PHYSICAL-LAYER SECURITY: WIDE-BAND COMMUNICATIONS & ROLE OF KNOWN INTERFERENCE

A Dissertation

by

MUSTAPHA EL-HALABI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Costas N. Georghiades
Co-Chair of Committee,	Tie Liu
Committee Members,	Krishna Narayanan
	Jim Ji
	Marcello Aguiar
	Joseph Jean Boutros
Head of Department,	Chanan Singh

December 2013

Major Subject: Electrical Engineering

ABSTRACT

Data security is of such paramount importance that security measures have been implemented across all layers of a communication network. One layer at which security has not been fully developed and studied is the physical layer, the lowest layer of the protocol stack. Towards establishing fundamental limits of secure communications at the physical layer, we address in this dissertation two main problems. First, we study secure communication in the wide-band regime, and second we study the role of known interference in secure communication.

The concept of channel capacity per unit cost was introduced by Verdú in 1990 to study the limits of cost-efficient wide-band communication. It was shown that orthogonal signaling can achieve the channel capacity per unit cost of memoryless stationary channels with a zero-cost input letter. The first part of this dissertation introduces the concept of secrecy capacity per unit cost to study cost-efficient wide-band secrecy communication. For degraded memoryless stationary wiretap channels, it is shown that an orthogonal coding scheme with randomized pulse position and constant pulse shape achieves the secrecy capacity per unit cost with a zero-cost input letter. For general memoryless stationary wiretap channels, the performance of orthogonal codes is studied, and the benefit of further randomizing the pulse shape is demonstrated via a simple example. Furthermore, the problem of secure communication in a MIMO setting is considered, and a single-letter expression for the secrecy capacity per unit cost is obtained for the MIMO wiretap channel.

Recently there has been a lot of success in using the deterministic approach to provide approximate characterization of Gaussian network capacity. The second part of this dissertation takes a deterministic view and revisits the problem of wiretap channel with side information. A precise characterization of the secrecy capacity is obtained for a linear deterministic model, which naturally suggests a coding scheme which we show to achieve the secrecy capacity of the degraded Gaussian model (dubbed as "secret writing on dirty paper") to within half a bit. The success of this approach allowed its application to the problem of "secret key agreement via dirty paper coding", where also a suggested coding scheme achieves the secret-key capacity to within half a bit.

DEDICATION

To the souls of my grandparents.

ACKNOWLEDGEMENTS

The completion of this dissertation was made possible due to the support of my advisor, Dr. Costas Georghiades, Professor and Associate Dean for Research, for his valuable guidance, scholarly inputs and his consistent encouragement throughout my research work. From our first meeting, he always triggered me to follow my intuition in research and gave me the freedom to pursue my interests. The many conferences I was able to attend around the world, contributed a lot to my professional and personal maturity, and that was not possible without his generosity and consistent support. My efforts towards this degree would similarly have been vane, without the contribution and the constant help of my co-advisor Dr. Tie Liu. I want to thank him for sharing his incredible mathematical skill and for the countless hours we have spent together trying to crack down problems. His vision and expertise shaped up the research included in this dissertation. I was lucky enough to have the opportunity to collaborate with a great mind like Shlomo Shamai (Shitz), a Shannon awardee with an encyclopedic knowledge of information theory. Our joint work will be a milestone in my future career.

I was very much privileged at Texas A&M to learn through some inspiring courses with great mentors. Thanks for Dr. Krishna Narayanan for Information Theory, Dr. P.R. Kumar for Engineering Optimization and Dr. Deepa Kundur for Cyber Security for the Smart Grid.

For the past five years, my life was made enjoyable due to the presence of valuable colleagues and friends, who supported me with their care and love. I feel very lucky for the unconditional friendship of my Siamese twin Salim El-Rouayheb, who opened

some closed doors for me and shared lot of good and bad moments. My Lebanese friends in Texas, Ali, Mary, Nataly, Sharbel, Rayya were my comfort bubble, and alleviated my homesickness with weekly dinners and outings. Also, my friends Hani and Milad always reached out to contact me, listen to my nagging and uplift my spirit.

I would also like to thank my collaborators on projects that did not appear in this dissertation: Deepa Kundur, Salim El-Rouayheb, Mohammad Ali Maddah-Ali, Abdallah Farraj, and Hung D. Ly, for valuable hours of discussion and work on problems related to the smart grid.

I am indebted to my office-mate Armin Banaei, with whom I had many fruitful discussions and who proved to be a true friend by reaching out to help me during tough times.

I would also like to thank the members of my dissertation committee, Dr. Jim Ji and Dr. Marcello Aguiar for their insights and assistance. A special thanks for Dr. Joseph Jean Boutros for making my defense at TAMU Qatar possible and for opening closed doors for me.

I owe a lot to my parents, who encouraged and helped me at every stage of my personal and academic life, and longed to see this dissertation come true. I want to thank them for their infinite support over an unbounded time interval and hope that I will be able to provide equal love and support for them.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Today, wireless networks continue to play an increasingly prominent role. The ubiquitous and pervasive use of internet services, along with the inherent openness of wireless communication channels, present inexhaustible reasons that call for serious interventions on the security level. The interception of data and the malicious use of wiretapped information induce harmful effects that lead to tremendous societal costs, and hence the continuous quest for effective solutions.

The importance of data security has called for implementing security measures at all layers of the standard network stack. However, the lowest layer, the *physical-layer*, where the information bits get modulated into signals, remained neglected when it comes to securing it. Several results from information theory suggest that the imperfections of communications channels (noise, fading, interference) can be harnessed with appropriate code designs to not only offer error correction for legitimate parties, but also provide provable confidentiality against passive eavesdroppers without the need for shared secret keys. With the advent of wireless networks and the growing concern for data confidentiality, the idea of exploiting the imperfections of the physical-layer as a first layer of defense has attracted much interest and is now colloquially known as physical-layer security.

Throughout this dissertation, I will be concerned with studying security technologies that are embedded at the *physical-layer* of the protocol architecture, a part of the communication system where little security exists today.

There exist fundamental differences between cryptography, based on computational complexities at the higher layers of the protocol stack, and physical-layer security which is based on information-theoretic principles. Hence, it is necessary to

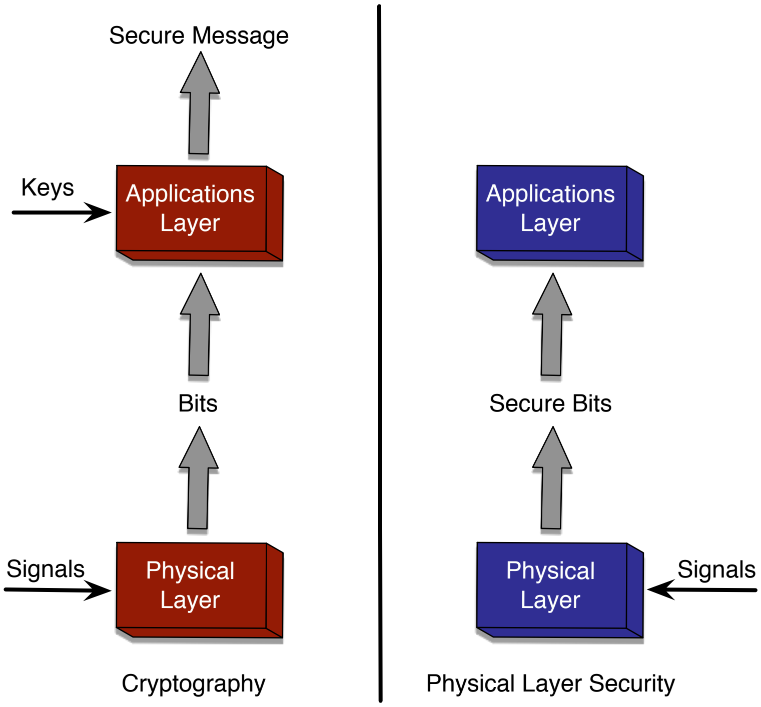

Figure 1.1: Cryptography versus physical layer security

understand the differences between both technologies, in order to assess which one to recur to, given a certain practical scenario. See Fig. 1.1.

A secure communication scenario basically involves three parties: a transmitter, a receiver and an eavesdropper; traditionally referred to as Alice, Bob and Eve, respectively. Alice is interested in confidentially sending a message to Bob with high efficiency and reliability through a channel, while keeping the message ideally unknown to Eve. The utmost goal of secrecy is to design an algorithm or a coding strategy that enable the optimal tradeoff between these objectives.

In 1948, Shannon [1] set the foundation of modern cryptography by introducing the idea of *entropy* as a measure of the amount of information associated with a message. By providing precise connections between provable security and the size of the key, plaintext, and ciphertext spaces, the first proofs of security using probability theory were established. It was shown [1] that the number of different keys must be at least as large as the number of messages to achieve *perfect secrecy.* A classical way to accomplish this objective is through the use of the well-known one-time-pad algorithm, which requires that the length of the key must be at least as large as the length of the message. Despite the impracticality of this requirement, it would still serve as an upper bound on the secrecy level, quantified by the *equivocation*, a measure of the eavesdropper's uncertainty about the message and the key.

Knowing that the exchange of the secret key has to be done publicly and sometimes through an insecure channel, one question that comes into play is how can this secret key be established. The answer to this question is provided by numerical techniques pertaining to cryptography, mainly through the use of public-key encryption algorithms

For a large number of applications, using famous algorithms such as (RSA and AES) has proved to be efficient, as no attacks succeeded in the past, and the algorithms which were compromised were consistently replaced by new ones. The technology is readily available and inexpensive. However, the computational model still suffers from a number of drawbacks. As the computational power keeps on increasing, some brute-force attacks, which were once considered as infeasible in the past, can be made possible in the near future. Also, the perfect security of public-key cryptographic techniques is still unproven from a mathematical point of view. Moreover, the strength of different ciphers can not be rigorously compared due to the lack of precise metrics.

As shown in the work of Shannon [1] and Wyner [2], the used cryptographic techniques fail to provide information-theoretic security, as the channel between the eavesdropper and the users is mostly noiseless and the secrecy capacity is thus null. In addition, most key-distribution schemes rely on the existence of a trusted third party and key management creates complexities due to the system architecture.

The main advantages of information-theoretic security over computational security come from the fact that no computational constraints are assumed in advance, and the fact that the information leaked to the adversaries can be quantified as a function of the channel quality. On the other hand, some disadvantages come into play, as information-theoretic security is based on average information measures, and confidentiality cannot be guaranteed with probability one. Also, some assumptions are made concerning the communication channels, which may not be accurate in practice. Physical-layer security has been deployed for some optical communication applications; however, the technology is still an early bloomer and relatively expensive. Also, it has been used in practice through quantum key distribution and, in theory, suitably long codes can come exponentially close to perfect secrecy. The system architecture for security is basically the same as the one for communication. Instead of distributing keys, it is possible to generate on-the-fly as many secret keys as desired.

Based on the comparison above, it is more likely that the incorporation of physical-layer security in classical communication systems will be part of a layered security solution rather than as a standalone solution. As a consequence, authentication and confidentiality will be provided at different layers, each with a targeted goal. Hence, physical-layer security provides an additional layer of security that does not yet exist in communication networks.

1.1 The Wiretap Channel

Let us consider an early example for the three-party network shown in Fig. 1.2 as a means of introducing some key concepts. Suppose Alice wants to communicate a secret message M to Bob, with the knowledge that another network participant Eve may be able to overhear the transmission. The channel that links the two

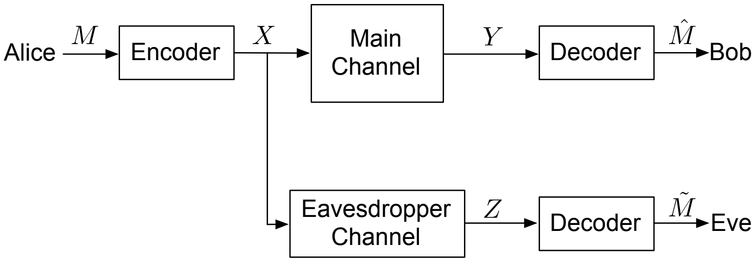

Figure 1.2: The wiretap channel model has three players. Alice is attempting to transmit a message to Bob over the main channel, while Eve is trying to eavesdrop on the transmitted data over a separate channel known as the eavesdropper's channel.

friendly parties is called the main channel, and Eve listens to the transmission via a parallel channel called the eavesdropper's channel. Alice transmits X, which she obtains by passing M through a prescribed encoder. Bob receives Y through the main channel, while Eve receives Z through the eavesdropper's channel. Bob and Eve then form their respective estimates of the message using a specified decoder or attack algorithm. This model for communication in an unreliable network is called the wiretap channel model [2].

1.2 Security Metrics

In this section, we review some of the standard metrics used to measure the secrecy guaranteed by error-control codes. Historically, the first metric was introduced by Claude Shannon together with the notion of perfect secrecy [1]. A system is said to operate with perfect secrecy if the message M and its corresponding encoder output X^n are statistically independent, so that revealing X^n does not provide any extra information to an attacker about M. In information-theoretic terms, this can be expressed as $I(M; X^n) = 0$, i.e., the mutual information between M and X^n is exactly zero. Unfortunately, Shannon also showed that perfect secrecy requires the use of a secret key at least as long as the message M, which limits the application of this result. Instead of requiring exact statistical independence, Aaron Wyner suggested the use of a weaker requirement for secrecy, as well as a new idea in channel modeling that allowed physical-layer security to take shape [2]. Specifically, the message M is encoded into a codeword X^n, of which the eavesdropper observes a possibly noisy version denoted by Z^n. A system is said to operate with weak secrecy if the rate of information leaked about the message M to the eavesdropper observing Z^n is asymptotically zero in the codeword length n, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} I(M;Z^n) = 0$$

Weak secrecy, however, does not prevent a few bits of M to be leaked through Z^n; to combat this shortcoming of weak secrecy, it has been advocated to use a stronger secrecy metric. A system is said to operate with strong secrecy if the total amount of information leaked about M through observing Z^n goes to zero as n goes to infinity, that is,

$$\lim_{n\to\infty} I(M;Z^n) = 0$$

By definition, note that: perfect secrecy$\Longrightarrow$strong secrecy$\Longrightarrow$weak secrecy. Secrecy code designs that seek to obtain strong secrecy are different in nature from those that are designed for weak secrecy only. Although in theory, we should be able to attain the same coding rates under both secrecy constraints, current designs tend to obtain strong secrecy at a cost of coding rate, and therefore, at a cost of information throughput in a communication system. Also, notice that these security measures require our code's blocklength n to go to infinity, and thus, implementation favors large blocklength in terms of secrecy. Although one should certainly strive to design schemes that provide strong or weak secrecy, the finite blocklength secrecy performance is often particularly hard to analyze. This issue can be circumvented by adopting other metrics that relate to the quality of the received messages and are easier to study. For instance, one could measure the eavesdropper's error rate with a prescribed decoder. Although there are possible pitfalls when only considering error rate as a measure of secrecy, as high error rate by itself does not guarantee secrecy, the error rate is a pointer to the performance of a particular code that may indicate good secrecy results, even in the information-theoretic sense.

Building on the notion of perfect secrecy, the information-theoretic foundations for a physical-layer approach to security were first laid by Wyner [2] and later by Csiszár and Körner [3], who proved in seminal papers that there exist channel codes guaranteeing robustness to transmission errors with a prescribed degree of data confidentiality. A single letter characterization of the *secrecy capacity* was attained through the use of two main techniques; *random binning* followed by *prefix coding.*

In this work, we first explore secrecy in the wideband regime, where to the secrecy constraint we add limitations on the power availability, but not on the bandwidth.The communication limits in wideband were investigated by Verdú [4], and the special case of a zero-cost input letter at the transmitter's end was considered, to give the name *capacity per unit cost* for the optimal limit of communication under such restrictions. Applying Verdú's formulation to the wiretap problem, a complete characterization for the maximum achievable secrecy rate per unit cost, mainly *the secrecy capacity per unit cost*, was presented in [5] for the degraded wiretap channel. Afterward, and due to the recent success of using the deterministic approach to provide approximate characterization of Gaussian network capacity, we take a deterministic view and revisit the problem of wiretap channel with side information. A precise characterization of the secrecy capacity is obtained for a linear deterministic model, which naturally suggests a coding scheme which we show to achieve the secrecy capacity of the Gaussian model (dubbed as "secret writing on dirty paper") to within 1/2 a bits [6]. However, our method is somewhat different from most of the practices along this line of research. In literature, a common practice has been to first gain "insight" from the capacity-achieving scheme for the linear deterministic model and then translate the success to the Gaussian model at the *scheme* level. To the best of our understanding, this translation is more art than science. For the problems that we considered in this paper, the translation of success from the linear deterministic model to the Gaussian model was done at the level of *single-letter descriptions* of network capacity and hence was much more systematic.

