5}$$

where $H_{\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}| \times 1}$, $H_{\mathcal{U}} \in \mathbb{R}^{|\mathcal{U}| \times 1}$, $W_{Y_{\mathcal{A}}} \sim \mathcal{N}(0, I_{|\mathcal{A}|})$, and $W_{Y_{\mathcal{U}}} \sim \mathcal{N}(0, I_{|\mathcal{U}|})$.

## 2.4 Main Results

### 2.4.1 Results for general access structures

For a given access structure $\mathbb{A}$, define $\mathcal{A}^\star \in \arg\min_{\mathcal{A} \in \mathbb{A}} H_{\mathcal{A}}^T H_{\mathcal{A}}$ and $\mathcal{U}^\star \in \arg\max_{\mathcal{U} \in \mathbb{U}} H_{\mathcal{U}}^T H_{\mathcal{U}}$.

**Theorem 2.1.** *For any access structure $\mathbb{A}$ and public communication rate $R_p \geq 0$, the secret capacity $C_s(\mathbb{A}, R_p)$ is*

$$C_s(\mathbb{A}, R_p) = \left[ \frac{1}{2} \log \frac{\sigma_X^2 H_{\mathcal{U}^\star}^T H_{\mathcal{U}^\star} 2^{-2R_p} + \sigma_X^2 H_{\mathcal{A}^\star}^T H_{\mathcal{A}^\star} (1 - 2^{-2R_p}) + 1}{\sigma_X^2 H_{\mathcal{U}^\star}^T H_{\mathcal{U}^\star} + 1} \right]^+ .$$

*Proof.* The converse and achievability are proved in Sections 2.5 and 2.6, respectively. □

From Theorem 2.1, we obtain the following corollary when the public communication is rate-unlimited.

**Corollary 2.1.** *For any access structure $\mathbb{A}$, and an unlimited public communication rate, the secret capacity is given by*

$$C_s(\mathbb{A}, R_p = +\infty) \triangleq \lim_{R_p \to +\infty} C_s(\mathbb{A}, R_p) = \left[ \frac{1}{2} \log \frac{\sigma_X^2 H_{\mathcal{A}^\star}^T H_{\mathcal{A}^\star} + 1}{\sigma_X^2 H_{\mathcal{U}^\star}^T H_{\mathcal{U}^\star} + 1} \right]^+ .$$

Note that in Theorem 2.1 and Corollary 2.1, the length of the public communication scales linearly with the length of the secret by construction and corresponds to a compressed version of the $n$ source observations of the dealer via a compound version of Wyner-Ziv coding. Hence, the size of the share of each participant, which comprises the public communication and $n$ quantized observations of a Gaussian random variable, scales linearly with the length of the secret – as explained in the proof of Theorem 2.1, the number of bits needed to store quantized realizations of Gaussian random variables is negligible compared to the number of source observations $n$ in our achievability scheme. Note that, unlike traditional secret-sharing models, which separately consider the share-creation phase and the share-distribution phase, we allow a joint design of these two phases in our setting. This is made possible by considering correlated random variables (at the participants and the dealer) and public communication instead of information-theoretically secure channels as in traditional secret-sharing models.

The following example illustrates Theorem 2.1 and Corollary 2.1.

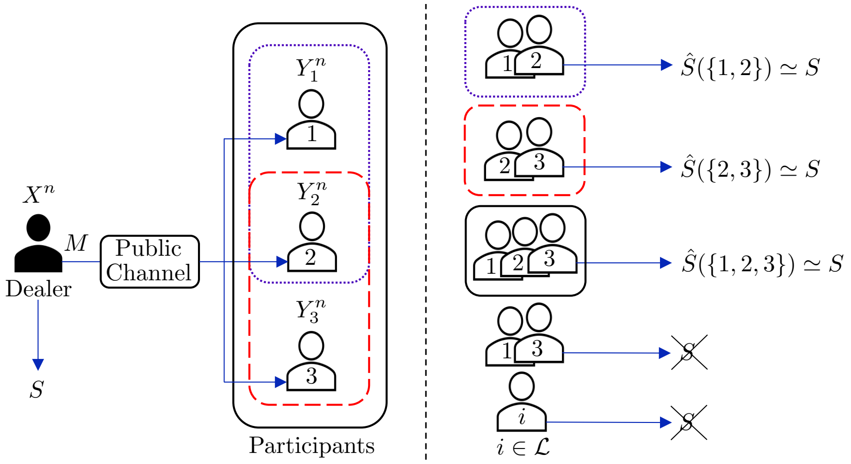
**Example 2.1.** *Consider a dealer and three participants who observe independently and identically distributed realizations of correlated Gaussian random variables as depicted in Figure 2.1. Define the access structure* $\mathbb{A} \triangleq \{\{1,2\},\{2,3\},\{1,2,3\}\}$ *and define* $\mathbb{U} \triangleq \{\{1,3\},\{1\},\{2\},\{3\}\}$ *such that (i) the sets of participants in* $\mathbb{A}$ *can recover the secret using their observations and the public message* $M$*, and (ii) the sets of participants in* $\mathbb{U}$ *cannot learn information about the secret. For* $\sigma_X^2 \triangleq 2$ *and* $H_{\mathcal{L}} \triangleq [0.5, 1, 0.8]^T$*, one can compute the secret capacity using Theorem 2.1 and Corollary 2.1, as shown in Figure 2.2.*

### 2.4.2 Results for threshold access structures

We now consider a special kind of access structure called a threshold access structure [5]. A threshold access structure with threshold $t \in [\![1, L]\!]$ is defined as

$$\mathbb{A}_t \triangleq \{\mathcal{A} \subseteq \mathcal{L} : |\mathcal{A}| \geq t\}.$$

The complement of $\mathbb{A}_t$ is defined as $\mathbb{U}_t \triangleq 2^{\mathcal{L}} \backslash \mathbb{A}_t = \{\mathcal{A} \subseteq \mathcal{L} : |\mathcal{A}| < t\}$. In other words, the threshold access structure is defined such that any set of $t$ participants can reconstruct the secret,



(a) Distribution phase (b) Reconstruction phase

Figure 2.1: Secret-sharing setting when $\mathbb{A} = \{\{1,2\},\{2,3\},\{1,2,3\}\}$ and $\mathbb{U} = \{\{1,3\},\{1\},\{2\},\{3\}\}$. Dashed, dotted, and solid contour lines represent the subsets of participants that are authorized to reconstruct the secret.

but no set of fewer than $t$ participants can learn information about the secret.

The following result provides necessary and sufficient conditions to determine whether the secret capacity increases or decreases as the threshold $t$ increases.

**Theorem 2.2.** *For any $t \in [\![1, L]\!]$, consider $\mathcal{A}_t^\star \in \arg\min_{\mathcal{A}\in\mathbb{A}_t} H_{\mathcal{A}}^T H_{\mathcal{A}}$, and $\mathcal{U}_t^\star \in \arg\max_{\mathcal{U}\in\mathbb{U}_t} H_{\mathcal{U}}^T H_{\mathcal{U}}$. For any communication rate $R_p \geq 0$, for any $t \in [\![1, L]\!]$, we have*

$$C_s(\mathbb{A}_1, R_p) \geq C_s(\mathbb{A}_t, R_p),$$

*and for any $t \in [\![1, L]\!]$ and $i \in [\![1, L-t]\!]$,*

$$C_s(\mathbb{A}_t, R_p) \geq C_s(\mathbb{A}_{t+i}, R_p) \iff \frac{H_{\mathcal{U}_{t+i}^\star}^T H_{\mathcal{U}_{t+i}^\star} - H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star}}{H_{\mathcal{A}_{t+i}^\star}^T H_{\mathcal{A}_{t+i}^\star} - H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star}} \geq \frac{1 + \sigma_X^2 H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star}}{1 + \sigma_X^2 H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star}}.$$

*Proof.* See Appendix B. □

Theorem 2.2 illustrates the fact that the secret capacity is not necessarily a monotonic decreasing function of the threshold $t$.

**Example 2.2.** *Consider a dealer and five participants. For $\sigma_X^2 \triangleq 2$, $H_{\mathcal{L}} \triangleq$*

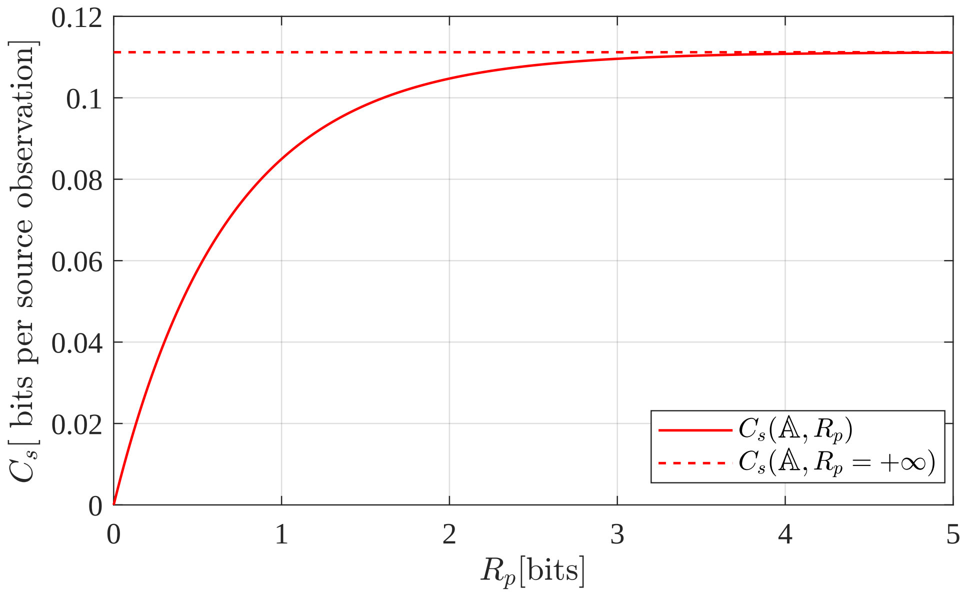


Figure 2.2: Secret capacity for Example 2.1.

$[1, 0.85, 0.9, 0.95, 0.75]^T$*, one can compare the secret capacities for different thresholds using Theorem 2.2, as shown in Figure 2.3.*

*From the definition of* $\mathcal{A}_t^\star$ *and* $\mathcal{U}_t^\star$*, we have* $H_{\mathcal{A}_1^\star} = [0.75]^T$*,* $H_{\mathcal{A}_2^\star} = [0.75, 0.85]^T$*,* $H_{\mathcal{A}_3^\star} = [0.75, 0.85, 0.9]^T$*,* $H_{\mathcal{A}_4^\star} = [0.75, 0.85, 0.9, 0.95]^T$*,* $H_{\mathcal{A}_5^\star} = [0.75, 0.85, 0.9, 0.95, 1]^T$*,* $H_{\mathcal{U}_2^\star} = [1]^T$*,* $H_{\mathcal{U}_3^\star} = [1, 0.95]^T$*,* $H_{\mathcal{U}_4^\star} = [1, 0.95, 0.9]^T$*, and* $H_{\mathcal{U}_5^\star} = [1, 0.95, 0.9, 0.85]^T$*.*

*For example, putting* $H_{\mathcal{A}_4^\star}^T H_{\mathcal{A}_4^\star} = 2.9975$*,* $H_{\mathcal{U}_4^\star}^T H_{\mathcal{U}_4^\star} = 2.7125$*,* $H_{\mathcal{A}_5^\star}^T H_{\mathcal{A}_5^\star} = 3.9975$*, and* $H_{\mathcal{U}_5^\star}^T H_{\mathcal{U}_5^\star} = 3.4350$ *in Theorem 2.2 with* $t = 4$ *and* $i = 1$*, we get* $C_s(\mathbb{A}_4, R_p) \leq C_s(\mathbb{A}_5, R_p)$ *for any* $R_p \geq 0$*.*

## 2.5 Converse Proof of Theorem 2.1

To prove the converse, we first derive an upper bound on the secret capacity $C_s(\mathbb{A}, R_p)$ by considering a worst-case scenario in terms of a secret-key generation problem. This upper bound takes the form of a minimax optimization problem. We then derive a closed-form expression of this upper bound by proving a minimax theorem.

Define for $\mathcal{A} \in \mathbb{A}, \mathcal{U} \in \mathbb{U}$, $O_\mathcal{A} \triangleq H_\mathcal{A}^T H_\mathcal{A}$, and $O_\mathcal{U} \triangleq H_\mathcal{U}^T H_\mathcal{U}$. Consider $V$ an auxiliary random variable jointly Gaussian with $X$, and let $\sigma_{X|V}^2$ be the conditional variance of $X$ given $V$. Consider

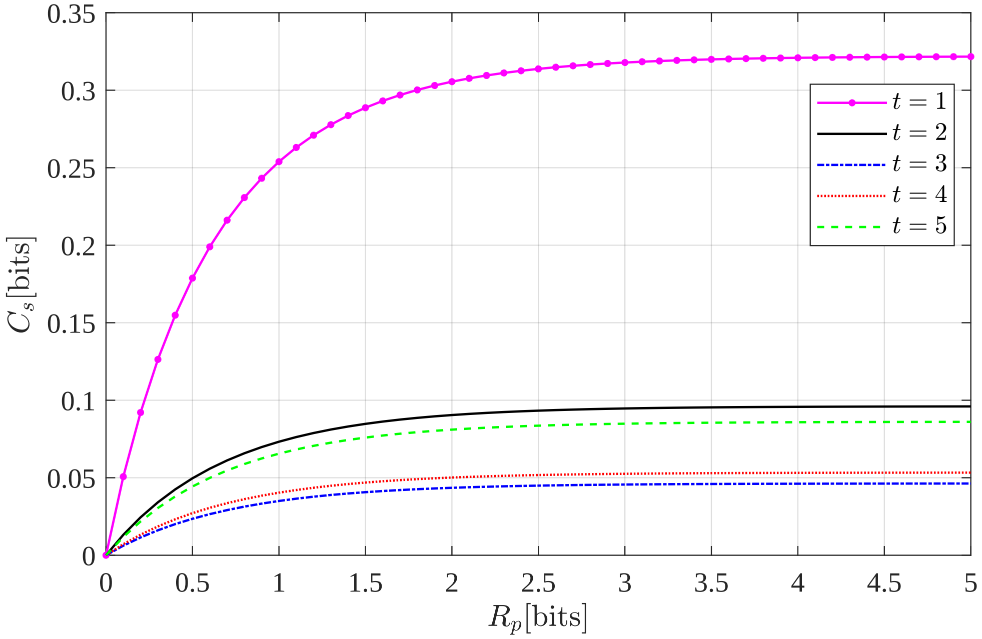


Figure 2.3: Secret capacity for threshold access structure.

also $\mathcal{A}^\star \in \arg\min_{\mathcal{A}\in\mathbb{A}} O_\mathcal{A}$ and $\mathcal{U}^\star \in \arg\max_{\mathcal{U}\in\mathbb{U}} O_\mathcal{U}$. Provided that $\sigma^2_{X|V} \neq 0$, for $\mathcal{A} \in \mathbb{A}, \mathcal{U} \in \mathbb{U}$, define

$$\begin{aligned} I_p(\sigma^2_{X|V}, \mathcal{A}) &\triangleq \frac{1}{2}\log\frac{\sigma^2_X}{\sigma^2_{X|V}} - \frac{1}{2}\log\frac{\sigma^2_X O_\mathcal{A}+1}{\sigma^2_{X|V}O_\mathcal{A}+1}, \\ I_s(\sigma^2_{X|V}, \mathcal{A}, \mathcal{U}) &\triangleq \frac{1}{2}\log\frac{\sigma^2_X O_\mathcal{A}+1}{\sigma^2_{X|V}O_\mathcal{A}+1} - \frac{1}{2}\log\frac{\sigma^2_X O_\mathcal{U}+1}{\sigma^2_{X|V}O_\mathcal{U}+1}. \end{aligned}$$

We will also use the following lemmas.

**Lemma 2.1** (Weinstein–Aronszajn identity, e.g., [27, Appendix B])**.** *For any $\sigma^2 \in \mathbb{R}^+$ and $A \in \mathbb{R}^{q\times 1}$, we have*

$$\det(A\sigma^2 A^T + I_q) = A^T A\sigma^2 + 1.$$

**Lemma 2.2.** *Let $c, d \in \mathbb{R}_+$ such that $c \geq d$. Then, the function $f_{c,d}$ is non-decreasing, where*

$$f_{c,d} : \mathbb{R}_+ \to \mathbb{R}$$

$$x \mapsto \frac{1}{2} \log \frac{cx+1}{dx+1}.$$

*Proof.* The derivative of $f_{c,d}$ at $x \in \mathbb{R}_+$ is $f'_{c,d}(x) = \frac{1}{2 \ln 2} \frac{c-d}{(cx+1)(dx+1)} \geq 0$. □

We now prove the converse of Theorem 2.1 through a series of lemmas.

**Lemma 2.3.** *Let $R_p \in \mathbb{R}_+$. An upper bound on the secret capacity $C_s(\mathbb{A}, R_p)$ for the Gaussian source model $(\mathcal{X} \times \mathcal{Y}_{\mathcal{L}}, p_{XY_{\mathcal{L}}})$ is given by*

$$C_s(\mathbb{A}, R_p) \leq \min_{\mathcal{A} \in \mathbb{A}} \min_{\mathcal{U} \in \mathbb{U}} \max_{\substack{0 < \sigma^2_{X|V} \leq \sigma^2_X \\ s.t.\ I_p(\sigma^2_{X|V}, \mathcal{A}) \leq R_p}} I_s(\sigma^2_{X|V}, \mathcal{A}, \mathcal{U}). \tag{2.6}$$

*Proof.* Fix $\mathcal{A} \in \mathbb{A}$, $\mathcal{U} \in \mathbb{U}$. We first consider the secret-key generation model in [21] consisting of a transmitter (Alice), a receiver (Bob), and an eavesdropper (Eve), who observe $X^n$, $Y^n$, and $Z^n$, respectively, independently and identically distributed according to a Gaussian source $((\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}), p_{XYZ})$, where $\mathcal{X} \triangleq \mathbb{R}$, $\mathcal{Y} \triangleq \mathbb{R}^{|\mathcal{A}|}$, $\mathcal{Z} \triangleq \mathbb{R}^{|\mathcal{U}|}$. In this model, a secret-key rate $R_k$ is achievable if after the transmission from Alice to Bob of message $M$ such that $H(M) \leq nR_p$ over an authenticated noiseless public channel, a secret key $K \in [\![1, 2^{nR_k}]\!]$ is generated by Alice, and an estimate $\widehat{K}$ of $K$ is generated by Bob such that (i) $\lim_{n\to\infty} \mathbb{P}[K \neq \widehat{K}] = 0$ (reliability), (ii) $\lim_{n\to\infty} I(K; Z^n M) = 0$ (security), and $\lim_{n\to\infty} \log\lceil 2^{nR_k} \rceil - H(K) = 0$ (uniformity). Moreover, the capacity region of this model is defined as $\mathcal{R}(p_{XYZ}, \mathcal{A}, \mathcal{U}) \triangleq \{(R_p, R_k) : (R_p, R_k) \text{ is achievable}\}$.

Consider now the secret-sharing problem described in Section 2.3 and the rate pair $(R_p, R_s) \in \mathcal{R}(p_{XY_{\mathcal{L}}}, \mathbb{A})$. Then, by conditions (2.1), (2.2), and (2.3), the rate pair $(R_p, R_s)$ also belongs to $\mathcal{R}(p_{XYZ}, \mathcal{A}, \mathcal{U})$ for any $\mathcal{A} \in \mathbb{A}, \mathcal{U} \in \mathbb{U}$. Therefore, by [21, Theorem 2], we have for any $\mathcal{A} \in \mathbb{A}$, $\mathcal{U} \in \mathbb{U}$,

$$\begin{aligned} R_s &\leq \frac{1}{2} \log \frac{\det(H_{\mathcal{A}} \sigma^2_X H^T_{\mathcal{A}} + I)}{\det(H_{\mathcal{A}} \sigma^2_{X|V} H^T_{\mathcal{A}} + I)} - \frac{1}{2} \log \frac{\det(H_{\mathcal{U}} \sigma^2_X H^T_{\mathcal{U}} + I)}{\det(H_{\mathcal{U}} \sigma^2_{X|V} H^T_{\mathcal{U}} + I)}, \\ R_p &\geq \frac{1}{2} \log \frac{\sigma^2_X}{\sigma^2_{X|V}} - \frac{1}{2} \log \frac{\det(H_{\mathcal{A}} \sigma^2_X H^T_{\mathcal{A}} + I)}{\det(H_{\mathcal{A}} \sigma^2_{X|V} H^T_{\mathcal{A}} + I)}, \end{aligned}$$

for some $\sigma^2_{X|V} \in (0, \sigma^2_X]$. Finally, using Lemma 2.1 and the definition of $O_{\mathcal{A}}$, $\mathcal{A} \in \mathbb{A}$ and $O_{\mathcal{U}}$, $\mathcal{U} \in \mathbb{U}$, we have (2.6). □

**Lemma 2.4.** *Let $R_p \in \mathbb{R}_+$. Let $\mathcal{A} \in \mathbb{A}$, $\mathcal{U} \in \mathbb{U}$, and assume that $O_{\mathcal{A}} \geq O_{\mathcal{U}}$. Then, we have*

$$\max_{\substack{0<\sigma^2_{X|V}\leq\sigma^2_X \\ s.t.\ I_p(\sigma^2_{X|V},\mathcal{A})\leq R_p}} I_s(\sigma^2_{X|V}, \mathcal{A}, \mathcal{U}) = \frac{1}{2}\log\frac{\sigma^2_X O_{\mathcal{U}} 2^{-2R_p} + \sigma^2_X O_{\mathcal{A}}(1-2^{-2R_p}) + 1}{\sigma^2_X O_{\mathcal{U}} + 1}. \tag{2.7}$$

*Proof.* Fix $\mathcal{A} \in \mathbb{A}$ and $\mathcal{U} \in \mathbb{U}$. Let $\sigma^{2\star}_{X|V}(\mathcal{A}, \mathcal{U})$ be an optimal solution on the left-hand side of (2.7). By writing $I_s(\sigma^2_{X|V}, \mathcal{A}, \mathcal{U})$ as

$$I_s(\sigma^2_{X|V}, \mathcal{A}, \mathcal{U}) = \frac{1}{2}\log\frac{\sigma^2_X O_{\mathcal{A}} + 1}{\sigma^2_X O_{\mathcal{U}} + 1} - \frac{1}{2}\log\frac{\sigma^2_{X|V} O_{\mathcal{A}} + 1}{\sigma^2_{X|V} O_{\mathcal{U}} + 1},$$

we have that $I_s(\sigma^2_{X|V}, \mathcal{A}, \mathcal{U})$ is a non-increasing function of $\sigma^2_{X|V}$ by Lemma 2.2 because $O_{\mathcal{A}} \geq O_{\mathcal{U}}$. Hence, $\sigma^{2\star}_{X|V}(\mathcal{A}, \mathcal{U})$ must be the smallest $\sigma^2_{X|V} \in (0, \sigma^2_X]$ that satisfies the constraint $I_p(\sigma^2_{X|V}, \mathcal{A}) \leq R_p$. However, $I_p(\sigma^2_{X|V}, \mathcal{A})$ is a non-increasing function of $\sigma^2_{X|V}$; thus, we must have $I_p(\sigma^{2\star}_{X|V}(\mathcal{A}, \mathcal{U}), \mathcal{A}) = R_p$, i.e.,

$$R_p = \frac{1}{2}\log\frac{\sigma^2_X}{\sigma^{2\star}_{X|V}(\mathcal{A}, \mathcal{U})} - \frac{1}{2}\log\frac{\sigma^2_X O_{\mathcal{A}} + 1}{\sigma^{2\star}_{X|V}(\mathcal{A}, \mathcal{U}) O_{\mathcal{A}} + 1},$$

which gives

$$\sigma^{2\star}_{X|V}(\mathcal{A}, \mathcal{U}) = \frac{\sigma^2_X}{\sigma^2_X O_{\mathcal{A}}(2^{2R_p} - 1) + 2^{2R_p}}. \tag{2.8}$$

Plugging in this value for $\sigma^{2\star}_{X|V}(\mathcal{A}, \mathcal{U})$ in $I_s(\sigma^{2\star}_{X|V}(\mathcal{A}, \mathcal{U}), \mathcal{A}, \mathcal{U})$ gives (2.7).

□

**Lemma 2.5.** *Assume that for any $\mathcal{A} \in \mathbb{A}$, $\mathcal{U} \in \mathbb{U}$, we have $O_{\mathcal{A}} \geq O_{\mathcal{U}}$. Let $R_p \in \mathbb{R}_+$. Then, we* *have*

$$\min_{\mathcal{A}\in\mathbb{A}}\min_{\mathcal{U}\in\mathbb{U}}\max_{\substack{0<\sigma^2_{X|V}\leq\sigma^2_X\\ \textit{s.t. } I_p(\sigma^2_{X|V},\mathcal{A})\leq R_p}} I_s(\sigma^2_{X|V},\mathcal{A},\mathcal{U}) = \max_{\substack{0<\sigma^2_{X|V}\leq\sigma^2_X\\ \textit{s.t. } I_p(\sigma^2_{X|V},\mathcal{A}^\star)\leq R_p}}\min_{\mathcal{A}\in\mathbb{A}}\min_{\mathcal{U}\in\mathbb{U}} I_s(\sigma^2_{X|V},\mathcal{A},\mathcal{U}). \tag{2.9}$$

*Proof.* By Lemma 2.2, we have for any $\sigma^2_{X|V}\in(0,\sigma^2_X]$, $\mathcal{A}\in\mathbb{A},\mathcal{U}\in\mathbb{U}$,

$$\begin{aligned}\frac{1}{2}\log\frac{\sigma^2_X O_{\mathcal{A}}+1}{\sigma^2_{X|V}O_{\mathcal{A}}+1} &\geq \frac{1}{2}\log\frac{\sigma^2_X O_{\mathcal{A}^\star}+1}{\sigma^2_{X|V}O_{\mathcal{A}^\star}+1},\\ -\frac{1}{2}\log\frac{\sigma^2_X O_{\mathcal{U}}+1}{\sigma^2_{X|V}O_{\mathcal{U}}+1} &\geq -\frac{1}{2}\log\frac{\sigma^2_X O_{\mathcal{U}^\star}+1}{\sigma^2_{X|V}O_{\mathcal{U}^\star}+1};\end{aligned}$$

hence, $I_s(\sigma^2_{X|V},\mathcal{A},\mathcal{U})\geq I_s(\sigma^2_{X|V},\mathcal{A}^\star,\mathcal{U}^\star)$, and we conclude that for any $\sigma^2_{X|V}\in(0,\sigma^2_X]$,

$$\min_{\mathcal{A}\in\mathbb{A}}\min_{\mathcal{U}\in\mathbb{U}} I_s(\sigma^2_{X|V},\mathcal{A},\mathcal{U}) = I_s(\sigma^2_{X|V},\mathcal{A}^\star,\mathcal{U}^\star). \tag{2.10}$$

Then, we have

$$\begin{aligned}\min_{\mathcal{A}\in\mathbb{A}}\min_{\mathcal{U}\in\mathbb{U}}\max_{\substack{0<\sigma^2_{X|V}\leq\sigma^2_X\\ \text{s.t. } I_p(\sigma^2_{X|V},\mathcal{A})\leq R_p}} I_s(\sigma^2_{X|V},\mathcal{A},\mathcal{U}) &\overset{(a)}{=} \min_{\mathcal{A}\in\mathbb{A}}\min_{\mathcal{U}\in\mathbb{U}} I_s(\sigma^{2\star}_{X|V}(\mathcal{A},\mathcal{U}),\mathcal{A},\mathcal{U})\\ &\overset{(b)}{=} I_s(\sigma^{2\star}_{X|V}(\mathcal{A}^\star,\mathcal{U}^\star),\mathcal{A}^\star,\mathcal{U}^\star)\\ &= \max_{\substack{0<\sigma^2_{X|V}\leq\sigma^2_X\\ \text{s.t. } I_p(\sigma^2_{X|V},\mathcal{A}^\star)\leq R_p}} I_s(\sigma^2_{X|V},\mathcal{A}^\star,\mathcal{U}^\star)\\ &\overset{(c)}{=} \max_{\substack{0<\sigma^2_{X|V}\leq\sigma^2_X\\ \text{s.t. } I_p(\sigma^2_{X|V},\mathcal{A}^\star)\leq R_p}}\min_{\mathcal{A}\in\mathbb{A}}\min_{\mathcal{U}\in\mathbb{U}} I_s(\sigma^2_{X|V},\mathcal{A},\mathcal{U}),\end{aligned}$$

where in $(a)$ we have defined $\sigma^{2\star}_{X|V}(\mathcal{A},\mathcal{U})\triangleq \operatorname*{arg\,max}_{\substack{0<\sigma^2_{X|V}\leq\sigma^2_X\\ \text{s.t. } I_p(\sigma^2_{X|V},\mathcal{A})\leq R_p}} I_s(\sigma^2_{X|V},\mathcal{A},\mathcal{U})$ for $\mathcal{A}\in\mathbb{A},\mathcal{U}\in\mathbb{U}$, $(b)$ holds because for any $\mathcal{A}\in\mathbb{A},\mathcal{U}\in\mathbb{U}$, we have $I_s(\sigma^{2\star}_{X|V}(\mathcal{A},\mathcal{U}),\mathcal{A},\mathcal{U})\geq I_s(\sigma^{2\star}_{X|V}(\mathcal{A},\mathcal{U}),\mathcal{A}^\star,\mathcal{U}^\star)\geq I_s(\sigma^{2\star}_{X|V}(\mathcal{A}^\star,\mathcal{U}^\star),\mathcal{A}^\star,\mathcal{U}^\star)$, where the first inequality holds by (2.10), and the second inequality holds because $I_s(\sigma^{2\star}_{X|V}(\mathcal{A},\mathcal{U}),\mathcal{A}^\star,\mathcal{U}^\star)$ is a non-increasing function of $\sigma^{2\star}_{X|V}(\mathcal{A},\mathcal{U})$ by Lemma 2.2, and $\sigma^{2\star}_{X|V}(\mathcal{A}^\star,\mathcal{U}^\star) \geq \sigma^{2\star}_{X|V}(\mathcal{A},\mathcal{U})$ by (2.8) in the proof of Lemma 2.4, and $(c)$ holds by (2.10). □

Next, we remark that if there exist $\mathcal{A} \in \mathbb{A}$ and $\mathcal{U} \in \mathbb{U}$ such that $O_{\mathcal{A}} < O_{\mathcal{U}}$, then $C_s(\mathbb{A}, R_p) = 0$ by Lemma 2.3 and Lemma 2.2 applied to $f_{\sigma^2_X,\sigma^2_{X|V}}$. Thus, we obtain the converse of Theorem 2.1 by combining Lemmas 2.3, 2.4, and 2.5.