2. SECURE COMMUNICATION IN THE WIDE-BAND REGIME

2.1 Introduction

In classical Shannon theory [1], communication over a noisy medium is modeled as a communication channel with discrete-time input and output. An important objective is to understand the *channel capacity*, which is defined as the maximum number of bits that can be reliably transmitted per channel use for a given constraint on the average transmission cost per input symbol. This formulation is suited for studying *band-limited* communication, where spectrum is the most valuable resource.

A different scenario emerges in the context of deep-space communication, where there are virtually no limitations on the available bandwidth. Instead, energy becomes the most valuable resource considering the prohibitively high cost for replacing satellite batteries. This communication scenario was abstracted by Verdú [4] using the concept of *channel capacity per unit cost*, which is defined as the maximum number of bits that can be reliably transmitted per transmission cost. Since there are no limitations on the number of channel uses, this formulation is tailored for *wide-band* communication.

Verdú's formulation of channel capacity per unit cost [4] can be viewed as a *relaxed* setting of the classical Shannon formulation of channel capacity, in the sense that there are no limitations imposed on the number of channel uses for communication. Therefore, it is not surprising that the channel capacity per unit cost of a memoryless stationary channel can be derived from its capacity-cost function. For the special case where there is a *zero-cost* letter in the input alphabet, however, Verdú provided an alternative characterization of the channel capacity per unit cost which does *not* depend on the notion of channel capacity. There are two main advantages for this new characterization:

1) Compared with the classical single-letter characterization of channel capacity, Verdú's characterization of channel capacity per unit cost is much *easier* to compute as it only involves an optimization over the input letters as opposed to the input distributions for channel capacity [1].

2) Even though *structured* codes that can achieve the channel capacity are difficult to construct, Verdú's characterization is strongly tied to the fact that with a zero-cost input letter, the channel capacity per unit cost can be achieved by highly structured *orthogonal* codes.

The above results were later extended by Liu and Viswanath [7] to memoryless stationary state-dependent channels, where the channel states are non-causally known at the transmitter as side information.

The main aim of this section is to extend Verdú's formulation of channel capacity per unit cost from regular communication (without any secrecy constraints) to *secrecy* communication, and to understand to what extent Verdú's results [4] can be extended to memoryless stationary *wiretap* channels. We note here that in the spirit of the classical Shannon formulation [1], the *secrecy capacity* of memoryless stationary wiretap channels was characterized by Wyner [2] for the degraded case and by Csiszár and Körner [3] for the general case.

2.2 Model, Notations and Definitions

As illustrated in Fig 2.1, a memoryless stationary wiretap channel consists of an input alphabet $\mathcal{X}$, two output alphabets $\mathcal{Y}$ and $\mathcal{Z}$ at the legitimate receiver and the eavesdropper respectively, and a conditional probability distribution $P_{Y,Z|X}$. An $(n, w_0, \nu, \epsilon_d, \epsilon_e)$ secrecy code consists of:

- a message W uniformly drawn from $\{1, \ldots, w_0\}$;

- a *stochastic* encoder which maps the message W to a length-n codeword $X^n = (X_1, \ldots, X_n) \in \mathcal{X}^n$ such that

$$\sum_{i=1}^{n} b(X_i) \leq \nu \tag{2.1}$$

  where $b : \mathcal{X} \to \mathbb{R}^+ = [0, +\infty)$ is a function that assigns a nonnegative cost to each letter in the input alphabet $\mathcal{X}$. Also, the encoder must be designed such that the mutual information between the message W and the received vector Z^n at the eavesdropper

$$I(W; Z^n) < \epsilon_e \tag{2.2}$$

- a decoder which maps the received vector $Y^n \in \mathcal{Y}^n$ at the legitimate receiver to an estimated message $\hat{W} \in \{1, \ldots, w_0\}$ such that the average probability of error

$$\Pr(\hat{W} \neq W) < \epsilon_d. \tag{2.3}$$

Following the classical Shannon formulation [1], the secrecy capacity of a memoryless stationary wiretap channel can be defined as follows.

Definition 1 (Secrecy capacity [2, 3]). *Given $0 < \epsilon < 1$ and $\beta > 0$, a nonnegative real R_s is an ϵ-achievable secrecy rate with cost per symbol not exceeding β if there exists a positive integer n_0 such that for any integer $n \geq n_0$ an $(n, w_0, n\beta, \epsilon, n\epsilon)$ code*

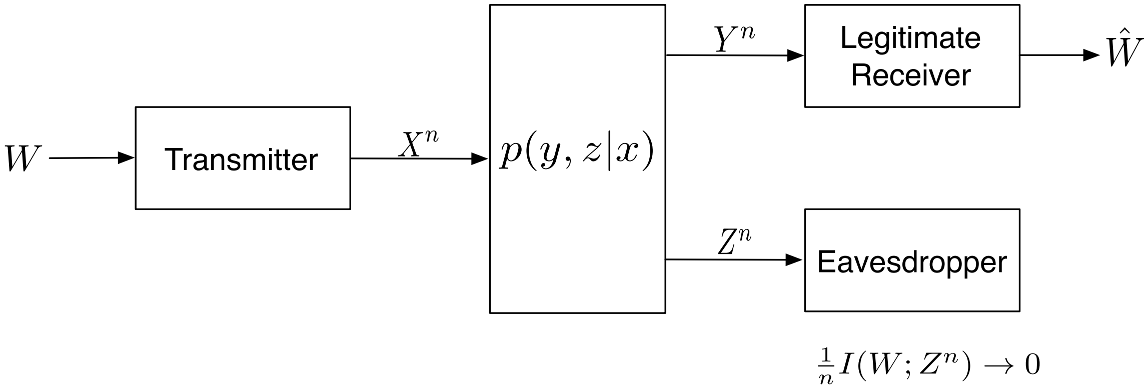

Figure 2.1: Memoryless stationary wiretap channel.

can be found for which

$$\frac{\log w_0}{n} > \mathsf{R}_s - \delta \tag{2.4}$$

for any $\delta > 0$. Furthermore, R_s is said to be achievable if it is ϵ-achievable for all $0 < \epsilon < 1$. The supreme of all achievable secrecy rates with cost per symbol not exceeding β is the secrecy capacity denoted by $\mathsf{C}_s(\beta)$. The secrecy capacity $\mathsf{C}_s(\beta)$ as a function of the cost per symbol β is formally referred to as the secrecy capacity-cost function.

For the case where the memoryless stationary wiretap channel $(\mathcal{X}, (\mathcal{Y}, \mathcal{Z}), P_{Y,Z|X})$ is degraded, i.e, $X \to Y \to Z$ forms a Markov chain in that order, Wyner [2] showed that the secrecy capacity-cost function $\mathsf{C}_s(\beta)$ is given by

$$\mathsf{C}_s(\beta) = \sup_{E[b(X)]\leq\beta} \left(I(X;Y) - I(X;Z)\right). \tag{2.5}$$

The secrecy capacity-cost function of a general memoryless stationary wiretap channel $(\mathcal{X}, (\mathcal{Y}, \mathcal{Z}), P_{Y,Z|X})$ was characterized by Csiszár and Körner [3] and can be written as

$$\mathsf{C}_s(\beta) = \sup_{E[b(X)]\leq\beta} \left(I(V;Y|U) - I(V;Z|U)\right) \tag{2.6}$$

where U and V are auxiliary random variables satisfying the Markov chain $U \rightarrow V \rightarrow X \rightarrow (Y, Z)$.

2.3 Secrecy Capacity per Unit Cost

We use the following definition for secrecy capacity per unit cost.

Definition 2 (Secrecy capacity per unit cost)**.** *Given $0 < \epsilon < 1$, a nonnegative real R_s is an ϵ-achievable secrecy rate per unit cost if there exists a positive real ν_0 such that for any $\nu \geq \nu_0$ an $(n, w_0, \nu, \epsilon, \nu\epsilon)$ code can be found for which*

$$\frac{\ln w_0}{\nu} > R_s - \delta \tag{2.7}$$

for any $\delta > 0$. Furthermore, R_s is an achievable secrecy rate per unit cost if it is ϵ-achievable for all $0 < \epsilon < 1$, and the secrecy capacity per unit cost C_s is the supreme of all achievable secrecy rates per unit cost.

Remark 1. *Note that in Definition 2, the secrecy requirement at the eavesdropper is that the mutual information $I(W; Z^n)$ normalized by the total transmission cost ν can be made arbitrarily small for sufficiently large ν. Unlike for secrecy capacity, the length of the codewords n does* not *play a fundamental role in the definition of secrecy capacity per unit cost. Therefore, normalizing the mutual information $I(W; Z^n)$ by the block-length n, instead of the total transmission cost ν, will only trivialize the secrecy constraint.*

Remark 2. *By Definition 2, to show that that R_s is an ϵ-achievable secrecy rate per* *unit cost, one has to consider all real $\nu \geq \nu_0$. From the proof viewpoint, however, it is sufficient to consider $\nu_k = k\beta$ for some $\beta > 0$ and show that there exists a positive integer k_0 such that for any integer $k \geq k_0$ an $(n, w_0, \nu_k, \epsilon, \nu_k \epsilon)$ code can be found for which*

$$\frac{\ln w_0}{\nu_k} > R_s - \frac{\delta}{2} \tag{2.8}$$

for any $\delta > 0$.

For completeness, a short proof of Remark 2 is provided in Appendix A. Similar to that between the Shannon capacity-cost function and Verdú's channel capacity per unit cost [4, Th. 2], we have the following simple relationship between the secrecy capacity-cost function and the secrecy capacity per unit cost.

Theorem 1. *The secrecy capacity per unit cost C_s of the memoryless stationary wiretap channel $(\mathcal{X}, (\mathcal{Y}, \mathcal{Z}), P_{Y,Z|X})$ is given by*

$$C_s = \sup_{\beta > 0} \frac{\mathsf{C}_s(\beta)}{\beta} \tag{2.9}$$

where $\mathsf{C}_s(\beta)$ is the secrecy capacity-cost function of the channel.

Proof. Let us first show that for any given $\beta > 0$, $\mathsf{C}_s(\beta)/\beta$ is an ϵ-achievable secrecy rate per unit cost for any $0 < \epsilon < 1$. Fix $\delta > 0$ and let $\epsilon' = \min(1, \beta)\epsilon$. Since $\mathsf{C}_s(\beta)$ is the secrecy capacity-cost function, there exists a positive integer n_0 such that for any integer $n \geq n_0$ an $(n, w_0, n\beta, \epsilon', n\epsilon')$ code can be found for which

$$\frac{\ln w_0}{n} > \mathsf{C}_s(\beta) - \frac{\beta\delta}{2}. \tag{2.10}$$

This immediately gives an $(n, w_0, n\beta, \epsilon, n\beta\epsilon)$ code for which

$$\frac{\ln w_0}{n\beta} > \frac{\mathsf{C}_s(\beta)}{\beta} - \frac{\delta}{2}. \tag{2.11}$$

Per Remark 2, this proves that $\mathsf{C}_s(\beta)/\beta$ is an ϵ-achievable secrecy rate per unit cost.

To prove the converse part of the theorem, let R_s be an achievable secrecy rate per unit cost. By definition, for any $0 < \epsilon < 1$ there exists a positive real ν_0 such that for any $\nu \geq \nu_0$ an $(n, w_0, \nu, \epsilon, \nu\epsilon)$ code can be found for which

$$\frac{\ln w_0}{\nu} > R_s - \delta \tag{2.12}$$

for any $\delta > 0$. By Fano's inequality, we have

$$H(W|Y^n) \leq \epsilon H(W) + \ln 2. \tag{2.13}$$

It follows that

$$(1-\epsilon)H(W) \leq I(W;Y^n) + \ln 2 \tag{2.14}$$

and hence

$$\begin{aligned} & R_s - \delta \\ & \quad < \frac{\ln w_0}{\nu} = \frac{H(W)}{\nu} && (2.15) \\ & \quad \leq \frac{1}{1-\epsilon}\left(\frac{I(W;Y^n)}{\nu} + \frac{\ln 2}{\nu}\right) && (2.16) \\ & \quad \leq \frac{1}{1-\epsilon}\left(\frac{I(W;Y^n) - I(W;Z^n)}{\nu} + \epsilon + \frac{\ln 2}{\nu}\right). && (2.17) \end{aligned}$$

For memoryless stationary wiretap channels, Csiszár and Körner [3] showed that

$$\begin{aligned} I(W;Y^n) - I(W;Z^n) \leq& \\ & n \sup_{E[b(X)] \leq \nu/n} \left(I(V;Y|U) - I(V;Z|U) \right) \end{aligned} \tag{2.18}$$

where U and V are auxiliary random variables satisfying the Markov chain $U \to V \to X \to (Y, Z)$. Substituting (2.18) into (2.17) gives

$$\begin{aligned} (1-\epsilon)(R_s - \delta) - \left(\epsilon + \frac{\ln 2}{\nu}\right)& \\ &< \frac{n}{\nu} \sup_{E[b(X)] \leq \nu/n} \left(I(V;Y|U) - I(V;Z|U) \right) &(2.19)\\ &\leq \sup_{\beta > 0} \left(\frac{1}{\beta} \sup_{E[b(X)] \leq \beta} \left(I(V;Y|U) - I(V;Z|U) \right) \right) &(2.20)\\ &= \sup_{\beta > 0} \frac{\mathsf{C}_s(\beta)}{\beta}. &(2.21) \end{aligned}$$

Letting δ and ϵ go to zero and ν go to infinity, we conclude that $R_s \leq \sup_{\beta > 0} \frac{\mathsf{C}_s(\beta)}{\beta}$ for any achievable secrecy rate per unit cost R_s. This completes the proof of the converse part and hence the entire theorem. □

From Theorem 1 we see that at least in theory, the secrecy capacity per unit cost can be calculated from the secrecy capacity-cost function. However, as shown in (2.5) and (2.6), calculating the secrecy capacity-cost function usually involves finding an optimal distribution for the input/auxiliary random variables and hence is highly nontrivial in general. Next, following [4], we shall focus on the case where there is a *zero-cost* letter in the input alphabet $\mathcal{X}$ (labeled as "0" throughout the rest of the paper, i.e., $b(0) = 0$) and look for a more direct way to calculate the secrecy capacity per unit cost *without* resorting to the secrecy capacity-cost function.

2.3.1 Degraded Wiretap Channel

When there is a zero-cost letter in the input alphabet $\mathcal{X}$, the secrecy capacity per unit cost of a degraded wiretap channel can be calculated without resorting to the secrecy capacity-cost function. The result is summarized in the following theorem.

Theorem 2. *The secrecy capacity per unit cost C_s of the memoryless stationary wiretap channel $(\mathcal{X}, (\mathcal{Y}, \mathcal{Z}), P_{Y,Z|X})$ under the degradedness assumption $X \to Y \to Z$ and with a zero-cost input letter "0" is given by*

$$C_s = \sup_{x \in \mathcal{X}} \frac{N(x)}{b(x)} \tag{2.22}$$

where

$$N(x) := D(P_{Y|X=x} \| P_{Y|X=0}) - D(P_{Z|X=x} \| P_{Z|X=0}) \tag{2.23}$$

and $D(P\|Q)$ denotes the (Kullback-Leibler) divergence between two generic probability distributions P and Q. Furthermore, the secrecy capacity per unit cost of the channel can be achieved by an orthogonal coding scheme.