## 2.6 Achievability Proof of Theorem 2.1

To prove the achievability part of Theorem 2.1, we first prove an achievability result for discrete random variables in Section 2.6.1 and then extend our result to Gaussian random variables by a quantization argument in Section 2.6.2.

### 2.6.1 Discrete case

Our coding scheme decouples the requirements (2.1) (reliability) and (2.2) (security with respect to unauthorized groups of colluding users). Specifically, as described next, we repeat $q \in \mathbb{N}$ times a reconciliation step to handle (2.1) via a compound version of Wyner-Ziv coding and then perform a privacy amplification step to handle (2.2) via universal hashing implemented with extractors. Note that Wyner-Ziv coding is a key component to handle rate-limited communication constraints as in rate-limited secret-key generation [28] and biometric secrecy system models, e.g., [29–33], which relies on rate-limited secret-key generation. Here, unlike in [29–33], we employ a compound version of Wyner-Ziv coding because unlike in [29–33], we simultaneously consider multiple reliability constraints due to the presence of an access structure.

#### Reconciliation step

Let $n \in \mathbb{N}$ and $\epsilon > 0$. For a probability mass function $p_X$, denote the set of $\epsilon$-letter typical sequences [34] (see also [35]) with respect to $p_X$ by $\mathcal{T}^n_\epsilon(X)$, and define $supp(p_X) \triangleq \{x \in \mathcal{X} : p_X(x) > 0\}$ and $\mu_X \triangleq \min_{x \in supp(p_X)} p_X(x)$. Define $\epsilon_1 \triangleq \frac{1}{2}\epsilon$.

**Code construction:** Fix a joint probability distribution $p_{VXY_{\mathcal{L}}}$ on $\mathcal{V} \times \mathcal{X} \times \mathcal{Y}_{\mathcal{L}}$, where $V$ is an auxiliary random variable such that $V - X - Y_\mathcal{L}$ forms a Markov chain. Define $R_v \triangleq \max_{\mathcal{A}\in\mathbb{A}} H(V|Y_\mathcal{A}) - H(V|X) + 6\epsilon H(V)$, $R'_v \triangleq H(V) - \max_{\mathcal{A}\in\mathbb{A}} H(V|Y_\mathcal{A}) - 3\epsilon H(V)$. Generate $2^{n(R_v+R'_v)}$ codewords, labeled $v^n(\omega,\nu)$ with $(\omega,\nu) \in [\![1, 2^{nR_v}]\!] \times [\![1, 2^{nR'_v}]\!]$, by generating the symbols $v_i(\omega,\nu)$ for $i \in [\![1,n]\!]$ and $(\omega,\nu) \in [\![1, 2^{nR_v}]\!] \times [\![1, 2^{nR'_v}]\!]$ independently according to $p_V$.

**Encoding:** Given $x^n$, find a pair $(\omega,\nu)$ such that $(x^n, v^n(\omega,\nu)) \in \mathcal{T}_\epsilon^n(XV)$. If there are several pairs, choose one (according to the lexicographic order); otherwise, set $(\omega,\nu) = (1,1)$. Define $v^n \triangleq v^n(\omega,\nu)$, and transmit $m \triangleq \omega$.

**Decoding:** Let $\mathcal{A} \in \mathbb{A}$. Given $y^n_\mathcal{A}$ and $m$, find $\tilde{\nu}_\mathcal{A}$ such that $(y^n_\mathcal{A}, v^n(\omega,\tilde{\nu}_\mathcal{A})) \in \mathcal{T}_\epsilon^n(Y_\mathcal{A}V)$. If there is one or more $\tilde{\nu}_\mathcal{A}$, then choose the smallest; otherwise, set $\tilde{\nu}_\mathcal{A} = 1$. Define $\widehat{v}^n_\mathcal{A} \triangleq v^n(\omega,\nu_\mathcal{A})$.

**Probability of error:** The random variable that represents the randomly generated code is denoted by $C_n$. As shown in Appendix C, there exists a codebook $\mathcal{C}_n^\star$ such that

$$\max_{\mathcal{A}\in\mathbb{A}} \mathbb{P}[V^n \neq \widehat{V}^n_\mathcal{A}] \leq |\mathbb{A}| \max_{\mathcal{A}\in\mathbb{A}} \delta(n,\epsilon,\mathcal{A}), \tag{2.11}$$

where $\delta(n,\epsilon,\mathcal{A}) \triangleq 2|\mathcal{X}||\mathcal{Y}_\mathcal{A}|e^{-n\epsilon_1^2\mu_{XY_\mathcal{A}}} + \exp(-(1-2|\mathcal{V}||\mathcal{X}|e^{-n\frac{(\epsilon-\epsilon_1)^2}{1+\epsilon_1}\mu_{VX}})2^{\epsilon n H(V)}) + 2^{-n\epsilon H(V)} + 2|\mathcal{V}||\mathcal{X}||\mathcal{Y}_\mathcal{A}|\exp(-n\frac{(\epsilon-\epsilon_1)^2}{1+\epsilon_1}\mu_{VXY_\mathcal{A}})$.

### Privacy amplification step

Let $q, n \in \mathbb{N}$, and define $N \triangleq nq$. The reconciliation step is repeated $q$ times such that the dealer has $V^N = (V^n)^q$ and the participants in $\mathcal{A} \in \mathbb{A}$ have $(\widehat{V}^n_\mathcal{A})^q$. Note that the total public communication $M \in \mathcal{M}$ is such that $\frac{H(M)}{N} \leq \frac{\log|\mathcal{M}|}{N} = \max_{\mathcal{A}\in\mathbb{A}} I(X;V|Y_\mathcal{A}) + 6\epsilon H(V)$. Next, another round of reconciliation with negligible communication is performed to ensure that $\max_{\mathcal{A}\in\mathbb{A}} \mathbb{P}[(V^n)^q \neq (\widehat{V}^n_\mathcal{A})^q] \leq \delta(q)$, where $\lim_{q\to\infty}\delta(q) = 0$ when $n$ is fixed. Finally, the dealer computes $S = g(V^N, U_d)$, while the participants in $\mathcal{A} \in \mathbb{A}$ compute $\widehat{S}(\mathcal{A}) = g(\widehat{V}^N_\mathcal{A}, U_d)$, where $U_d$ is a sequence of $d$ (to be defined later) uniformly distributed random bits, and $g : \{0,1\}^N \times \{0,1\}^d \to \{0,1\}^k$ is to be defined later.

**Analysis of reliability**

The secrets computed by the dealer and the participants in $\mathcal{A} \in \mathbb{A}$ are asymptotically the same for a fixed $n$ as $q$ goes to infinity.

$$\mathbb{P}[\widehat{S}(\mathcal{A}) \neq S] \leq \mathbb{P}[(\widehat{V}_{\mathcal{A}}^n)^q \neq (V^n)^q] \leq \delta(q).$$

**Analysis of security**

Let the min-entropy of a discrete random variable $X$, defined over $\mathcal{X}$ with probability mass function $p_X$, be denoted by $H_\infty(X) \triangleq -\log\left(\max_{x\in\mathcal{X}} p_X(x)\right)$. We will use the following lemmas:

**Lemma 2.6** (Adapted from [36])**.** *Let $E_{\mathcal{U}}$ be the random variable that represents the total knowledge about $V^N$ available to participants in $\mathcal{U} \in \mathbb{U}$. Let $e_{\mathcal{U}}$ be a particular realization of $E_{\mathcal{U}}$. If $H_\infty(V^N|E_{\mathcal{U}} = e_{\mathcal{U}}) \geq \gamma N$, for some $\gamma \in [0,1]\backslash\{0,1\}$, then there exists an extractor $g : \{0,1\}^N \times \{0,1\}^d \to \{0,1\}^k$ with $d \leq N\delta(N)$ and $k \geq N(\gamma - \delta(N))$, where $\delta(N)$ is such that* $\lim_{N\to+\infty} \delta(N) = 0$*. Moreover,*

$$H(S|U_d, E_{\mathcal{U}} = e_{\mathcal{U}}) \geq k - \delta^\star(N), \text{ with } \delta^\star(N) = 2^{-\sqrt{N}/\log N}(k + \sqrt{N}/\log N).$$

**Lemma 2.7** ([36], see also [37])**.** *Consider a discrete memoryless source $(\mathcal{X}\times\mathcal{Y}, p_{XY})$ and define the random variable $\Theta$ as*

$$\Theta \triangleq \mathbb{1}\{(X^q, Y^q) \in \mathcal{T}_{2\epsilon}^q(XY)\}\mathbb{1}\{Y^q \in \mathcal{T}_\epsilon^q(Y)\}.$$

*Then, $\mathbb{P}[\Theta = 1] \geq 1 - (2|S_X|e^{-\epsilon^2 q\mu_X/3} + 2|S_{XY}|e^{-\epsilon^2 q\mu_{XY}/3})$, with $S_{XY} \triangleq supp(p_{XY})$ and $S_Y \triangleq supp(p_Y)$. Moreover, if $y^q \in \mathcal{T}_\epsilon^q(Y)$, then*

$$H_\infty(X^q|Y^q = y^q, \Theta = 1) \geq q(1-\epsilon)H(X|Y) + \log(1 - 2|S_{XY}|e^{-\epsilon^2 q\mu_{XY}/6}).$$

Define for any $\mathcal{U} \in \mathbb{U}$, the random variables

$$\Theta_{\mathcal{U}} \triangleq \mathbb{1}\{(V^N, Y_{\mathcal{U}}^N) \in \mathcal{T}_{2\epsilon}^q(V^n Y_{\mathcal{U}}^n)\}\mathbb{1}\{Y_{\mathcal{U}}^N \in \mathcal{T}_{\epsilon}^q(Y_{\mathcal{U}}^n)\}, \tag{2.12}$$

$$\begin{aligned}\Upsilon_{\mathcal{U}} \triangleq \mathbb{1}\{&H_{\infty}(V^N|Y_{\mathcal{U}}^N = y_{\mathcal{U}}^N, \Theta_{\mathcal{U}} = 1) - H_{\infty}(V^N|Y_{\mathcal{U}}^N = y_{\mathcal{U}}^N, M = m, \Theta_{\mathcal{U}} = 1)\\ &\leq \log|\mathcal{M}| + \sqrt{N}\}.\end{aligned} \tag{2.13}$$

For any $\mathcal{U} \in \mathbb{U}$, $\mathbb{P}[\Theta_{\mathcal{U}} = 1] \geq 1 - \delta_{\epsilon}^0(n,\mathcal{U})$, where $\delta_{\epsilon}^0(n,\mathcal{U}) \triangleq 2|S_{V^n}|e^{-\epsilon^2 q\mu_{V^n}/3} + 2|S_{V^n Y_{\mathcal{U}}^n}|e^{-\epsilon^2 q\mu_{V^n Y_{\mathcal{U}}^n}/3}$ by Lemma 2.7 applied to the discrete memoryless source model $(\mathcal{V}^n \times \mathcal{Y}_{\mathcal{U}}^n, p_{V^n Y_{\mathcal{U}}^n})$, and $\mathbb{P}[\Upsilon_{\mathcal{U}} = 1] \geq 1 - 2^{-\sqrt{N}}$ by [36, Lemma 10]. Hence,

$$\mathbb{P}[\Upsilon_{\mathcal{U}} = 1, \Theta_{\mathcal{U}} = 1] \geq 1 - \delta_{\epsilon}^0(n,\mathcal{U}) - 2^{-\sqrt{N}}. \tag{2.14}$$

Then, for any $\mathcal{U} \in \mathbb{U}$, we have

$$\begin{aligned}H(S|U_d Y_{\mathcal{U}}^N M) &\overset{(a)}{\geq} H(S|U_d Y_{\mathcal{U}}^N M \Theta_{\mathcal{U}} \Upsilon_{\mathcal{U}})\\ &\geq \min_{\mathcal{U}\in\mathbb{U}} H(S|U_d Y_{\mathcal{U}}^N M \Theta_{\mathcal{U}} \Upsilon_{\mathcal{U}})\\ &\geq \min_{\mathcal{U}\in\mathbb{U}} \mathbb{P}[\Theta_{\mathcal{U}} = 1, \Upsilon_{\mathcal{U}} = 1] H(S|U_d Y_{\mathcal{U}}^N M, \Theta_{\mathcal{U}} = 1, \Upsilon_{\mathcal{U}} = 1)\\ &\geq \min_{\mathcal{U}\in\mathbb{U}} \mathbb{P}[\Theta_{\mathcal{U}} = 1, \Upsilon_{\mathcal{U}} = 1] \min_{\mathcal{U}\in\mathbb{U}} H(S|U_d Y_{\mathcal{U}}^N M, \Theta_{\mathcal{U}} = 1, \Upsilon_{\mathcal{U}} = 1)\\ &\overset{(b)}{\geq} \left(1 - \max_{\mathcal{U}\in\mathbb{U}} \delta_{\epsilon}^0(n,\mathcal{U}) - 2^{-\sqrt{N}}\right) \min_{\mathcal{U}\in\mathbb{U}} H(S|U_d Y_{\mathcal{U}}^N M, \Theta_{\mathcal{U}} = 1, \Upsilon_{\mathcal{U}} = 1),\end{aligned} \tag{2.15}$$

where $(a)$ holds because conditioning reduces entropy and $(b)$ holds by (2.14). To lower bound $\min_{\mathcal{U}\in\mathbb{U}} H(S|U_d Y_{\mathcal{U}}^N M, \Theta_{\mathcal{U}} = 1, \Upsilon_{\mathcal{U}} = 1)$ in (2.15) with Lemma 2.6, we now lower bound $\min_{\mathcal{U}\in\mathbb{U}} H_{\infty}(V^N|Y_{\mathcal{U}}^N = y_{\mathcal{U}}^N, M = m, \Theta_{\mathcal{U}} = 1, \Upsilon_{\mathcal{U}} = 1)$. We have for any $\mathcal{U} \in \mathbb{U}$,

$$\begin{aligned}&H_{\infty}(V^N|Y_{\mathcal{U}}^N = y_{\mathcal{U}}^N, M = m, \Theta_{\mathcal{U}} = 1, \Upsilon_{\mathcal{U}} = 1)\\ &\overset{(a)}{\geq} H_{\infty}(V^N|Y_{\mathcal{U}}^N = y_{\mathcal{U}}^N, \Theta_{\mathcal{U}} = 1) - \log|\mathcal{M}| - \sqrt{N}\\ &\overset{(b)}{\geq} q(1-\epsilon)H(V^n|Y_{\mathcal{U}}^n) - \delta_{\epsilon}^1(q,n,\mathcal{U}) - N(\max_{\mathcal{A}\in\mathbb{A}} I(V;X|Y_{\mathcal{A}}) + 6\epsilon H(V)) - \sqrt{N}\end{aligned}$$

$$
\begin{aligned}
&\overset{(c)}{\geq} N[I(X;V|Y_{\mathcal{U}}) - \max_{\mathcal{A}\in\mathbb{A}} I(V;X|Y_{\mathcal{A}}) - \delta_\epsilon^2(q,n,\mathcal{U})] \\
&\geq N[\min_{\mathcal{U}\in\mathbb{U}} I(X;V|Y_{\mathcal{U}}) - \max_{\mathcal{A}\in\mathbb{A}} I(V;X|Y_{\mathcal{A}}) - \max_{\mathcal{U}\in\mathbb{U}} \delta_\epsilon^2(q,n,\mathcal{U})] \\
&\overset{(d)}{=} N[\min_{\mathcal{A}\in\mathbb{A}} I(V;Y_{\mathcal{A}}) - \max_{\mathcal{U}\in\mathbb{U}} I(V;Y_{\mathcal{U}}) - \max_{\mathcal{U}\in\mathbb{U}} \delta_\epsilon^2(q,n,\mathcal{U})],
\end{aligned} \tag{2.16}
$$

where $(a)$ holds by (2.13), $(b)$ holds by Lemma 2.7 with $\delta_\epsilon^1(q,n,\mathcal{U}) \triangleq -\log(1 - 2|S_{V^nY_{\mathcal{U}}^n}|e^{-\epsilon^2 q\mu_{V^nY_{\mathcal{U}}^n}/6})$, $(c)$ holds with $\delta_\epsilon^2(q,n,\mathcal{U}) \triangleq \epsilon I(X;V|Y_{\mathcal{U}}) + (1-\epsilon)[2\epsilon H(X|Y_{\mathcal{U}}V) + 2n^{-1} + \log|\mathcal{X}|(4|\mathcal{V}||\mathcal{X}|e^{-n\epsilon^2\mu_{XV}} + 2|\mathcal{V}||\mathcal{X}||\mathcal{Y}_{\mathcal{U}}|e^{-\epsilon^2 n\mu_{VXY_{\mathcal{U}}}/2})] + N^{-1}\delta_\epsilon^1(q,n,\mathcal{U}) + 6\epsilon H(V) + N^{-1/2}$ because, as shown in Appendix D, we have

$$
\begin{aligned}
H(V^n|Y_{\mathcal{U}}^n) &\geq n(H(X|Y_{\mathcal{U}}) - H(X|Y_{\mathcal{U}}V)(1+2\epsilon)) \\
&\quad - 2 - n\log|\mathcal{X}|(4|\mathcal{V}||\mathcal{X}|e^{-n\epsilon^2\mu_{XV}} + 2|\mathcal{V}||\mathcal{X}||\mathcal{Y}_{\mathcal{U}}|e^{-\epsilon^2 n\mu_{VXY_{\mathcal{U}}}/2}),
\end{aligned} \tag{2.17}
$$

and $(d)$ holds because $V - X - (Y_{\mathcal{A}}, Y_{\mathcal{U}})$.

Next, we set the output size $k$ of the extractor to be less than the lower bound in (2.16) by $\sqrt{N}$, i.e.,

$$
k \triangleq \lfloor N[\min_{\mathcal{A}\in\mathbb{A}} I(V;Y_{\mathcal{A}}) - \max_{\mathcal{U}\in\mathbb{U}} I(V;Y_{\mathcal{U}}) - \max_{\mathcal{U}\in\mathbb{U}} \delta_\epsilon^2(q,n,\mathcal{U}) - N^{-1/2}]\rfloor, \tag{2.18}
$$

Finally, we have

$$
\begin{aligned}
\max_{\mathcal{U}\in\mathbb{U}} I(S;U_dY_{\mathcal{U}}^NM) &= H(S) - \min_{\mathcal{U}\in\mathbb{U}} H(S|U_dY_{\mathcal{U}}^NM) \\
&\overset{(a)}{\leq} k - \left(1 - \max_{\mathcal{U}\in\mathbb{U}} \delta_\epsilon^0(n,\mathcal{U}) - 2^{-\sqrt{N}}\right)(k - \delta^\star(N)) \\
&\overset{(b)}{\leq} \delta_\epsilon^3(N),
\end{aligned} \tag{2.19}
$$

where $(a)$ holds by (2.15), (2.16) (valid for any $\mathcal{U} \in \mathbb{U}$), (2.18), and Lemma 2.6 with $\delta^\star(N) \triangleq 2^{-\sqrt{N}/\log N}\left(k + \sqrt{N}/\log N\right)$ and $(b)$ holds with $\delta_\epsilon^3(N) \triangleq \delta^\star(N) + \left(\max_{\mathcal{U}\in\mathbb{U}} \delta_\epsilon^0(n,\mathcal{U}) + 2^{-\sqrt{N}}\right)k$.

#### Analysis of uniformity

Similar to (2.19), we have

$$\begin{aligned} H(S) &\geq \min_{\mathcal{U}\in\mathbb{U}} H(S|U_d Y_{\mathcal{U}}^N M) \\ &\geq k - \delta_\epsilon^3(N). \end{aligned} \tag{2.20}$$

#### Public communication rate

The public communication rate corresponds to the rate of $M$ plus the rate of $U_d$, i.e.,

$$\lim_{N\to\infty} R_p = \max_{\mathcal{A}\in\mathbb{A}} I(X;V|Y_{\mathcal{A}}) + 6\epsilon H(V).$$

#### Achievable secret rate

The secret rate $R_s \triangleq k/N$ satisfies

$$R_s \geq \min_{\mathcal{A}\in\mathbb{A}} I(V;Y_{\mathcal{A}}) - \max_{\mathcal{U}\in\mathbb{U}} I(V;Y_{\mathcal{U}}) - \max_{\mathcal{U}\in\mathbb{U}} \delta_\epsilon^2(q,n,\mathcal{U}) - N^{-1/2} - N^{-1}. \tag{2.21}$$

### 2.6.2 Continuous case

In this section, we extend the achievability result of Section 2.6.1 for discrete random variables to Gaussian random variables by means of quantization. Quantization also allows us to show that the size of the shares linearly scales with the length of the secret. The main issue with quantization is that it might lead to an underestimation of the information that unauthorized sets of participants may learn about the secret. We will, however, show that this issue can be overcome provided that the quantization is fine enough.

We now build upon Section 2.6.1 to show that $(R_p, R_s) \in \mathcal{R}(p_{XY_{\mathcal{L}}}, \mathbb{A})$, where

$$R_p = \frac{1}{2}\log\frac{\sigma_X^2}{\sigma_{X|V}^2} - \frac{1}{2}\log\frac{\sigma_X^2 O_{\mathcal{A}^\star}+1}{\sigma_{X|V}^2 O_{\mathcal{A}^\star}+1}, \tag{2.22}$$

$$R_s \;=\; \frac{1}{2}\log\frac{\sigma_X^2 O_{\mathcal{A}^\star}+1}{\sigma_{X|V}^2 O_{\mathcal{A}^\star}+1} - \frac{1}{2}\log\frac{\sigma_X^2 O_{\mathcal{U}^\star}+1}{\sigma_{X|V}^2 O_{\mathcal{U}^\star}+1}. \tag{2.23}$$

We use the following lemma to extend Section 2.6.1 to the continuous case by means of quantization.

**Lemma 2.8** ([38–40])**.** *Let $X$ and $Y$ be two real-valued random variables with probability distribution $\mathbb{P}_X$ and $\mathbb{P}_Y$, respectively. Let $\mathcal{C}_{\Delta_1} = \{C_i\}_{i\in\mathcal{I}}$, $\mathcal{D}_{\Delta_2} = \{D_j\}_{j\in\mathcal{J}}$ be two partitions of the real line for $X$ and $Y$ such that for any $i \in \mathcal{I}$, $\mathbb{P}_X[C_i] = \Delta_1$, for any $j \in \mathcal{J}$, $\mathbb{P}_Y[D_j] = \Delta_2$, where $\Delta_1, \Delta_2 > 0$. Let $X_{\Delta_1}, Y_{\Delta_2}$ be the quantized version of $X, Y$ with respect to the partitions $\mathcal{C}_{\Delta_1}, \mathcal{D}_{\Delta_2}$, respectively. Then, we have*

$$I(X,Y) = \lim_{\Delta_1,\Delta_2\to 0} I(X_{\Delta_1}, Y_{\Delta_2}).$$

We first show that a quantization does not affect the security requirement (2.2).

**Proposition 2.1.** *A quantization of $Y_{\mathcal{U}}^n$, $\mathcal{U} \in \mathbb{U}$, might lead to an underestimation of $I(S; M, Y_{\mathcal{U}}^n)$. However, if the quantized version $Y_{\mathcal{U},\Delta}^n$ of $Y_{\mathcal{U}}^n$, $\mathcal{U} \in \mathbb{U}$, is fine enough, then for any $\delta > 0$*

$$\max_{\mathcal{U}\in\mathbb{U}} I(S; MY_{\mathcal{U}}^n) \leq \max_{\mathcal{U}\in\mathbb{U}} I(S; MY_{\mathcal{U},\Delta}^n) + \delta. \tag{2.24}$$

*Proof.* For any $\delta > 0$, for any $\mathcal{U} \in \mathbb{U}$, we have

$$\begin{aligned} I(S; MY_{\mathcal{U}}^n) &\leq |I(S; MY_{\mathcal{U}}^n) - I(S; MY_{\mathcal{U},\Delta}^n)| + I(S; MY_{\mathcal{U},\Delta}^n) \\ &\leq \max_{\mathcal{U}\in\mathbb{U}} |I(S; MY_{\mathcal{U}}^n) - I(S; MY_{\mathcal{U},\Delta}^n)| + \max_{\mathcal{U}\in\mathbb{U}} I(S; MY_{\mathcal{U},\Delta}^n) \\ &\leq \delta + \max_{\mathcal{U}\in\mathbb{U}} I(S; MY_{\mathcal{U},\Delta}^n), \end{aligned} \tag{2.25}$$

where the last inequality holds by Lemma 2.8, if the quantized version $Y_{\mathcal{U},\Delta}^n$ of $Y_{\mathcal{U}}^n$, $\mathcal{U} \in \mathbb{U}$, is fine enough. Since (2.25) is valid for any $\mathcal{U} \in \mathbb{U}$, we obtain (2.24). □

For $\mathcal{A} \in \mathbb{A}$ and $\mathcal{U} \in \mathbb{U}$, we quantize $X, Y_{\mathcal{A}}, Y_{\mathcal{U}},$ and $V$ as in Lemma 2.8 to form $X_\Delta, Y_{\mathcal{A},\Delta}, Y_{\mathcal{U},\Delta}$, and $V_\Delta$ such that $\Delta = l^{-1}$ and $|\mathcal{X}_\Delta| = |\mathcal{Y}_{\mathcal{A},\Delta}| = |\mathcal{Y}_{\mathcal{U},\Delta}| = |\mathcal{V}_\Delta| = l$ with $l > 0$. Next, we apply the proof for the discrete case to the random variables $X_\Delta, Y_{\mathcal{A},\Delta}, Y_{\mathcal{U},\Delta}$, $V_\Delta$. By Lemma 2.8, we can fix $l$ large enough such that, for any $\mathcal{A} \in \mathbb{A}$, $|I(V_\Delta; Y_{\mathcal{A},\Delta}) - I(V; Y_\mathcal{A})| < \delta/2$, for any $\mathcal{U} \in \mathbb{U}$, $|I(V_\Delta; Y_{\mathcal{U},\Delta}) - I(V; Y_\mathcal{U})| < \delta/2$, such that (2.21) becomes

$$R_s \geq \min_{\mathcal{A}\in\mathbb{A}} I(V; Y_\mathcal{A}) - \max_{\mathcal{U}\in\mathbb{U}} I(V; Y_\mathcal{U}) - \max_{\mathcal{U}\in\mathbb{U}} \delta_\epsilon^2(q, n, \mathcal{U}) - N^{-1/2} - N^{-1} - \delta.$$

Note that $\delta_\epsilon^2(q, n, \mathcal{U}), \mathcal{U} \in \mathbb{U}$, in the above equation hides the terms $2\epsilon(1-\epsilon)H(X_\Delta|Y_{\mathcal{U},\Delta}V_\Delta)$ and $6\epsilon H(V_\Delta)$, which do not go to zero as $l$ goes to infinity. Consequently, we choose $\epsilon = n^{-\alpha}$, where $\alpha \in [0, 1/2]\backslash\{0, 1/2\}$, such that if we choose $l$ large enough, then $n$ large enough, and finally $q$ large enough, then the asymptotic secret rate is as close as desired to

$$\min_{\mathcal{A}\in\mathbb{A}} I(V; Y_\mathcal{A}) - \max_{\mathcal{U}\in\mathbb{U}} I(V; Y_\mathcal{U}), \tag{2.26}$$

$\delta_\epsilon^3(N)$ vanishes to zero in (2.19), (2.20), and the asymptotic public communication rate is as close as desired to

$$\max_{\mathcal{A}\in\mathbb{A}} I(V; X|Y_\mathcal{A}). \tag{2.27}$$

By taking the auxiliary random variable $V$ jointly Gaussian with $X$ in (2.26) and (2.27), we obtain (2.22) and (2.23), as shown in Appendix E.

**Remark 2.2.** *We observe that the size of the shares scales linearly with the secret size. First, note that the size of each share is the sum of the length of the public communication, i.e., $NR_p$ bits, and the length of $N$ quantized observations of a Gaussian random variable. Then, since we achieve the secret rate in* (2.26) *by making the quantization parameter $l$ fixed when $N$ grows to infinity, we conclude that the size of the shares scales linearly with $N$, which is also the case for the length of the generated secret.*

## 2.7 Concluding Remarks

We studied information-theoretic secret sharing from Gaussian correlated sources over a one-way rate-limited public channel and characterized its secret capacity, which provides a closed-form expression of the trade-off between public communication and the secret rate. By contrast with a traditional secret-sharing protocol, our setting does not require information-theoretically secure channels between the dealer and participants, and provides information-theoretic security during the distribution phase, where the dealer distributes shares of the secret to the participants. Moreover, we have shown that the size of the shares scales linearly with the size of the secret for any access structure. We also characterized the secret capacity for threshold access structures and showed that the secret capacity is, in general, not a monotone function of the threshold.

While explicit and low-complexity coding schemes have been proposed for information-theoretic secret sharing that rely on discrete channel models [11, 41] and discrete source models [42], developing low-complexity coding schemes that achieve the limits derived in this chapter for Gaussian sources remains an open problem.

# CHAPTER 3

# DESIGN OF SHORT BLOCKLENGTH WIRETAP CHANNEL CODES: DEEP LEARNING AND CRYPTOGRAPHY WORKING HAND IN HAND

## 3.1 Introduction

The wiretap channel [19] is a basic model to account for eavesdroppers in wireless communication. In this model, a sender (Alice) encodes a confidential message $M$ into a codeword $X^n$ and transmits it to a legitimate receiver (Bob) over $n$ uses of a channel in the presence of an external eavesdropper (Eve). Bob's estimate of $M$ from his channel output observations is denoted by $\hat{M}$, and Eve's channel output observations are denoted by $Z^n$. In [43], the constraints are that Bob must be able to recover $M$, i.e., $\lim_{n\to\infty} \mathbb{P}[M \neq \hat{M}] = 0$, and the leakage about $M$ at Eve, quantified by $I(M; Z^n)$, is not too large in the sense that $\lim_{n\to\infty} \frac{1}{n} I(M; Z^n) = 0$. Note that the stronger security requirement $\lim_{n\to\infty} I(M; Z^n) = 0$ can also be considered [36], meaning that Eve's observations $Z^n$ are almost independent of $M$ for large $n$. The secrecy capacity has been characterized for degraded discrete memoryless channels in [19], for arbitrary discrete memoryless channels in [44], and for Gaussian channels in [45].

While [19,44,45] provide non-constructive achievability schemes for the wiretap channel, constructive coding schemes have also been proposed. Specifically, coding schemes based on low-density parity-check (LDPC) codes [46–48], polar codes [49–52], and invertible extractors [53,54] have been constructed for degraded or symmetric wiretap channel models. Moreover, the method in [53,54] has been extended to the Gaussian wiretap channel [55]. Coding schemes based on random lattice codes have also been proposed for the Gaussian wiretap channel [56]. Subsequently, constructive [57–59] and random [60] polar coding schemes have been proposed to achieve the secrecy capacity of non-degraded discrete wiretap channels. Coding schemes that combine polar codes and invertible extractors have also been proposed to avoid the need for a pre-shared secret under strong secrecy [11, 61]. All the references above consider the asymptotic regime, i.e., the regime where $n$ approaches infinity. However, many practical applications require short packet lengths or low latency [62]. To fulfill this need, non-asymptotic and second-order asymptotics achievability and converse bounds on the secrecy capacity of discrete and Gaussian wiretap channels have been established in [63–65]. Note that [63–65] focus on deriving fundamental limits and not on code constructions. We will review the works that are most related to our study and focus on code constructions at finite blocklength for the wiretap channel in Section 3.2.

In this chapter, we propose to design short blocklength codes (smaller than or equal to $128$) for the Gaussian wiretap channel under information-theoretic security guarantees. Such an information-theoretic approach enables coding solutions robust against computationally unbounded adversaries, and are thus technology independent and, in particular, quantum proof. Specifically, we quantify security in terms of the leakage $I(M; Z^n)$, i.e., the mutual information between the confidential message and the eavesdropper's channel observations. The main idea of our approach is to decouple the reliability and secrecy constraints. Specifically, we use a deep learning approach based on a feed-forward neural network autoencoder [66] to handle the reliability constraint and cryptographic tools, namely, hash functions [67], to handle the secrecy constraint.[1] Then, to evaluate the performance of our constructed code, we empirically estimate the leakage $I(M; Z^n)$. Note that even for small values of $n$ this estimation is challenging with standard techniques such as binning of the probability space [68], $k$-nearest neighbor statistics [69], or maximum likelihood estimation [70]. Unlike [63–65], which analytically derive upper bounds on

[1] Note that a coding strategy that separately handles the reliability and secrecy constraints with two separate coding layers is also used for the discrete wiretap channel in [53, 54], and for the Gaussian wiretap channel in [55]. In these works, an asymptotic regime is considered, i.e., the blocklength $n$ tends to infinity. Further, in [53–55], the security layer relies on the random choice of a hash function in a family of universal hash functions, and therefore, the coding scheme is non-constructive. In this chapter, we also consider a family of hash functions for the security layer but only select a specific function in this family. This choice is deterministic and part of the coding scheme design, thus making it constructive, as elaborated on in our simulation results.

the leakage, we consider a practical approach to estimate the leakage via the mutual information neural estimator (MINE) from [71], which is provably consistent and offers better performances than other known mutual information estimators in high dimension. We also compare the performances of our codes with the best-known achievability and converse bounds on optimal secrecy rates for the Gaussian wiretap channel [63].

Our main contributions are as follows.

1. We propose a framework based on neural networks that enables a flexible design of finite blocklength codes for the Gaussian wiretap channel. Additionally, as seen in our simulations, our code design provides examples of wiretap codes that outperform the best known achievable secrecy rates from [63] obtained non-constructively for the Gaussian wiretap channel.

2. We demonstrate that our proposed framework is also able to handle compound [72, 73] and arbitrarily varying [74, 75] settings, when uncertainty holds on both the legitimate users' channel and the eavesdropper's channel, as demonstrated by our simulations results in Section 3.5. These models are particularly useful to capture uncertainty about the channel statistics of the eavesdropper channel or to model an active eavesdropper who can influence its channel statistics by changing its location.

3. We propose a coding scheme design able to precisely control the level of information leakage at the eavesdropper through the independent design of a reliability coding layer and a secrecy coding layer. By contrast, as elaborated on in Section 3.2, deep learning approaches that seek to simultaneously design codes for reliability and secrecy do not seem to offer good control over the information leakage at the eavesdropper.

Additionally, our proposed code design offers the following features.

- A modular approach that separates the code design into a secrecy layer and a reliability layer. The secrecy layer only deals with the secrecy constraint and only depends on the statistics of the eavesdropper's channel, whereas the reliability layer only deals with the reliability constraint and only depends on the statistics of the legitimate receiver's channel. This approach allows a simplified code design, for instance, if only one of the two layers needs to be (re)designed.

- A universal way of dealing with the secrecy constraint through the use of hash functions. This is beneficial, for instance, for compound [9,73] and arbitrarily varying [74,75] settings, as our results show that it becomes sufficient to design our code with respect to the best eavesdropper's channel.
- A method that can be applied to an arbitrary channel model as the conditional probability distribution that defines the channel is not needed and only input and output channel samples are needed to design the reliability and secrecy layers.

Note that it is difficult to analytically characterize optimal secrecy rates for the Gaussian wiretap channel in the finite blocklength regime. In this study, we adopt a practical approach based on deep learning to better understand this regime.

The remainder of the chapter is organized as follows. Section 3.2 reviews related works. Section 3.3 introduces the Gaussian wiretap channel model. Section 3.4 describes our proposed code design and our simulation results for the Gaussian wiretap channel model. Section 3.5 discusses the compound and arbitrarily varying Gaussian wiretap channel models and presents our simulation results. Finally, Section 3.6 provides concluding remarks.

## 3.2 Related Works

As elaborated on in the introduction, several code constructions have already been proposed for Gaussian wiretap channel coding in the asymptotic blocklength regime. Another challenging task is code designs in the finite blocklength regime. Next, we review known finite-length code constructions based on coding theoretic tools and deep learning tools in Sections 3.2.1 and 3.2.2, respectively.

### 3.2.1 Works based on coding theory

In the following, we distinguish the works that consider a non-information-theoretic secrecy metric from the works that consider an information-theoretic secrecy metric.

#### Non-information-theoretic secrecy metric

A non-information-theoretic security metric called security gap, which is based on an error probability analysis at the eavesdropper, is used to evaluate the secrecy performance in [76–80]. Specifically, randomized convolutional codes for Gaussian and binary symmetric wiretap channels are studied in [76], and randomized turbo codes for the Gaussian wiretap channel are investigated in [77]. Coding schemes for the Gaussian wiretap channel based on LDPC codes are proposed in [78, 79]. Additionally, another non-information-theoretic security approach called practical secrecy is investigated in [81], where a leakage between Alice's message and an estimate of the message at Eve is estimated.

#### Information-theoretic secrecy metric

Next, we review works that consider the leakage $I(M; Z^n)$ as a secrecy metric. In [82], punctured systematic irregular LDPC codes are proposed for the binary phase-shift-keyed-constrained Gaussian wiretap channel, and a leakage as low as $11$ percent of the message length has been obtained for a blocklength $n = 10^6$. In [83], LDPC codes for the Gaussian wiretap channel have also been developed, and a leakage as low as $20$ percent of the message length has been obtained for a blocklength $n = 50,000$. Most recently, in [84], randomized Reed-Muller codes are developed for the Gaussian wiretap channel, and a leakage as low as $0.2$ percent of the message length has been obtained for a blocklength $n = 16$.

### 3.2.2 Works based on deep learning

Artificial neural networks have gained attention in communication system design because they approach the performance of state-of-the-art channel coding solutions. In [85,86], neural networks (autoencoder) are used to learn the encoder and decoder for a channel coding task without secrecy constraints. Other machine learning approaches for channel coding without secrecy constraints have also been investigated in [87,88] with reinforcement learning, in [89] with mutual information estimators, and in [90] with generative adversarial networks.

Recently, deep learning approaches for channel coding have been extended to wiretap channel coding. In [91, 92], a coding scheme that imitates coset coding by clustering learned signal constellations is developed for the Gaussian wiretap channel under a non-information-theoretic secrecy metric, which relies on a cross-entropy loss function. In [93], neural networks are used to learn optimal precoding for the MIMO Gaussian wiretap channel. In [94], a coding scheme for the Gaussian wiretap channel is developed under the information-theoretic leakage $I(M; Z^n)$ with an autoencoder approach that seeks to simultaneously optimize the reliability and secrecy constraints. A leakage as low as $15$ percent of the message length is obtained in [94] for a blocklength $n = 16$. It seems that precisely controlling and minimizing the leakage is challenging with such an approach. By contrast, in this chapter, we propose an approach that separates the code design into a part that only deals with the reliability constraint (by means of an autoencoder) and another part that only deals with the secrecy constraint (by means of hash functions). As supported by our simulation results, one of the advantages of our approach is a better control of how small the leakage can be made.

## 3.3 Gaussian wiretap channel model

Notation: Unless specified otherwise, capital letters represent random variables, whereas lowercase letters represent realizations of associated random variables, e.g., $x$ is a realization of the random variable $X$. $|\mathcal{X}|$ denotes the cardinality of the set $\mathcal{X}$. $\|\cdot\|_2$ denotes the Euclidean norm. GF($2^q$) denotes a finite field of order $2^q$, $q \in \mathbb{N}^*$.

For $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$, consider a memoryless Gaussian wiretap channel $(\mathcal{X}, P_{YZ|X}, \mathcal{Y} \times \mathcal{Z})$

defined by

$$Y \triangleq X + N_Y, \tag{3.1}$$

$$Z \triangleq X + N_Z, \tag{3.2}$$

where $N_Y$ and $N_Z$ are zero-mean Gaussian random variables with variances $\sigma_Y^2$ and $\sigma_Z^2$, respectively. As formalized next, the objective of the sender is to transmit a confidential message $M$ to a legitimate receiver by encoding it into a sequence $X^n$, which is then sent over $n$ uses of the channels (3.1), (3.2) and yields the channel observations $Y^n$ and $Z^n$ at the legitimate receiver and eavesdropper, respectively.

**Definition 3.1.** *Let $\mathbb{B}_0^n(\sqrt{nP})$ be the ball of radius $\sqrt{nP}$ centered at the origin in $\mathbb{R}^n$ under the Euclidian norm. An $(n, k, P)$ code consists of*

- *a message set $\{0, 1\}^k$;*
- *an encoder $e : \{0, 1\}^k \to \mathbb{B}_0^n(\sqrt{nP})$, which, for a message $M \in \{0, 1\}^k$, forms the codeword $X^n \triangleq e(M)$;*
- *a decoder $d : \mathbb{R}^n \to \{0, 1\}^k$, which, from the channel observations $Y^n$, forms an estimate of the message as $d(Y^n)$.*

*The codomain of the encoder $e$ reflects the power constraint $\|e(m)\|_2^2 \leq nP,\ \forall m \in \{0, 1\}^k$.*

The performance of an $(n, k, P)$ code is measured in terms of

1. The average probability of error $\mathbf{P_e} \triangleq \frac{1}{2^k} \sum_{m=1}^{2^k} \mathbb{P}[d(Y^n) \neq m | m \text{ is sent}]$;
2. The leakage at the eavesdropper $\mathbf{L_e} \triangleq I(M; Z^n)$.

**Definition 3.2.** *An $(n, k, P)$ code is $\epsilon$-reliable if $\mathbf{P_e} \leq \epsilon$ and $\delta$-secure if $\mathbf{L_e} \leq \delta$. Moreover, a secrecy rate $\frac{k}{n}$ is $(\epsilon, \delta)$-achievable with power constraint $P$ if there exists an $\epsilon$-reliable and $\delta$-secure $(n, k, P)$ code.*

### 3.4 Coding scheme

We first describe, at a high level, our coding scheme in Section 3.4.1. Specifically, our coding approach consists of two coding layers, one reliability layer, whose design is described in Section 3.4.2, and one security layer, whose design is described in Section 3.4.3. We then comment on the communication rate of our proposed coding scheme when considering multiplexing of protected and unprotected messages in Section 3.4.4. Finally, we provide simulation results and examples of our code design for the Gaussian wiretap channel in Sections 3.4.5 and 3.4.6.

#### 3.4.1 High-level description of our coding scheme

Our code construction consists of (i) a reliability layer with an $\epsilon$-reliable $(n, q, P)$ code, described by the encoder/decoder pair $(e_0, d_0)$ (this code is designed without any security requirement, i.e., its performance is solely measured in terms of average probability of error), and (ii) a security layer implemented with hash functions. We design the encoder/decoder pair $(e_0, d_0)$ of the reliability layer using a deep learning approach based on neural network autoencoders as described in Section 3.4.2. We will then design two functions $\varphi_s$ and $\psi_s$ in Section 3.4.3 to perform the encoding and decoding, respectively, at the secrecy layer. The encoder/decoder pair $(e, d)$ for the encoding and decoding process of the reliability and secrecy layers considered jointly is described as follows:

*Encoding*: Assume that a fixed sequence of bits $s \in \mathcal{S} \triangleq \{0,1\}^q \backslash \{\mathbf{0}\}$, called seed, is known to all parties. Alice generates a sequence $B$ of $q-k$ bits uniformly at random in $\{0,1\}^{q-k}$ (this sequence represents local randomness used to randomize the output of the function $\varphi_s$) and encodes the message $M \in \{0,1\}^k$ as $e_0(\varphi_s(M, B))$, where $\varphi_s(M, B) \in \{0,1\}^q$. The overall encoding map $e$ that describes the encoding at the secrecy and reliability layers is described by

$$\begin{aligned} e : \{0,1\}^k \times \{0,1\}^{q-k} &\to \mathbb{B}_0^n(\sqrt{nP}) \\ (m, b) &\mapsto e_0(\varphi_s(m, b)). \end{aligned}$$

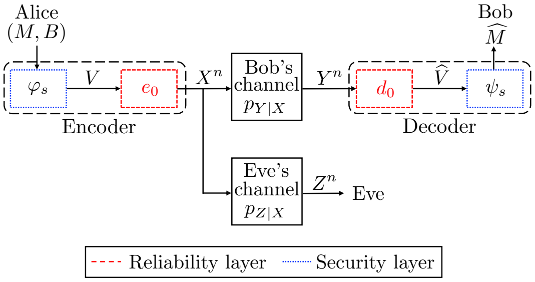


Figure 3.1: Our code design consists of a reliability layer and a security layer. The reliability layer is implemented using an autoencoder $(e_0, d_0)$ described in Section 3.4.2, and the security layer is implemented using the functions $\varphi_s$ and $\psi_s$ described in Section 3.4.3.

*Decoding*: Given $Y^n$ and $s$, Bob decodes the message as $\psi_s(d_0(Y^n))$. The overall decoding map $d$ that describes the decoding at the reliability and secrecy layers is

$$\begin{aligned} d : \mathbb{R}^n &\to \{0,1\}^k \\ y^n &\mapsto \psi_s(d_0(y^n)). \end{aligned}$$

For a given code design, described by the encoder/decoder pair $(e, d)$, we will then evaluate the performance of this code by empirically measuring the leakage using a neural network-based mutual information estimator as described in Section 3.4.3. Our code design is summarized in Figure 3.1.

### 3.4.2 Design of the reliability layer $(e_0, d_0)$

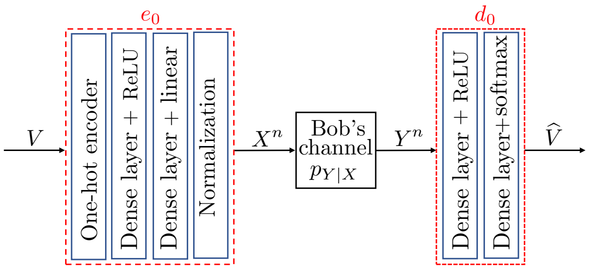


Figure 3.2: Architecture of the autoencoder $(e_0, d_0)$ via feed-forward neural networks.

The design of the reliability layer consists in designing an $\epsilon$-reliable $(n, q, P)$ code described by the encoder/decoder pair $(e_0, d_0)$ for the channel (3.1). Define $Q \triangleq 2^q$ and let $\mathcal{V} \triangleq \{1, 2, \ldots, Q\}$

be the message set of this code. $(e_0, d_0)$ is obtained with an autoencoder as in [85]. Specifically, the goal of the autoencoder is here to learn a representation of the encoded message that is robust to the channel noise, so that the received message at Bob can be reconstructed from its noisy channel observations with a small probability of error. In other words, the encoding part (denoted by $e_0$) of this autoencoder adds redundancy to the message to ensure recoverability by Bob in the presence of noise. As depicted in Figure 3.2, the encoder consists of (i) a one-hot encoder where the input $v \in \mathcal{V}$ is mapped to a one-hot vector $\mathbf{1}_v \in \mathbb{R}^Q$, i.e., a vector whose components are all equal to zero except the $v$-th component which is equal to one, followed by (ii) dense hidden layers (with rectified linear unit (ReLU) or linear activation functions [85]) that take $v$ as input and return an $n$-dimensional vector, followed by (iii) a normalization layer that ensures that the average power constraint $\frac{1}{n}\|e_0(v)\|_2^2 \leq P$ is met for the codeword $e_0(v)$. Note that, without loss of generality, one can assume that $P = 1$ since one can rewrite $\frac{1}{n}\|e_0(v)\|_2^2 \leq P$ as $\frac{1}{n}\|\tilde{e}_0(v)\|_2^2 \leq 1$, where $\tilde{e}_0(v) \triangleq e_0(v)/\sqrt{P}$. As depicted in Figure 3.2, the decoder consists of dense hidden layers and a softmax layer. More specifically, let $\mu^{|\mathcal{V}|}$ be the output of the last dense layer in the decoder. The softmax layer takes $\mu^{|\mathcal{V}|}$ as input and returns a vector of probabilities $p^{|\mathcal{V}|} \in [0,1]^{|\mathcal{V}|}$, whose components $p_v$, $v \in \mathcal{V}$, are $p_v \triangleq \exp(\mu_v)\left(\sum_{i=1}^{|\mathcal{V}|}\exp(\mu_i)\right)^{-1}$. Finally, the decoded message $\hat{v}$ corresponds to the index of the component of $p^{|\mathcal{V}|}$ associated with the highest probability, i.e., $\hat{v} \in \arg\max_{v\in\mathcal{V}} p_v$. The autoencoder is trained over all possible messages $v \in \mathcal{V}$ using a stochastic gradient descent (SGD) [95] and the categorical cross-entropy loss function.

### 3.4.3 Design of the security layer $(\varphi_s, \psi_s)$

The objective is now to design $(\varphi_s, \psi_s)$ such that the total amount of leaked information about the original message is small in the sense that $I(M; Z^n) \leq \delta$, for some $\delta > 0$. For a given choice of $(\varphi_s, \psi_s)$, the performance of our code construction will be evaluated using a mutual information neural estimator (MINE) [71]. Before we describe the construction of $(\varphi_s, \psi_s)$, we review the definition of 2-universal hash functions.

**Definition 3.3.** *[96] Given two finite sets $\mathcal{X}$ and $\mathcal{Y}$, a family $\mathcal{G}$ of functions from $\mathcal{X}$ to $\mathcal{Y}$ is 2-*

*universal if* $\forall x_1, x_2 \in \mathcal{X},\ x_1 \neq x_2 \implies \mathbb{P}[G(x_1) = G(x_2)] \leq |\mathcal{Y}|^{-1}$*, where* $G$ *is the random variable that represents the choice of a function* $g \in \mathcal{G}$ *uniformly at random in* $\mathcal{G}$*.*

**Design of** $(\varphi_s, \psi_s)$

Let $\mathcal{S} \triangleq \{0,1\}^q \backslash \{\mathbf{0}\}$. For $k \leq q$, consider the 2-universal hash family of functions $\mathcal{G} \triangleq \{\psi_s\}_{s \in \mathcal{S}}$, where for $s \in \mathcal{S}$,

$$\begin{aligned} \psi_s : \{0,1\}^q &\to \{0,1\}^k \\ v &\mapsto (s \odot v)_k, \end{aligned}$$

where $\odot$ is the multiplication in GF($2^q$) and $(\cdot)_k$ selects the $k$ most significant bits. In our proposed code design, the security layer is handled via a specific function $\psi_s \in \mathcal{G}$ indexed by the seed $s \in \mathcal{S}$. Then, we define

$$\begin{aligned} \varphi_s : \{0,1\}^k \times \{0,1\}^{q-k} &\to \{0,1\}^q \\ (m, b) &\mapsto s^{-1} \odot (m \| b), \end{aligned} \tag{3.3}$$

where $(\cdot\|\cdot)$ denotes the concatenation of two strings.

When the secrecy layer is combined with the reliability layer, our coding scheme can be summarized as follows. The input of the encoder $e_0$ is obtained by computing $V \triangleq \varphi_s(M, B)$, where $M \in \{0,1\}^k$ is the confidential message, and $B \in \{0,1\}^{q-k}$ is a sequence of $q - k$ random bits generated uniformly at random. After computing $V$, the encoder $e_0$, trained as described in Section 3.4.2, generates the codeword $X^n \triangleq e_0(V)$, which is sent over the channel by Alice. Bob and Eve observe $Y^n$ and $Z^n$, respectively, as described by (3.1) and (3.2). The decoder $d_0$, trained as described in Section 3.4.2, decodes $Y^n$ as $\widehat{V} \triangleq d_0(Y^n)$. Then, the receiver performs the multiplication of $\widehat{V}$ and $s$, which is followed by a selection of the $k$ most significant bits to create an estimate $\widehat{M}$ of $M$, i.e., $\widehat{M} \triangleq \psi_s(\widehat{V})$.

**Leakage evaluation via Mutual Information Neural Estimator (MINE) [71]**

Let $\{T_\theta : \{0,1\}^k \times \mathbb{R}^n \to \mathbb{R}\}_{\theta \in \Theta}$ be the set of functions parameterized by a deep neural network with parameters $\theta \in \Theta$. Define

$$I_\Theta(p_{MZ^n}) \triangleq \sup_{\theta \in \Theta} \mathbb{E}_{p_{MZ^n}} T_\theta(M, Z^n) - \log \mathbb{E}_{p_M p_{Z^n}} e^{T_\theta(M,Z^n)},$$

where $p_{MZ^n}$ is the joint probability distribution of $(M, Z^n)$. By [71], $I_\Theta(p_{MZ^n})$ can approximate the mutual information $I(M; Z^n)$ with arbitrary accuracy. Note that because the true distribution $p_{MZ^n}$ is unknown, one cannot directly use $I_\Theta(p_{MZ^n})$ to estimate $I(M; Z^n)$. However, by estimating the expectations in $I_\Theta(p_{MZ^n})$ with samples from $p_{MZ^n}$ and $p_M$ and $p_{Z^n}$, one can rewrite $I_\Theta(p_{MZ^n})$ as

$$\begin{aligned}\widehat{I}(M; Z^n) \triangleq \sup_{\theta \in \Theta} \frac{1}{l} \sum_{i=1}^{l} & T_\theta(m(i), z^n(i)) \\ & - \log \left[ \frac{1}{l} \sum_{i=1}^{l} e^{T_\theta(\bar{m}(i), \bar{z}^n(i))} \right],\end{aligned}$$

where the term $\frac{1}{l} \sum_{i=1}^{l} T_\theta(m(i), z^n(i))$ represents a sample mean using $l$ samples $(m(i), z^n(i))_{i \in \{1,\ldots,l\}}$ from $p_{MZ^n}$, and the term $\frac{1}{l} \sum_{i=1}^{l} e^{T_\theta(m(i), z^n(i))}$ represents a sample mean using $l$ samples $(\bar{m}(i), \bar{z}^n(i))_{i \in \{1,\ldots,l\}}$ from $p_M p_{Z^n}$.

The goal of MINE, whose architecture is depicted in Figure 3.3, is to design $T_\theta$ such that $\widehat{I}(M; Z^n)$ approaches the mutual information $I(M; Z^n)$. By [71], the estimator $\widehat{I}(M; Z^n)$ converges to $I(M; Z^n)$ when the number of samples is sufficiently large [71]. Guidelines to implement the estimator $\widehat{I}(M; Z^n)$ are also provided in [71].

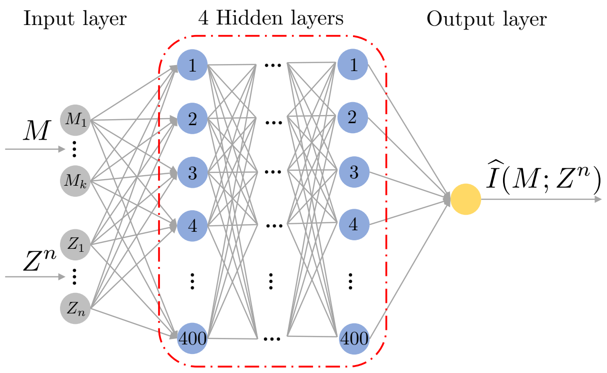


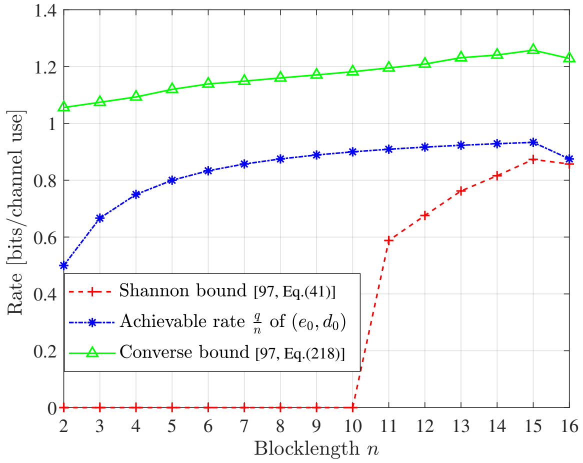
Figure 3.3: The security performance is evaluated in terms of the leakage $I(M; Z^n)$ via the mutual information estimator described in Section 3.4.3, where $M \triangleq (M_i)_{i\in\{1,2,\ldots,k\}}$, $Z^n \triangleq (Z_j)_{j\in\{1,2,\ldots,n\}}$.