Proof. Let us first prove (2.22). The fact that

$$C_s \geq \sup_{x \in \mathcal{X}} \frac{N(x)}{b(x)} \tag{2.24}$$

does not depend on the assumption that the channel is degraded and can be inferred from a stronger lower bound on the secrecy capacity per unit cost of the general wiretap channel provided in Sec. 2.3.2. Next, we show that for degraded wiretap channels, we also have the reversed inequality

$$C_s \le \sup_{x \in \mathcal{X}} \frac{N(x)}{b(x)}. \tag{2.25}$$

By the degradedness assumption $X \to Y \to Z$, the probability distributions $P_{Z|X=x}$, $P_{Z|X=0}$ and P_Z can be obtained from $P_{Y|X=x}$, $P_{Y|X=0}$ and P_Y respectively via the *same* "processing" $P_{Z|Y}$. By the data-processing inequality for divergence [8], we have

$$D(P_{Y|X=x} \| P_{Y|X=0}) \ge D(P_{Z|X=x} \| P_{Z|X=0}) \tag{2.26}$$

$$D(P_Y \| P_{Y|X=0}) \ge D(P_Z \| P_{Z|X=0}). \tag{2.27}$$

By (2.26), we have $N(x) \ge 0$ for any $x \in \mathcal{X}$. Let

$$\mathcal{X}' := \{x \in \mathcal{X} : b(x) > 0\}. \tag{2.28}$$

If there exists an $x \in \mathcal{X} \setminus \mathcal{X}'$ such that $N(x) > 0$, then $N(x)/b(x) = \infty$ and there is nothing to prove from the converse point of view. Therefore, without loss of generality we may assume that $N(x) = 0$ for all $x \in \mathcal{X} \setminus \mathcal{X}'$.

Now consider the following relations between mutual information and divergence:

$$\begin{aligned} I(X;Y) = &\int_{\mathcal{X}} D(P_{Y|X=x} \| P_{Y|X=0}) dP_X(x) - \\ &D(P_Y \| P_{Y|X=0}) \end{aligned} \tag{2.29}$$

$$\begin{aligned} I(X;Z) = &\int_{\mathcal{X}} D(P_{Z|X=x} \| P_{Z|X=0}) dP_X(x) - \\ &D(P_Z \| P_{Z|X=0}). \end{aligned} \tag{2.30}$$

We thus have

$$\begin{aligned}
&I(X;Y) - I(X;Z) \\
&\quad = \int_{\mathcal{X}} N(x) dP_X(x) - \\
&\qquad \big(D(P_Y \| P_{Y|X=0}) - D(P_Z \| P_{Z|X=0})\big) && (2.31) \\
&\quad \leq \int_{\mathcal{X}} N(x) dP_X(x) && (2.32) \\
&\quad = \int_{\mathcal{X}'} N(x) dP_X(x) && (2.33) \\
&\quad = \int_{\mathcal{X}'} \frac{N(x)}{b(x)} b(x) dP_X(x) && (2.34) \\
&\quad \leq \left(\sup_{x \in \mathcal{X}'} \frac{N(x)}{b(x)}\right) \int_{\mathcal{X}'} b(x) dP_X(x) && (2.35) \\
&\quad = \left(\sup_{x \in \mathcal{X}'} \frac{N(x)}{b(x)}\right) \int_{\mathcal{X}} b(x) dP_X(x) && (2.36) \\
&\quad = \left(\sup_{x \in \mathcal{X}'} \frac{N(x)}{b(x)}\right) E[b(X)] && (2.37) \\
&\quad \leq \left(\sup_{x \in \mathcal{X}} \frac{N(x)}{b(x)}\right) E[b(X)] && (2.38)
\end{aligned}$$

where (2.32) follows from (2.27), and (2.33) follows from the assumption that $N(x) = 0$ for all $x \in \mathcal{X} \setminus \mathcal{X}'$, and (2.36) follows from the definition of $\mathcal{X}'$. Substituting (2.5) and (2.38) into (2.9) gives

$$\begin{aligned}
C_s &= \sup_{\beta > 0} \sup_{E[b(X)] \leq \beta} \frac{I(X;Y) - I(X;Z)}{\beta} && (2.39) \\
&\leq \sup_{x \in \mathcal{X}} \frac{N(x)}{b(x)} \cdot \sup_{\beta > 0} \sup_{E[b(X)] \leq \beta} \frac{E[b(X)]}{\beta} && (2.40) \\
&= \sup_{x \in \mathcal{X}} \frac{N(x)}{b(x)}. && (2.41)
\end{aligned}$$

This completes the proof of (2.22).

To show that for any given $x \in \mathcal{X}$, the secrecy rate per unit cost $N(x)/b(x)$

can be achieved by an orthogonal coding scheme, let us first consider $x \in \mathcal{X}'$. Fix $0 < \delta < 2N(x)/b(x)$, $0 < \epsilon < 1$, and k to be a sufficiently large positive integer. Let $m = w_0 l_0$ for some integers w_0 and l_0 such that

$$\exp\left(k\left(N(x) - \frac{\delta b(x)}{2}\right)\right) < \\ w_0 < \exp\left(k\left(N(x) - \frac{\delta b(x)}{3}\right)\right) \tag{2.42}$$

and

$$\exp\left(k\left(D(P_{Z|X=x}\|P_{Z|X=0}) + \frac{\delta b(x)}{12}\right)\right) < \\ l_0 < \exp\left(k\left(D(P_{Z|X=x}\|P_{Z|X=0}) + \frac{\delta b(x)}{6}\right)\right). \tag{2.43}$$

Codebook. Each codeword is identified by an integer pair (w, l), where $w \in \{1, \ldots, w_0\}$ and $l \in \{1, \ldots, l_0\}$, and corresponds to an $m \times k$ matrix

$$\{x_{i,j} : 1 \leq i \leq m, 1 \leq j \leq k\}.$$

Denote by x_i^k the ith row of the codeword matrix $\{x_{i,j}\}$. For codeword (w, l),

$$x_i^k = \begin{cases} (x, \ldots, x) & \text{if } i = (w-1)l_0 + l \\ (0, \ldots, 0) & \text{otherwise.} \end{cases} \tag{2.44}$$

Thus, the block length of this code $n = mk$, and the cost of each codeword is $kb(x)$. In this paper, the nonzero row of a codeword matrix is referred to as a "pulse". Note that the pulses for different codewords are non-overlapping, so the codewords are orthogonal to each other.

Encoding. Given message W, randomly and uniformly choose an integer $L \in$

$\{1, \ldots, l_0\}$, and send codeword (W, L) through the channel. The randomness used for choosing L is intrinsic to the transmitter and is *not* shared with either the legitimate receiver or the eavesdropper.

Decoding at the legitimate receiver. Given the matrix of observations

$$\{Y_{i,j} : 1 \leq i \leq m, 1 \leq j \leq k\}$$

the decoder performs m independent binary hypothesis tests, one on each row of the transmitted codeword matrix:

$$\begin{aligned} H_{i,0}: &\quad x_i^k = (0, \ldots, 0) \\ H_{i,1}: &\quad x_i^k = (x, \ldots, x) \end{aligned}$$

for $i = 1, \ldots, m$. The conditional error probabilities of these tests are denoted by

$$\alpha_i^{(k)} = \Pr\{\hat{H}_{i,1} | H_{i,0}\} \tag{2.45}$$

$$\beta_i^{(k)} = \Pr\{\hat{H}_{i,0} | H_{i,1}\} \tag{2.46}$$

and the decision rule is set so that $\beta_i^{(k)} < \epsilon/2$. If one and only one $H_{i,1}$ was claimed (denoted by $H_{\hat{i},1}$), we declare the transmitted codeword to be $(\hat{w}, \hat{i} - (\hat{w} - 1)l_0)$ (and the transmitted message to be $\hat{w}$) where $\hat{w}$ is given by the smallest integer greater than or equal to $\hat{i}/l_0$. Otherwise, an error is declared.

Obviously, the probability $P_{w,l}$ for erroneously decoding the transmitted codeword conditioned on codeword (w, l) being sent is independent of the value of (w, l) and can be bounded from above as

$$P_{1,1} \leq \beta_i^{(k)} + (m - 1)\alpha_i^{(k)}. \tag{2.47}$$

Denote by Y_i^k the ith row of the matrix of observations $\{Y_{i,j}\}$. Note that Y_i^k is i.i.d. according to $P_{Y|X=0}$ under $H_{i,0}$ and i.i.d. according to $P_{Y|X=x}$ under $H_{i,1}$. By the Chernoff-Stein lemma [9], since $\beta_i^{(k)} < \epsilon/2$, we can achieve

$$\alpha_i^{(k)} < \exp\left(-k\left(D(P_{Y|X=x}\|P_{Y|X=0}) - \frac{\delta b(x)}{12}\right)\right) \tag{2.48}$$

for sufficiently large k. Consequently,

$$\begin{aligned} P_{1,1} &< \epsilon/2 + m\exp\left(-k\left(D(P_{Y|X=x}\|P_{Y|X=0}) - \frac{\delta b(x)}{12}\right)\right) &(2.49)\\ &< \epsilon/2 + \exp(-k\delta b(x)/12) &(2.50)\\ &< \epsilon &(2.51)\end{aligned}$$

for sufficiently large k, where (2.50) follows from the fact that

$$m = w_0 l_0 < \exp\left(k\left(D(P_{Y|X=x}\|P_{Y|X=0}) - \frac{\delta b(x)}{6}\right)\right). \tag{2.52}$$

Confidentiality at the eavesdropper. The mutual information between the transmitted message W and the matrix of observations

$$\{Z_{i,j} : 1 \le i \le m, 1 \le j \le k\}$$

is given by

$$\begin{aligned}
&I(W;\{Z_{i,j}\}) && (2.53)\\
&\quad = H(W) - H(W|\{Z_{i,j}\}) && (2.54)\\
&\quad = H(W) - H(W,L|\{Z_{i,j}\}) + H(L|W,\{Z_{i,j}\}) && (2.55)\\
&\quad = H(W) - H(W,L) + I(W,L;\{Z_{i,j}\}) + H(L|W,\{Z_{i,j}\}) && (2.56)\\
&\quad = -H(L) + I(\{X_{i,j}\};\{Z_{i,j}\}) + H(L|W,\{Z_{i,j}\}). && (2.57)
\end{aligned}$$

Induced by the random selection of (W, L), the transmitted codeword entries $X_{i,j}$ are identically distributed according to

$$P_X(x) = 1 - P_X(0) = 1/m. \qquad (2.58)$$

Note from (2.42) and (2.43) that $m \to \infty$ in the limit as $k \to \infty$. By [10, Eq. (2.13)], for any $(i, j) \in \{1, \ldots, m\} \times \{1, \ldots, k\}$ we have

$$\lim_{k\to\infty} (mI(X_{i,j}; Z_{i,j})) = D(P_{Z|X=x} \| P_{Z|X=0}). \qquad (2.59)$$

Thus for sufficiently large k,

$$\begin{aligned}
I(\{X_{i,j}\};\{Z_{i,j}\}) &\le \sum_{i=1}^{m}\sum_{j=1}^{k} I(X_{i,j};Z_{i,j}) && (2.60)\\
&< k\left(D(P_{Z|X=x}\|P_{Z|X=0}) + \frac{\epsilon b(x)}{2}\right) && (2.61)
\end{aligned}$$

where (2.60) follows from the fact that the channel is memoryless, and (2.61) follows from (2.59). We also have the following lemma, whose proof is provided in Appendix B.

Lemma 1.

$$\begin{aligned} &H(L|W,\{Z_{i,j}\}) \\ &\qquad < H(L) - k\left(D(P_{Z|X=x}\|P_{Z|X=0}) - \frac{\epsilon b(x)}{2}\right) \end{aligned} \tag{2.62}$$

for sufficiently large k.

Substituting (2.61) and (2.62) into (2.57) gives

$$I(W;\{Z_{i,j}\}) < kb(x)\epsilon. \tag{2.63}$$

Combing (2.51) and (2.63), we have successfully constructed for any $0 \leq \epsilon < 1$, any $0 < \delta < 2N(x)/b(x)$, and any sufficiently large k, an $(mk, w_0, kb(x), \epsilon, kb(x)\epsilon)$ code for which

$$\frac{\log w_0}{kb(x)} > \frac{N(x)}{b(x)} - \frac{\delta}{2}. \tag{2.64}$$

Per Remark 2, this proves that for any $x \in \mathcal{X}'$, the secrecy rate per unit cost $N(x)/b(x)$ can be achieved by the proposed orthogonal coding scheme.

For $x \in \mathcal{X} \setminus \mathcal{X}'$, we have $b(x) = 0$. If $N(x) = 0$, by our convention $N(x)/b(x) = 0$, and there is nothing to prove from the achievability point of view. If, on the other hand, $N(x) > 0$, replace $b(x)$ by some positive real b in the previous analysis. Then, the same orthogonal coding scheme can achieve the secrecy capacity per unit cost $N(x)/b$ for *any* $b > 0$. Letting $b \to 0$ proves that in this case, the proposed orthogonal coding scheme can achieve an infinite secrecy rate per unit cost.

Combining the above two cases proves that for any $x \in \mathcal{X}$, the secrecy rate per unit cost $N(x)/b(x)$ can be achieved by the proposed orthogonal coding scheme. This completes the proof of the entire theorem. □

Compared with the expression (2.5) for the secrecy capacity-cost function, the expression (2.22) for the secrecy capacity per unit cost involves only an optimization over the input letter rather than the input distribution. For many specific channels, this represents a significant reduction of the computational complexity.

Example 1. *Consider the Gaussian wiretap channel*

$$\begin{aligned} Y &= X + N_1 \\ Z &= X + N_2 \end{aligned} \tag{2.65}$$

with a real channel input X and a quadratic cost function $b(x) = x^2$, where N_1 and N_2 are additive Gaussian noise with zero means and variance σ_1^2 and σ_2^2, respectively. Assume that $\sigma_1^2 \leq \sigma_2^2$. Just like the secrecy capacity, the secrecy capacity per unit cost of the channel depends on the joint distribution of the additive noise (N_1, N_2) only through its marginals. Thus, for the purpose of calculating the secrecy capacity per unit cost we can write $N_2 = N_1 + \tilde{N}_2$ where $\tilde{N}_2$ is Gaussian with zero mean and variance $\tilde{\sigma}_2^2 = \sigma_2^2 - \sigma_1^2$ and is independent of N_1. It follows that the Gaussian wiretap channel (2.65) *can be equivalently written as*

$$\begin{aligned} Y &= X + N_1 \\ Z &= X + N_1 + \tilde{N}_2 \end{aligned} \tag{2.66}$$

which satisfies the Markov relation $X \to Y \to Z$. The divergence between two

Gaussian random variables is given by

$$D(\mathcal{N}(\mu_1,\sigma_1^2)\|\mathcal{N}(\mu_0,\sigma_0^2)) \\ = \ln\frac{\sigma_0}{\sigma_1} + \frac{\sigma_1^2-\sigma_0^2+(\mu_1-\mu_0)^2}{2\sigma_0^2}. \qquad (2.67)$$

Therefore, for any $x \neq 0$ we have

$$\frac{N(x)}{x^2} = \frac{1}{2}\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right). \qquad (2.68)$$

Thus, by (2.22) *and* without *any optimization, we may conclude that the secrecy capacity per unit cost of the Gaussian wiretap channel* (2.65) *under the quadratic cost function $b(x) = x^2$ is given by*

$$C_s = \frac{1}{2}\left(\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right). \qquad (2.69)$$

Under some mild regularity conditions on the family of distributions $\{P_{Y|X=x} : x \in \mathbb{R}\}$, the following asymptotic result on divergence is known [11, Ch. 2.6]:

$$\lim_{x\downarrow 0}\frac{D(P_{Y|X=x}\|P_{Y|X=0})}{x^2} = \frac{1}{2}J_0(P_{Y|X}) \qquad (2.70)$$

where $J_0(P_{Y|X})$ is the Fisher information over the parameter family $\{P_{Y|X=x} : x \in \mathbb{R}\}$ evaluated at $x = 0$. This result can be used to obtain a simple lower bound on the secrecy capacity per unit cost.

Theorem 3. *The secrecy capacity per unit cost C_s of the memoryless stationary wiretap channel $(\mathcal{X},(\mathcal{Y},\mathcal{Z}),P_{Y,Z|X})$ with an input alphabet $\mathcal{X} = \mathbb{R}$ and a quadratic* *cost function $b(x) = x^2$ can be bounded from below as*

$$C_s \geq \frac{1}{2}\left(J_0(P_{Y|X}) - J_0(P_{Z|X})\right). \tag{2.71}$$

Proof. As mentioned previously in the proof of Theorem 2, for any $x \in \mathcal{X}$ the secrecy rate per unit cost $N(x)/b(x)$ (when it is positive) is achievable for *any* memoryless stationary wiretap channels (not necessarily degraded) with a zero-cost input letter. Under the quadratic cost function $b(x) = x^2$, "0" is a zero-cost input letter, i.e., $b(0) = 0$. Thus, (2.71) can be proved by letting $x \downarrow 0$ in $N(x)/b(x)$ and applying the asymptotic result (2.70) to both $D(P_{Y|X=x} \| P_{Y|X=0})$ and $D(P_{Z|X=x} \| P_{Z|X=0})$. □

With the help of Theorem 3, we can prove a *worst-noise* property for the Gaussian wiretap channel (2.65).

Theorem 4. *Consider the memoryless stationary wiretap channel* (2.66) *with an input alphabet $\mathcal{X} = \mathbb{R}$, a quadratic cost function $b(x) = x^2$, and independent additive noise N_1 and $\tilde{N}_2$. While $\tilde{N}_2$ is assumed to be Gaussian with zero mean and variance $\tilde{\sigma}_2^2 = \sigma_2^2 - \sigma_1^2$, N_1 is possibly non-Gaussian. The secrecy capacity per unit cost of the channel*

$$C_s \geq \frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) \tag{2.72}$$

for any distribution of N_1 with zero mean and variance σ_1^2. The equality holds when N_1 is also Gaussian.

Proof. By Theorem 3, the secrecy capacity per unit cost of the channel can be bounded from below as:

$$C_s \geq \frac{1}{2}\left(J(N_1) - J(N_1 + \tilde{N}_2)\right) \tag{2.73}$$

where $J(X)$ denotes the Fisher information of generic random variable X relative to a translation parameter. By the Fisher information inequality [12],

$$J(N_1 + \tilde{N}_2) \leq \frac{J(N_1)J(\tilde{N}_2)}{J(N_1) + J(\tilde{N}_2)} = \frac{J(N_1)}{\tilde{\sigma}_2^2 J(N_1) + 1} \tag{2.74}$$

where the last equality follows from the fact that $\tilde{N}_2$ is $\mathcal{N}(0, \tilde{\sigma}_2^2)$ so $J(\tilde{N}_2) = 1/\tilde{\sigma}_2^2$. Substituting (2.74) into (2.73), we have

$$C_s \geq \frac{1}{2}\frac{\tilde{\sigma}_2^2\left(J(N_1)\right)^2}{\tilde{\sigma}_2^2 J(N_1) + 1} \tag{2.75}$$

Note that the right-hand side of (2.75) is monotonically increasing with $J(N_1)$. By the well-known Cramér-Rao inequality,

$$J(N_1) \geq \frac{1}{\sigma_1^2}. \tag{2.76}$$

Substituting (2.76) into (2.75) gives

$$C_s \geq \frac{1}{2}\frac{\tilde{\sigma}_2^2}{\sigma_1^2(\sigma_1^2 + \tilde{\sigma}_2^2)} = \frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right). \tag{2.77}$$

This completes the proof of the theorem. □

2.3.2 General Wiretap Channel

For general memoryless stationary wiretap channels with a zero-cost input letter, the following secrecy rates per unit cost can be achieved by an orthogonal coding scheme.

Theorem 5. *The secrecy capacity per unit cost C_s of the memoryless stationary wiretap channel $(\mathcal{X}, (\mathcal{Y}, \mathcal{Z}), P_{Y,Z|X})$ with a zero-cost input letter "0" can be bounded from below as*

$$C_s \geq \sup_{P_X} \frac{D(P_Y \| P_{Y|X=0}) - D(P_Z \| P_{Z|X=0})}{E[b(X)]} \tag{2.78}$$

where

$$P_Y = \int_{\mathcal{X}} P_{Y|X=x} dP_X(x) \tag{2.79}$$

$$P_Z = \int_{\mathcal{X}} P_{Z|X=x} dP_X(x) \tag{2.80}$$

and

$$E[b(X)] = \int_{\mathcal{X}} b(x) dP_X(x). \tag{2.81}$$

Furthermore, for any P_X over $\mathcal{X}$ such that both

$$D(P_Y \| P_{Y|X=0}) - D(P_Z \| P_{Z|X=0})$$

and $E[b(X)]$ are positive, the secrecy rate per unit cost

$$R_s = \frac{D(P_Y \| P_{Y|X=0}) - D(P_Z \| P_{Z|X=0})}{E[b(X)]} \tag{2.82}$$

can be achieved by an orthogonal coding scheme.

Proof. Let us first prove (2.78). For general memoryless stationary wiretap channels, by (2.6) and (2.9) we have

$$C_s \geq \frac{I(V;Y) - I(V;Z)}{E[b(X)]} \tag{2.83}$$

for any joint distribution $P_{Y,Z,X,V} = P_{Y,Z|X} P_{X|V} P_V$. Consider the following distribution for (V, X): V is a binary random variable such that

$$P_V(1) = 1 - P_V(0) = 1/m \tag{2.84}$$

for some positive integer m, and $X \sim P_{X|V=1}$ if $V = 1$ and $X = 0$ with probability one if $V = 0$. For this particular choice of distribution for (V, X), we have

$$P_{Y|V=0} = P_{Y|X=0} \tag{2.85}$$

$$P_{Z|V=0} = P_{Z|X=0} \tag{2.86}$$

$$P_{Y|V=1} = \int_{\mathcal{X}} P_{Y|X=x} dP_{X|V=1}(x) \tag{2.87}$$

$$P_{Z|V=1} = \int_{\mathcal{X}} P_{Z|X=x} dP_{X|V=1}(x) \tag{2.88}$$

and

$$E[b(X)] = \frac{1}{m} E[b(X)|V = 1]. \tag{2.89}$$

By [10, Eq. (2.13)], we also have

$$\lim_{m \to \infty} (mI(V;Y)) = D(P_{Y|V=1} \| P_{Y|V=0}) \tag{2.90}$$

$$\lim_{m \to \infty} (mI(V;Z)) = D(P_{Z|V=1} \| P_{Z|V=0}). \tag{2.91}$$

Substituting (2.89)–(2.91) into (2.83) gives

$$
\begin{aligned}
C_s &\geq \lim_{m\to\infty} \frac{m\left(I(V;Y)-I(V;Z)\right)}{E[b(X)|V=1]} && (2.92)\\
&= \frac{D(P_{Y|V=1}\|P_{Y|V=0})-D(P_{Z|V=1}\|P_{Z|V=0})}{E[b(X)|V=1]} && (2.93)\\
&= \frac{D(P_{Y|V=1}\|P_{Y|X=0})-D(P_{Z|V=1}\|P_{Z|X=0})}{E[b(X)|V=1]} && (2.94)
\end{aligned}
$$

for any $P_{X|V=1}$ over $\mathcal{X}$. Renaming $P_{X|V=1}$, $P_{Y|V=1}$ and $P_{Z|V=1}$ as P_X, P_Y and P_Z respectively completes the proof of (2.78).

To prove that (2.82) can be achieved by an orthogonal coding scheme, we shall consider the following modification of the orthogonal coding scheme proposed for the degraded case.

Let $m = w_0 l_0$ for some integers w_0 and l_0 such that

$$
\begin{aligned}
w_0 &\approx \exp\left(k\left(D(P_Y\|P_{Y|X=0})-D(P_Z\|P_{Z|X=0})\right)\right) && (2.95)\\
l_0 &\approx \exp\left(kD(P_Z\|P_{Z|X=0})\right) && (2.96)
\end{aligned}
$$

and t_0 be an integer such that

$$t_0 \approx \exp\left(kI(X;Z)\right). \tag{2.97}$$

Let $\mathcal{C} = \{c^k(1), \ldots, c^k(t_0)\}$ be a collection of t_0 length-k vectors from $\mathcal{X}^k$.

Codebook. Each codeword is identified by an integer triple (w, l, t), where $w \in \{1, \ldots, w_0\}$, $l \in \{1, \ldots, l_0\}$, and $t \in \{1, \ldots, t_0\}$, and corresponds to an $m \times k$ matrix

$$\{x_{i,j} : 1 \leq i \leq m, 1 \leq j \leq k\}.$$

Denote by x_i^k the ith row of the codeword matrix $\{x_{i,j}\}$. For codeword (w, l, t),

$$x_i^k = \begin{cases} c^k(t) & \text{if } \ i = (w-1)l_0 + l \\ (0, \ldots, 0) & \text{otherwise.} \end{cases} \tag{2.98}$$

Encoding. Given message W, randomly, uniformly, and independently choose an integer $L \in \{1, \ldots, l_0\}$ and an integer $T \in \{1, \ldots, t_0\}$, and send codeword (W, L, T) through the channel. The randomness used for choosing L and T is intrinsic to the transmitter and is *not* shared with neither the legitimate receiver nor the eavesdropper. Note that:

1) even though the codewords are not necessarily orthogonal to each other, the codewords representing *different* messages remain orthogonal to each other; and

2) compared with the orthogonal coding scheme proposed for the degrade case, additional randomization on the "shape" of the pulse is used in the modified scheme.

Decoding at the legitimate receiver. Given the matrix of observations

$$\{Y_{i,j} : 1 \leq i \leq m, 1 \leq j \leq k\}$$

the decoder performs m independent binary hypothesis tests, one on each row of the transmitted codeword matrix:

$$\begin{aligned} H_{i,0} : &\quad x_i^k = (0, \ldots, 0) \\ H_{i,1} : &\quad x_i^k \text{ uniformly drawn from } \mathcal{C} \end{aligned}$$

for $i = 1, \ldots, m$. If one and only one $H_{i,1}$ was claimed (denoted by $H_{\hat{\imath},1}$), we declare the transmitted message to be the smallest integer greater than or equal to $\hat{i}/l_0$. Otherwise, an error is declared.

Performance analysis. Using a random-coding argument for which the entries of the vectors from $\mathcal{C}$ are independently generated according to P_X, it can be shown that for sufficiently large k:

1) the cost associated with each vector from $\mathcal{C}$ is approximately $kE[b(X)]$; and
2) x_i^k, when uniformly drawn from $\mathcal{C}$, has approximately i.i.d. entries according to P_X.

Then, following the same footsteps as those for the degraded case, it can be shown that the secrecy rate per unit cost (2.82) can be achieved by the proposed coding scheme. This completes the proof of the theorem. □

Note that if we choose $X = x$ with probability one, the right-hand side of (2.78) reduces to (2.22) which was shown to be the secrecy capacity per unit cost when the wiretap channel is degraded. The following example, however, shows that further randomization of the pulse shape can strictly improve the achievable secrecy rate per unit cost when the wiretap channel is *not* degraded.

Example 2. *Consider a binary memoryless stationary wiretap channel with* $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. *The marginal channel transition probabilities are given by*

$$P_{Y|X=1} = (0.4, 0.6), \quad P_{Y|X=0} = (0.6, 0.4)$$
$$P_{Z|X=1} = (0.3, 0.7), \quad P_{Z|X=0} = (0.5, 0.5).$$

Obviously, the channel is not degraded. The cost function is given by $b(0) = 0$ *and*

$b(1) = 1$. *Simple calculations give*

$$\frac{D(P_{Y|X=1} \| P_{Y|X=0}) - D(P_{Z|X=1} \| P_{Z|X=0})}{b(1)} \approx -0.0012.$$

Therefore, orthogonal codes with constant pulse shape cannot achieve any positive secrecy rate per unit cost. On the other hand, let $P_X = (0.5, 0.5)$ *and we get*

$$P_Y = (0.5, 0.5) \quad \textit{and} \quad P_Z = (0.4, 0.6).$$

By Theorem 5, the following secrecy rate per unit cost is achievable by an orthogonal code with randomized pulse shape:

$$\frac{D(P_Y \| P_{Y|X=0}) - D(P_Z \| P_{Z|X=0})}{E[b(X)]} \approx 0.0006.$$

So at least for this simple example, further randomization of the pulse shape can strictly improve the achievable secrecy rate per unit cost.

2.3.3 MIMO Wiretap Channel

We consider the model of a MIMO wiretap channel with t transmit antennas, r receive antennas at the legitimate receiver's terminal and e receive antennas at the eavesdropper's terminal:

$$\begin{aligned} \mathbf{Y} &= \mathbf{H}_r \mathbf{X} + \mathbf{N}_r \\ \mathbf{Z} &= \mathbf{H}_e \mathbf{X} + \mathbf{N}_e \end{aligned} \tag{2.99}$$

where $\mathbf{H}_r \in \mathbb{R}^{r \times t}$ and $\mathbf{H}_e \in \mathbb{R}^{e \times t}$ are the channel gain matrices associated respectively with the legitimate receiver and the eavesdropper, and $\mathbf{N}_r \in \mathbb{R}^{r \times 1}$ and $\mathbf{N}_e \in \mathbb{R}^{e \times 1}$ are independent additive Gaussian noise vectors with zero mean and identity covariance matrices $\mathbf{I}_r$ and $\mathbf{I}_e$.

The secrecy capacity of the MIMO Gaussian wiretap channel was characterized in [13,14] under the average total power constraint and in [15] under the more general matrix constraint. Here, we show that the secrecy capacity per unit cost under the quadratic cost function $b(\mathbf{x}) = \|\mathbf{x}\|^2$ can be achieved by orthogonal codes with constant pulse shape.

We make the following assumptions:

- the transmitted signal vector $\mathbf{X}$ with covariance matrix $\mathbf{K}_X \succeq 0$, satisfies the power constraint $\mathsf{tr}(\mathbf{K}_X) \leq \beta$.
- $\mathbf{H}_r$ and $\mathbf{H}_e$ are fixed channel gain matrices such that $\mathbf{H}_r^t\mathbf{H}_r \succ 0$, $\mathbf{H}_e^t\mathbf{H}_e \succ 0$.The channel gain matrices considered in this section are assumed to be fixed and known to all three terminals.

Theorem 6. *The secrecy capacity per unit cost of the Gaussian MIMO wiretap channel as given by (2.99), with a zero-cost input letter "0" is bounded from below as*

$$C_s \geq \lambda_t \tag{2.100}$$

where λ_t is the maximum eigenvalue of of the Matrix $\mathbf{M} := \mathbf{H}_r^t\mathbf{H}_r - \mathbf{H}_e^t\mathbf{H}_e$

Proof. A natural lower bound for C_s is given by (2.71), since the assumption of a degraded version of the wiretap channel gives a weaker channel as compared to the general wiretap channel:

$$C_s \geq \sup_{\mathbf{x}} \frac{D(P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}\|P_{\mathbf{Y}|\mathbf{X}=\mathbf{0}}) - D(P_{\mathbf{Z}|\mathbf{X}=\mathbf{x}}\|P_{\mathbf{Z}|\mathbf{X}=\mathbf{0}})}{\|\mathbf{x}\|^2} \tag{2.101}$$

The expression for the divergence between two Gaussian distributions is well known [16]

$$
\begin{aligned}
D(N(\mu_1, \mathbf{\Sigma}_1) \| N(\mu_0, \mathbf{\Sigma}_0)) = (\mu_1 - \mu_0)^t \mathbf{\Sigma}_0^{-1} (\mu_1 - \mu_0) \\
+ \log\det \mathbf{\Sigma}_0 - \log\det \mathbf{\Sigma}_1 \qquad (2.102) \\
+ \mathsf{tr}(\mathbf{\Sigma}_0^{-1} \mathbf{\Sigma}_1 - I)
\end{aligned}
$$

Use (2.102) to evaluate the right-hand side of (2.101)

$$
\begin{aligned}
\mathbf{C} &\geq \sup_{\mathbf{x}} \frac{\|\mathbf{H}_r \mathbf{x}\|^2 - \|\mathbf{H}_e \mathbf{x}\|^2}{\|\mathbf{x}\|^2} \\
&= \sup_{\mathbf{x}} \frac{\mathbf{x}^t (\mathbf{H}_r^t \mathbf{H}_r - \mathbf{H}_e^t \mathbf{H}_e) \mathbf{x}}{\|\mathbf{x}\|^2} \\
&= \sup_{\mathbf{x}} \frac{\mathbf{x}^t \mathbf{M} \mathbf{x}}{\|\mathbf{x}\|^2} \qquad (2.103)
\end{aligned}
$$

where $\mathbf{M} := \mathbf{H}_r^t \mathbf{H}_r - \mathbf{H}_e^t \mathbf{H}_e$ is a real and symmetric matrix of dimension t. The right-hand side of (2.103) is a maximization of a *Rayleigh quotient* which is achieved whenever $\mathbf{x}$ is proportional to the unit eigenvector associated with the maximum eigenvalue λ_t of the matrix $\mathbf{M}$ [17]. As a result:

$$
\mathbf{C} \geq \lambda_t \qquad (2.104)
$$

□

3. SECURE COMMUNICATION WITH KNOWN INTERFERENCE

3.1 Introduction

In information theory, an interesting and useful communication model is a state-dependent channel where the channel states are non-causally known at the transmitter as side information. Of particular importance is a discrete-time channel with real input and additive white Gaussian noise and interference, where the interference is non-causally known at the transmitter as side information.