### 3.4.4 Discussion on the communication rate when multiplexing protected and unprotected messages

Note that our approach incurs no rate loss compared to a traditional channel code. Our proposed design of an $(n, k, P)$ code with power constraint $P$, blocklength $n$, and secret message length $k$ consists of two layers: (i) a reliability layer implemented with a $(n, q, P)$ channel code $(e_0, d_0)$, and (ii) a secrecy layer. As described in Section 3.4.3, the secrecy layer takes as input a sequence of $q$ bits, out of which $k$ bits correspond to the secret message $M$ and $q - k$ bits correspond to random bits (denoted by $B$ in Section 3.4.3). By construction, the sequence $B$ can be reconstructed at Bob with an average probability of error $\mathbf{P}_e(e_0, d_0)$. However, the security constraint only holds on $M$ and not on $B$. To summarize, our code design transforms a channel code with rate $\frac{q}{n}$ into a wiretap code able to transmit a confidential message $M$ with rate $\frac{k}{n}$ and an unprotected message $B$ with rate $\frac{q-k}{n}$. Hence, there is no loss compared to a channel code, as the overall transmission rate is $\frac{q}{n}$.

### 3.4.5 Simulations and examples of code designs for $n \leq 16$

We now provide examples of code designs that follow the guidelines described in Sections 3.4.2, 3.4.3, and evaluate their performance in terms of average probability of error at Bob and leakage at Eve. The neural networks are implemented in Python 3.7 using Tensorflow 2.5.0.



(a)

| $n$ | $\mathbf{P}_e(e_0, d_0)$ |
|---|---|
| 2 | $3.26 \cdot 10^{-5}$ |
| 3 | $1.040 \cdot 10^{-4}$ |
| 4 | $2.042 \cdot 10^{-4}$ |
| 5 | $3.620 \cdot 10^{-4}$ |
| 6 | $5.280 \cdot 10^{-4}$ |
| 7 | $6.442 \cdot 10^{-4}$ |
| 8 | $7.710 \cdot 10^{-4}$ |
| 9 | $8.982 \cdot 10^{-4}$ |
| 10 | $1.0438 \cdot 10^{-3}$ |
| 11 | $1.2410 \cdot 10^{-3}$ |
| 12 | $1.4864 \cdot 10^{-3}$ |
| 13 | $2.0256 \cdot 10^{-3}$ |
| 14 | $2.2706 \cdot 10^{-3}$ |
| 15 | $2.8636 \cdot 10^{-3}$ |
| 16 | $1.7192 \cdot 10^{-3}$ |

(b)

Figure 3.4: Figure 3.4a shows the rate versus the blocklength $n$ obtained with $\epsilon \triangleq \mathbf{P}_e(e_0, d_0)$ listed in Figure 3.4b when $\text{SNR}_B = 9$dB.

### Autoencoder training for the design of the reliability layer $(e_0, d_0)$

We consider the channel model (3.1) with $\sigma_Y^2 \triangleq 10^{-\text{SNR}_B/10}$ and $\text{SNR}_B = 9$dB, where, as explained in Section 3.4.2, without loss of generality, we choose $P = 1$. The autoencoder is trained for $q = n - 1$ using SGD with the Adam optimizer [95] at a learning rate of $0.001$ over $600$ epochs of $10^5$ random encoder input messages with a batch size of $1000$. Due to the exponential growth of the complexity with $q$, we changed the value of $q$ to $n - 2$ when $n = 16$. Specifically, to evaluate $\mathbf{P}_e(e_0, d_0)$, we first generate the input $V \in \{0, 1\}^q$. Then, $V$ is passed through the trained encoder $e_0$, which generates the codewords $X^n$ and the channel output $Y^n$. Finally, the trained decoder $d_0$ forms an estimate of $V$ from $Y^n$.

Figure 3.4a compares the achievable rate $\frac{q}{n}$ of our reliability layer $(e_0, d_0)$ with the best known achievability and converse bounds [97, Section III.J] for channel coding. We observe that the rate of our reliability layer outperforms the achievability bounds from [97] for blocklengths smaller than or equal to 16 when $\text{SNR}_B = 9$dB. Note that for each value of $n$, this comparison is made for a given average probability of error $\mathbf{P}_e(e_0, d_0)$ as specified in Figure 3.4b.

### Design of the secrecy layer and leakage evaluation

The seeds selected for the simulations are given in Table 3.1. All possible seeds have been tested for the values of $n$ smaller than or equal to eight to minimize the leakage, and only one seed is tested for the values of $n$ greater than eight. The leakage is evaluated using MINE as follows. We used a fully connected feed-forward neural network with $4$ hidden layers, each having $400$ neurons, and a ReLU as activation function. The input layer has $k + n$ neurons, and the Adam optimizer with a learning rate of $0.001$ is used for the training. The samples of the joint distribution $p_{MZ^n}$ are produced via uniform generation of messages $M \in \{0, 1\}^k$ that are fed to the encoder $e = e_0 \circ \varphi_s$, whose output $X^n$ produces the channel output $Z^n$ at Eve. The samples of the marginal distributions are generated by dropping either $m$ or $z^n$ from the joint samples $(m, z^n)$. We have trained the neural network over $10000$ epochs of $20,000$ messages with a batch size of $2500$. Figure 3.5 shows the leakage versus the blocklength $n$ for different values of $k$ and SNR$_E = -5$dB. We observe that the leakage increases as $k$ increases for fixed $n$ and SNR$_E$, which is also supported by the following inequality on the leakage. When $k = 2$, if we write $M = (M_1, M_2)$, where $M_1, M_2 \in \{0, 1\}$, then by the chain rule and nonnegativity of the mutual information, we have

$$I(M; Z^n) = I(M_1; Z^n) + I(M_2; Z^n|M_1) \geq I(M_1; Z^n),$$

where $I(M_1; Z^n)$ is interpreted as the leakage of a code with secrecy rate $\frac{1}{n}$ by considering that $M_2$ is a random bit part of $B$ in (3.3).

**Remark 3.1.** *We observe a significant improvement in the leakage for a channel code coupled with our secrecy layer compared to the same channel code without any secrecy layer. For instance, for the blocklength $n = 8$ when $q = n - 1$ and SNR$_E$ $= -5$dB, the estimated mutual information between the input message of length $q$ to the encoder $e_0$ and the eavesdropper's channel observations is $\widehat{I}(V; Z^n) = 1.55$ bits. Therefore, for a one-bit input, the leakage is 0.2214 bits on average. Also, for $n = 8$, $q = n - 1$, $k = 1$, $s = 0001101$, and SNR$_E$ $= -5$dB, the estimated mutual information between the one-bit confidential message and the eavesdropper's channel observations is*

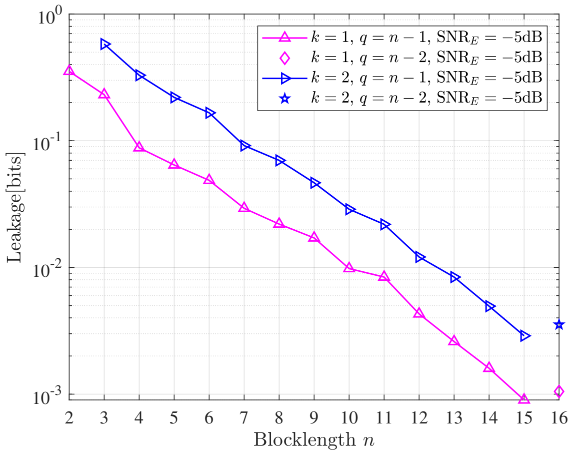


Figure 3.5: Leakage $\widehat{I}(M; Z^n)$ versus blocklength. When $n \in \{2, 3, 4 \ldots, 15\}$, $q = n - 1$, and when $n = 16$, $q = n - 2$.

$\widehat{I}(M; Z^n) = 0.022$ *bits. Hence, in this example, without our secrecy layer, a leakage as low as* 22 *percent is obtained per information bit on average, while with our secrecy layer, a leakage as low as* 2.2 *percent is obtained per information bit.*

### Average probability of error analysis

To evaluate $\mathbf{P}_e(e, d)$, the trained encoder $e_0$ encodes the message $M \in \{0, 1\}^k$ as $e_0(\varphi_s(M, B))$, as described in Section 3.4.3, where $B \in \{0, 1\}^{q-k}$ is a sequence of $q - k$ bits generated uniformly at random. The trained decoder $d_0$ forms $\widehat{M} \triangleq \psi_s(d_0(Y^n))$, as described in Section 3.4.3. Figure 3.6 shows $\mathbf{P}_e(e, d)$ versus the blocklength $n$. Note that we only plotted $\mathbf{P}_e(e, d)$ when $k = 1$ and $k = 2$ as an example, as one will always have $\mathbf{P}_e(e, d) \leq \mathbf{P}_e(e_0, d_0)$ for any value of $k$ by construction. From Figure 3.6, we also observe that, for fixed $n, q$, and $\mathbf{SNR}_B = 9$dB, the probability of error decreases as $k$ decreases, which is also supported by the following inequality

$$\mathbb{P}[(\widehat{M_1}, \widehat{M_2}) \neq (M_1, M_2)] \geq \mathbb{P}[\widehat{M_1} \neq M_1],$$

where we have used the union bound and $\mathbb{P}[\widehat{M_1} \neq M_1]$ is interpreted as the probability of error of a code with secrecy rate $\frac{1}{n}$ by considering that $M_2$ is a random bit part of $B$ in (3.3).

Table 3.1: Selected seeds for the security layer.

| $n$ | seed $s$<br>$(k = 1)$ | seed $s$<br>$(k = 2)$ |
|---|---|---|
| 2 | 1 | - |
| 3 | 11 | 11 |
| 4 | 010 | 010 |
| 5 | 1100 | 1100 |
| 6 | 00010 | 00011 |
| 7 | 001001 | 001001 |
| 8 | 0001101 | 0001101 |
| 9 | 10000000 | 10000000 |
| 10 | 100000000 | 100000000 |
| 11 | 1000000000 | 1000000000 |
| 12 | 10000000000 | 10000000000 |
| 13 | 100000000000 | 100000000000 |
| 14 | 1000000000000 | 1000000000000 |
| 15 | 10000000000000 | 10000000000000 |
| 16 | 10000000000000 | 10000000000000 |

### Discussion

From Figures 3.5 and 3.6, we, for instance, see that for $\text{SNR}_B = 9\text{dB}$ and $\text{SNR}_E = -5\text{dB}$, we have designed codes that show that the secrecy rate $\frac{1}{10}$ is $(\epsilon = 5.330 \cdot 10^{-4}, \delta = 9.80 \cdot 10^{-3})$-achievable with blocklength $n = 10$, and the secrecy rate $\frac{2}{13}$ is $(\epsilon = 1.5194 \cdot 10^{-3}, \delta = 8.40 \cdot 10^{-3})$-achievable with blocklength $n = 13$.

Figure 3.7a compares the achievable secrecy rate $\frac{k}{n}$ of our code design $(e, d)$ with the best known achievability [63, Theorem 7] and converse bounds [63, Theorem 12] for the Gaussian wiretap channel, which are reviewed in the Appendix F for convenience. We observe that the rate of our code outperforms the best known achievability bounds for blocklengths smaller than or equal to 16 when $k = 1$, $\text{SNR}_B = 9\text{dB}$, and $\text{SNR}_E = -5\text{dB}$. Note that the best known upper bounds from [63] may not be optimal for small blocklengths, and improving them is an open problem. Note also that for each value of $n$, the comparison is made for a given average probability of error $\mathbf{P}_e(e, d)$ and leakage $\widehat{I}(M; Z^n)$ as specified in Figure 3.7b.

In Figure 3.8, we also plotted $\epsilon \triangleq \mathbf{P}_e(e, d)$ versus $\delta \triangleq \widehat{I}(M; Z^n)$ obtained from Figure 3.7b.

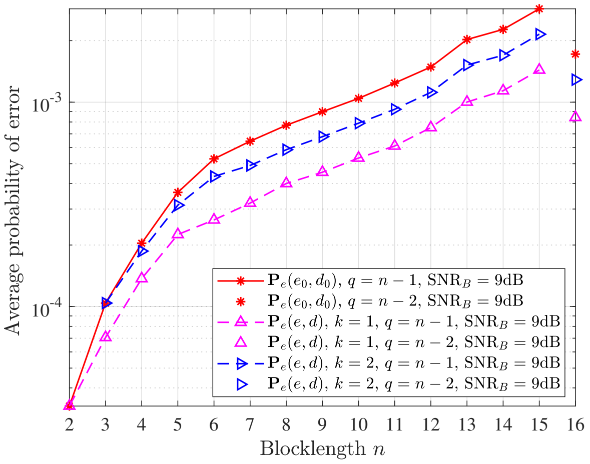
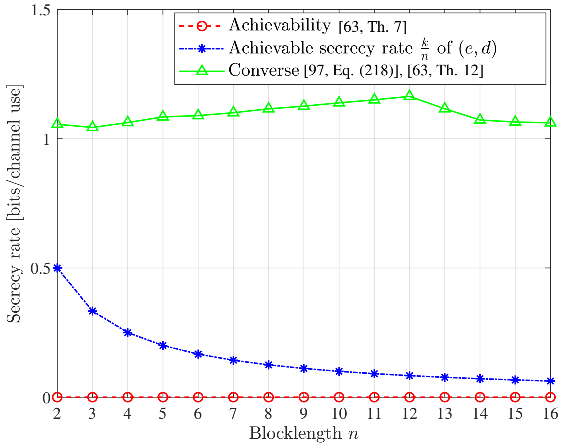


Figure 3.6: Average probability of error versus blocklength. When $n \in \{2, 3, 4 \ldots, 15\}$, $q = n-1$, and when $n = 16$, $q = n - 2$. During the training, the signal-to-noise ratio is $\text{SNR}_B = 9\text{dB}$.

### 3.4.6 Simulations and examples of code designs for $n \leq 128$

We consider the channel model (3.1) with $\sigma_Y^2 \triangleq 10^{-\text{SNR}_B/10}$ and $\text{SNR}_B = 0\text{dB}$. For the design of the reliability layer $(e_0, d_0)$, the autoencoder is trained for $(n, q) = (32, 8)$, $(n, q) = (64, 8)$, $(n, q) = (96, 12)$, and $(n, q) = (128, 12)$ at a learning rate of $0.0001$ over 600 epochs of $4 \times 10^5$ random encoder input messages with a batch size of $5000$. Then, $50 \times 10^6$ random messages are used to evaluate the average probability of errors $\mathbf{P}_e(e_0, d_0)$ and $\mathbf{P}_e(e, d)$ as described in Sections 3.4.5 and 3.4.5. Figure 3.9 shows $\mathbf{P}_e(e, d)$ versus the blocklength $n$. As expected, we observe that, for fixed $k$ and $q$, the probability of error decreases as $n$ increases.

The secrecy layer is implemented similarly to Section 3.4.3 with $k = 1$, and we compute the leakage $I(M; Z^n)$ as in Section 3.4.5. We consider the model in (3.2) with $\sigma_Z^2 \triangleq 10^{-\text{SNR}_E/10}$ and $\text{SNR}_E = -15\text{dB}$. Additionally, in our simulations, for blocklengths $n = 32$ and $n = 64$, the seed is chosen as $s = 00000001$, and for blocklengths $n = 96$ and $n = 128$, the seed is chosen as $s = 000010000000$. Figure 3.10 shows $\widehat{I}(M; Z^n)$ versus the blocklength $n$. As expected, we observe that, for fixed $k$ and $q$, the leakage increases as $n$ increases. Finally, we observe in Figure 3.11a that the rate of our code outperforms the best known achievability bounds [63, Theorem 7] for blocklengths smaller than or equal to 128 when $k = 1$, $\text{SNR}_B = 0\text{dB}$, and $\text{SNR}_E = -15\text{dB}$.



(a)

| $n$ | $\mathbf{P}_e(e,d)$<br>$(k=1)$ | $\widehat{I}(M;Z^n)$<br>$(k=1)$ |
|---|---|---|
| 2 | $3.26 \cdot 10^{-5}$ | $3.5300 \cdot 10^{-1}$ |
| 3 | $7.06 \cdot 10^{-5}$ | $2.3100 \cdot 10^{-1}$ |
| 4 | $1.370 \cdot 10^{-4}$ | $8.800 \cdot 10^{-2}$ |
| 5 | $2.252 \cdot 10^{-4}$ | $6.440 \cdot 10^{-2}$ |
| 6 | $2.654 \cdot 10^{-4}$ | $4.870 \cdot 10^{-2}$ |
| 7 | $3.208 \cdot 10^{-4}$ | $2.930 \cdot 10^{-2}$ |
| 8 | $4.000 \cdot 10^{-4}$ | $2.200 \cdot 10^{-2}$ |
| 9 | $4.542 \cdot 10^{-4}$ | $1.710 \cdot 10^{-2}$ |
| 10 | $5.330 \cdot 10^{-4}$ | $9.80 \cdot 10^{-3}$ |
| 11 | $6.104 \cdot 10^{-4}$ | $8.40 \cdot 10^{-3}$ |
| 12 | $7.510 \cdot 10^{-4}$ | $4.30 \cdot 10^{-3}$ |
| 13 | $1.0014 \cdot 10^{-3}$ | $2.60 \cdot 10^{-3}$ |
| 14 | $1.1376 \cdot 10^{-3}$ | $1.16 \cdot 10^{-3}$ |
| 15 | $1.4368 \cdot 10^{-3}$ | $9.0 \cdot 10^{-4}$ |
| 16 | $8.426 \cdot 10^{-4}$ | $1.05 \cdot 10^{-3}$ |

(b)

Figure 3.7: Figure 3.7a shows the secrecy rate versus the blocklength $n$ obtained from $\epsilon \triangleq \mathbf{P}_e(e,d)$ and $\delta \triangleq \widehat{I}(M;Z^n)$ listed in Figure 3.7b when $\text{SNR}_B = 9\text{dB}$ and $\text{SNR}_E = -5\text{dB}$. The converse bound is obtained as the minimum between [97, Eq. (218)] and [63, Th. 12].

## 3.5 Compound and arbitrarily varying wiretap channel models

We first motivate the compound and arbitrarily varying wiretap channel models in Section 3.5.1. We then formally describe these two models in Section 3.5.2. We present our coding scheme design in Section 3.5.3. Finally, we evaluate the performances of our code design through simulations for the compound and arbitrarily varying Gaussian wiretap channels in Sections 3.5.4 and 3.5.5, respectively.

### 3.5.1 Background

In the setting of Section 3.3, the channel statistics are assumed to be perfectly known to Alice and Bob and fixed during the entire transmission. However, in practice, the channel statistics may not be perfectly known due to the nature of the wireless channel and inaccuracy in estimating channel statistics. Further, in some scenarios, eavesdroppers could be active and influence their own channel statistics by changing their location, or the statistics of Bob's channel through jamming. To

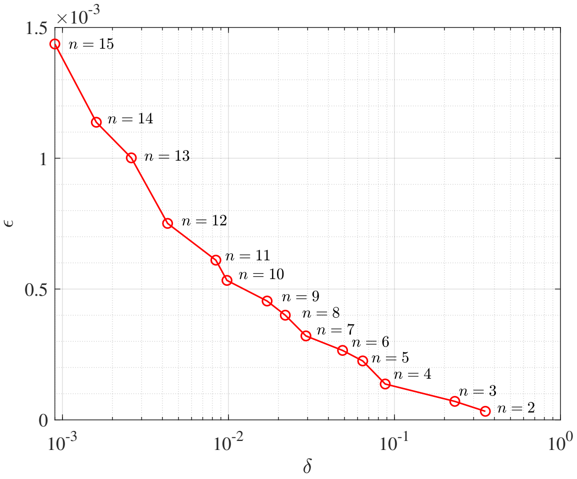


Figure 3.8: $\epsilon \triangleq \mathbf{P}_e(e,d)$ versus $\delta \triangleq \widehat{I}(M;Z^n)$ obtained from Figure 3.7b. When $n \in \{2,3,4\ldots,15\}$, $q = n-1$, and the secrecy rate is $\frac{1}{n}$.

model such scenarios, two types of models have been introduced: compound wiretap channels and arbitrarily varying wiretap channels. For compound models, e.g., [9,73,98,99], the channel statistics are fixed for all channel uses. Whereas for arbitrarily varying models, e.g., [74,75,98–101], the channel statistics may change in an unknown and arbitrary manner from channel use to channel use. Constructive coding schemes have also been proposed in [11,61] for discrete compound and arbitrarily varying wiretap channels. While all the works above consider the asymptotic regime, in this section, we design short blocklength codes for the compound and arbitrarily varying wiretap channel models.

### 3.5.2 Models

For $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathbb{R}$, a compound or arbitrarily varying memoryless Gaussian wiretap channel $\left(\mathcal{X}, (p_{Y_i Z_j|X})_{i\in\mathcal{I}, j\in\mathcal{J}}, \mathcal{Y}\times\mathcal{Z}\right)$ is defined for $i \in \mathcal{I}$, $j \in \mathcal{J}$, by

$$Y_i \triangleq X + N_{Y_i}, \quad Z_j \triangleq X + N_{Z_j},$$

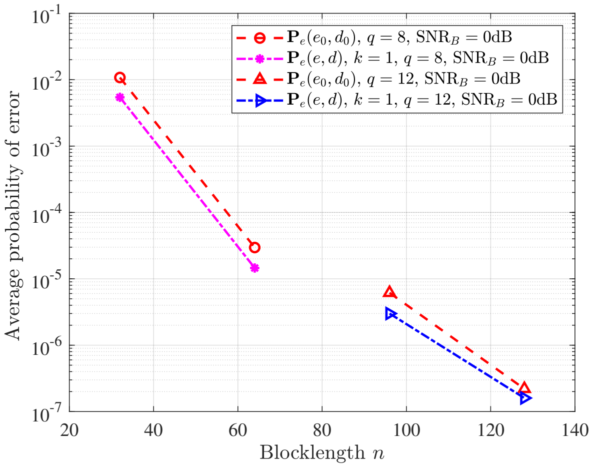


Figure 3.9: Average probability of error versus blocklength. When $n \in \{32, 64\}$, $q = 8$, and when $n \in \{96, 128\}$, $q = 12$.

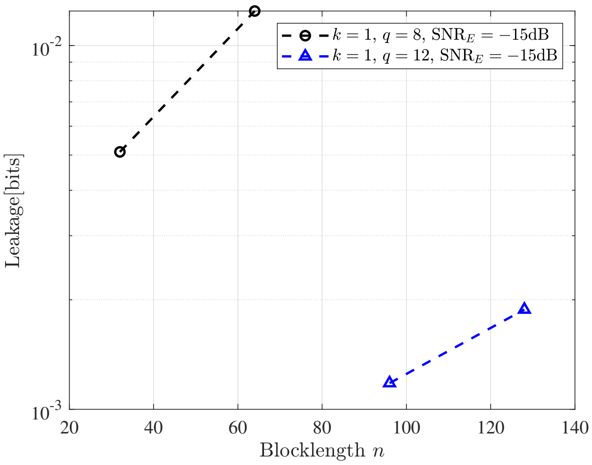
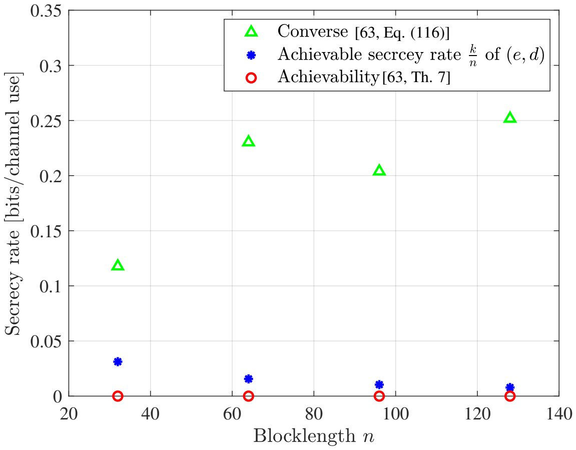


Figure 3.10: Leakage $\widehat{I}(M; Z^n)$ versus blocklength. When $n \in \{32, 64\}$, $q = 8$, and when $n \in \{96, 128\}$, $q = 12$.



(a)

| $n$ | $\mathbf{P}_e(e,d)$<br>$(k=1)$ | $\widehat{I}(M;Z^n)$<br>$(k=1)$ |
|---|---|---|
| 32 | $5.430199 \cdot 10^{-3}$ | $5.10 \cdot 10^{-3}$ |
| 64 | $1.4549 \cdot 10^{-5}$ | $1.244 \cdot 10^{-2}$ |
| 96 | $2.999 \cdot 10^{-6}$ | $1.18 \cdot 10^{-3}$ |
| 128 | $1.59 \cdot 10^{-7}$ | $1.88 \cdot 10^{-3}$ |

(b)

Figure 3.11: Figure 3.11a shows the secrecy rate versus the blocklength $n$ obtained from $\epsilon \triangleq \mathbf{P}_e(e,d)$ and $\delta \triangleq \widehat{I}(M;Z^n)$ listed in Figure 3.11b when SNR$_B = 0$dB and SNR$_E = -15$dB.

where $N_{Y_i}$ and $N_{Z_j}$ are zero mean Gaussian random variables with variances $\sigma^2_{Y_i}$ and $\sigma^2_{Z_j}$, respectively.

For the compound wiretap channel model, the channel statistics are constant throughout the transmission and are known to belong to given uncertainty sets $\mathcal{I}$, $\mathcal{J}$. The confidential message $M$ is encoded into a transmitted sequence $X^n$, and $Y_i^n$ and $Z_j^n$ represent the corresponding received sequence at the legitimate receiver and eavesdropper, respectively, for some $i \in \mathcal{I}$ and $j \in \mathcal{J}$.

**Definition 3.4.** *A secrecy rate $\frac{k}{n}$ is $(\epsilon, \delta)$-achievable with power constraint $P$ for the compound wiretap channel if there exists a $(n, k, P)$ code such that*

$$\max_{i \in \mathcal{I}} \mathbf{P}^i_e(e,d) \leq \epsilon, \tag{3.4}$$

$$\max_{j \in \mathcal{J}} I(M; Z^n_j) \leq \delta, \tag{3.5}$$

*where $\mathbf{P}^i_e(e,d) \triangleq \frac{1}{2^k}\sum_{m=1}^{2^k} \mathbb{P}[d(Y^n_i) \neq m | m \text{ is sent}]$.*

In contrast to the compound wiretap channel, the channel statistics in arbitrarily varying wiretap channel models may vary in an unknown and arbitrary manner from channel use to channel use. Specifically, for the arbitrarily varying wiretap channel model, let $Y^n_{\mathbf{i}}$ and $Z^n_{\mathbf{j}}$ represent the corresponding received sequence at the legitimate receiver and eavesdropper, respectively, for some $\mathbf{i} \in \mathcal{I}^n$ and $\mathbf{j} \in \mathcal{J}^n$.

**Definition 3.5.** *A secrecy rate $\frac{k}{n}$ is $(\epsilon, \delta)$-achievable with power constraint $P$ for the arbitrarily varying wiretap channel if there exists a $(n, k, P)$ code such that*

$$\max_{\mathbf{i} \in \mathcal{I}^n} \mathbf{P}_e^{\mathbf{i}}(e, d) \leq \epsilon,$$

$$\max_{\mathbf{j} \in \mathcal{J}^n} I(M; Z_{\mathbf{j}}^n) \leq \delta,$$

*where $\mathbf{P}_e^{\mathbf{i}}(e, d) \triangleq \frac{1}{2^k} \sum_{m=1}^{2^k} \mathbb{P}[d(Y_{\mathbf{i}}^n) \neq m | m$ is sent$]$.*

### 3.5.3 Coding scheme design

For the compound and the arbitrarily varying wiretap channels, the design of $(e_0, d_0)$ for the reliability layer and $(\varphi_s, \psi_s)$ for the secrecy layer is similar to Sections 3.4.2 and 3.4.3, respectively. Specifically, we train the encoder/decoder pair for Bob's channel with noise variance $\sigma_{Y_{i^*}}^2 \triangleq \max_{i \in \mathcal{I}} \sigma_{Y_i}^2$, where $\sigma_{Y_i}^2 \triangleq 10^{-\text{SNR}_B(i)/10}$, $i \in \mathcal{I}$. In other words, the reliability layer is designed for the worse, in terms of signal-to-noise ratio, Bob's channel. Note also that, during the training phase, the noise variance is fixed for all the channel uses. Then, we optimize the seed $s$ by minimizing the leakage for Eve's channel with noise variance $\sigma_{Z_{j^*}}^2 \triangleq \min_{j \in \mathcal{J}} \sigma_{Z_j}^2$, where $\sigma_{Z_j}^2 \triangleq 10^{-\text{SNR}_E(j)/10}$, $j \in \mathcal{J}$. In other words, the secrecy layer is designed for the best, in terms of signal-to-noise ratio, Eve's channel. This optimized seed $s$ is then used by the encoder/decoder pair $(e, d) = (e_0 \circ \varphi_s, \psi_s \circ d_0)$, from which we evaluate $(\mathbf{P}_e^i(e, d), I(M; Z_j^n))$, $i \in \mathcal{I}$, $j \in \mathcal{J}$, and $(\mathbf{P}_e^{\mathbf{i}}(e, d), I(M; Z_{\mathbf{j}}^n))$, $\mathbf{i} \in \mathcal{I}^n$, $\mathbf{j} \in \mathcal{J}^n$ in Sections 3.5.4, 3.5.5.

### 3.5.4 Simulations and examples of code designs for the compound wiretap channel

#### Average probability of error analysis

In our simulations, we consider $\mathcal{I} = \{1, 2\}$ and $\text{SNR}_B(i) \in \{9, 10\}$dB, $i \in \mathcal{I}$. We evaluate the average probability of error $\mathbf{P}_e^i(e, d)$ for $i \in \mathcal{I}$ as follows. The autoencoder is trained at

$\text{SNR}_B(i^*) = 9\text{dB}$, where $i^* = 1$. The message $M \in \{0,1\}^k$ generated uniformly at random is passed through the trained encoder $e_0$, which generates the codewords $X^n$ and the channel output $Y_i^n \triangleq X^n + N_{Y_i}^n$, $i \in \mathcal{I}$, where $N_{Y_i} \sim \mathcal{N}(0, \sigma_{Y_i}^2)$. Then, the trained decoder $d_0$ forms an estimate $\widehat{M_i}, i \in \mathcal{I}$. Here, $\sigma_{Y_i}^2$ is fixed for the entire duration of the transmission. We use $5 \times 10^6$ random messages to evaluate the average probability of error. Figure 3.12 shows the average probability of error $\mathbf{P}_e^i(e,d)$ when $k = 1$ for $\text{SNR}_B(i^* = 1) = 9\text{dB}$ and $\text{SNR}_B(i = 2) = 10\text{dB}$. We observe from Figure 3.12 that it is sufficient to design our code for the worst signal-to-noise ratio for Bob, i.e., $\text{SNR}_B(i^*) = 9\text{dB}$. In particular, we observe that, irrespective of what the actual channel is, Bob is able to decode the message with a probability of error smaller than or equal to $\mathbf{P}_e^{i^*=1}(e,d)$.

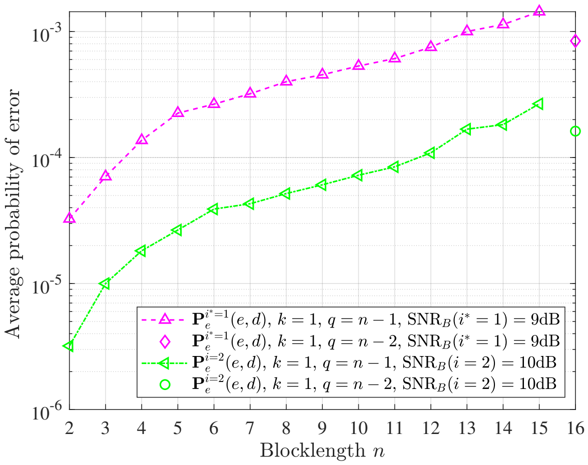


Figure 3.12: Average probability of error versus blocklength $n$.

### Leakage evaluation

For the simulations, we consider $\mathcal{J} = \{1,2,3\}$ and $\text{SNR}_E(j) \in \{-8, -6.5, -5\}\text{dB}$, $j \in \mathcal{J}$. We compute the leakage $I(M; Z_j^n)$ for $j \in \mathcal{J}$ as in Section 3.4.5. The message $M \in \{0,1\}^k$ is passed through the trained encoder $e_0$, which generates the codewords $X^n$, and the channel output at Eve is $Z_j^n \triangleq X^n + N_{Z_j}^n, j \in \mathcal{J}$, where $N_{Z_j} \sim \mathcal{N}(0, \sigma_{Z_j}^2)$. The noise variance $\sigma_{Z_j}^2$ is fixed for the entire duration of the transmission. Figure 3.13 shows the estimated leakage $\widehat{I}(M; Z_j^n)$,

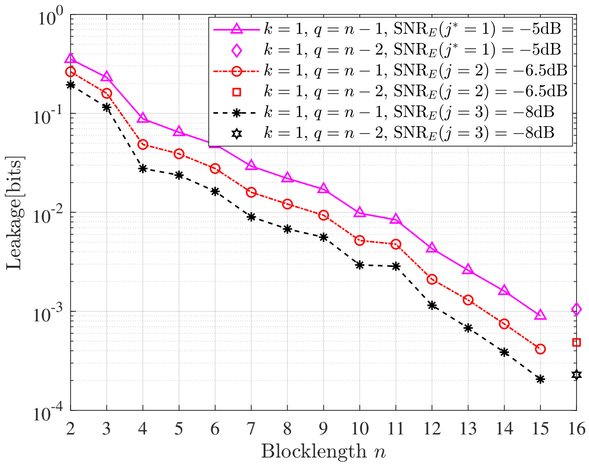


Figure 3.13: Leakage $\widehat{I}(M; Z_j^n)$ versus blocklength $n$.

at $\mathrm{SNR}_E(j = 1) = -8\mathrm{dB}$, $\mathrm{SNR}_E(j = 2) = -6.5\mathrm{dB}$, and $\mathrm{SNR}_E(j^* = 3) = -5\mathrm{dB}$. From Figure 3.13, we observe that it is sufficient to design our code for the best signal-to-noise ratio, i.e., $\mathrm{SNR}_E(j^*) = -5\mathrm{dB}$. In particular, we see that, irrespective of what the actual eavesdropper's channel is, we always achieve a leakage smaller than or equal to $\widehat{I}(M; Z_{j^*}^n)$, which is also supported by the following inequality on the leakage. For $j \in \mathcal{J}$ and $j^* \in \arg\min_{j \in \mathcal{J}} \sigma_{Z_j}^2$, we have

$$
\begin{aligned}
I(M; Z_j^n) &\leq I(M; Z_j^n Z_{j^*}^n) \\
&= I(M; Z_{j^*}^n) + I(M; Z_j^n | Z_{j^*}^n) \\
&= I(M; Z_{j^*}^n),
\end{aligned}
$$

where the first inequality holds by the chain rule and nonnegativity of the mutual information, the first equality holds by the chain rule, and the last equality holds because, without loss of generality, one can redefine $Z_j$ such that $Z_j = Z_{j^*} + N$, where $N \sim \mathcal{N}(0, \sigma_{Z_j}^2 - \sigma_{Z_{j^*}}^2)$, since the distributions $p_{Z_j|X}$ and $p_{Z_{j^*}|X}$ are preserved and the constraints (3.4) and (3.5) of the problem do not depend on the joint distributions $p_{Z_j Z_{j^*}}$.

As an example, from Figures 3.12 and 3.13, we see that for $\mathrm{SNR}_B(i) \in \{9, 10\}\mathrm{dB}$, $i \in \mathcal{I}$, and $\mathrm{SNR}_E(j) \in \{-8, -6.5, -5\}\mathrm{dB}$, $j \in \mathcal{J}$, we have designed codes that show that the secrecy rate $\frac{1}{8}$

is ($\epsilon = 4.0 \cdot 10^{-4}, \delta = 2.2 \cdot 10^{-2}$)-achievable with blocklength $n = 8$.

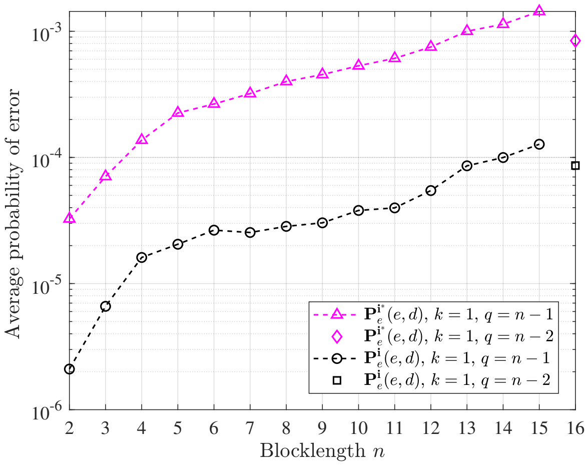


Figure 3.14: Average probability of error versus blocklength $n$ when $k = 1$ and training $\text{SNR}_B(\mathbf{i}^*) = 9\text{dB}$.

### 3.5.5 Simulations and examples of code designs for the arbitrarily varying wiretap channel

#### Average probability of error analysis

For the arbitrarily varying channel, we evaluate the probability of error $\mathbf{P}_e^{\mathbf{i}}(e, d)$, $\mathbf{i} \in \mathcal{I}^n$, for $k = 1$ in Figure 3.14 as follows. We consider $\mathcal{I} = \{1, 2, 3, \dots, 31\}$ and $\text{SNR}_B(i) \in \{9, 9.1, 9.2, \dots, 12\}$, $i \in \mathcal{I}$. The autoencoder is trained at $\text{SNR}_B(i^* = 1) = 9\text{dB}$, where the noise variance is fixed for the entire duration of the transmission. The message $M \in \{0, 1\}^k$ generated uniformly at random is passed through the trained encoder $e_0$, which generates the codewords $X^n$ and the channel output at Bob $Y_{\mathbf{i}}^n \triangleq X^n + N_{Y_{\mathbf{i}}^n}$, $\mathbf{i} \in \mathcal{I}^n$. Then, the trained decoder $d_0$ forms an estimate $\widehat{M_{\mathbf{i}}}$, $\mathbf{i} \in \mathcal{I}^n$. Here, $N_{Y_{\mathbf{i}}^n}$ is a length $n$ vector whose variance of each component is picked uniformly at random from the known uncertainty set $\{10^{-12\text{dB}/10}, 10^{-11.9\text{dB}/10}, 10^{-11.8\text{dB}/10}, \dots, 10^{-9\text{dB}/10}\}$. For our simulations, the variance of the noise vector $N_{Y_{\mathbf{i}}^n}$ is fixed for every $50,000$ codewords. The autoencoder is tested with $5 \times 10^6$ random messages for $n > 10$ and with $10^7$ random messages for $n \leq 10$. Figure 3.14 shows that even though there is a mismatch between the training signal-to-noise ratio of the encoder/decoder pair and the actual channel, Bob is still able to decode the message with a probability of error smaller than or equal to $\mathbf{P}_e^{\mathbf{i}^*}(e,d)$, where $\mathbf{i}^*$ is a vector made of $n$ repetitions of $i^*$.

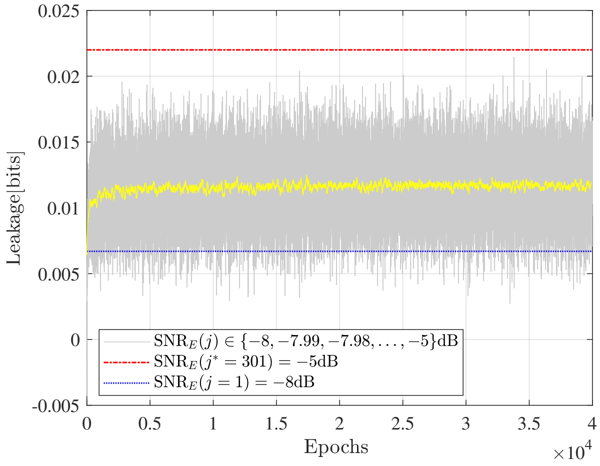


Figure 3.15: Example of leakage $\widehat{I}(M; Z_{\mathbf{j}}^n)$ versus epochs when $k = 1$, $n = 8$, $q = n - 1$, $s = 001101$, and $j \in \{1, 2, 3, \ldots, 301\}$. The grey curve represents the estimated leakage by a mutual information neural estimator and the yellow curve represents the 100-sample moving average of the estimated leakage. The red and blue curves represent the estimated leakage for $\mathrm{SNR}_E(j^* = 301)$ and $\mathrm{SNR}_E(j = 1)$, respectively, after convergence.

### Leakage evaluation

For the arbitrarily varying channel, we evaluate the leakage $I(M; Z_{\mathbf{j}}^n)$, for $\mathbf{j} \in \mathcal{J}^n$, as in Section 3.4.5. The channel output at Eve is $Z_{\mathbf{j}}^n \triangleq X^n + N_{Z_{\mathbf{j}}^n}, \mathbf{j} \in \mathcal{J}^n$. In Figure 3.15, we consider $\mathcal{J} = \{1, 2, 3, \ldots, 301\}$ and $\mathrm{SNR}_E(j) \in \{-8, -7.99, -7.98, \ldots, -5\}$dB, $j \in \mathcal{J}$. Figure 3.15 shows the estimated mutual information $\widehat{I}(M; Z_{\mathbf{j}}^n)$ when $k = 1$ and $n = 8$, where the variance of the noise vector $N_{Z_{\mathbf{j}}^n}$ is fixed for $20,000$ codewords per epoch. Here, $N_{Z_{\mathbf{j}}^n}$ is a length $n$ vector whose variance of each component is picked uniformly at random from the known uncertainty set $\{10^{5\mathrm{dB}/10}, 10^{5.01\mathrm{dB}/10}, 10^{5.02\mathrm{dB}/10}, \ldots, 10^{8\mathrm{dB}/10}\}$. We can see from Figure 3.15 that it is sufficient to design our code for the best signal-to-noise ratio, i.e., $\mathrm{SNR}_E(j^* = 301) = -5$dB. In particular, we observe that, regardless of what the actual channel is, we always achieve a leakage smaller than or equal to $\widehat{I}(M; Z_{\mathbf{j}^*}^n)$, where $\mathbf{j}^*$ is a vector made of $n$ repetitions of $j^*$, which is also supported by the following inequality on the leakage.

For any $\mathbf{j} \in \mathcal{J}^n$, we have $I(M; Z_{\mathbf{j}}^n) \leq I(M; Z_{\mathbf{j}^*}^n)$, because, similar to the compound setting, without loss of generality, one can redefine $Z_{\mathbf{j}}^n$ such that $M - Z_{\mathbf{j}}^n - Z_{\mathbf{j}^*}^n$ forms a Markov chain.

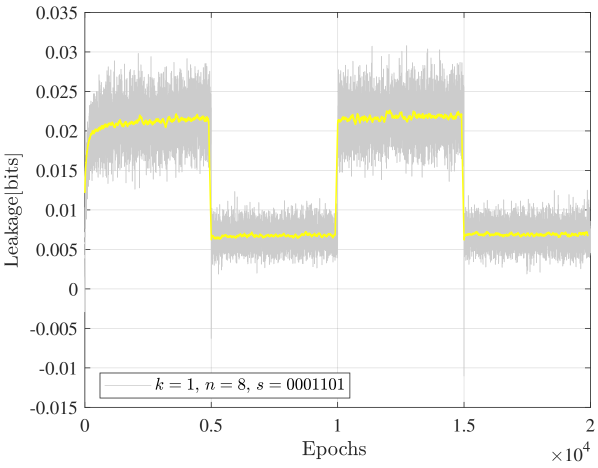


Figure 3.16: Example of leakage $\widehat{I}(M; Z_{\mathbf{j}}^n)$ versus epochs when $q = n - 1$ and $\text{SNR}_E(j) \in \{-8, -5\}$, $j \in \{1, 2\}$. The yellow curve represents the 100-sample moving average of the estimated leakage.

Figure 3.16 shows the estimated mutual information $\widehat{I}(M; Z_{\mathbf{j}}^n)$ for $k = 1$ and $n = 8$, when $\text{SNR}_E \in \{-8, -5\}$dB. Here, each component of the noise vector $N_{Z_{\mathbf{j}}^n}$ has fixed variance, which alternatively takes the values $10^{5\text{dB}/10}$ and $10^{8\text{dB}/10}$, for every $5000$ epochs with $20,000$ codewords per epoch. In other words, each component of the noise vector has variance $10^{5\text{dB}/10}$ for the first $5000$ epochs, then $10^{8\text{dB}/10}$ for the next $5000$ epochs, and so on. Again, we observe that it is sufficient to consider the worst case for the code design, since the leakage is always upper bounded by the leakage obtained for the best eavesdropper's signal-to-noise ratio, i.e., when $\text{SNR}_E = -5$dB.

### 3.6 Concluding Remarks

We designed short blocklength codes for the Gaussian wiretap channel under an information-theoretic secrecy metric. Our approach consisted in decoupling the reliability and secrecy constraints to offer a simple and modular code design, and to precisely control how small the leakage is. Specifically, we handled the reliability constraint via an autoencoder and the secrecy constraint via hash functions. We evaluated the performance of our code design through simulations in terms of probability of error at the legitimate receiver and leakage at the eavesdropper for blocklengths smaller than or equal to $128$. Our results provide examples of code designs that outperform the best known achievable secrecy rates obtained non-constructively for the Gaussian wiretap channel. We highlight that our code design method can be applied to any channels since it does not require the knowledge of the channel model but only the knowledge of input and output channel samples. We also showed that our code design is applicable to settings where uncertainty holds on the channel statistics, e.g., compound wiretap channels, and arbitrarily varying wiretap channels.

# CHAPTER 4

# SECRET SHARING OVER A GAUSSIAN BROADCAST CHANNEL: OPTIMAL CODING SCHEME DESIGN AND DEEP LEARNING APPROACH AT SHORT BLOCKLENGTH

## 4.1 Introduction

[5] and [6] pioneered the secret sharing model where a dealer distributes a secret among a set of participants with the requirement that only pre-defined sets of participants can recover the secret, while any other sets of colluding participants cannot learn any information about the secret. In such traditional secret sharing models (we refer to [23] for an excellent literature review of the subject), it is assumed that the dealer and each participant can communicate over an individual and perfectly secure channel at no cost. Subsequently, with the goal to avoid this assumption, secret sharing schemes from noisy resources have been studied for channel models [8] and source models [2, 10, 11], where no secure communication links are available between the dealer and the participants. Another feature of secret sharing from noisy resources is that, unlike traditional secret sharing schemes, the creation and distribution phases of the protocol are no longer restricted to being independently designed but can now be jointly designed.

### 4.1.1 Overview of the model studied in this chapter

We consider the secret sharing model of [8] where noisy resources, in the form of a Gaussian channel, are available between the dealer and participants. Specifically, the dealer can communicate with the participants over a Gaussian broadcast channel where each participant observes scalar

Gaussian channel outputs. The dealer transmits a secret message by encoding it into a codeword, which is then sent over $n$ uses of the channel and yields the channel output observations at the participants. In this setting, a reliability constraint ensures that any subset of participants with size $t$ can recover the secret from their vector of $t$ channel outputs, and a security constraint ensures that any subset of participants with size $z < t$ cannot learn any information about the secret from their vector of $z$ channel outputs. These two constraints define a threshold access structure similar to traditional secret sharing models as in [5, 6, 102, 103].

### 4.1.2 Contributions

The secret sharing capacity has been established in [8] with a random coding argument that jointly considers the reliability and secrecy constraints. We consider a coding scheme that relies on two coding layers, namely, a reliability layer and a secrecy layer, to achieve the secret sharing capacity. Specifically, the secrecy layer can be implemented with hash functions, and the reliability layer can be implemented with a channel code for a compound channel without any security constraint.

The main insights and benefits of our approach are $(a)$ a modular design that allows a simplified code design, for instance, if only one of the two layers need to be (re)designed, $(b)$ a code design that offers a universal way of dealing with the secrecy constraint through the use of hash functions, which is particularly useful to handle an access structure and, in particular, to ensure security against multiple subsets of colluding users that are associated with different channel statistics, $(c)$ guidelines for a practical code design at finite blocklength as discussed next in Section 4.1.2.

#### Coding scheme design at finite blocklength

Following the two-layer coding approach described above, we design secret sharing schemes at finite blocklength. Specifically, we use a neural network-based autoencoder to design the reliability layer, and hash functions to design the security layer.

To evaluate the performance of our constructed code, we empirically evaluate the probability of error and estimate the information leakage for blocklengths $n$ at most $20$. Specifically, information leakage is defined as the mutual information between the secret and the channel output observations for unauthorized sets of users. Note that, even for small values of $n$, this information leakage estimation is challenging with standard techniques such as binning of the probability space [68], $k$-nearest neighbor statistics [69], or maximum likelihood estimation [70], we thus evaluate the information leakage by using the mutual information neural estimator (MINE) from [71]. Our simulation results demonstrate a precise control of the probability of error and leakage thanks to the two separate coding layers.

### 4.1.3 Related works

**Related works on compound wiretap channels**

a) *Theoretical and non-constructive results*: The compound wiretap channel [9, 73] is a generalization of Wyner's wiretap channel [19] to the case of multiple unknown channel states, where secure and reliable communication needs to be guaranteed regardless of the channel state. The connection between compound wiretap channels and secret sharing over noisy channels has been established in [8] by remarking that, similar to compound settings, an access structure for secret sharing yields multiple security constraints and multiple reliability constraints that must be ensured simultaneously.

b) *Results on code constructions*: Explicit compound wiretap codes have been proposed for discrete memoryless channels for the asymptotic blocklength regime in [61]. Finite-length code constructions for scalar Gaussian compound channels have been studied in [4]. In contrast, in this chapter, our contributions (see Section 4.1.2) is to not only design finite-length compound wiretap codes but to also consider vector Gaussian channel observations, which is needed in the context of the secret sharing problem we consider.

### Related works on wiretap channels

While several wiretap code designs have been proposed for various channel models under non-information-theoretic security metrics, e.g., [76–81, 89, 104], we focus our discussion on works that, similar to our work, consider an information-theoretic security metric.

a) *Results for the asymptotic blocklength regime*: A coding strategy that separately handles the reliability and secrecy constraints with two separate coding layers has previously been used for the discrete wiretap channels in [53], [54], and the Gaussian wiretap channel in [55]. The above references consider the asymptotic blocklength regime.

   Our contribution is to not only generalize the result in [55] to a compound setting but also to generalize it to vector, rather than scalar, Gaussian channel outputs.

   Note that other coding schemes based on LDPC codes [46–48], polar codes [49–52], or random lattice codes [56] have been proposed for degraded or symmetric wiretap channel models, and constructive [57–59] and random [60] polar coding schemes have been proposed to achieve the secrecy capacity of non-degraded discrete wiretap channels. Coding schemes that combine polar codes and invertible extractors have also been proposed to avoid the need for a pre-shared secret under strong secrecy [11, 61]. However, only [61] considers a compound setting (for discrete memoryless wiretap channels), and none of the references above consider compound Gaussian wiretap channels, as it is needed for the secret sharing model we consider.

b) *Code construction in the non-asymptotic regime*: Punctured systematic irregular LDPC codes have been proposed for the binary phase-shift-keyed-constrained Gaussian wiretap channel in [82], and LDPC codes for the Gaussian wiretap channel have also been developed in [83]. Randomized Reed-Muller codes have been proposed for the Gaussian wiretap channel in [84]. Coding scheme designs based on feed-forward neural network autoencoders have also been proposed in [94], where the security and reliability constraints are handled jointly, and in [4], where the security and reliability constraints are handled separately. As discussed in Section 4.1.2, one of our contributions is to design a short blocklength coding scheme that, unlike the above references, handles $(a)$ multiple reliability constraints and multiple security constraints simultaneously, i.e., a compound model, and $(b)$ vector Gaussian channel outputs.

The chapter is organized as follows. The problem statement is provided in Section 4.2. The main results are provided in Section 4.3. In Section 4.5, a finite blocklength implementation of the coding scheme is proposed using a deep learning approach. Finally, Section 4.6 provides concluding remarks.

### 4.2 Problem statement

Notation: For $a, b \in \mathbb{R}$, define $[a] \triangleq [1, \lceil a \rceil] \cap \mathbb{N}$, $[\![a, b]\!] \triangleq [\lfloor a \rfloor, \lceil b \rceil] \cap \mathbb{N}$. The components of a vector $X^n$ of length $n \in \mathbb{N}$ are denoted with subscripts, i.e., $X^n \triangleq (X_1, X_2, \dots, X_n)$. $\|.\|$ denotes the Euclidean norm and $\|.\|_1$ denotes the $L^1$ norm.

Consider a dealer and $J$ participants index in $\mathcal{J} \triangleq [J]$. We assume that the dealer can communicate with the participants over a Gaussian broadcast channel defined as

$$Y_j^n \triangleq X^n + N_j^n, \quad \forall j \in \mathcal{J}, \tag{4.1}$$

where $X^n$ is the signal transmitted by the dealer with the power constraint

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \leq P, \tag{4.2}$$

$Y_j^n$ is the channel observation at Participant $j \in \mathcal{J}$, and $N_j^n$ is a random vector of length $n$ with components independent and identically distributed according to a zero-mean Gaussian random variable with variance $\sigma_j^2$. The noise vectors $N_j^n$, $j \in \mathcal{J}$, are assumed independent.

For $t \in [J]$, the set of authorized sets of users is defined as

$$\mathbb{A}_t \triangleq \{\mathcal{A} \subseteq \mathcal{J} : |\mathcal{A}| \geq t\},$$

and, for $z \in [t-1]$, the set of unauthorized sets of users is defined as

$$\mathbb{U}_z \triangleq \{\mathcal{U} \subseteq \mathcal{J} : |\mathcal{U}| \leq z\}.$$

The parameters $t$ and $z$ are chosen by the system designer. In the following, for any $\mathcal{U} \in \mathbb{U}_z$ and $\mathcal{A} \in \mathbb{A}_t$, we use the notation $Y_{\mathcal{U}}^n \triangleq (Y_j^n)_{j\in\mathcal{U}}$, $N_{\mathcal{U}}^n \triangleq (N_j^n)_{j\in\mathcal{U}}$, $Y_{\mathcal{A}}^n \triangleq (Y_j^n)_{j\in\mathcal{A}}$, and $N_{\mathcal{A}}^n \triangleq (N_j^n)_{j\in\mathcal{A}}$ such that

$$Y_{\mathcal{A}}^n \triangleq \mathbf{1}_t X^n + N_{\mathcal{A}}^n, \quad \mathcal{A} \in \mathbb{A}_t, \tag{4.3}$$

$$Y_{\mathcal{U}}^n \triangleq \mathbf{1}_z X^n + N_{\mathcal{U}}^n, \quad \mathcal{U} \in \mathbb{U}_z, \tag{4.4}$$

where $\mathbf{1}_t$ and $\mathbf{1}_z$ are all-ones column vectors of size $t$ and $z$, respectively, and the covariance matrices of $N_{\mathcal{A}}$ and $N_{\mathcal{U}}$ are $\Sigma_{\mathcal{A}} \triangleq \text{diag}((\sigma_j^2)_{j\in\mathcal{A}})$ and $\Sigma_{\mathcal{U}} \triangleq \text{diag}((\sigma_j^2)_{j\in\mathcal{U}})$, respectively.