Costa [18] was the first to study this communication scenario, which he whimsically coined as "writing on dirty paper." Based on an earlier result of Gel'fand and Pinsker [19], Costa [18] proved the surprising result that the capacity of writing on dirty paper is the *same* as that of writing on clean paper without interference. Since [18], dirty-paper coding has found a wide range of applications in digital watermarking and network communications, particularly involving broadcast scenarios.

Recent works [20] and [21] studied the problem of dirty-paper coding in the presence of an additional eavesdropper, which is a natural extension of Costa's dirty-paper channel to the secrecy communication setting. In this scenario, which we dub as "secret writing on dirty paper", the legitimate receiver channel is a dirty-paper channel of Costa. The signal received at the eavesdropper, on the other hand, is assumed to be a *degraded* version of the signal received at the legitimate receiver. An achievable secrecy rate was established based on a double-binning scheme and was shown to be the secrecy capacity of the channel under some channel parameter configurations [20, 21]. For the *general* channel parameter configuration, however, the secrecy capacity of the channel remains *unknown.*

In facing some challenging Gaussian network communication problems, recent advances [22, 23] in network information theory advocate a *deterministic* approach and seeks *approximate* characterization of the network capacity to within *finite* bits (regardless of the received signal-to-noise ratios). Motivated by the success of [22] and [23], we take a deterministic view and revisit the problem of wiretap channel with side information. A precise characterization of the secrecy capacity is obtained for a linear deterministic model, which naturally suggests a coding scheme which we show to achieve the secrecy capacity of the degraded Gaussian model to within half a bit.

3.2 Writing on Dirty Paper

3.2.1 Gaussian Model

Consider the dirty-paper channel of Costa [18], where the received signal $Y[i]$ at time index i is given by

$$Y[i] = hX[i] + gS[i] + N[i]. \tag{3.1}$$

Here, $X[i]$ is the channel input which is subject to a *unit* average power constraint, $N[i]$ and $S[i]$ are independent *standard* Gaussian noise and interference and are independently identically distributed (i.i.d.) across the time index i, and h and g are the (real) channel coefficients corresponding to the channel input and interference, respectively. The interference $S[i]$ is assumed to be non-causally known at the transmitter as side information. The channel coefficients h and g are fixed during communication and are assumed to be known at both transmitter and receiver.

The channel capacity, as shown by Costa [18], is given by

$$C = I(U; Y) - I(U; S)$$

where the input variable X is standard Gaussian and independent of the known interference S, and U is an auxiliary variable chosen as

$$U = hX + \frac{h^2}{h^2+1} gS. \tag{3.2}$$

For this choice of auxiliary-input variable pair (U, X),

$$I(U;Y) - I(U;S) = \frac{1}{2}\log(1+h^2)$$

which equals the capacity of the channel (3.1) when the interference $S[i]$ is also known at the receiver.

3.2.2 Linear Deterministic Model

Consider the linear deterministic model [23] for Costa's dirty-paper channel (3.1), where the received signal $Y[i]$ at time index i is given by

$$Y[i] = D^{q-n}X[i] \oplus D^{q-m}S[i]. \tag{3.3}$$

Here, $X[i]$ is the binary input vector of length $q = \max\{n, m\}$, $S[i]$ is the i.i.d. interference vector whose elements are i.i.d. Bernoulli-1/2, $D = [d_{j,k}]$ is the $q \times q$ down-shift matrix with elements

$$d_{j,k} = \begin{cases} 1 & \text{if } 2 \le j = k+1 \le q \\ 0 & \text{otherwise} \end{cases}$$

and n and m are the integer channel gains corresponding to the channel input and interference, respectively. The vector interference $S[i]$ is assumed to be non-causally known at the transmitter as side information. The channel gains n and m are fixed during communication and are assumed to be known at both transmitter and receiver.

Following the result of Gel'fand and Pinsker [19], the capacity of the linear deterministic dirty-paper channel (3.3) is given by

$$C = I(U;Y) - I(U;S)$$

where the input variable X is an i.i.d. Bernoulli-1/2 random vector and independent of S, and U is an auxiliary variable chosen as

$$U = Y = D^{q-n}X \oplus D^{q-m}S. \tag{3.4}$$

For this choice of the auxiliary-input variable pair (U, X),

$$\begin{aligned} I(U;Y) - I(U;S) &= I(Y;Y) - I(Y;S) \\ &= H(Y) - I(Y;S) \\ &= H(Y|S) \\ &= H(D^{q-n}X) \\ &= rank(D^{q-n}) \\ &= n \end{aligned}$$

which equals the capacity of the channel (3.3) when the interference $S[i]$ is also known at the receiver.

We emphasize that in (3.4), we may choose $U = Y$ only because Y here is a *deterministic* function of X and S. In fact, for *any* deterministic Gel'fand-Pinsker channel (not necessarily linear) where the channel output Y is a deterministic (bivariate) function of the channel input X and state S, $\max_{p(x|s)} H(Y|S)$ is the capacity of the channel when the channel state S is also known at the receiver. Thus, $U = Y$ is always an optimal choice for deterministic Gel'fand-Pinsker channels, a fact which was also observed in [24] recently.

3.2.3 Connections between the Gaussian and the Linear Deterministic Model

A quick comparison between the Gaussian (3.1) and the linear deterministic (3.3) models reveals the following equivalence relationship between these two models:

$$h \longleftrightarrow D^{q-n} \quad \text{and} \quad g \longleftrightarrow D^{q-m}. \tag{3.5}$$

Given this equivalence relationship, the optimal choice (3.4) of auxiliary variable U for the linear deterministic model (3.3) naturally suggests the following choice of auxiliary variable U for the Gaussian model (3.1):

$$U = hX + gS \tag{3.6}$$

where X is standard Gaussian and independent of S. Compared with the optimal choice (3.2), the choice (3.6) of auxiliary variable U is *suboptimal.* However, for this suboptimal choice of auxiliary-input variable pair (U, X),

$$\begin{aligned} I(U;S) &= \frac{1}{2}\log\left(1+\frac{g^2}{h^2}\right) \\ \text{and} \quad I(U;Y) &= \frac{1}{2}\log(1+h^2+g^2) \end{aligned}$$

giving an achievable rate

$$\begin{aligned} R &= [I(U;Y) - I(U;S)]^+ \\ &= \left[\frac{1}{2}\log\frac{(1+h^2+g^2)h^2}{h^2+g^2}\right]^+ \\ &\geq \left[\frac{1}{2}\log(h^2)\right]^+ \end{aligned}$$

which is always within half a bit of the actual channel capacity $C = \frac{1}{2}\log(1+h^2)$. Here, we denote $x^+ := \max\{0, x\}$ so that the achievable rates are always nonnegative.

The fact that the choice (3.6) of auxiliary variable U leads to an achievable rate which is always within half a bit of the dirty-paper channel capacity is well known (see [25] for example). However, it is interesting to see that such a choice comes up *naturally* in the context of the deterministic approach.

3.3 Secret Writing on Dirty Paper

Having understood how the linear deterministic model of [23] may be used to obtain an approximate characterization of the capacity of Costa's dirty-paper channel, next we shall extend the deterministic approach to the problem of secret writing on dirty paper.

3.3.1 Discrete Memoryless Model

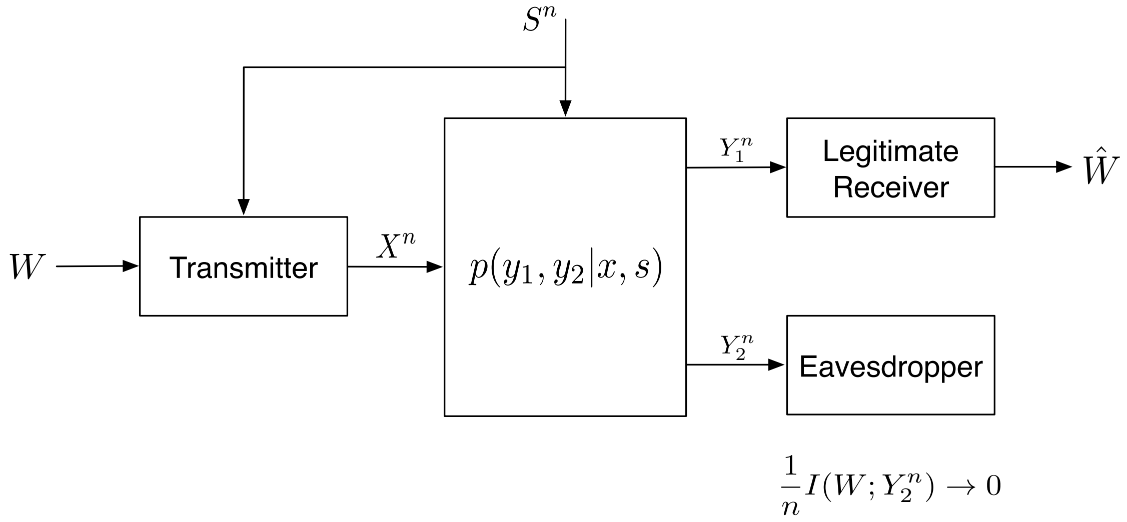

Figure 3.1: Wiretap channel with side information.

As illustrated in Fig. 3.1, consider a discrete-time memoryless wiretap channel with transition probability $p(y_1, y_2|x, s)$, where $X[i]$ is the channel input (at time index i), $S[i]$ is the channel state, and $Y_1[i]$ and $Y_2[i]$ are the received signals at the legitimate receiver and the eavesdropper, respectively. The channel state $S[i]$ is i.i.d. across the time index i and is assumed to be non-causally known at the transmitter as side information. The transmitter has a message W, which is intended for the legitimate receiver but needs to be kept asymptotically perfectly secret from the eavesdropper. Following the classical works [2] and [3], it is required that

$$\frac{1}{n}I(W; Y_2^n) \to 0 \tag{3.7}$$

in the limit as the block length $n \to \infty$, where $Y_2^n := (Y_2[1], \dots, Y_2[n])$. The secrecy capacity C_s is defined as the *largest* secrecy rate that can be achieved by a coding scheme.

Chen and Vinck [21] derived a single-letter lower bound on the secrecy capacity (an achievable secrecy rate), which can be written as

$$\begin{aligned} C_s \geq \max\nolimits_{p(u,x|s)} \min \{ & I(U;Y_1) - I(U;S), \\ & I(U;Y_1) - I(U;Y_2)\} \end{aligned} \tag{3.8}$$

where U is an auxiliary variable satisfying the Markov chain $U \rightarrow (X,S) \rightarrow (Y_1,Y_2)$.

We also have the following simple upper bound on the secrecy capacity.

Proposition 1. *The secrecy capacity C_s of a discrete memoryless wiretap channel $p(y_1,y_2|x,s)$ with channel state S non-causally known at the transmitter as side information can be bounded from above as*

$$C_s \leq \max_{p(x|s)} \min \{I(X;Y_1|S), I(X,S;Y_1|Y_2)\}\,. \tag{3.9}$$

Note that $\max_{p(x|s)} I(X;Y_1|S)$ is an upper bound on the Shannon capacity of the legitimate receiver channel by giving the channel state S to the legitimate receiver, and $\max_{p(x|s)} I(X,S;Y_1|Y_2)$ is an upper bound on the secrecy capacity of the wiretap channel by allowing the transmit message W to be encoded by the channel state S (i.e., fully action-dependent state [26]) and by giving the received signal Y_2 to the legitimate receiver. Here, a simple single-letterization technique of Willems [27] allows the maximizations to be moved outside the minimization. See Appendix C for the details of the proof.

For *semi-deterministic* channels where the channel output at the legitimate receiver is a deterministic (bivariate) function of the channel input and state, the lower (3.8) and the upper (3.9) bounds coincide, leading to a *precise* characterization of the secrecy capacity. The result is summarized in the following theorem.

Theorem 7. *Consider a discrete memoryless wiretap channel $p(y_1, y_2|x, s)$ with channel state S non-causally known at the transmitter as side information. If the received signal Y_1 at the legitimate receiver is a* deterministic *function of the channel input X and state S, i.e., $Y_1 = f(X, S)$ for some bivariate function f, the secrecy capacity C_s of the channel is given by*

$$C_s = \max_{p(x|s)} \min \left\{ H(Y_1|S), H(Y_1|Y_2) \right\}. \tag{3.10}$$

Proof. The fact that

$$C_s \geq \max_{p(x|s)} \min \left\{ H(Y_1|S), H(Y_1|Y_2) \right\}$$

follows from the lower bound (3.8) by setting $U = Y_1$ (we may do so only because here Y_1 is a deterministic function of X and S), which gives

$$I(U; Y_1) - I(U; S) = H(Y_1) - I(Y_1; S) = H(Y_1|S)$$

and similarly

$$I(U; Y_1) - I(U; Y_2) = H(Y_1) - H(Y_1|Y_2) = H(Y_1|Y_2).$$

The converse part of the theorem follows from the upper bound (3.9) and the fact that Y_1 is a deterministic function of (X, S), so we have

$$I(X; Y_1|S) = H(Y_1|S) - H(Y_1|X, S) = H(Y_1|S)$$

and

$$I(X, S; Y_1|Y_2) = H(Y_1|Y_2) - H(Y_1|X, S, Y_2) = H(Y_1|Y_2).$$

This completes the proof of the theorem. □

Note that when the channel state S is deterministic, a semi-deterministic wiretap channel with side information reduces to a regular semi-deterministic wiretap channel without side information. In this case, let S be a constant in (3.10) and we have

$$C_s = \max_{p(x)} \min \{H(Y_1), H(Y_1|Y_2)\} = \max_{p(x)} H(Y_1|Y_2)$$

which recovered the result of [28] on the secrecy capacity of the semi-deterministic wiretap channel (without side information).

3.3.2 Linear Deterministic Model

Next, let us use the result of Theorem 7 to determine the secrecy capacity of a linear deterministic wiretap channel with side information. In this model, the received signals (at time index i) at the legitimate receiver and the eavesdropper are given by

$$\begin{aligned} Y_1[i] &= D^{q-n_1}X[i] \oplus D^{q-m_1}S[i] \\ Y_2[i] &= D^{q-n_2}X[i] \oplus D^{q-m_2}S[i] \end{aligned} \tag{3.11}$$

where $X[i]$ is the binary input vector of length $q = \max\{n_1, n_2, m_1, m_2\}$, $S[i]$ is the i.i.d. vector interference whose elements are i.i.d. Bernoulli-1/2, D is the $q \times q$ down-shift matrix, and n_1, n_2, m_1 and m_2 are the integer channel gains. The vector interference $S[i]$ is assumed to be non-causally known at the transmitter as side information. The channel gains n_1, n_2, m_1 and m_2 are fixed during communication and are assumed to be known at all terminals.

The following theorem provides an *explicit* characterization of the secrecy capacity of the linear deterministic wiretap channel (3.11) with side information.