**Definition 4.1.** *A $(2^{nR_s}, t, z, n)$ secret sharing strategy consists of*

- *A secret $S$ uniformly distributed over $\mathcal{S} \triangleq [2^{nR_s}]$;*
- *An encoding function $f : \mathcal{S} \to \mathbb{R}^n$;*
- *A decoding function $g_{\mathcal{A}} : \mathbb{R}^{nt} \to \mathcal{S}$ for each qualified set $\mathcal{A} \in \mathbb{A}_t$;*

*and operates as follows:*

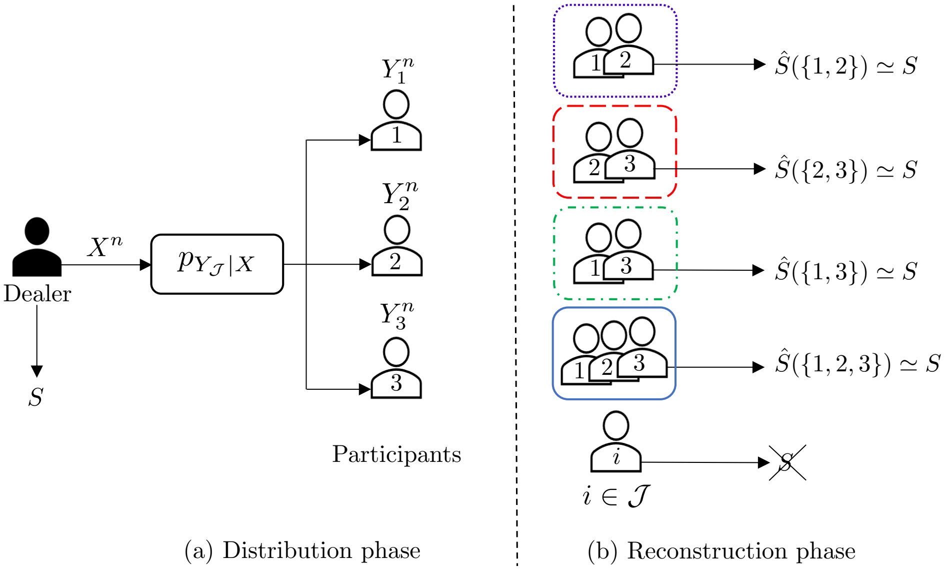
1. *The dealer encodes the secret $S \in \mathcal{S}$ as $X^n$;*
2. *The dealer sends $X^n$ over the channel and Participant $j \in \mathcal{J}$ observes $Y_j^n$;*
3. *Any subset of participants $\mathcal{A} \in \mathbb{A}_t$ can form an estimate of $S$ as $\hat{S}(\mathcal{A}) \triangleq g_{\mathcal{A}}(Y_{\mathcal{A}}^n)$.*

**Definition 4.2.** *A secret sharing rate $R_s$ is said to be $\epsilon$-reliable and $\delta$-secure if there exists a sequence of $(2^{nR_s}, t, z, n)$ secret sharing strategies such that*

$$\max_{\mathcal{A}\in\mathbb{A}_t} \mathbb{P}[\hat{S}(\mathcal{A}) \neq S] \leq \epsilon, \qquad (\textit{Reliability}) \tag{4.5}$$

$$\max_{\mathcal{U}\in\mathbb{U}_z} I(S;Y^n_{\mathcal{U}}) \leq \delta. \qquad (\textit{Security}) \tag{4.6}$$

$(4.5)$ *means that any subset of at least t participants is able to recover the secret, and* $(4.6)$ *means that any subset of at most $z$ participants cannot learn any information about the secret.*



(a) Distribution phase

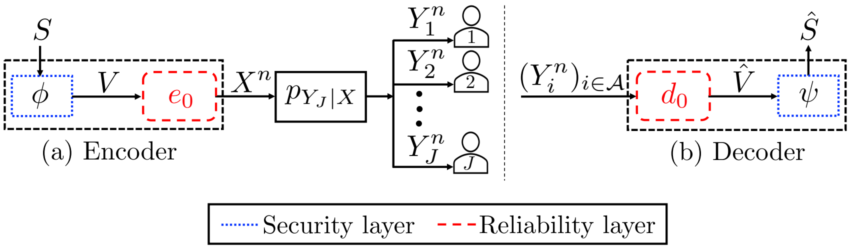
(b) Reconstruction phase

Figure 4.1: The access structure is defined by $\mathbb{A}_{t=2} \triangleq \{\{1,2\},\{2,3\},\{1,3\},\{1,2,3\}\}$ and $\mathbb{U}_{z=1} \triangleq \{\{1\},\{2\},\{3\}\}$, meaning that any two participants can reconstruct the secret, and any individual participant cannot learn any information about the secret.

The setting is illustrated in Figure 4.1 for the special case $\mathcal{J} = [3]$, $t = 2$, and $z = 1$.

## 4.3 Main results

We first introduce two-layer coding schemes in Definition 4.3.



(a) Encoder

(b) Decoder

Figure 4.2: Two-layer code design. The reliability layer is implemented using a channel code $(e_0, d_0)$ without any security constraints, and the security layer is implemented using the functions $\psi$ and $\phi$.

**Definition 4.3.** *A two-layer secret sharing coding scheme consists of*

- *A reliability layer defined by an encoder/decoder pair* $(\mathrm{e}_0, \mathrm{d}_0)$ *without any security constraints for the compound channel defined in* (4.3) *such that if* $X^n \triangleq \mathrm{e}_0(V)$ *is the channel input, where* $V$ *is uniformly distributed over* $\{0,1\}^r$ *and* $\|X^n\|^2 \leq nP$*, and* $Y^n_{\mathcal{A}}$*,* $\mathcal{A} \in \mathbb{A}_t$*, are the channel outputs, then* $\max_{\mathcal{A} \in \mathbb{A}_t} \mathbb{P}[d_0(Y^n_{\mathcal{A}}) \neq V] \leq \epsilon$*;*
- *A secrecy layer is defined by two functions* $\psi : \{0,1\}^r \to \{0,1\}^k$ *and* $\phi : \{0,1\}^k \to \{0,1\}^r$*, for some* $k \in \mathbb{N}$*;*

*and operates as follows:*

1. *Encoding: The dealer encodes a secret message* $S$ *that is uniformly distributed over* $\mathcal{S} \triangleq \{0,1\}^k$ *as* $\mathrm{e}_0(\phi(S))$*. Hence, the encoder* $\mathrm{e}$ *of a two-layer secret sharing coding scheme is*

$$e : \mathcal{S} \to \mathbb{R}^n$$
$$s \mapsto e_0\left(\phi(s)\right).$$

2. *Decoding: From the channel output observations* $Y^n_{\mathcal{A}}$*,* $\mathcal{A} \in \mathbb{A}_t$*, the participants in* $\mathcal{A}$ *estimate* $S$ *as* $\hat{S}(\mathcal{A}) \triangleq \psi(\mathrm{d}_0(Y^n_{\mathcal{A}}))$*. Hence, the decoder* $\mathrm{d}$ *of a two-layer secret sharing coding*

*scheme is*

$$
\begin{aligned}
d : \mathbb{R}^{nt} &\to \mathcal{S} \\
y^{nt} &\mapsto \psi(d_0(y^{nt})).
\end{aligned}
$$

The architecture of a two-layer coding scheme, as described in Definition 4.3, is depicted in Figure 4.2.

### 4.3.1 Design of a secret sharing coding scheme at short blocklength and performance evaluation

We design a two-layer coding scheme at finite blocklength, as defined in Definition 4.3. As detailed in Section 4.5.1, we design the reliability layer using an autoencoder $(e_0, d_0)$ and the secrecy layer via a two-universal hash function $\psi$.

In our simulations, detailed in Section 4.5.2, we consider $J = 200$ participants, $t = 100$, $z \in [10]$, and $\sigma_j^2 \triangleq 10^{-\text{SNR}/10}$, $j \in \mathcal{J}$, with a signal-to-noise ratio (SNR), $\text{SNR} = -16\text{dB}$.

Figure 4.3 shows the information leakage $\max\limits_{\mathcal{U} \in \mathbb{U}_z} I(S; Y_{\mathcal{U}}^n)$ with respect to the secret sharing rate $R_s = \frac{k}{n}$ when $z$ varies in $[10]$, $k = 1$, and $n \in \{5, 10, 15, 20\}$. Figure 4.3 confirms the intuition that the information leakage increases as the secret rate increases. Figure 4.4 shows the average probability of error $\max\limits_{\mathcal{A} \in \mathbb{A}_t} \mathbb{P}[\hat{S}(\mathcal{A}) \neq S]$ with respect to the secret sharing rate $R_s = \frac{k}{n}$ when $k = 1$ and $n \in \{5, 10, 15, 20\}$. We see the average probability of error decreases as the secret rate increases for fixed $k$ and SNR by our code construction.

## 4.4 Sufficient statistics

In this section, we reduce the Gaussian vector channel outputs to Gaussian scalar channel outputs using sufficient statistics [38].

Using Lemma 4.1 below, there is no loss of generality in considering the following Equations (4.7) and (4.8) as channel model instead of (4.3) and (4.4). Hence, in this section, we consider the

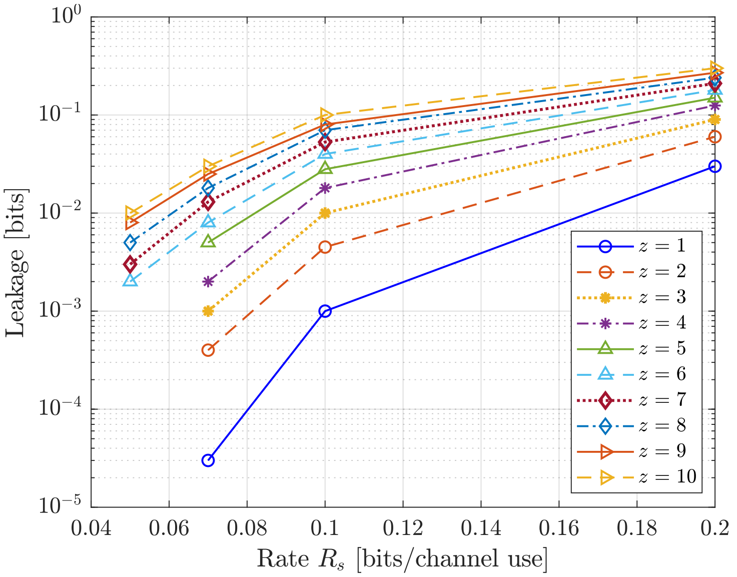


Figure 4.3: Information leakage $\max\limits_{\mathcal{U} \in \mathbb{U}_z} I(S; Y_{\mathcal{U}}^n)$ versus secret sharing rate $R_s = \frac{k}{n}$ for $z \in [10]$.

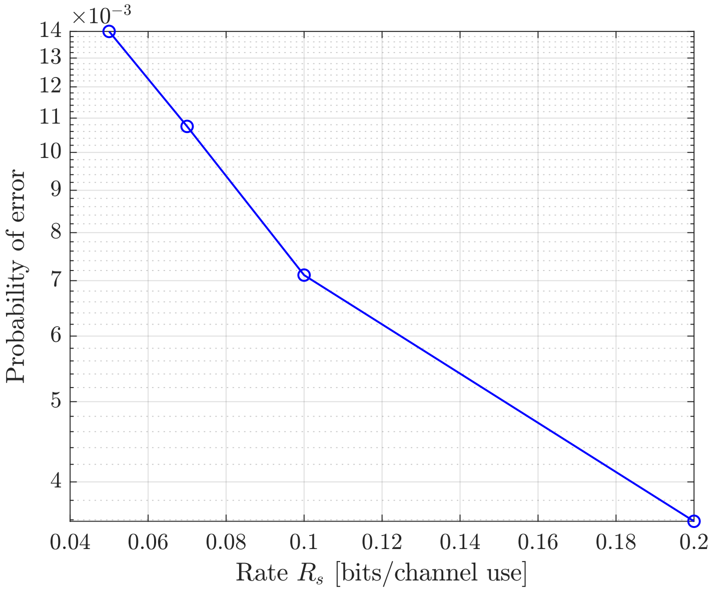


Figure 4.4: Probability of error $\max\limits_{\mathcal{A} \in \mathbb{A}_t} \mathbb{P}[\hat{S}(\mathcal{A}) \neq S]$ versus secret sharing rate $R_s = \frac{k}{n}$.

following channel model: For any $i \in [n]$,

$$\tilde{Y}_{\mathcal{A},i} \triangleq \sigma^2_{\tilde{Y}_{\mathcal{A}}} X_i + \tilde{N}_{\mathcal{A},i}, \quad \mathcal{A} \in \mathbb{A}_t, \tag{4.7}$$

$$\tilde{Y}_{\mathcal{U},i} \triangleq \sigma^2_{\tilde{Y}_{\mathcal{U}}} X_i + \tilde{N}_{\mathcal{U},i}, \quad \mathcal{U} \in \mathbb{U}_z, \tag{4.8}$$

where

$$\sigma^2_{\tilde{Y}_{\mathcal{A}}} \triangleq \mathbf{1}_t^T \Sigma_{\mathcal{A}}^{-1} \mathbf{1}_t, \tag{4.9}$$

$$\sigma^2_{\tilde{Y}_{\mathcal{U}}} \triangleq \mathbf{1}_z^T \Sigma_{\mathcal{U}}^{-1} \mathbf{1}_z, \tag{4.10}$$

$$\tilde{N}_{\mathcal{A},i} \triangleq \mathbf{1}_t^T \Sigma_{\mathcal{A}}^{-1} N_{\mathcal{A},i} \sim \mathcal{N}(0, \sigma^2_{\tilde{Y}_{\mathcal{A}}}), \tag{4.11}$$

$$\tilde{N}_{\mathcal{U},i} \triangleq \mathbf{1}_z^T \Sigma_{\mathcal{U}}^{-1} N_{\mathcal{U},i} \sim \mathcal{N}(0, \sigma^2_{\tilde{Y}_{\mathcal{U}}}). \tag{4.12}$$

**Lemma 4.1** ( [105, Lemma 3.1])**.** *Consider the channel model*

$$Y_{\mathcal{S}} = \mathbf{1}_{|\mathcal{S}|} X + N_{\mathcal{S}}, \quad \mathcal{S} \subset \mathcal{J},$$

*where $N_{\mathcal{S}}$ is a Gaussian vector of length $|\mathcal{S}|$ with zero mean and covariance matrix $\Sigma_{\mathcal{S}}$. A sufficient statistic to correctly determine $X$ from $Y_{\mathcal{S}}$ is the scalar*

$$\tilde{Y}_{\mathcal{S}} = \mathbf{1}_{|\mathcal{S}|}^T \Sigma_{\mathcal{S}}^{-1} Y_{\mathcal{S}}, \quad \mathcal{S} \subset \mathcal{J}.$$

## 4.5 Secret sharing scheme at finite blocklength

We design a secret sharing scheme at finite blocklength in Section 4.5.1 and evaluate its performance through simulations in Section 4.5.2.

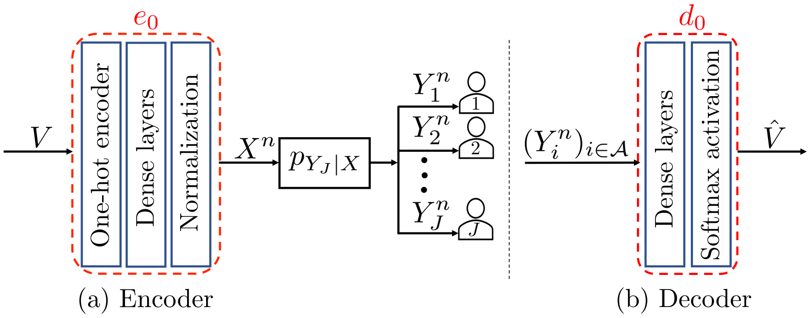


Figure 4.5: Architecture of the autoencoder $(e_0, d_0)$ via feed-forward neural networks.

### 4.5.1 Secret sharing scheme design

#### Reliability layer design

The design of the reliability layer consists in designing an encoder/decoder pair $(e_0, d_0)$ as described in Definition 4.3. Let $\mathcal{V} \triangleq [2^r]$ be the message set. $(e_0, d_0)$ is implemented with an autoencoder as in [85]. The goal of the autoencoder is here to learn a representation of the encoded message that is robust to the channel noise so that the authorized participants can reconstruct the message from their noisy channel observations with a small probability of error. As depicted in Figure 4.5, the encoder $e_0$ consists of three layers. An embedding layer, where the input $v \in \mathcal{V}$ is mapped to a one-hot vector $1_v \in \mathbb{R}^{2^r}$, which is a vector whose components are all zeros except the $v$-th component which is one. Dense hidden layers that take $v$ as input and return an $n$-dimensional vector. And, a normalization layer that ensures that the codeword $e_0(v)$, $v \in \mathcal{V}$, meets the average power constraint

$$\frac{1}{n}\|e_0(v)\|^2 \leq P.$$

As depicted in Figure 4.5, the decoder consists of dense hidden layers and a softmax layer. More specifically, let $\mu^{|\mathcal{V}|}$ be the output of the last dense layer in the decoder. The softmax layer takes $\mu^{|\mathcal{V}|}$ as input and returns a vector of probabilities $p^{|\mathcal{V}|} \in [0,1]^{|\mathcal{V}|}$, whose components $p_v$, $v \in \mathcal{V}$, are $p_v \triangleq \exp(\mu_v)\left(\sum_{i=1}^{|\mathcal{V}|}\exp(\mu_i)\right)^{-1}$. Finally, the decoded message $\hat{v}$ corresponds to the index of the component of $p^{|\mathcal{V}|}$ associated with the highest probability, i.e., $\hat{v} \in \arg\max_{v\in\mathcal{V}} p_v$.

The autoencoder is trained over all possible messages $v \in \mathcal{V}$ using a stochastic gradient descent (SGD) as in [95] and the categorical cross-entropy loss function.

### Secrecy layer design

We first review the definition of two-universal hash functions.

**Definition 4.4** ( [67], [106]). *A family $\mathcal{F}$ of two-universal hash functions $\mathcal{F} = \{f : \{0,1\}^n \to \{0,1\}^k\}$ is such that*

$$\forall x, x' \in \{0,1\}^n, x \neq x' \Rightarrow \mathbb{P}[F(x) = F(x')] \leq 2^{-k}, \tag{4.13}$$

*where $F$ is a function uniformly chosen in $\mathcal{F}$.*

Define $\mathcal{L} \triangleq \{0,1\}^r \backslash \{\mathbf{0}\}$, and consider the two-universal hash family of functions $\mathcal{P} \triangleq \{v \mapsto \psi(\lambda, v)\}_{\lambda \in \mathcal{L}}$, where $\psi$ is defined as

$$\begin{aligned} \psi : \mathcal{L} \times \{0,1\}^r &\to \{0,1\}^k \\ (\lambda, v) &\mapsto (\lambda \odot v)_k, \end{aligned} \tag{4.14}$$

where $\odot$ is the multiplication in GF($2^r$), and $(\cdot)_k$ selects the $k$ most significant bits. Define also the mapping $\phi$

$$\begin{aligned} \phi : \mathcal{L} \times \{0,1\}^k \times \{0,1\}^{r-k} &\to \{0,1\}^r \\ (\lambda, s, b) &\mapsto \lambda^{-1} \odot (s \| b), \end{aligned} \tag{4.15}$$

where $(\cdot\|\cdot)$ represents concatenation of two sequences of bits.

We implement the secrecy layer with the functions $\psi$ and $\phi$ defined in Equations (4.14) and (4.15), respectively.

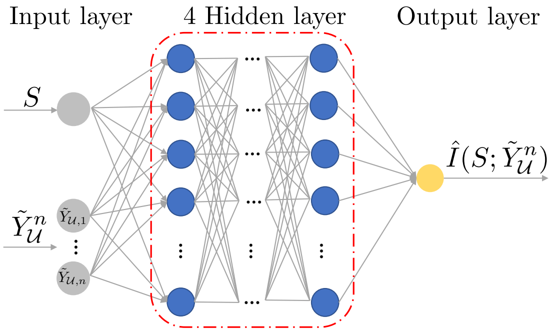

Input layer
4 Hidden layer
Output layer
$S$
$\tilde{Y}^n_{\mathcal{U}}$
$\tilde{Y}_{\mathcal{U},1}$
$\tilde{Y}_{\mathcal{U},n}$
$\hat{I}(S;\tilde{Y}^n_{\mathcal{U}})$

Figure 4.6: The security performance is evaluated in terms of the leakage $I(S;\tilde{Y}^n_{\mathcal{U}})$, $\mathcal{U} \in \mathbb{U}_z$, via the mutual information estimator where $\tilde{Y}^n_{\mathcal{U}} \triangleq (\tilde{Y}_{\mathcal{U},i})_{i\in[n]}$.

### Encoding and decoding for secret sharing scheme

The idea of the coding scheme below is to repeat $\gamma \in \mathbb{N}$ times the coding scheme described in Section 4.5.1. Hence, after the $\gamma$ repetitions, $\gamma$ secrets have been shared. With the objective to reduce the information leakage, the dealer and the participants extract another secret from these $\gamma$ secrets with a two-universal hash function. The price paid for this reduced leakage is a decrease of the secret sharing rate.

Fix a seed $\lambda \in \mathcal{L}$, a seed $\alpha \in \{0,1\}^{\gamma}\backslash\{\mathbf{0}\}$, and set the length of the secret to $k = 1$.

*Encoding*: For $i \in [\gamma]$, the dealer generates $r - k$ bits, denoted by $B_i$, uniformly at random from $\{0,1\}^{r-k}$, and a bit $M_i$ uniformly at random in $\{0,1\}$. The dealer sends the codeword $X^n_i \triangleq e_0(\phi(\lambda, M_i, B_i))$, $i \in [\gamma]$, such that the channel observations of the participants in $\mathcal{A} \in \mathbb{A}_t$ are $\tilde{Y}^n_{\mathcal{A},i} \triangleq \mathbf{1}^T_t \Sigma^{-1}_{\mathcal{A}} Y^n_{\mathcal{A},i}$. Then, the dealer forms the secret $S \triangleq \psi(\alpha, M_{1:\gamma})$, where $M_{1:\gamma} \triangleq (M_i)_{i\in[\gamma]}$.
*Decoding*: From $\tilde{Y}^n_{\mathcal{A},i}$, $i \in [\gamma]$, the participants in $\mathcal{A} \in \mathbb{A}_t$ estimate $V_i \triangleq \phi(\lambda, S, B_i)$ as $\hat{V}_i \triangleq d_0(\tilde{Y}^n_{\mathcal{A},i})$, $M_i$ as $\hat{M}_i \triangleq \psi(\lambda, \hat{V}_i)$, and the secret $S$ as $\hat{S}(\mathcal{A}) \triangleq \psi(\alpha, \hat{M}_{1:\gamma})$.

### 4.5.2 Performance evaluation

### Simulation parameters

In our simulations, we consider $J = 200$ participants, $t = 100$, $z \in [10]$, and $\sigma^2_j \triangleq 10^{-\text{SNR}/10}$, $j \in \mathcal{J}$, with SNR $= -16$dB. We also consider a secret of length $k = 1$ and a power $P = 1$. Note that since the SNR is the same for all the participants, we have for $\mathcal{A}^* \triangleq [t]$ and $\mathcal{U}^* \triangleq [z]$

$$\max_{\mathcal{A}\in\mathbb{A}_t} \mathbb{P}[\hat{S}(\mathcal{A}) \neq S] = \mathbb{P}[\hat{S}(\mathcal{A}^*) \neq S],$$

$$\max_{\mathcal{U}\in\mathbb{U}_z} I(S; Y^n_{\mathcal{U}}) = I(S; Y^n_{\mathcal{U}^*}).$$

For the autoencoder training and neural network implementation, we use Python 3.7 and Tensorflow 2.3.

### Performance evaluation of the reliability layer

For the parameters defined in Section 4.5.2, we train the autoencoder for $(n, r) = (5, 2)$ using SGD with the Adam optimizer [95] at a learning rate of 0.0001 over 100,000 random encoder input messages. To evaluate the performance of $(e_0, d_0)$, we first generate the input $V \in \{0, 1\}^r$. Then, $V$ is passed through the trained encoder $e_0$, which generates the codewords $X^n$ and the channel output $Y^n_{\mathcal{A}^*}$. By Lemma 4.1, without loss of generality, we consider $\tilde{Y}^n_{\mathcal{A}^*} \triangleq \mathbf{1}^T_t \Sigma^{-1}_{\mathcal{A}^*} Y^n_{\mathcal{A}^*}$, where $\Sigma_{\mathcal{A}^*} \triangleq \text{diag}((\sigma^2_j)_{j\in\mathcal{A}^*})$. Finally, the trained decoder $d_0$ forms an estimate of $V$, $\hat{V}(\mathcal{A}^*) \triangleq d_0(\tilde{Y}^n_{\mathcal{A}^*})$. Figure 4.7 shows the average probability of error $\mathbb{P}[\hat{V}(\mathcal{A}^*)\neq V]$.

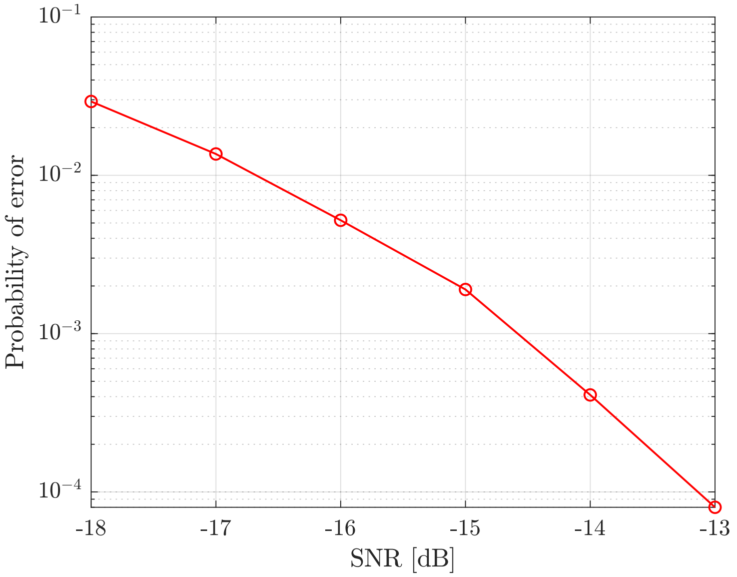


Figure 4.7: Probability of error $(e_0, d_0)$ at SNR $= -16$dB.

### Information leakage

Consider $\phi$ and $\psi$ with $n = 5$ and $r = 2$. Generate uniformly at random $M_{1:\gamma} \in \{0,1\}^{\gamma}$ and $B_{1:\gamma} \in \{0,1\}^{(r-k)\gamma}$. For $i \in [\gamma]$, generate

$$X_i^n \triangleq e_0(\phi(\lambda, M_i, B_i)), \tag{4.16}$$

such that the channel observations of the participants in $\mathcal{U}^*$ are $Y^n_{\mathcal{U}^*,i} \triangleq X_i^n + N^n_{\mathcal{U}^*,i}$. Using Lemma 4.1, there is no loss of generality in considering $\tilde{Y}^n_{\mathcal{U}^*,i} \triangleq \mathbf{1}_z^T \Sigma_{\mathcal{U}^*}^{-1} Y^n_{\mathcal{U}^*,i}$ instead of $Y^n_{\mathcal{U}^*,i}$. Finally, generate the secret as

$$S \triangleq \psi(\alpha, M_{1:\gamma}). \tag{4.17}$$

All possible combinations of $\lambda$ and $\alpha$ are tested to minimize the leakage. The optimal seeds found are $\lambda = 11$, $\alpha = 10$ when $\gamma = 2$, $\alpha = 110$ when $\gamma = 3$, and $\alpha = 1110$ when $\gamma = 4$. Next, we concatenate all the observations $\tilde{Y}^n_{\mathcal{U}^*,i}$, $i \in [\gamma]$, as $\tilde{Y}^n_{\mathcal{U}^*,1:\gamma}$ and to evaluate the leakage $I(S; \tilde{Y}^n_{\mathcal{U}^*,1:\gamma})$, we use the MINE from [71] based on neural networks, whose architecture is depicted in Figure 4.6. Specifically, we use a fully connected feed-forward neural network with $4$ hidden layers, each having $400$ neurons, and used rectified linear unit (ReLU) as an activation function. The input layer has $k + n$ neurons, and the Adam optimizer with a learning rate of $0.0001$ is used for the training. The samples of the joint distribution $p_{S\tilde{Y}^n_{\mathcal{U}^*,1:\gamma}}$ are produced as described above. The samples of the marginal distributions are generated by dropping either $s$ or $y^n_{\mathcal{U}^*,1:\gamma}$, from the joint samples $(s, y^n_{\mathcal{U}^*,1:\gamma})$. We trained the neural network over $40000$ epochs of $20,000$ messages with a batch size of $2500$. Figure 4.3 shows the information leakage $I(S; \tilde{Y}^n_{\mathcal{U}^*,1:\gamma})$ with respect to the secret sharing rate $R_s = \frac{k}{n}$ when $z$ varies in [10].

**Probability of error**

To evaluate the probability of error between $S$ and $\hat{S}(\mathcal{A}^*)$, generate uniformly at random $M_{1:\gamma} \in \{0,1\}^\gamma$ and $B_{1:\gamma} \in \{0,1\}^{(r-k)\gamma}$. Then, for $i \in [\gamma]$, generate the codeword $X_i^n$ as in (4.16) so that the channel outputs at the participants in $\mathcal{A}^*$ are $Y_{\mathcal{A}^*,i}^n \triangleq X_i^n + N_{\mathcal{A}^*,i}^n$. Using Lemma 4.1, there is no loss of generality in considering $\tilde{Y}_{\mathcal{A}^*,i}^n \triangleq \mathbf{1}_t^T \Sigma_{\mathcal{A}^*}^{-1} Y_{\mathcal{A}^*,i}^n$ instead of $Y_{\mathcal{A}^*,i}^n$. Finally, generate the secret $S$ as in (4.17).

At the participants in $\mathcal{A}^*$, for $i \in [\gamma]$, $M_i$ is estimated from $\tilde{Y}_{\mathcal{A}^*,i}^n$ as $\hat{M}_i \triangleq \psi(\lambda, d_0(\tilde{Y}_{\mathcal{A}^*,i}^n))$. Then, the secret is estimated as $\hat{S}(\mathcal{A}^*) \triangleq \psi(\alpha, \hat{M}_{1:\gamma})$. Figure 4.4 shows the average probability of error $\mathbb{P}[\hat{S}(\mathcal{A}^*) \neq S]$ with respect to the secret sharing rate $R_s = \frac{k}{n}$.

## 4.6 Concluding remarks

We considered a secret sharing model where a dealer can communicate with participants over a Gaussian broadcast channel. We proposed a coding approach that consists in separating the code design into a secrecy layer and a reliability layer. Our contribution was to design a two-layer coding scheme at finite blocklength, where we implemented the reliability layer with an autoencoder and the secrecy layer with two-universal hash functions. We empirically evaluated the probability of error and estimated the leakage for blocklength at most $20$ with a neural network-based mutual information estimator. Our simulation results demonstrated a precise control of the probability of error and leakage thanks to the two separate coding layers.

# CHAPTER 5

# CONCLUSIONS

Security is a major concern in wireless communication. The most common security issues in communication networks are (i) confidentiality, (ii) authenticity, (iii) integrity, and (iv) non-repudiation. With the increasing amount of information being transmitted over wireless networks, attacks on the communication network are very common. The information-theoretic approach can provide solutions to such problems. In information-theoretically secure models, the security derives from information theory and are secure even when the adversary has unbounded computational power.

We have discussed the problem of information-theoretic secret sharing with Gaussian sources, where we have considered rate-limited public communication to account for bandwidth constraints. Further, we have designed short blocklength codes for the Gaussian wiretap channel to account for practical applications that require low latency. Specifically, we have opted for deep learning, which provides us with a practical approach to design codes and better understand the finite blocklength regime as it is difficult to analytically characterize the optimal secrecy rates for the Gaussian wiretap channel in the finite blocklength regime. We proposed a framework that separates the code design into two layers: a reliability layer and a secrecy layer. We implemented the reliability layer with an autoencoder and the secrecy layer with universal hash functions. We have also discussed the problem of secret sharing over a broadcast channel, where we designed a coding scheme at a finite blocklength using an approach that separates the code design into a reliability layer and a secrecy layer. Our simulation results have demonstrated a precise control of the probability of error and leakage by decoupling two layers in our code design.

# CHAPTER 6

# FUTURE WORKS

There are several directions over which this work can be further studied. We list below some of them.

*Multiple users setting*: Developing finite-length codes for the multiple access wiretap channel (MAC-WT) under information-theoretic security guarantees will be an interesting problem. In other words, the setting where multiple users communicate with the legitimate receiver in the presence of an eavesdropper.

*Complexity based on users*: The complexity of the deep learning-based model increases as the number of users increases. A design based on a modular approach for complex models might be an effective way to tackle complexity-based issues.

*Complexity based on blocklength*: The complexity of the neural network-based autoencoder increases as blocklength $n$ increases. For blocklengths $n \leq 128$, we used a regular autoencoder in this thesis. For blocklengths $n \geq 128$, using a convolutional autoencoder will be a better choice. This implementation certainly reduces the complexity but at the expense of performance losses in terms of error probability for the message.

*Active eavesdropper*: In this thesis, we considered a passive eavesdropper in which the eavesdropper only listens to the transmission but does not try to modify the transmission. In a more practical scenario, an eavesdropper can modify or delete some bits from its observations. To take such an adversary into account in a finite-blocklength regime will be challenging and exciting.

# REFERENCES

# REFERENCES

[1] V. Rana, R. Chou, and H. Kwon, "Secret sharing from correlated Gaussian random variables and public communication," in *IEEE Inf. Theory Workshop*, (Riva del Garda, Italy), pp. 1–5, April 2021.

[2] V. Rana, R. A. Chou, and H. Kwon, "Information-theoretic secret sharing from correlated Gaussian random variables and public communication.," in *IEEE Trans. Inf. Theory*, vol. 68, pp. 549–559, 2022.

[3] V. Rana and R. Chou, "Design of short blocklength wiretap channel codes: Deep learning and cryptography working hand in hand," in *IEEE Inf. Theory Workshop*, pp. 1–6, 2021.

[4] V. Rana and R. A. Chou, "Short blocklength wiretap channel codes via deep learning: Design and performance evaluation," *IEEE Trans. Commun.*, vol. 71, no. 3, pp. 1462–1474, 2023.

[5] A. Shamir, "How to share a secret," *Commun. of the ACM*, vol. 22, no. 11, pp. 612–613, November 1979.

[6] G. Blakley, "Safeguarding cryptographic keys," in *Proc. of the National Computer Conference*, (New York, U.S.A.), pp. 313–317, June 1979.

[7] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. 22, no. 6, pp. 644–654, November 1976.

[8] S. Zou, Y. Liang, L. Lai, and S. Shamai, "An information theoretic approach to secret sharing," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3121–3136, June 2015.

[9] Y. Liang, G. Kramer, H. Poor, and S. Shamai, "Compound wiretap channels," *EURASIP J. Wireless Commun. and Networking*, no. 142374, October 2009.

[10] I. Csiszár and P. Narayan, "Capacity of a shared secret key," in *IEEE Int. Symp. Inf. Theory*, pp. 2593–2596, June 2010.

[11] R. Chou, "Secret sharing over a public channel from correlated random variables," in *IEEE Int. Symp. Inf. Theory*, (Colorado, U.S.A.), pp. 991–995, June 2018.

[12] R. A. Chou, "Distributed secret sharing over a public channel from correlated random variables," *arXiv preprint arXiv:2110.10307*, 2021.

[13] N. Tavangaran, H. Boche, and R. Schaefer, "Secret-key generation using compound sources and one-way public communication," *IEEE Trans. Inf. Forensics and Security*, vol. 12, no. 1, pp. 227–241, January 2017.

[14] M. Bloch, "Channel intrinsic randomness," in *IEEE Int. Symp. Inf. Theory*, (Texas, U.S.A.), pp. 2607–2611, June 2010.

[15] H. ZivariFard and R. A. Chou, "Secure data storage resilient against compromised users via an access structure," in *IEEE Inf. Theory Workshop*, pp. 464–469, 2022.

[16] R. Chou, “Biometric systems with multiuser access structures,” in *IEEE Int. Symp. Inf. Theory*, (Paris, France), pp. 807–811, July 2019.

[17] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. B. Mandayam, “Information-theoretically secret key generation for fading wireless channels,” *IEEE Trans. Inf. Forensics and Security*, vol. 5, no. 2, pp. 240–254, June 2010.

[18] A. Pierrot, R. Chou, and M. Bloch, “Experimental aspects of secret key generation in indoor wireless environments,” in *Signal Processing Advances in Wireless Commun.*, pp. 669–673, June 2013.

[19] A. Wyner, “The wire-tap channel,” *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, October 1975.

[20] S. Vadhan, “Extracting all the randomness from a weakly random source,” *Electronic Colloquium on Computational Complexity, Technical Report*, 1998.

[21] S. Watanabe and Y. Oohama, “Secret key agreement from vector Gaussian sources by rate limited public communication,” *IEEE Trans. Inf. Forensics and Security*, vol. 6, no. 3, pp. 541–550, September 2011.

[22] S. Watanabe and Y. Oohama, “Secret key agreement from correlated Gaussian sources by rate limited public communication,” *IEICE Transactions on Fundamentals of Electronics, Commun. and Computer Sciences*, vol. E93-A, no. 11, pp. 1–8, November 2010.

[23] A. Beimel, “Secret-sharing schemes: A survey,” in *Int. Conf. Coding and Cryptology*, (Qingdao, China), pp. 11–46, May-June 2011.

[24] J. Benaloh and J. Leichter, “Generalized secret sharing and monotone functions,” *Advances in Cryptology – CRYPTO*, vol. 403, no. 88, pp. 27–35, 1990.

[25] U. Maurer, “Secret key agreement by public discussion from common information,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[26] R. Ahlswede and I. Csiszar, “Common randomness in information theory and cryptography. I. secret sharing,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, July 1993.

[27] C. Pozrikidis, *An Introduction to Grids, Graphs, and Networks*. Oxford University Press, 2014.

[28] I. Csiszar and P. Narayan, “Common randomness and secret key generation with a helper,” *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 344–366, March 2000.

[29] R. Chou, M. Bloch, and E. Abbe, “Polar coding for secret-key generation,” *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6213–6237, November 2015.

[30] T. Ignatenko and F. M. J. Willems, *Biometric Security from an Information-Theoretical Perspective*. 2012.

[31] T. Ignatenko and F. Willems, “Privacy leakage in binary biometric systems: From Gaussian to binary data,” *Security and Privacy in Biometrics. Springer*, pp. 105–122, 2013.

[32] O. Günlü, “Multi-entity and multi-enrollment key agreement with correlated noise,” *IEEE Trans. Inf. Forensics and Security*, vol. 16, pp. 1190–1202, October 2021.

[33] O. Günlü, O. İşcan, V. Sidorenko, and G. Kramer, “Code constructions for physical unclonable functions and biometric secrecy systems,” *IEEE Trans. Inf. Forensics and Security*, vol. 14, no. 11, pp. 2848–2858, April 2019.

[34] A. Orlitsky and J. Roche, “Coding for computing,” *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, March 2001.

[35] G. Kramer, “Topics in multi-user information theory,” *Foundation Trends in Communication and Information Theory*, vol. 4, pp. 265–444, 2008.

[36] U. Maurer and S. Wolf, “Information-theoretic key agreement: From weak to strong secrecy for free,” in *Lecture Notes in Computer Science*, pp. 351–368, 2000.

[37] R. Chou and M. Bloch, “Separation of reliability and secrecy in rate-limited secret-key generation,” *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4941–4957, August 2014.

[38] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 2nd ed., 2006.

[39] M. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holden-Day, 1964.

[40] R. Fano, *Transmission of Information: A Statistical Theory of Communications*. MIT Press, 1961.

[41] R. Chou, “Unified framework for polynomial-time wiretap channel codes,” *arXiv preprint arXiv:2002.01924*, 2020.

[42] R. Sultana and R. Chou, “Low-complexity secret sharing schemes using correlated random variables and rate-limited public communication,” in *IEEE Int. Symp. Inf. Theory*, 2021.

[43] A. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, January, 1976.

[44] I. Csiszar and J. Korner, “Broadcast channels with confidential messages,” *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[45] S. Cheong and M. Hellman, “The Gaussian wire-tap channel,” *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, 1978.

[46] A. Thangaraj, S. Dihidar, A. Calderbank, S. McLaughlin, and J. Merolla, “Applications of LDPC codes to the wiretap channel,” *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2933–2945, 2007.

[47] A. Subramanian, A. Thangaraj, M. Bloch, and S. McLaughlin, “Strong secrecy on the binary erasure wiretap channel using large-girth LDPC codes,” *IEEE Trans. Inf. Forensics and Security*, vol. 6, no. 3, pp. 585–594, 2011.

[48] V. Rathi, R. Urbanke, M. Andersson, and M. Skoglund, “Rate-equivocation optimal spatially coupled LDPC codes for the BEC wiretap channel,” in *IEEE Int. Symp. Inf. Theory*, pp. 2393–2397, 2011.

[49] H. Mahdavifar and A. Vardy, “Achieving the secrecy capacity of wiretap channels using polar codes,” *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, 2011.

[50] E. Şaşoğlu and A. Vardy, “A new polar coding scheme for strong security on wiretap channels,” in *IEEE Int. Symp. Inf. Theory*, pp. 1117–1121, 2013.

[51] M. Andersson, R. Schaefer, T. Oechtering, and M. Skoglund, “Polar coding for bidirectional broadcast channels with common and confidential messages,” *IEEE J. Selected Areas Commun.*, vol. 31, no. 9, pp. 1901–1908, 2013.

[52] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, “Nested polar codes for wiretap and relay channels,” *IEEE Commun. Letters*, vol. 14, no. 8, pp. 752–754, 2010.

[53] M. Hayashi and R. Matsumoto, “Construction of wiretap codes from ordinary channel codes,” in *IEEE Int. Symp. Inf. Theory*, pp. 2538–2542, 2010.

[54] M. Bellare, S. Tessaro, and A. Vardy, “Semantic security for the wiretap channel,” in *Advances of Cryptology. Springer*, pp. 294–311, 2012.

[55] H. Tyagi and A. Vardy, “Explicit capacity-achieving coding scheme for the Gaussian wiretap channel,” in *IEEE Int. Symp. Inf. Theory*, pp. 956–960, 2014.

[56] C. Ling, L. Luzzi, J. Belfiore, and D. Stehlé, “Semantically secure lattice codes for the Gaussian wiretap channel,” *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, 2014.

[57] Y. Wei and S. Ulukus, “Polar coding for the general wiretap channel with extensions to multiuser scenarios,” *IEEE J. Selected Areas Commun.*, vol. 34, no. 2, pp. 278–291, 2016.

[58] R. Chou and M. Bloch, “Polar coding for the broadcast channel with confidential messages: A random binning analogy,” *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2410–2429, 2016.

[59] J. Renes, R. Renner, and D. Sutter, “Efficient one-way secret-key agreement and private channel coding via polarization,” in *Advances in Cryptology. Springer*, pp. 194–213, 2013.

[60] T. Gulcu and A. Barg, “Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component,” *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 1311–1324, 2017.

[61] R. A. Chou, “Explicit wiretap channel codes via source coding, universal hashing, and distribution approximation, when the channels’ statistics are uncertain,” *IEEE Trans. Inf. and Forensics Security*, 2022.

[62] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 124–130, 2018.

[63] W. Yang, R. Schaefer, and H. Poor, "Wiretap channels: Nonasymptotic fundamental limits," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4069–4093, 2019.

[64] M. Hayashi, "Tight exponential analysis of universally composable privacy amplification and its applications," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7728–7746, 2013.

[65] V. Tan, "Achievable second-order coding rates for the wiretap channel," in *IEEE Int. Conf. Commun. Systems*, pp. 65–69, 2012.

[66] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[67] M. Carter, *Foundations of Mathematical Economics*. MIT Press, 2001.

[68] G. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1315–1321, 1999.

[69] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, no. 6, pp. 066138(1–16), 2004.

[70] T. Suzuki, M. Sugiyama, J. Sese, and T. Kanamori, "Approximating mutual information by maximum likelihood density ratio estimation," in *Workshop on New Challenges for Feature Selection in Data Mining and Knowledge Discovery*, pp. 5–20, 2008.

[71] M. Belghazi, A. Baratin, S. Rajeswar, S. Ozair, Y. Bengio, A. Courville, and R. Hjelm, "MINE: Mutual information neural estimation," *arXiv preprint arXiv:1801.04062*, 2018.

[72] Y. Liang, G. Kramer, H. Poor, and S. Shamai, "Compound wiretap channels," *EURASIP J. Wirel. Commun. Netw.*, no. 142374, 2009.

[73] I. Bjelakovic, H. Boche, and J. Sommerfeld, "Secrecy results for compound wiretap channels," *Problems Inf. Transmission*, vol. 49, no. 1, pp. 73–98, 2013.

[74] I. Bjelakovic, H. Boche, and J. Sommerfeld, "Capacity results for arbitrarily varying wiretap channels," in *Inf. Theory, Combinatorics, and Search Theory*, pp. 123–144, 2013.

[75] E. MolavianJazi, M. Bloch, and J. Laneman, "Arbitrary jamming can preclude secure communication," in *Annual Allerton Conf. Commun., Control, and Computing*, pp. 1069–1075, 2009.

[76] A. Nooraiepour and T. Duman, "Randomized convolutional codes for the wiretap channel," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3442–3452, 2017.

[77] A. Nooraiepour and T. Duman, "Randomized turbo codes for the wiretap channel," in *IEEE Global Commun. Conf.*, pp. 1–6, 2017.

[78] D. Klinc, J. Ha, S. McLaughlin, J. Barros, and B. Kwak, “LDPC codes for the Gaussian wiretap channel,” *IEEE Trans. Inf. Forensics and Security*, vol. 6, no. 3, pp. 532–540, 2011.

[79] M. Baldi, M. Bianchi, and F. Chiaraluce, “Non-systematic codes for physical layer security,” in *IEEE Inf. Theory Workshop*, pp. 1–5, 2010.

[80] M. Baldi, F. Chiaraluce, N. Laurenti, S. Tomasin, and F. Renna, “Secrecy transmission on parallel channels: Theoretical limits and performance of practical codes,” *IEEE Trans. Inf. Forensics and Security*, vol. 9, no. 11, pp. 1765–1779, 2014.

[81] W. Harrison, E. Beard, S. Dye, E. Holmes, K. Nelson, M. Gomes, and J. Vilela, “Implications of coding layers on physical-layer security: A secrecy benefit approach,” *Entropy*, vol. 21, no. 8, p. 755, 2019.

[82] C. W. Wong, T. F. Wong, and J. M. Shea, “LDPC code design for the BPSK-constrained Gaussian wiretap channel,” in *IEEE Globecom Workshops*, pp. 898–902, 2011.

[83] M. Baldi, G. Ricciutelli, N. Maturo, and F. Chiaraluce, “Performance assessment and design of finite length LDPC codes for the Gaussian wiretap channel,” in *IEEE Int. Conf. Commun. Workshop*, pp. 435–440, 2015.

[84] A. Nooraiepour, S. Aghdam, and T. Duman, “On secure communications over Gaussian wiretap channels via finite-length codes,” *IEEE Commun. Letters*, vol. 24, no. 9, pp. 1904–1908, 2020.

[85] T. O’Shea and J. Hoydis, “An introduction to deep learning for the physical layer,” *IEEE Trans. Cognitive Commun. Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[86] S. Dörner, S. Cammerer, J. Hoydis, and S. Brink, “Deep learning based communication over the air,” *IEEE J. Selected Topics Signal Processing*, vol. 12, no. 1, pp. 132–143, 2018.

[87] F. Aoudia and J. Hoydis, “End-to-end learning of communications systems without a channel model,” in *Asilomar Conf. Signals, Systems, and Computers*, pp. 298–303, 2018.

[88] M. Goutay, F. Aoudia, and J. Hoydis, “Deep reinforcement learning autoencoder with noisy feedback,” pp. 1–6, 2019.

[89] R. Fritschek, R. Schaefer, and G. Wunder, “Deep learning for channel coding via neural mutual information estimation,” in *IEEE Int. Workshop Signal Processing Advances Wirel. Commun.*, pp. 1–5, 2019.

[90] H. Ye, G. Li, F. Juang, and K. Sivanesan, “Channel agnostic end-to-end learning based communication systems with conditional GAN,” in *IEEE Globecom Workshops*, pp. 1–5, 2018.

[91] R. Fritschek, R. Schaefer, and G. Wunder, “Deep learning for the Gaussian wiretap channel,” in *IEEE Int. Conf. Commun.*, pp. 1–6, 2019.

[92] R. Fritschek, R. Schaefer, and G. Wunder, “Deep learning based wiretap coding via mutual information estimation,” in *ACM Workshop on Wireless Security and Machine Learning*, pp. 74–79, 2020.

[93] X. Zhang and M. Vaezi, “Deep learning based precoding for the MIMO Gaussian wiretap channel,” in *IEEE Globecom Workshops*, pp. 1–6, 2019.

[94] K. Besser, P. Lin, C. Janda, and E. Jorswieck, “Wiretap code design by neural network autoencoders,” *IEEE Trans. Inf. Forensics and Security*, vol. 15, pp. 3374–3386, 2020.

[95] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[96] J. Carter and M. Wegman, “Universal classes of hash functions,” *J. Computer and System Sciences*, vol. 18, no. 2, pp. 143–154, 1979.

[97] Y. Polyanskiy, H. Poor, and S. Verdu, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[98] R. Schaefer, H. Boche, and H. Poor, “Secure communication under channel uncertainty and adversarial attacks,” *Proc. of the IEEE*, vol. 103, no. 10, pp. 1796–1813, 2015.

[99] H. Boche, R. Schaefer, and H. Poor, “On the continuity of the secrecy capacity of compound and arbitrarily varying wiretap channels,” *IEEE Trans. Inf. Forensics and Security*, vol. 10, no. 12, pp. 2531–2546, 2015.

[100] R. Chou and A. Yener, “The Gaussian multiple access wiretap channel when the eavesdropper can arbitrarily jam,” in *IEEE Int. Symp. Inf. Theory*, pp. 1958–1962, 2017.

[101] R. A. Chou and A. Yener, “Gaussian multiuser wiretap channels in the presence of a jammer-aided eavesdropper,” *Entropy*, vol. 24, no. 11, p. 1595, 2022.

[102] H. Yamamoto, “Secret sharing system using (k, L, n) threshold scheme,” *Electron. Commun. Japan*, vol. 69, no. 9, pp. 46–54, 1986.

[103] G. R. Blakley and C. Meadows, “Security of ramp schemes,” in *Workshop on the Theory and Application of Cryptographic Techniques*, pp. 242–268, 1985.

[104] R. Fritschek, R. F. Schaefer, and G. Wunder, “Deep learning based wiretap coding via mutual information estimation,” in *Proc. of ACM Workshop on Wireless Security and Machine Learning*, pp. 74–79, 2020.

[105] P. Parada and R. Blahut, “Secrecy capacity of SIMO and slow fading channels,” in *Proc. of IEEE Int. Symp. Inf. Theory*, pp. 2152–2155, 2005.

[106] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, “Generalized privacy amplification,” *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1915–1923, 1995.

[107] R. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013.

[108] M. Eaton, *Multivariate Statistics: A Vector Space Approach*. John Wiley and Sons, 1983.

# APPENDICES

# Appendix A

## Derivation of (2.4), (2.5)

Let $Z$ and $Z'$ be zero-mean jointly Gaussian and jointly non-singular random vectors with covariance matrices $\Sigma_Z$ and $\Sigma'_Z$, respectively. By [107, Theorem 3.5.2], we have

$$Z' = PZ + W, \tag{1}$$

where $P \triangleq \Sigma_{Z'Z}\Sigma_Z^{-1}$ and $W$ is independent of $Z$ with covariance $\Sigma_W \triangleq \Sigma_{Z'} - \Sigma_{Z'Z}\Sigma_Z^{-1}\Sigma_{Z'Z}^T$. Hence, by (1), we have

$$Y_{\mathcal{L}} = \Sigma_{Y_{\mathcal{L}}X}\sigma_X^{-2}X + W_{Y_{\mathcal{L}}}, \tag{2}$$

where $\Sigma_{W_{Y_{\mathcal{L}}}} \triangleq \Sigma_{Y_{\mathcal{L}}} - \Sigma_{Y_{\mathcal{L}}X}\sigma_X^{-2}\Sigma_{Y_{\mathcal{L}}X}^T$. Then, we normalize (2) as follows. By Cholesky decomposition, there exists an invertible matrix $B \in \mathbb{R}^{L\times L}$ such that $\Sigma_{W_{Y_{\mathcal{L}}}} = BB^T$. Hence, (2) can be rewritten as

$$Y'_{\mathcal{L}} \;=\; H_{\mathcal{L}}X + W'_{Y_{\mathcal{L}}},$$

where $Y'_{\mathcal{L}} \triangleq B^{-1}Y_{\mathcal{L}}$, $H_{\mathcal{L}} = B^{-1}\Sigma_{Y_{\mathcal{L}}X}\sigma_X^{-2}$, and $W'_{Y_{\mathcal{L}}} \sim \mathcal{N}(0, I_L)$.

# Appendix B

## Proof of Theorem 2.2

To prove Theorem 2.2, we proceed as follows. For a threshold access structure $\mathbb{A}_t$, we first prove that there exist sets of authorized and unauthorized participants $\mathcal{A}_t^\star \in \arg\min_{\mathcal{A}\in\mathbb{A}_t} H_\mathcal{A}^T H_\mathcal{A}$ and $\mathcal{U}_t^\star \in \arg\max_{\mathcal{U}\in\mathbb{U}_t} H_\mathcal{U}^T H_\mathcal{U}$, respectively, such that for any $t \in [\![1, L-1]\!]$, $\mathcal{A}_t^\star \subset \mathcal{A}_{t+1}^\star$, $\mathcal{U}_t^\star \subset \mathcal{U}_{t+1}^\star$. Then, by Theorem 1, we remark that $\mathcal{A}_t^\star$ and $\mathcal{U}_t^\star$ also correspond to the sets that appear in the expression of the secret capacity for the threshold access structure $\mathbb{A}_t$. Finally, using the monotonicity property (with respect to $t$) of the sets $(\mathcal{A}_t^\star)_{t\in[\![1,L]\!]}$ and $(\mathcal{U}_t^\star)_{t\in[\![1,L]\!]}$ and Theorem 1, we derive necessary and sufficient conditions to determine whether the secret capacity increases or decreases as the threshold $t$ increases.

We will need the following lemma.

**Lemma .1.** *Let $a, c \in \mathbb{R}_+$ and $R_p \in \mathbb{R}_+$. The function $f_{a,c,R_p}$ is non-increasing*

$$\begin{aligned} f_{a,c,R_p} &: \mathbb{R}_+ \to \mathbb{R} \\ & y \mapsto \frac{1}{2}\log\frac{cy2^{-2R_p} + ca(1-2^{-2R_p}) + 1}{cy+1}. \end{aligned}$$

*Proof.* The derivative of $f_{a,c,R_p}$ at $y \in \mathbb{R}_+$ is $f'_{a,c,R_p} = \frac{1}{2\ln 2}\frac{c(1+ca)(2^{-2R_p}-1)}{(cy+1)\left(cy2^{-2R_p}+ca(1-2^{-2R_p})+1\right)} \leq 0.$ □

Using Lemma .1, we obtain the following result:

**Lemma .2.** *One can find sets $(\mathcal{A}_t^\star)_{t\in[\![1,L]\!]}$ and $(\mathcal{U}_t^\star)_{t\in[\![1,L]\!]}$ such that for any $t \in [\![1, L-1]\!]$, we have $\mathcal{A}_t^\star \subset \mathcal{A}_{t+1}^\star$, $\mathcal{U}_t^\star \subset \mathcal{U}_{t+1}^\star$, and for any $t \in [\![1, L]\!]$,*

$$\{\mathcal{A}_t^\star, \mathcal{U}_t^\star\} \in \underset{\mathcal{A}\in\mathbb{A}_t, \mathcal{U}\in\mathbb{U}_t}{\arg\min} \left[f_{H_\mathcal{A}^T H_\mathcal{A}, \sigma_X^2, R_p}(H_\mathcal{U}^T H_\mathcal{U})\right]^+, \tag{3}$$

*where we have used the notation of Lemma .1.*

*Proof.* For $t \in [\![1, L]\!]$, remark that

$$\underset{\mathcal{A}\in\mathbb{A}_t,\mathcal{U}\in\mathbb{U}_t}{\arg\min} \left[ f_{H_{\mathcal{A}}^T H_{\mathcal{A}},\sigma_X^2,R_p}(H_{\mathcal{U}}^T H_{\mathcal{U}}) \right]^+ = \left\{ \underset{\mathcal{A}\in\mathbb{A}_t}{\arg\min}\, H_{\mathcal{A}}^T H_{\mathcal{A}}, \underset{\mathcal{U}\in\mathbb{U}_t}{\arg\max}\, H_{\mathcal{U}}^T H_{\mathcal{U}} \right\}, \tag{4}$$

because $f_{H_{\mathcal{A}}^T H_{\mathcal{A}},\sigma_X^2,R_p}(H_{\mathcal{U}}^T H_{\mathcal{U}})$ is an increasing function of $H_{\mathcal{A}}^T H_{\mathcal{A}}$ and is a decreasing function of $H_{\mathcal{U}}^T H_{\mathcal{U}}$ by Lemma .1. Next, write the vector $H_{\mathcal{L}}$ as $H_{\mathcal{L}} = [H_{\mathcal{L}}(1), H_{\mathcal{L}}(2), \ldots, H_{\mathcal{L}}(L)]^T$. By relabelling the participants, if necessary, assume that $|H_{\mathcal{L}}(1)| \leq |H_{\mathcal{L}}(2)| \leq \cdots \leq |H_{\mathcal{L}}(L)|$. For $t \in [\![1, L]\!]$, choose $\mathcal{A}_t^\star \triangleq [\![1, t]\!]$ and $\mathcal{U}_t^\star \triangleq [\![L-t+2, L]\!]$. Clearly, for any $t \in [\![1, L-1]\!]$, we have $\mathcal{A}_t^\star \subset \mathcal{A}_{t+1}^\star, \mathcal{U}_t^\star \subset \mathcal{U}_{t+1}^\star$, and by (4), we have that (3) holds for any $t \in [\![1, L]\!]$. □

By Theorem 2.1 and (3), we have

$$C_s(\mathbb{A}_1, R_p) = \left[ \frac{1}{2} \log \left( \sigma_X^2 H_{\mathcal{A}_1^\star}^T H_{\mathcal{A}_1^\star}(1 - 2^{-2R_p}) + 1 \right) \right]^+, \tag{5}$$

and for $t \in [\![2, L]\!]$, we have

$$C_s(\mathbb{A}_t, R_p) = \left[ \frac{1}{2} \log \frac{\sigma_X^2 H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} 2^{-2R_p} + \sigma_X^2 H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star}(1 - 2^{-2R_p}) + 1}{\sigma_X^2 H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} + 1} \right]^+. \tag{6}$$

Using (5) and (6), we easily obtain for any $t \in [\![1, L]\!]$

$$\begin{aligned} &C_s(\mathbb{A}_1, R_p) \geq C_s(\mathbb{A}_t, R_p) \\ \iff\ & \sigma_X^2 H_{\mathcal{A}_1^\star}^T H_{\mathcal{A}_1^\star} H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} + H_{\mathcal{A}_1^\star}^T H_{\mathcal{A}_1^\star} + H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} - H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star} \geq 0. \end{aligned}$$

From the proof of Lemma .2, there exists $O \geq 0$ such that $O \leq H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star}$ and $H_{\mathcal{A}_1^\star}^T H_{\mathcal{A}_1^\star} + O = H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star}$. Therefore, $H_{\mathcal{A}_1^\star}^T H_{\mathcal{A}_1^\star} + H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} \geq H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star}$, and $C_s(\mathbb{A}_1, R_p) \geq C_s(\mathbb{A}_t, R_p)$.