Theorem 8. *The secrecy capacity C_s of the linear deterministic wiretap channel* (3.11) *with side information is given by*

$$C_s = \begin{cases} n_1, & \text{if } n_1 - m_1 \neq n_2 - m_2, \\ & \quad n_1 \leq m_1 \text{ or } n_2 \leq m_2 \\ \max\{m_1, n_1 - n_2 + m_2\}, & \\ & \text{if } n_1 - m_1 \neq n_2 - m_2, \\ & \quad n_1 > m_1 \text{ and } n_2 > m_2 \\ (n_1 - n_2)^+, & \text{if } n_1 - m_1 = n_2 - m_2. \end{cases} \tag{3.12}$$

To prove Theorem 8, let us first prove the following proposition.

Proposition 2. *The secrecy capacity C_s of the linear deterministic wiretap channel* (3.11) *with side information is given by*

$$C_s = \min\left\{ n_1, rank\left(\begin{bmatrix} A \\ B \end{bmatrix} \right) - rank(B) \right\} \tag{3.13}$$

where

$$\begin{aligned} A &:= \begin{bmatrix} D^{q-n_1} & D^{q-m_1} \end{bmatrix} \\ \text{and} \quad B &:= \begin{bmatrix} D^{q-n_2} & D^{q-m_2} \end{bmatrix}. \end{aligned} \tag{3.14}$$

Proof. To prove (3.13), we shall show that for the linear deterministic model (3.11), both $H(Y_1|S)$ and $H(Y_1|Y_2)$ are simultaneously maximized when X is an i.i.d. Bernoulli-1/2 random vector and independent of S.

First,

$$\begin{aligned} H(Y_1|S) &= H(D^{q-n_1}X|S) \\ &\leq H(D^{q-n_1}X) \\ &\leq rank(D^{q-n_1}) \\ &= n_1 \end{aligned} \tag{3.15}$$

where the equalities hold when X is an i.i.d. Bernoulli-$1/2$ random vector and independent of S. To show that $H(Y_1|Y_2)$ is also maximized when X is an i.i.d. Bernoulli-$1/2$ random vector and independent of S, we shall need the following technical lemma, which can be proved using a counting argument as provided in Appendix D.

Lemma 2. *For any matrices A and B in $\mathbb{F}_2$ (Galois field of size 2) that have the same number of columns,*

$$\max H(AZ|BZ) = rank\left(\begin{bmatrix} A \\ B \end{bmatrix}\right) - rank(B) \tag{3.16}$$

where the maximization is over all possible binary random vector Z. The maximum is achieved when Z is an i.i.d. Bernoulli-$1/2$ random vector.

Now let

$$Z := \begin{bmatrix} X \\ S \end{bmatrix}.$$

By Lemma 2,

$$
\begin{aligned}
H(Y_1|Y_2) &= H(AZ|BZ) \\
&\leq rank\left(\begin{bmatrix} A \\ B \end{bmatrix}\right) - rank(B)
\end{aligned} \tag{3.17}
$$

where the equality holds when X is an i.i.d. Bernoulli-1/2 random vector and independent of S.

Substituting (3.15) and (3.17) into (3.10) completes the proof of the proposition.

□

Given Proposition 2, the explicit characterization (3.12) of the secrecy capacity C_s can be obtained from (3.13) by evaluating the rank of the matrices

$$
\begin{bmatrix} A \\ B \end{bmatrix}
$$

and B. The details of the evaluation process are provided in Appendix E.

3.3.3 Degraded Gaussian Model

Finally, let us consider the Gaussian wiretap channel where the received signals (at time index i) at the legitimate receiver and the eavesdropper are given by

$$
\begin{aligned}
Y_1[i] &= h_1 X[i] + g_1 S[i] + N_1[i] \\
Y_2[i] &= h_2 X[i] + g_2 S[i] + N_2[i].
\end{aligned} \tag{3.18}
$$

Here, $X[i]$ is the channel input which is subject to a *unit* average power constraint, $N_k[i]$, $k = 1, 2$ and $S[i]$ are independent *standard* Gaussian noise and interference and are i.i.d. across the time index i, and h_1, h_2, g_1 and g_2 are the (real) channel coefficients. The interference $S[i]$ is assumed to be non-causally known at the transmitter as side information. The channel coefficients h_1, h_2, g_1 and g_2 are fixed during communication and are assumed to be known at all terminals.

A single-letter expression for an achievable secrecy rate was given in (3.8), which involves an auxiliary variable U. However, it is *not* clear what would be a reasonable choice of U, letting alone finding an optimal one that maximizes the achievable secrecy rate expression (3.8). On the other hand, for the linear deterministic model (3.11), it is clear from Theorem 7 and Proposition 2 that the following choice of auxiliary variable U is optimal:

$$U = Y_1 = D^{q-n_1}X \oplus D^{q-m_1}S \tag{3.19}$$

where X is an i.i.d. Bernoulli-1/2 random vector and independent of S.

Based on the equivalence relationship (3.5) between the Gaussian and the linear deterministic model and the success of Sec. 3.2 for Costa's dirty-paper channel, the optimal choice (3.19) of auxiliary variable U for the linear deterministic model (3.11) suggests the following choice of auxiliary variable U for the Gaussian model (3.18):

$$U = h_1X + g_1S \tag{3.20}$$

where X is standard Gaussian and independent of S. For this choice of auxiliary-input variable pair (U, X),

$$\begin{aligned} I(U;S) &= \frac{1}{2}\log\left(1+\frac{g_1^2}{h_1^2}\right) && (3.21) \\ I(U;Y_1) &= \frac{1}{2}\log(1+h_1^2+g_1^2) && (3.22) \\ \text{and} \quad I(U;Y_2) &= \frac{1}{2}\log\frac{(h_1^2+g_1^2)(1+h_2^2+g_2^2)}{h_1^2+g_1^2+(h_1g_2-h_2g_1)^2} && (3.23) \end{aligned}$$

giving

$$I(U;Y_1) - I(U;S) = \frac{1}{2}\log\frac{(1+h_1^2+g_1^2)h_1^2}{h_1^2+g_1^2}$$

and

$$\begin{aligned} & I(U;Y_1) - I(U;Y_2) \\ & \quad = \frac{1}{2}\log\frac{(1+h_1^2+g_1^2)[h_1^2+g_1^2+(h_1g_2-h_2g_1)^2]}{(h_1^2+g_1^2)(1+h_2^2+g_2^2)}. \end{aligned}$$

By the single-letter achievable secrecy rate expression (3.8),

$$\begin{aligned} R_s = \Bigg(\min\Bigg\{ & \tfrac{1}{2}\log\tfrac{(1+h_1^2+g_1^2)h_1^2}{h_1^2+g_1^2}, \\ & \tfrac{1}{2}\log\tfrac{(1+h_1^2+g_1^2)[h_1^2+g_1^2+(h_1g_2-h_2g_1)^2]}{(h_1^2+g_1^2)(1+h_2^2+g_2^2)}\Bigg\}\Bigg)^+ \end{aligned} \tag{3.24}$$

is an achievable secrecy rate for the Gaussian wiretap channel (3.18) with side information.

Following the works [20] and [21], below we focus on the special case where

$$h_2 = \beta h_1 \quad \text{and} \quad g_2 = \beta g_1 \tag{3.25}$$

for some $|\beta| \leq 1$. Note that the secrecy capacity of the channel (3.18) does *not* depend on the correlation between the additive Gaussian noise $N_1[i]$ and $N_2[i]$, so we may write

$$N_2[i] = \beta N_1[i] + N[i]$$

where $N[i]$ is Gaussian with zero mean and variance $1-\beta^2$ and independent of $N_1[i]$. Thus, for the special case of (3.25), the channel (3.18) can be equivalently written as

$$\begin{aligned} Y_1[i] &= h_1X[i] + g_1S[i] + N_1[i] \\ Y_2[i] &= \beta Y_1[i] + N[i] \end{aligned} \tag{3.26}$$

i.e., the received signal $Y_2[i]$ at the eavesdropper is *degraded* with respect to the the received signal $Y_1[i]$ at the legitimate receiver.

Following [18], an interesting interpretation of the degraded Gaussian model (3.26) is “secret writing on dirty paper.” In this scenario, a user intends to convey (to a legitimate receiver) a confidential message on a piece of paper with preexisting dirt on it. The legitimate receive has access to the *original* paper with the message written on it and hence can decode the intended message. On the other hand, the eavesdropper can only access a *noisy* copy of the original paper, from which essentially no information on the conveyed message can be inferred.

Next, we show that for the degraded Gaussian model (3.26), the achievable secrecy rate (3.24) is always within half a bit of the secrecy capacity. The result is summarized in the following theorem.

Theorem 9. *For the degraded Gaussian wiretap channel* (3.26) *with side information, the secrecy capacity C_s can be bounded as*

$$\begin{gathered} \left(\min\left\{\tfrac{1}{2}\log\tfrac{(1+h_1^2+g_1^2)h_1^2}{h_1^2+g_1^2}, \tfrac{1}{2}\log\tfrac{1+h_1^2+g_1^2}{1+\beta^2(h_1^2+g_1^2)}\right\}\right)^+ \leq \\ C_s \leq \min\left\{\tfrac{1}{2}\log(1+h_1^2), \tfrac{1}{2}\log\tfrac{2(h_1^2+g_1^2)+1}{2\beta^2(h_1^2+g_1^2)+1}\right\}. \end{gathered} \tag{3.27}$$

Moreover, the lower bound here is always within half a bit of the upper bound.

Proof. The lower bound in (3.27) follows from (3.24) and the degradedness assumption (2.22). To prove the upper bound, note that for any input variable X such that

$\mathbb{E}[X^2] \leq 1$ we have

$$\begin{aligned} I(X;Y_1|S) &= h(Y_1|S) - h(Y_1|X,S) \\ &= h(h_1X + N_1|S) - h(N_1) \\ &\leq h(h_1X + N_1) - h(N_1) \\ &\leq \frac{1}{2}\log(1 + h_1^2\mathrm{Var}(X)) \\ &\leq \frac{1}{2}\log(1 + h_1^2). \end{aligned} \tag{3.28}$$

Furthermore,

$$\begin{aligned} & I(X,S;Y_1|Y_2) \\ &\quad = h(Y_1|Y_2) - h(Y_1|X,S,Y_2) \\ &\quad = h(Y_1|\beta Y_1 + N) - h(N_1|\beta N_1 + N) \\ &\quad = h(Y_1|\beta Y_1 + N) - \frac{1}{2}\log\left(2\pi e(1-\beta^2)\right). \end{aligned} \tag{3.29}$$

By an inequality of Thomas [29, Lemma 1] and the independence between Y_1 and N,

$$h(Y_1|\beta Y_1 + N) \leq \frac{1}{2}\log\frac{2\pi e\mathrm{Var}(Y_1)(1-\beta^2)}{\beta^2\mathrm{Var}(Y_1) + (1-\beta^2)}. \tag{3.30}$$

Note that the right-hand side of (3.30) is a monotone increasing function of $\mathrm{Var}(Y_1)$, which can be bounded from above as

$$\begin{aligned} \mathrm{Var}(Y_1) &= \mathrm{Var}(h_1X + g_1S + N_1) \\ &= \mathrm{Var}(h_1X + g_1S) + 1 \\ &\leq 2\left(\mathrm{Var}(h_1X) + \mathrm{Var}(g_1S)\right) + 1 \\ &\leq 2h_1^2 + 2g_1^2 + 1. \end{aligned}$$

Hence,

$$\begin{aligned} & h(Y_1|\beta Y_1 + N) \\ & \quad \leq \frac{1}{2}\log\frac{2\pi e(2h_1^2+2g_1^2+1)(1-\beta^2)}{\beta^2(2h_1^2+2g_1^2+1)+(1-\beta^2)} \\ & \quad = \frac{1}{2}\log\frac{2\pi e(2h_1^2+2g_1^2+1)(1-\beta^2)}{2\beta^2(h_1^2+g_1^2)+1}. \end{aligned} \tag{3.31}$$

Substituting (3.31) into (3.29), we have

$$I(X,S;Y_1|Y_2) \leq \frac{1}{2}\log\frac{2(h_1^2+g_1^2)+1}{2\beta^2(h_1^2+g_1^2)+1}. \tag{3.32}$$

Further substituting (3.28) and (3.32) into (3.9) establishes the upper bound in (3.27).

To show that the lower bound is always within half a bit of the upper bound, let us define

$$\begin{aligned} a &:= \frac{1}{2}\log(1+h_1^2) \\ b &:= \frac{1}{2}\log\frac{2(h_1^2+g_1^2)+1}{2\beta^2(h_1^2+g_1^2)+1} \\ c &:= \frac{1}{2}\log\frac{(1+h_1^2+g_1^2)h_1^2}{h_1^2+g_1^2} \\ \text{and} \quad d &:= \frac{1}{2}\log\frac{1+h_1^2+g_1^2}{1+\beta^2(h_1^2+g_1^2)}. \end{aligned}$$

We shall consider the following two cases separately.

Case 1: $h_1^2 < 1$. In this case,

$$a = \frac{1}{2}\log(1+h_1^2) < \frac{1}{2}$$

and the gap between the upper and the lower bound can be bounded from above as

$$\min\{a,b\} - (\min\{c,d\})^+ \leq \min\{a,b\} \leq a < \frac{1}{2}. \tag{3.33}$$

Case 2: $h_1^2 \geq 1$. In this case,

$$\begin{aligned}
a - c &= \frac{1}{2}\log(1+h_1^2) - \frac{1}{2}\log\frac{(1+h_1^2+g_1^2)h_1^2}{h_1^2+g_1^2} \\
&\leq \frac{1}{2}\log(1+h_1^2) - \frac{1}{2}\log(h_1^2) \\
&= \frac{1}{2}\log\left(1+\frac{1}{h_1^2}\right) \\
&\leq \frac{1}{2}.
\end{aligned} \tag{3.34}$$

Note that for any channel parameters h_1, g_1 and β,

$$\begin{aligned}
b - d &= \frac{1}{2}\log\frac{2(h_1^2+g_1^2)+1}{2\beta^2(h_1^2+g_1^2)+1} - \frac{1}{2}\log\frac{1+h_1^2+g_1^2}{1+\beta^2(h_1^2+g_1^2)} \\
&= \frac{1}{2}\log\left[\frac{2(h_1^2+g_1^2)+1}{1+h_1^2+g_1^2}\cdot\frac{1+\beta^2(h_1^2+g_1^2)}{2\beta^2(h_1^2+g_1^2)+1}\right] \\
&\leq \frac{1}{2}\log\frac{2(h_1^2+g_1^2)+1}{1+h_1^2+g_1^2} \\
&\leq \frac{1}{2}\log\frac{2(h_1^2+g_1^2)+2}{1+h_1^2+g_1^2} \\
&= \frac{1}{2}
\end{aligned} \tag{3.35}$$

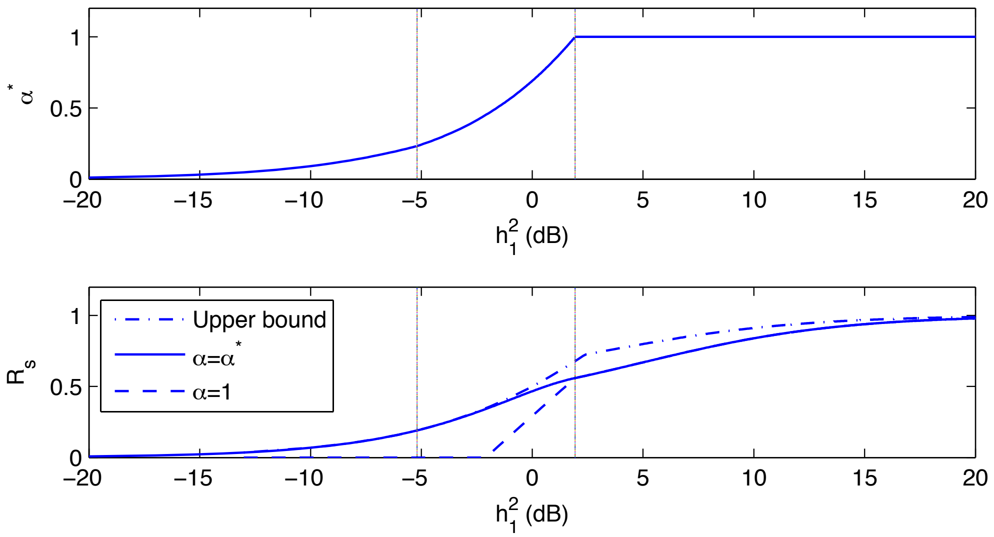

Figure 3.2: A numerical comparison between the achievable secrecy rates for choosing $\alpha = \alpha^*$ and $\alpha = 1$ in (3.38). Both α^* and the achievable secrecy rate R_s are plotted as a function of h_1^2, while g_1 and β are fixed to be 1 and 0.5, respectively.