Next, for any $\mathcal{S} \subseteq \mathcal{L}$, for $s \in [\![1, |\mathcal{S}|]\!]$, let $H_{\mathcal{S}}(s)$ denote the $s$-th component of $H_{\mathcal{S}}$. We have for $i \in [\![1, L-t]\!]$,

$$\begin{aligned}
&C_s(\mathbb{A}_t, R_p) \geq C_s(\mathbb{A}_{t+i}, R_p)\\
&\iff \sigma_X^2 H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star} H_{\mathcal{U}_{t+i}^\star}^T H_{\mathcal{U}_{t+i}^\star} + H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star} + H_{\mathcal{U}_{t+i}^\star}^T H_{\mathcal{U}_{t+i}^\star}\\
&\qquad \geq \sigma_X^2 H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} H_{\mathcal{A}_{t+i}^\star}^T H_{\mathcal{A}_{t+i}^\star} + H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} + H_{\mathcal{A}_{t+i}^\star}^T H_{\mathcal{A}_{t+i}^\star}\\
&\iff \sigma_X^2 H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star} (H_{\mathcal{U}_{t+i}^\star}^T H_{\mathcal{U}_{t+i}^\star} - H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} + H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star}) + H_{\mathcal{U}_{t+i}^\star}^T H_{\mathcal{U}_{t+i}^\star} - H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star}\\
&\qquad \geq \sigma_X^2 H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} (H_{\mathcal{A}_{t+i}^\star}^T H_{\mathcal{A}_{t+i}^\star} - H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star} + H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star}) + H_{\mathcal{A}_{t+i}^\star}^T H_{\mathcal{A}_{t+i}^\star} - H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star}\\
&\iff (1 + \sigma_X^2 H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star})(H_{\mathcal{U}_{t+i}^\star}^T H_{\mathcal{U}_{t+i}^\star} - H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star})\\
&\qquad \geq (1 + \sigma_X^2 H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star})(H_{\mathcal{A}_{t+i}^\star}^T H_{\mathcal{A}_{t+i}^\star} - H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star}),
\end{aligned}$$

where the first equivalence is obtained using (6). Note that, by Lemma .2, one can choose $\mathcal{A}_t^\star \subset \mathcal{A}_{t+i}^\star$ and $\mathcal{U}_t^\star \subset \mathcal{U}_{t+i}^\star$, hence, $H_{\mathcal{A}_{t+i}^\star}^T H_{\mathcal{A}_{t+i}^\star} - H_{\mathcal{A}_t^\star}^T H_{\mathcal{A}_t^\star} \geq 0$ and $H_{\mathcal{U}_{t+i}^\star}^T H_{\mathcal{U}_{t+i}^\star} - H_{\mathcal{U}_t^\star}^T H_{\mathcal{U}_t^\star} \geq 0$.

# Appendix C

## Proof of (2.11)

The probability of error averaged over $C_n$, i.e., $\mathbb{E}_{C_n}\left[\mathbb{P}[V^n \neq \widehat{V}^n_{\mathcal{A}}]\right]$ for any $\mathcal{A} \in \mathbb{A}$ can be upper bounded via the union bound by the four following terms:

1. The probability that $(x^n, y^n_{\mathcal{A}}) \notin \mathcal{T}^n_{\epsilon_1}(XY_{\mathcal{A}})$, which is upper bounded by $2|\mathcal{X}||\mathcal{Y}_{\mathcal{A}}|\exp(-n\epsilon_1^2\mu_{XY_{\mathcal{A}}})$ [35, Page 272 Equation (1.12)].

2. The probability that the encoder cannot find $(\omega,\nu)$ such that $(x^n, v^n(\omega,\nu)) \in \mathcal{T}^n_{\epsilon}(XV)$, given that $(x^n, y^n_{\mathcal{A}}) \in \mathcal{T}^n_{\epsilon_1}(XY_{\mathcal{A}})$, which is upper bounded by

$$\begin{aligned}
&\mathbb{E}_{C_n}\left[\sum_{x^n,y^n_{\mathcal{A}}} p_{X^nY^n_{\mathcal{A}}}(x^n,y^n_{\mathcal{A}})\mathbb{1}\{\forall(\omega,\nu),(v^n(\omega,\nu),x^n)\notin\mathcal{T}^n_{\epsilon}(VX)\text{ and }(x^n,y^n_{\mathcal{A}})\in\mathcal{T}^n_{\epsilon_1}(XY_{\mathcal{A}})\}\right]\\
&=\sum_{(x^n,y^n_{\mathcal{A}})\in T^n_{\epsilon_1}(XY_{\mathcal{A}})} p_{X^nY^n_{\mathcal{A}}}(x^n,y^n_{\mathcal{A}})\mathbb{P}[\forall(\omega,\nu),(V^n(\omega,\nu),x^n)\notin\mathcal{T}^n_{\epsilon}(VX)]\\
&=\sum_{(x^n,y^n_{\mathcal{A}})\in T^n_{\epsilon_1}(XY_{\mathcal{A}})} p_{X^nY^n_{\mathcal{A}}}(x^n,y^n_{\mathcal{A}})(1-\mathbb{P}[(V^n(\omega,\nu),x^n)\in\mathcal{T}^n_{\epsilon}(VX)])^{2^{n(R_v+R'_v)}}\\
&\overset{(a)}{\leq}\sum_{(x^n,y^n_{\mathcal{A}})\in T^n_{\epsilon_1}(XY_{\mathcal{A}})} p_{X^nY^n_{\mathcal{A}}}(x^n,y^n_{\mathcal{A}})\exp(-2^{n(R_v+R'_v)}\mathbb{P}[(V^n(\omega,\nu),x^n)\in\mathcal{T}^n_{\epsilon}(VX)])\\
&\overset{(b)}{\leq}\sum_{(x^n,y^n_{\mathcal{A}})\in T^n_{\epsilon_1}(XY_{\mathcal{A}})} p_{X^nY^n_{\mathcal{A}}}(x^n,y^n_{\mathcal{A}})\exp\Big(-2^{n(R_v+R'_v)}\left(1-\delta^{(2)}_{\epsilon_1,\epsilon}(n)\right)2^{-n(I(V;X)+2\epsilon H(V))}\Big)\\
&\leq\exp\Big(-\left(1-\delta^{(2)}_{\epsilon_1,\epsilon}(n)\right)2^{\epsilon nH(V)}\Big),
\end{aligned}$$

where $(a)$ holds because for any $x \geq 0$ and any $p \in [0,1]$, $(1-p)^x \leq e^{-px}$, and in $(b)$ we have defined $\delta^{(2)}_{\epsilon_1,\epsilon}(n) \triangleq 2|\mathcal{V}||\mathcal{X}|\exp\left(-n\frac{(\epsilon-\epsilon_1)^2}{1+\epsilon_1}\mu_{VX}\right)$.

3. The probability that the decoder finds $\tilde{\nu}_{\mathcal{A}} \neq \nu$ such that $(y^n_{\mathcal{A}}, v^n(\omega,\tilde{\nu}_{\mathcal{A}})) \in \mathcal{T}^n_{\epsilon}(Y_{\mathcal{A}}V)$, given that $(x^n, y^n_{\mathcal{A}}) \in \mathcal{T}^n_{\epsilon_1}(XY_{\mathcal{A}})$ and the encoder found $(\omega,\nu)$ such that $(x^n, v^n(\omega,\nu)) \in \mathcal{T}^n_{\epsilon}(XV)$,

which is upper bounded by

$$
\begin{aligned}
&\sum_{\omega,\nu} p(\omega,\nu) \sum_{\nu'_{\mathcal{A}}\neq\nu} \mathbb{E}_{C_n} \sum_{(x^n,y^n_{\mathcal{A}})\in T^n_{\epsilon_1}(XY_{\mathcal{A}})} p_{X^nY^n_{\mathcal{A}}}(x^n,y^n_{\mathcal{A}})\mathbb{1}\{y^n_{\mathcal{A}},(v^n(\omega,\nu'_{\mathcal{A}}))\in\mathcal{T}^n_{\epsilon}(Y_{\mathcal{A}}V)\}\\
&=\sum_{\omega,\nu} p(\omega,\nu) \sum_{\nu'_{\mathcal{A}}\neq\nu} \sum_{(x^n,y^n_{\mathcal{A}})\in T^n_{\epsilon_1}(XY_{\mathcal{A}})} p_{X^nY^n_{\mathcal{A}}}(x^n,y^n_{\mathcal{A}})\mathbb{P}[(y^n_{\mathcal{A}},(V^n(\omega,\nu'_{\mathcal{A}}))\in\mathcal{T}^n_{\epsilon}(Y_{\mathcal{A}}V)]\\
&\leq\sum_{\omega,\nu} p(\omega,\nu) \sum_{\nu'_{\mathcal{A}}\neq\nu} \sum_{(x^n,y^n_{\mathcal{A}})\in T^n_{\epsilon_1}(XY_{\mathcal{A}})} p_{X^nY^n_{\mathcal{A}}}(x^n,y^n_{\mathcal{A}})2^{-n(I(V;Y_{\mathcal{A}})-2\epsilon H(V))}\\
&\leq 2^{n(R'_v-I(V;Y_{\mathcal{A}})+2\epsilon H(V))}\\
&\leq 2^{-n\epsilon H(V)}.
\end{aligned}
$$

4. The probability that the decoder cannot find $\tilde{\nu}_{\mathcal{A}}$ such that $(y^n_{\mathcal{A}}, v^n(\omega,\tilde{\nu}_{\mathcal{A}}))\in\mathcal{T}^n_{\epsilon}(Y_{\mathcal{A}}V)$, given that $(x^n,y^n_{\mathcal{A}})\in\mathcal{T}^n_{\epsilon_1}(XY_{\mathcal{A}})$ and the encoder found $(\omega,\nu)$ such that $(x^n,v^n(\omega,\nu))\in\mathcal{T}^n_{\epsilon}(XV)$, which is upper bounded with Markov lemma [35, Page 319 Equation (5.1)] by $2|\mathcal{V}||\mathcal{X}||\mathcal{Y}_{\mathcal{A}}|\exp\left(-n\frac{(\epsilon-\epsilon_1)^2}{1+\epsilon_1}\mu_{VXY_{\mathcal{A}}}\right)$.

Hence, for any $\mathcal{A}\in\mathbb{A}$, we have $\mathbb{E}_{C_n}[\mathbb{P}[V^n\neq\widehat{V}^n_{\mathcal{A}}]]\leq\delta(n,\epsilon,\mathcal{A})$. Next, we have

$$
\begin{aligned}
\mathbb{E}_{C_n}\left[\max_{\mathcal{A}\in\mathbb{A}}\mathbb{P}[\widehat{V}^n_{\mathcal{A}}\neq V^n]\right] &\leq \mathbb{E}_{C_n}\left[\sum_{\mathcal{A}\in\mathbb{A}}\mathbb{P}[\widehat{V}^n_{\mathcal{A}}\neq V^n]\right]\\
&= \sum_{\mathcal{A}\in\mathbb{A}}\mathbb{E}_{C_n}\left[\mathbb{P}[\widehat{V}^n_{\mathcal{A}}\neq V^n]\right]\\
&\leq \sum_{\mathcal{A}\in\mathbb{A}}\delta(n,\epsilon,\mathcal{A})\\
&\leq |\mathbb{A}|\max_{\mathcal{A}\in\mathbb{A}}\delta(n,\epsilon,\mathcal{A}).
\end{aligned}
$$

By Markov's inequality, we conclude that there exists a codebook such that $\max_{\mathcal{A}\in\mathbb{A}}\mathbb{P}[\widehat{V}^n_{\mathcal{A}}\neq V^n]\leq|\mathbb{A}|\max_{\mathcal{A}\in\mathbb{A}}\delta(n,\epsilon,\mathcal{A})$.

# Appendix D

## Proof of (2.17)

For any $\mathcal{U} \in \mathbb{U}$, we have

$$\begin{aligned} H(V^n|Y^n_{\mathcal{U}}) &\overset{(a)}{\geq} I(X^n; V^n|Y^n_{\mathcal{U}}) \\ &= H(X^n|Y^n_{\mathcal{U}}) - H(X^n|V^nY^n_{\mathcal{U}}) \\ &\overset{(b)}{=} nH(X|Y_{\mathcal{U}}) - H(X^n|V^nY^n_{\mathcal{U}}), \end{aligned} \tag{7}$$

where $(a)$ holds by definition of mutual information, and $(b)$ holds because the $X_i$'s and $(Y_{\mathcal{U}})_i$'s are independently and identically distributed. We now lower bound the term $-H(X^n|V^nY^n_{\mathcal{U}})$. Define for any $\mathcal{U} \in \mathbb{U}$,

$$\begin{aligned} \Gamma_{\mathcal{U}} &\triangleq \mathbb{1}\{(X^n, V^n, Y^n_{\mathcal{U}}) \in \mathcal{T}^n_{2\epsilon}(XVY_{\mathcal{U}})\}, \\ \Delta_{\mathcal{U}} &\triangleq \mathbb{1}\{(X^n, V^n) \in \mathcal{T}^n_{\epsilon}(XV)\}, \end{aligned}$$

so that,

$$\begin{aligned} &H(X^n|V^nY^n_{\mathcal{U}}) \\ &\leq H(X^n\Gamma_{\mathcal{U}}\Delta_{\mathcal{U}}|V^nY^n_{\mathcal{U}}) \\ &= H(\Gamma_{\mathcal{U}}\Delta_{\mathcal{U}}|V^nY^n_{\mathcal{U}}) + H(X^n|V^nY^n_{\mathcal{U}}\Gamma_{\mathcal{U}}\Delta_{\mathcal{U}}) \\ &\overset{(a)}{\leq} 2 + \sum_{\delta_{\mathcal{U}},\gamma_{\mathcal{U}}\in\{0,1\}} \mathbb{P}(\Gamma_{\mathcal{U}} = \gamma_{\mathcal{U}}|\Delta_{\mathcal{U}} = \delta_{\mathcal{U}})\mathbb{P}(\Delta_{\mathcal{U}} = \delta_{\mathcal{U}})H(X^n|V^nY^n_{\mathcal{U}}, \Gamma_{\mathcal{U}} = \gamma_{\mathcal{U}}, \Delta_{\mathcal{U}} = \delta_{\mathcal{U}}) \\ &\overset{(b)}{\leq} 2 + H(X^n|V^nY^n_{\mathcal{U}}, \Gamma_{\mathcal{U}} = 1, \Delta_{\mathcal{U}} = 1) + n(2\delta_{\epsilon}(n) + \delta^2_{\epsilon}(n, \mathcal{U})) \log|\mathcal{X}| \\ &= \sum_{y^n_{\mathcal{U}}, v^n} p(y^n_{\mathcal{U}}, v^n|1, 1)H(X^n|Y^n_{\mathcal{U}} = y^n_{\mathcal{U}}, V^n = v^n, \Gamma_{\mathcal{U}} = 1, \Delta_{\mathcal{U}} = 1) \end{aligned}$$

$$
\begin{aligned}
&+ 2 + (2\delta_\epsilon(n) + \delta_\epsilon^2(n,\mathcal{U})) \log |\mathcal{X}|^n \\
&\overset{(c)}{\le} \sum_{y_{\mathcal{U}}^n, v^n} p(y_{\mathcal{U}}^n, v^n|1,1) \log |T_{2\epsilon}^n(X|y_{\mathcal{U}}^n, v^n)| + 2 + (2\delta_\epsilon(n) + \delta_\epsilon^2(n,\mathcal{U})) \log |\mathcal{X}|^n \\
&\le \sum_{y_{\mathcal{U}}^n, v^n} p(y_{\mathcal{U}}^n, v^n|1,1) n H(X|Y_{\mathcal{U}}V)(1+2\epsilon) + 2 + (2\delta_\epsilon(n) + \delta_\epsilon^2(n,\mathcal{U})) \log |\mathcal{X}|^n \\
&\le nH(X|Y_{\mathcal{U}}V)(1+2\epsilon) + 2 + (2\delta_\epsilon(n) + \delta_\epsilon^2(n,\mathcal{U})) \log |\mathcal{X}|^n. \qquad (8)
\end{aligned}
$$

where $(a)$ holds because $(\Gamma_{\mathcal{U}}, \Delta_{\mathcal{U}})$ is defined over an alphabet of cardinality equal to four so that $H(\Gamma_{\mathcal{U}}\Delta_{\mathcal{U}}|V^n Y_{\mathcal{U}}^n) \le \log 4 = 2$, $(b)$ holds because $\mathbb{P}[\Delta_{\mathcal{U}} = 0] \le \delta_\epsilon(n) \triangleq 2|\mathcal{X}||\mathcal{V}|e^{-n\epsilon^2\mu_{XV}}$ and $\mathbb{P}[\Gamma_{\mathcal{U}} = 0|\Delta_{\mathcal{U}} = 1] \le \delta_\epsilon^2(n,\mathcal{U}) \triangleq 2|\mathcal{V}||\mathcal{X}||\mathcal{Y}_{\mathcal{U}}|e^{-\epsilon^2 n\mu_{VXY_{\mathcal{U}}}/2}$ by Markov Lemma [35, Page 319 Equation (5.1)], and $(c)$ holds because $H(X) \le \log |\mathcal{X}|$ for any discrete random variable $X$ defined over $|\mathcal{X}|$. Combining (7) and (8), we obtain (2.17).

# Appendix E

## Proof of (2.22) and (2.23)

We rewrite (2.26) and (2.27) as

$$\begin{aligned} R_p &= \max_{\mathcal{A}\in\mathbb{A}} \left(h(X) - h(X|V) - h(Y_{\mathcal{A}}) + h(Y_{\mathcal{A}}|V)\right), & (9)\\ R_s &= \min_{\mathcal{A}\in\mathbb{A}} \min_{\mathcal{U}\in\mathbb{U}} \left(h(Y_{\mathcal{A}}) - h(Y_{\mathcal{A}}|V) - h(Y_{\mathcal{U}}) + h(Y_{\mathcal{U}}|V)\right). & (10)\end{aligned}$$

Let $K_{XV} \triangleq \begin{bmatrix} \sigma_X^2 & \sigma_{XV} \\ \sigma_{VX} & \sigma_V^2 \end{bmatrix}$ be the covariance matrix of $(X, V)$. We have

$$\begin{aligned} h(X|V) &= h(X,V) - h(V) \\ &= \frac{1}{2}\log(2\pi e)^2 \det(K_{XV}) - \frac{1}{2}\log 2\pi e \sigma_V^2 \\ &= \frac{1}{2}\log 2\pi e(\sigma_X^2 - \sigma_{XV}\sigma_V^{-2}\sigma_{XV}) \\ &= \frac{1}{2}\log 2\pi e \sigma_{X|V}^2, & (11)\end{aligned}$$

where the last equality holds by [108, Proposition 3.13]. Next, for any $\mathcal{A} \in \mathbb{A}$, let $K_{Y_{\mathcal{A}}V} \triangleq \begin{bmatrix} \Sigma_{Y_{\mathcal{A}}} & \Sigma_{Y_{\mathcal{A}}V} \\ \Sigma_{Y_{\mathcal{A}}V}^T & \sigma_V^2 \end{bmatrix}$ be the covariance matrix of $(Y_{\mathcal{A}}, V)$. We have

$$\begin{aligned} h(Y_{\mathcal{A}}|V) &= \frac{1}{2}\log(2\pi e)^{|\mathcal{A}|}\frac{\det(K_{Y_{\mathcal{A}}V})}{\sigma_V^2} \\ &\overset{(a)}{=} \frac{1}{2}\log(2\pi e)^{|\mathcal{A}|}\frac{\sigma_V^2 \det(\Sigma_{Y_{\mathcal{A}}} - \Sigma_{Y_{\mathcal{A}}V}\sigma_V^{-2}\Sigma_{Y_{\mathcal{A}}V}^T)}{\sigma_V^2} \\ &\overset{(b)}{=} \frac{1}{2}\log(2\pi e)^{|\mathcal{A}|}\det(\Sigma_{Y_{\mathcal{A}}|V}), \\ &\overset{(c)}{=} \frac{1}{2}\log(2\pi e)^{|\mathcal{A}|}\det(H_{\mathcal{A}}\sigma_{X|V}^2 H_{\mathcal{A}}^T + I), & (12)\end{aligned}$$

where $(a)$ holds by the formula for the determinant of a block matrix, $(b)$ holds by [108, Proposition 3.13], $(c)$ holds by (2.4) and the definition of the conditional variance $\Sigma_{Y_{\mathcal{A}}|V} \triangleq \mathbb{E}\left[\left(Y_{\mathcal{A}} - \mathbb{E}[Y_{\mathcal{A}}|V]\right)\left(Y_{\mathcal{A}} - \mathbb{E}[Y_{\mathcal{A}}|V]\right)^T|V\right] = H_{\mathcal{A}}\mathbb{E}\left[\left(X - \mathbb{E}[X|V]\right)\left(X - \mathbb{E}[X|V]\right)^T|V\right]H_{\mathcal{A}}^T + \mathbb{E}\left[W_{Y_{\mathcal{A}}}W_{Y_{\mathcal{A}}}^T\right]$ and $W_{Y_{\mathcal{A}}}$ is a Gaussian noise vector with with identity covariance matrix. Similarly, for any $\mathcal{U} \in \mathbb{U}$, we have

$$h(Y_{\mathcal{U}}|V) \;=\; \frac{1}{2}\log(2\pi e)^{|\mathcal{U}|}\det(H_{\mathcal{U}}\sigma^2_{X|V}H_{\mathcal{U}}^T + I). \tag{13}$$

Thus, from (9), (10), (11), (12), and (13), we have

$$R_p \;=\; \max_{\mathcal{A}\in\mathbb{A}}\left(\frac{1}{2}\log\frac{\sigma_X^2}{\sigma^2_{X|V}} - \frac{1}{2}\log\frac{\det(H_{\mathcal{A}}\sigma_X^2H_{\mathcal{A}}^T + I)}{\det(H_{\mathcal{A}}\sigma^2_{X|V}H_{\mathcal{A}}^T + I)}\right), \tag{14}$$

$$R_s \;=\; \min_{\mathcal{A}\in\mathbb{A}}\min_{\mathcal{U}\in\mathbb{U}}\left(\frac{1}{2}\log\frac{\det(H_{\mathcal{A}}\sigma_X^2H_{\mathcal{A}}^T + I)}{\det(H_{\mathcal{A}}\sigma^2_{X|V}H_{\mathcal{A}}^T + I)} - \frac{1}{2}\log\frac{\det(H_{\mathcal{U}}\sigma_X^2H_{\mathcal{U}}^T + I)}{\det(H_{\mathcal{U}}\sigma^2_{X|V}H_{\mathcal{U}}^T + I)}\right). \tag{15}$$

Then, by Lemma 2.1 and the definition of $O_{\mathcal{A}}, \mathcal{A} \in \mathbb{A}$ and $O_{\mathcal{U}}, \mathcal{U} \in \mathbb{U}$, we can rewrite (14) and (15) as

$$R_p \;=\; \max_{\mathcal{A}\in\mathbb{A}}\left(\frac{1}{2}\log\frac{\sigma_X^2}{\sigma^2_{X|V}} - \frac{1}{2}\log\frac{\sigma_X^2O_{\mathcal{A}} + 1}{\sigma^2_{X|V}O_{\mathcal{A}} + 1}\right), \tag{16}$$

$$R_s \;=\; \min_{\mathcal{A}\in\mathbb{A}}\min_{\mathcal{U}\in\mathbb{U}}\left(\frac{1}{2}\log\frac{\sigma_X^2O_{\mathcal{A}} + 1}{\sigma^2_{X|V}O_{\mathcal{A}} + 1} - \frac{1}{2}\log\frac{\sigma_X^2O_{\mathcal{U}} + 1}{\sigma^2_{X|V}O_{\mathcal{U}} + 1}\right). \tag{17}$$

Finally, by Lemma 2.2, (16) and (17) become (2.22) and (2.23).

# Appendix F

## Achievability and converse bounds for the Gaussian wiretap channel

### F.1 Achievability bound for the Gaussian wiretap channel

The maximal secrecy rate $R(n, \epsilon, \delta)$ achievable by an $\epsilon$-reliable and $\delta$-secure $(n, k, P)$ code is lower bounded as [63, Theorem 7 and Section IV.C-1]

$$R(n, \epsilon, \delta) \geq \frac{1}{n} \log_2 \frac{M(\epsilon, n)}{L(n, \delta)},$$

with $M(\epsilon, n)$ the number of codewords for a probability of error $\epsilon$ and blocklength $n$ inferred by Shannon's channel coding achievability bound [97, Section III.J-4], and $L(n, \delta)$ such that

$$\sqrt{L(n, \delta)} \triangleq \min_{\gamma} \frac{\sqrt{\gamma \mathbb{E}[\exp(-|B_n - \log \gamma|)]}}{2(\delta + \mathbb{E}[\exp(-|B_n - \log \gamma|^+)] - 1)}, \tag{18}$$

where the minimization is over all $\gamma > 0$ such that the denominator is positive, and

$$B_n \triangleq \frac{n}{2} \log_2 \left(1 + \frac{P}{\sigma_Z^2}\right) + \frac{\log_2 e}{2} \sum_{t=1}^{n} \left(1 - \frac{(\sqrt{P} Z_t - \sqrt{\sigma_Z^2})^2}{P + \sigma_Z^2}\right),$$

where $Z_t$, $t \in \{1, \dots, n\}$, are i.i.d. according to the standard normal distribution.

### F.2 Converse bound for the Gaussian wiretap channel

An $\epsilon$-reliable and $\delta$-secure $(n, k, P)$ code for the wiretap channel $(\mathcal{X}, P_{YZ|X}, \mathcal{Y} \times \mathcal{Z})$ satisfies [63, Theorem 12 and Section IV.C-3]

$$2^k \leq \inf_{\tau \in (0, 1-\epsilon-\delta)} \frac{\tau + \delta}{\tau \beta_{1-\epsilon-\delta-\tau}(P_{X^n Y^n Z^n}, P_{X^n Z^n} Q_{Y^n|Z^n})}, \tag{19}$$

where $P_{X^nY^nZ^n}$ denotes the joint probability distribution induced by the code and for $Q_{Y^n|Z^n}$ as in [63, Eq. (129)]

$$\beta_{1-\epsilon-\delta-\tau}(P_{X^nY^nZ^n}, P_{X^nZ^n}Q_{Y^n|Z^n}) \geq \mathbb{P}[\bar{D}_{n+1} \geq \bar{\gamma}],$$

where

$$\begin{aligned}\bar{D}_{n+1} &\triangleq (n+1)C_s \\ &\quad + \frac{\log_2 e}{2}\sum_{t=1}^{n+1}\left(\frac{N_{Z_t}^2}{\sigma_Z^2} - \frac{(\bar{N}_{Z_t} - c_0(N_{Z_t} + \sqrt{P}))^2}{P+\sigma_Z^2}\right. \\ &\quad \left. + \frac{\bar{N}_{Z_t}^2}{P+\sigma_Y^2} - \frac{(c_1 N_{Z_t} + c_0\bar{N}_{Z_t} - c_0^2\sqrt{P})^2}{\sigma_Y^2}\right),\end{aligned}$$

with $N_{Z_t} \sim \mathcal{N}(0,\sigma_Z^2)$, $\bar{N}_{Z_t} \sim \mathcal{N}(0, P+\sigma_Y^2)$, $C_s \triangleq \frac{1}{2}\log_2\frac{1+P/\sigma_Y^2}{1+P/\sigma_Z^2}$, and $c_0 \triangleq \sqrt{\frac{\sigma_Z^2-\sigma_Y^2}{P+\sigma_Z^2}}$, $c_1 \triangleq \frac{P+\sigma_Y^2}{P+\sigma_Z^2}$, and the threshold $\bar{\gamma}$ satisfies $\mathbb{P}[\bar{B}_{n+1} \geq \bar{\gamma}] = 1-\epsilon-\delta-\tau$ with

$$\begin{aligned}\bar{B}_{n+1} \triangleq (n+1)C_s + \frac{\log_2 e}{2}\sum_{t=1}^{n+1}&\left(\frac{(N_{Y_t}+\bar{N}_{Y_t})^2}{\sigma_Z^2} - \frac{N_{Y_t}^2}{\sigma_Y^2}\right. \\ &\left. + \frac{(\sqrt{P}+N_{Y_t})^2}{P+\sigma_Y^2} - \frac{(\sqrt{P}+N_{Y_t}+\bar{N}_{Y_t})^2}{P+\sigma_Z^2}\right),\end{aligned}$$

where $N_{Y_t} \sim \mathcal{N}(0,\sigma_Y^2)$ and $\bar{N}_{Y_t} \sim \mathcal{N}(0,\sigma_Z^2-\sigma_Y^2)$, $t \in \{1,\ldots,n+1\}$, are independent and identically distributed.