and for any real scalers a, b, c and d,

$$\begin{aligned} & \min\{a,b\} - (\min\{c,d\})^+ \\ & \quad \leq \min\{a,b\} - \min\{c,d\} \\ & \quad = \max\left\{\min\{a,b\} - c, \min\{a,b\} - d\right\} \\ & \quad = \max\{a-c, b-d\}. \end{aligned} \tag{3.36}$$

Substituting (3.34) and (3.35) into (3.36), we have

$$\min\{a,b\} - (\min\{c,d\})^+ \leq \max\left\{\frac{1}{2}, \frac{1}{2}\right\} = \frac{1}{2}. \tag{3.37}$$

Combining the above two cases proves that the lower bound in (3.27) is always within half a bit of the upper bound. This completes the proof of the theorem. □

Finally, we note that the work [20] considered, as a *heuristic* choice, the auxiliary variable

$$U = h_1 X + \alpha g_1 S \tag{3.38}$$

where X is standard Gaussian and independent of S, and α is chosen to maximize the achievable secrecy rate. A closed-form expression for the maximizing α can be written as

$$\alpha^* = \begin{cases} \frac{h_1^2}{h_1^2+1}, & \text{if } 0 \leq h_1^2 < h_{1L}^2 \\ \frac{\beta^2 h_1^2\left(|g_1|+\sqrt{h_1^2+g_1^2+1/\beta^2}\right)}{|g_1|(1+\beta^2 h_1^2)}, & \text{if } h_{1L}^2 \leq h_1^2 < h_{1H}^2 \\ 1, & \text{if } h_1^2 \geq h_{1H}^2 \end{cases}$$

where

$$\begin{aligned} h_{1L}^2 &= \left(-\frac{g_1^2}{2} - 1 + \frac{|g_1|}{2}\sqrt{g_1^2 + \frac{4}{\beta^2} - 4}\right)^+ \\ \text{and} \quad h_{1H}^2 &= -\frac{g_1^2}{2} + \frac{|g_1|}{2}\sqrt{g_1^2 + \frac{4}{\beta^2}}. \end{aligned}$$

Thus, for $h_1^2 \geq h_{1H}^2$, the heuristic choice (3.38) with the maximizing α coincides with the choice $U = h_1 X + g_1 S$ suggested by the linear deterministic model.

A numerical comparison between the achievable secrecy rates for choosing $\alpha = \alpha^*$ and $\alpha = 1$ in (3.38) as well as the upper bound in (3.27) is provided in Figure 3.2. As we can see, when h_1^2 (which represents the received signal-to-noise ratio at the legitimate receiver) is small, the choice $\alpha = 1$ (as suggested by the linear deterministic model) can be very suboptimal in maximizing the achievable secrecy rate. However, in this case, the secrecy capacity of the channel is also small, so the achievable secrecy rate given by the suboptimal choice $\alpha = 1$ remains within half a bit of the secrecy capacity. For small h_1^2, substantial improvement to the achievable secrecy rate can be made by optimizing over α. In fact, when $h_1^2 \leq h_{1L}^2$, the achievable secrecy rate given $\alpha = \frac{h_1^2}{1+h_1^2}$ coincides with the upper bound and hence gives the *exact* secrecy capacity of the channel. When h_1^2 is large, the maximizing α approaches 1 (it is exactly equal to 1 when $h_1^2 \geq h_{1H}^2$), and both choices lead to achievable secrecy rates which are within half a bit of the secrecy capacity.

3.4 Secret-Key Agreement via Dirty-Paper Coding

A different but closely related communication scenario is *secret-key agreement* via dirty-paper coding, which was first considered in [30]. In this setting, the channel model is exactly the same as that for secret writing on dirty paper. The difference is in the goal of communication. For secret writing on dirty paper, the goal is to convey to the legitimate receiver a secret message W, which is pre-chosen and hence is *independent* of the known interference $\{S[i]\}$. For secret-key agreement, the goal is to establish, between the transmitter and the legitimate receiver, an agreement on a secret key K, which must be kept asymptotically perfectly secret from the eavesdropper, i.e.,

$$\frac{1}{n} I(K; Y_2^n) \to 0$$

in the limit as the block length $n \to \infty$. The secret-key capacity C_K is defined as the largest entropy rate $(1/n) \log H(K)$ that can be achieved by a coding scheme. Unlike the problem of secret writing on dirty paper, the secret key K can be potentially *correlated* with the known interference $\{S[i]\}$. Hence, the secret-key capacity C_K is at least as large as the secrecy capacity C_s for the same wiretap channel.

For a general discrete memoryless wiretap channel with side information, the secret-key capacity C_K is unknown. The following lower and upper bounds were established in [30]:

$$\max_{p(u,x|s)} [I(U;Y_1) - I(U;Y_2)] \leq C_K \leq \max_{p(x|s)} I(X,S;Y_1|Y_2) \tag{3.39}$$

where U is an auxiliary variable satisfying the Markov chain $U \to (X, S) \to (Y_1, Y_2)$ and such that

$$I(U; Y_1) - I(U; S) \geq 0. \tag{3.40}$$

For semi-deterministic wiretap channels where the received signal Y_1 at the legitimate receiver is a deterministic bivariate function of the channel input X and state S, the lower bound in (3.39) with the choice of auxiliary variable $U = Y_1$ coincides with the upper bound, giving an *exact* characterization of the secret-key capacity

$$C_K = \max_{p(x|s)} H(Y_1|Y_2). \tag{3.41}$$

Note here that the choice $U = Y_1$ always satisfies the constraint (3.40).

For the linear deterministic wiretap channel (3.11) with side information, by Lemma 2 the conditional entropy $H(Y_1|Y_2)$ is maximized when the input variable X is standard Gaussian and independent of S. By the equivalence relationship (3.5) between the linear deterministic and the Gaussian model, this suggests the following choice of auxiliary variable U for the degraded Gaussian model (3.26):

$$U = h_1 X + g_1 S \tag{3.42}$$

where X is standard Gaussian and independent of S, as long as (3.40) is satisfied. Substituting (3.42), (3.21)–(3.23), and the degradedness assumption (2.22) into (3.39) and (3.40), we have the following lower and upper bounds on the secret-key capacity C_K of the degraded Gaussian model (3.26):

$$\frac{1}{2} \log \frac{1 + h_1^2 + g_1^2}{1 + \beta^2(h_1^2 + g_1^2)} \leq C_K \leq \frac{1}{2} \log \frac{2(h_1^2 + g_1^2) + 1}{2\beta^2(h_1^2 + g_1^2) + 1} \tag{3.43}$$

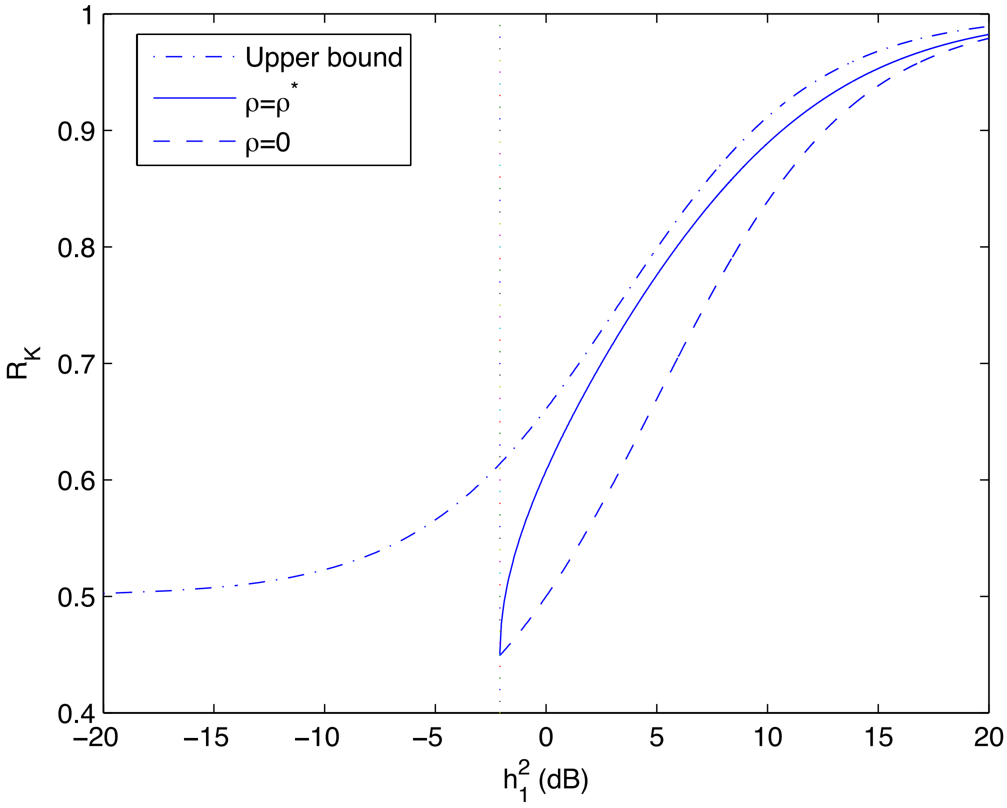

Figure 3.3: A numerical comparison between the achievable secret-key rates for choosing $\rho = \rho^*$ and $\rho = 0$ in (3.42). The achievable secret-key rate R_K are plotted as a function of h_1^2, while g_1 and β are fixed to be 1 and 0.5, respectively.

for all channel coefficients h_1 and g_1 such that[1]

$$h_1^2 \geq h_{1T}^2 := -\frac{g_1^2}{2} + \frac{|g_1|}{2}\sqrt{g_1^2 + 4}. \tag{3.44}$$

By (3.35), the lower bound in (3.43) is always within half a bit of the upper bound.

We mention here that [30] also considered, as a heuristic choice, the auxiliary variable U of form (3.42) where X is standard Gaussian. However, instead of choosing X to be independent of S as suggested by the linear deterministic model, [30] consid-

[1] The upper bound is valid for all channel parameters. Due to the constraint (3.40), when $h_1^2 < h_{1T}^2$ the linear deterministic model does not appear to provide any insight on how to choose the auxiliary variable U for the degraded Gaussian model.

ered X which is correlated with S and with correlation coefficient $\rho = \mathbb{E}[XS] = \rho^*$, where

$$\rho^* = \sqrt{1 - \frac{h_1^2 + g_1^2}{(1 + h_1^2 + g_1^2)h_1^2}} \cdot sgn(h_1 g_1). \tag{3.45}$$

Here, $sgn(x)$ denotes the sign of real scalar x. It is straightforward to verify that condition (3.44) guarantees the existence of ρ^*. A numerical comparison between the achievable secret-key rates for choosing the correlation coefficient $\rho = \rho^*$ and $\rho = 0$ as well as the upper bound in (3.43) is provided in Figure 3.3. As we can see, even though the choice $\rho = 0$ is suboptimal in maximizing the achievable secret-key rate, both choices lead to achievable secret-key rates that are within half a bit of the secret-key capacity for $h_1^2 \geq h_{1T}^2$.

4. CONCLUSION

The first part of this dissertation introduced the concept of secrecy capacity per unit cost to study cost-efficient wide-band secrecy communication. For degraded memoryless stationary wiretap channels with a zero-cost input letter, it was shown that an orthogonal coding scheme with randomized pulse position and constant pulse shape achieves the secrecy capacity per unit cost. For general memoryless stationary wiretap channels, orthogonal coding schemes with constant pulse shape may be suboptimal: further randomization of the pulse shape can strictly improve the achievable secrecy rate per unit cost. The problem whether orthogonal codes can achieve the secrecy capacity per unit cost of the *general* memoryless stationary wiretap channel with a zero-cost input letter remains open.

In the second part of this dissertation, a deterministic view was taken and used to revisit the problem of wiretap channel with side information. A precise characterization of the secrecy capacity was obtained for a linear deterministic model, which naturally suggests a coding scheme to achieve the secrecy capacity of the degraded Gaussian model (dubbed as "secret writing on dirty paper") to within half a bit. The linear deterministic model twas used o provide approximate characterization of Gaussian network capacity, an approach which has become increasingly popular in information theory literature. However, in this dissertation, the use of this method is somewhat different from most of the practices along this line of research. In literature, a common practice has been to first gain "insight" from the capacity-achieving scheme for the linear deterministic model and then translate the success to the Gaussian model at the *scheme* level. Such translations are more art than science. For the considered problems, the translation of success from the linear deterministic model to the Gaussian model was done at the level of a *single-letter description* of channel capacity and hence was much more systematic. A suggested line of research would focus on understanding to what extent this method can be applied to more complex network communication scenarios.

REFERENCES

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[2] A. D. Wyner, “The wire-tap channel,” *Bell Syst. Tech. Journal*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[3] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” *IEEE Trans. Inform. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.

[4] S. Verdú, “On channel capacity per unit cost,” *IEEE Trans. Inform. Theory*, vol. 36, no. 5, pp. 1019–1030, Sept. 1990.

[5] M. El-Halabi, T. Liu, and C. Georghiades, “On secrecy capacity per unit cost,” In Proceedings of *2009 IEEE International Symposium on Information Theory (ISIT)*, Seoul, Korea, July 2009.

[6] M. El-Halabi, T. Liu, C. Georghiades, and S. Shamai (Shitz), “Secret writing on dirty paper: A deterministic view,” *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3419–3429, June 2012.

[7] T. Liu and P. Viswanath, ”Opportunistic orthogonal writing on dirty paper,” *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1828–1846, May 2006.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York, NY: Academic, 1981.

[9] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations," *Ann. Math. Statistics*, vol. 23, no. 4, pp. 493–507, Dec. 1952.

[10] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Ann. Probability*, vol. 3, no. 1, pp. 146–158, Feb. 1975.

[11] S. Kullback, *Information Theory and Statistics*. New York, NY: Dover, 1968.

[12] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inform. Control*, vol. 2, no. 2, pp. 102–112, June 1959.

[13] A. Khisti and G. W. Wornell, "The MIMOME channel," in *Proc. 45th Annual Allerton Conf. Comm., Contr., Computing*, Monticello, IL, Sept. 2007.

[14] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," in *Proc. IEEE Int. Symp. Inform. Theory*, Toronto, Canada, July 2008, pp. 524–528.

[15] T. Liu and S. Shamai (Shitz), "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2547–2553, June 2009.

[16] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, pp. 1319–1343, June 2002.

[17] G. Golub and C. F. V. Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.

[18] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[19] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[20] C. Mitrpant, A. J. H. Vinck, and Y. Luo, "An achievable region for the Gaussian wiretap channel with side information," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2181–2190, May 2006.

[21] Y. Chen and A. J. H. Vinck, "Wiretap channel with side information," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 395–402, January 2008.

[22] R. H. Etkin, D. N. C. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Transactions on Information Theory*, vol. 54, no, 12, pp. 5534–5562, December 2008.

[23] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.

[24] R. Khosravi-Farsani and F. Marvast, "Capacity bounds for multiuser channels with non-causal channel state information at the transmitters," in *Proc. IEEE Int. Symp. Information Theory*, 2011, submitted for publication. Available at http://arxiv.org/abs/1102.3410.

[25] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[26] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5396–5411, November 2010.

[27] F. M. J. Willems, “An information theoretic approach to information embedding,” in *Proc. 21st Symp. Inf. Theory Benelux*, pp. 255–260, Wassenaar, The Netherlands, May 2000.

[28] J. Grubb, S. Vishwanath, Y. Liang, and H. V. Poor, “Secrecy capacity of semi-deterministic wire-tap channels,” in *Proc. IEEE Inf. Theory Workshop on Inf. Theory for Wireless Networks*, pp. 1–4, Bergen, Norway, July 2007.

[29] J. A. Thomas, “Feedback can at most double Gaussian multiple access channel capacity,” *IEEE Transactions on Information Theory*, vol. IT-33, no. 5, pp. 711–716, September 1987.

[30] A. Khisti, S. N. Diggavi, and G. W. Wornell, “Secret-key agreement over wiretap channels with random state parameters,” *IEEE Trans. Inf. For. Security*, vol. 6, no. 3, pp. 672–681, September 2011.

APPENDIX A

PROOF OF REMARK 2

Suppose that there exists a positive real β and a positive integer k_0 such that for any integer $k \geq k_0$ an $(n, w_0, k\beta, \epsilon, k\beta\epsilon)$ code can be found for which

$$\frac{\log w_0}{k\beta} > R_s - \frac{\delta}{2} \tag{A.1}$$

for any $\delta > 0$. Let

$$\nu_0 := \max\left(k_0, \frac{2R_s}{\delta}\right)\beta. \tag{A.2}$$

When $\nu = k\beta$ for some $k \geq n_0$, we have

$$\frac{\log w_0}{\nu} > R_s - \frac{\delta}{2} > R_s - \delta \tag{A.3}$$

for the $(n, w_0, k\beta, \epsilon, k\beta\epsilon)$ code.

When $k\beta < \nu < (k+1)\beta$ for some $k \geq k_0$ and such that $\nu \geq \nu_0$, we have

$$k > \frac{\nu}{\beta} - 1 \geq \frac{\nu_0}{\beta} - 1 \geq \frac{2R_s}{\delta} - 1. \tag{A.4}$$

In this case, the $(n, w_0, k\beta, \epsilon, k\beta\epsilon)$ is also an $(n, w_0, \nu, \epsilon, \nu\epsilon)$ code for which

$$\begin{aligned}
\frac{\log w_0}{\nu} &> \left(R_s - \frac{\delta}{2}\right)\frac{k\beta}{\nu} && \text{(A.5)}\\
&> \left(R_s - \frac{\delta}{2}\right)\frac{k}{k+1} && \text{(A.6)}\\
&> \left(R_s - \frac{\delta}{2}\right)\left(1 - \frac{\delta}{R_s}\right) && \text{(A.7)}\\
&> R_s - \delta && \text{(A.8)}
\end{aligned}$$

for any $\delta < R_s$.

Combining the above two cases completes the proof of the remark.

APPENDIX B

PROOF OF LEMMA 1

By the symmetry of the code construction, the value of the conditional entropy $H(L|W = w, \{Z_{i,j}\})$ does not depend on the realization w, so we have

$$H(L|W, \{Z_{i,j}\}) = H(L|W = 1, \{Z_{i,j}\}). \tag{B.1}$$

Given $W = 1$ and the matrix of observations $\{Z_{i,j}\}$, consider the following l_0 binary hypotheses, each on one of the first l_0 rows of the transmitted codeword metrix:

$$\begin{aligned} H_{l,0}: &\quad x_l^k = (0, \dots, 0) \\ H_{l,1}: &\quad x_l^k = (x, \dots, x) \end{aligned}$$

for $l = 1, \dots, l_0$. The conditional error probabilities are denoted by

$$\alpha_l^{(k)} = \Pr(\hat{H}_{l,1}|H_{l,0}) \tag{B.2}$$

$$\beta_l^{(k)} = \Pr(\hat{H}_{l,0}|H_{l,1}). \tag{B.3}$$

Note that Z_l^k is i.i.d. according to $P_{Z|X=0}$ under $H_{l,0}$ and i.i.d. according to $P_{Z|X=x}$ under $H_{l,1}$. Fix $\delta' > 0$. By the Chernoff-Stein lemma [9] a decision rule can be found such that $\beta_l^{(k)} \to 0$ in the limit as $k \to \infty$ and

$$\alpha_l^{(k)} \le e^{-k\left(D(P_{Z|X=x} \| P_{Z|X=0}) - \delta'\right)} \tag{B.4}$$

for sufficiently large k.

For $l = 1, \dots, l_0$, let $N_l = 1$ if $H_{l,1}$ is declared and $N_l = 0$ if $H_{l,0}$ is declared. Then, for all $l \neq L$ we have

$$E[N_l] = \alpha_l^{(k)} \tag{B.5}$$

$$\mathrm{Var}[N_l] \leq E[N_l^2] = \alpha_l^{(k)}. \tag{B.6}$$

Further let

$$N := \sum_{l \neq L} N_l. \tag{B.7}$$

Since N_l, $l \neq L$, are i.i.d., we have

$$\begin{aligned} E[N] &= \sum_{l \neq L} E[N_l] = (l_0 - 1)\alpha_l^{(k)} && \text{(B.8)} \\ &\leq l_0 e^{-k\left(D(P_{Z|X=x} \| P_{Z|X=0}) - \delta'\right)} && \text{(B.9)} \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}[N] &= \sum_{l \neq L} \mathrm{Var}[N_l] \leq (l_0 - 1)\alpha_l^{(k)} && \text{(B.10)} \\ &\leq l_0 e^{-k\left(D(P_{Z|X=x} \| P_{Z|X=0}) - \delta'\right)}. && \text{(B.11)} \end{aligned}$$

It follows that

$$\begin{aligned}
&\Pr\left\{N \geq 2l_0 e^{-k\left(D(P_{Z|X=x}\|P_{Z|X=0})-\delta'\right)}\right\} \\
&\leq \Pr\left\{N - E[N] \geq l_0 e^{-k\left(D(P_{Z|X=x}\|P_{Z|X=0})-\delta'\right)}\right\} && \text{(B.12)} \\
&\leq \Pr\left\{|N - E[N]| \geq l_0 e^{-k\left(D(P_{Z|X=x}\|P_{Z|X=0})-\delta'\right)}\right\} && \text{(B.13)} \\
&\leq \frac{\mathrm{Var}[N]}{\left(l_0 e^{-k\left(D(P_{Z|X=x}\|P_{Z|X=0})-\delta'\right)}\right)^2} && \text{(B.14)} \\
&\leq \frac{1}{l_0 e^{-k\left(D(P_{Z|X=x}\|P_{Z|X=0})-\delta'\right)}} && \text{(B.15)} \\
&< \frac{1}{e^{k\left(\frac{\delta b(x)}{12}+\delta'\right)}} && \text{(B.16)} \\
&\to 0 && \text{(B.17)}
\end{aligned}$$

in the limit as $k \to \infty$, where (B.12) follows from (B.9), (B.14) follows from the well-known Chebyshev's inequality, (B.15) follows from (B.11), and (B.16) follow from the assumption that

$$l_0 > \exp\left(k\left(D(P_{Z|X=x}\|P_{Z|X=0}) + \frac{\delta b(x)}{12}\right)\right). \tag{B.18}$$

Let E be a random variable such that $E = 1$ if

$$N < 2l_0 e^{-k\left(D(P_{Z|X=x}\|P_{Z|X=0})-\delta'\right)}$$

and $H_{L,1}$ is declared, and $E = 0$ otherwise. We have

$$\begin{aligned} & H(L|W=1,\{Z_{i,j}\}) && \text{(B.19)} \\ & \quad \leq H(L,E|W=1,\{Z_{i,j}\}) && \text{(B.20)} \\ & \quad = H(E|W=1,\{Z_{i,j}\}) + H(L|W=1,\{Z_{i,j}\},E) && \text{(B.21)} \\ & \quad \leq H(E) + \Pr\{E=0\} H(L) + \\ & \qquad H(L|W=1,\{Z_{i,j}\},E=1). && \text{(B.22)} \end{aligned}$$

Note that

$$\begin{aligned} H(E) &\leq \ln 2 && \text{(B.23)} \\ H(L) &= \ln l_0 < k\left(D(P_{Z|X=x}\|P_{Z|X=0}) + \frac{\delta b(x)}{6}\right). && \text{(B.24)} \end{aligned}$$

By the union bound, the probability $\Pr\{E=0\}$ can be bounded from above as

$$\begin{aligned} & \Pr\left\{\hat{H}_{L,0}\right\} + \Pr\left\{N \geq 2l_0 e^{-k\left(D(P_{Z|X=x}\|P_{Z|X=0})+\delta'\right)}\right\} \\ & \qquad \to 0 && \text{(B.25)} \end{aligned}$$

in the limit as $k \to \infty$. Furthermore,

$$\begin{aligned} & H(L|W=1,\{Z_{i,j}\},E=1) \\ & \qquad < \ln\left(2l_0 e^{-k\left(D(P_{Z|X=x}\|P_{Z|X=0})-\delta'\right)} + 1\right) && \text{(B.26)} \\ & \qquad = H(L) - k\left(D(P_{Z|X=x}\|P_{Z|X=0}) - \epsilon'\right) && \text{(B.27)} \end{aligned}$$

where

$$\epsilon' = \frac{1}{k} \ln \left(2e^{k\delta'} + \frac{e^{kD(P_{Z|X=x} \| P_{Z|X=0})}}{l_0} \right) \tag{B.28}$$

$$\to \delta' \tag{B.29}$$

in the limit as $k \to \infty$. We thus have

$$H(L|W = 1, \{Z_{i,j}\}) \tag{B.30}$$

$$< \ln 2 + k\Pr\{E = 0\} \left(D(P_{Z|X=x} \| P_{Z|X=0}) + \frac{\delta b(x)}{6} \right)$$

$$+ H(L) - k \left(D(P_{Z|X=x} \| P_{Z|X=0}) - \epsilon' \right) \tag{B.31}$$

$$= H(L) - k \left(D(P_{Z|X=x} \| P_{Z|X=0}) - \epsilon'' \right) \tag{B.32}$$

where

$$\epsilon'' := \epsilon' + \frac{\ln 2}{k} + \Pr\{E = 0\} \left(D(P_{Z|X=x} \| P_{Z|X=0}) + \frac{\delta b(x)}{6} \right) \tag{B.33}$$

$$\to \delta' \tag{B.34}$$

in the limit as $k \to \infty$. Letting $\delta' \to 0$ completes the proof of the lemma.

APPENDIX C

PROOF OF PROPOSITION 1

By Fano's inequality, any achievable secrecy rate R_s must satisfy

$$\begin{aligned}
n(R_s - \epsilon_n) &\leq I(W; Y_1^n) \\
&\leq I(W; Y_1^n, S^n) \\
&= I(W; Y_1^n | S^n) \\
&\leq I(X^n; Y_1^n | S^n) \\
&= H(Y_1^n | S^n) - H(Y_1^n | X^n, S^n) \\
&= H(Y_1^n | S^n) - \sum_{i=1}^{n} H(Y_1[i] | X[i], S[i]) \\
&\leq \sum_{i=1}^{n} H(Y_1[i] | S[i]) - \sum_{i=1}^{n} H(Y_1[i] | X[i], S[i]) \\
&= n[H(Y_{1,Q} | S_Q, Q) - H(Y_{1,Q} | X_Q, S_Q, Q)] \\
&= n[H(Y_{1,Q} | S_Q, Q) - H(Y_{1,Q} | X_Q, S_Q)] \\
&\leq n[H(Y_{1,Q} | S_Q) - H(Y_{1,Q} | X_Q, S_Q)] \\
&= n \cdot I(X_Q; Y_{1,Q} | S_Q)
\end{aligned}$$

where $\epsilon_n \to 0$ in the limit as $n \to \infty$, and Q is a standard time-sharing variable.

Similarly, for any achievable secrecy rate R_s we have

$$\begin{aligned}
&n(R_s - \epsilon_n) \\
&\quad \leq \; I(W; Y_1^n) - I(W; Y_2^n) \\
&\quad \leq \; I(W; Y_1^n, Y_2^n) - I(W; Y_2^n) \\
&\quad = \; I(W; Y_1^n | Y_2^n) \\
&\quad \leq \; I(X^n, S^n; Y_1^n | Y_2^n) \\
&\quad = \; H(Y_1^n | Y_2^n) - H(Y_1^n | X^n, S^n, Y_2^n) \\
&\quad = \; H(Y_1^n | Y_2^n) - \sum_{i=1}^{n} H(Y_1[i] | X[i], S[i], Y_2[i]) \\
&\quad \leq \; \sum_{i=1}^{n} H(Y_1[i] | Y_2[i]) - \sum_{i=1}^{n} H(Y_1[i] | X[i], S[i], Y_2[i]) \\
&\quad = \; n[H(Y_{1,Q} | Y_{2,Q}, Q) - H(Y_{1,Q} | X_Q, S_Q, Y_{2,Q}, Q)] \\
&\quad = \; n[H(Y_{1,Q} | Y_{2,Q}, Q) - H(Y_{1,Q} | X_Q, S_Q, Y_{2,Q})] \\
&\quad \leq \; n[H(Y_{1,Q} | Y_{2,Q}) - H(Y_{1,Q} | X_Q, S_Q, Y_{2,Q})] \\
&\quad = \; n \cdot I(X_Q, S_Q; Y_{1,Q} | Y_{2,Q}).
\end{aligned}$$

Note that the channel states are memoryless, so S_Q has the same distribution as $S[i]$ for any $i = 1, \ldots, n$. The channel is also memoryless, so the conditional distribution of $(Y_{1,Q}, Y_{2,Q})$ given (X_Q, S_Q) is given by the channel transition probability $p(y_1, y_2 | x, s)$. Letting $X_Q = X$, $S_Q = S$, $Y_{1,Q} = Y_1$, $Y_{2,Q} = Y_2$, and $n \to \infty$ completes the proof of the proposition.

APPENDIX D

PROOF OF LEMMA 2

Let Z be an i.i.d. Bernoulli-1/2 vector. We have

$$\begin{aligned} H(AZ|BZ) &= H\left(\begin{bmatrix} A \\ B \end{bmatrix} Z\right) - H(BZ) \\ &= rank\left(\begin{bmatrix} A \\ B \end{bmatrix}\right) - rank(B). \end{aligned}$$

We thus conclude that

$$\max H(AZ|BZ) \geq rank\left(\begin{bmatrix} A \\ B \end{bmatrix}\right) - rank(B). \tag{D.1}$$

To prove the reverse inequality, let us consider the null space of B and its coset partition based on the null space of

$$\begin{bmatrix} A \\ B \end{bmatrix}.$$

Fix $BZ = b$. Then, any solution Z can be written as the sum of a particular solution Z_p and a vector Z_h in the null space of B. Note that all vectors Z_h in the same coset of the null space of B relative to the null space of

$$\begin{bmatrix} A \\ B \end{bmatrix}$$

give the *same* value for AZ_h. Thus, the number of *different* values that AZ can take for any given value of b equals the number of cosets in the null space of B, which is given by

$$2^{nullitiy(B)-nullity\left(\begin{bmatrix} A \\ B \end{bmatrix}\right)} = 2^{rank\left(\begin{bmatrix} A \\ B \end{bmatrix}\right)-rank(B)}.$$

We thus conclude that

$$\max H(AZ|BZ) \leq rank\left(\begin{bmatrix} A \\ B \end{bmatrix}\right) - rank(B). \tag{D.2}$$

Combining (D.1) and (D.2) completes the proof of the lemma.

APPENDIX E

PROOF OF THEOREM 8

The matrix B is a horizontal stack of two down-shift matrices with rank n_2 and m_2, respectively. Since both sub-matrices are in reduced row-echelon form, it suffices to count the number of nonzero rows of B to find its rank:

$$\begin{aligned} rank(B) &= q - \min\{q - n_2, q - m_2\} \\ &= \max\{n_2, m_2\}. \end{aligned}$$

The matrix

$$G := \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} D^{q-n_1} & D^{q-m_1} \\ D^{q-n_2} & D^{q-m_2} \end{bmatrix}$$

is formed by vertically stacking two matrices A and B. Thus, evaluating the rank of G is equivalent to counting the number of zero rows along with the number of redundant nonzero rows between A and B (denoted by d_{AB}):

$$\begin{aligned} rank(G) &= 2q - \min\{q - n_1, q - m_1\} - \\ &\qquad \min\{q - n_2, q - m_2\} - d_{AB} \\ &= \max\{n_1, m_1\} + \max\{n_2, m_2\} - d_{AB}. \end{aligned}$$

To calculate d_{AB}, let us consider the following five cases separately:

Case 1: Either $n_1 \leq m_1$ and $n_2 > m_2$, or $n_2 \leq m_2$ and $n_1 > m_1$. In this case, all nonzero rows of G are independent so $d_{AB} = 0$. By Proposition 2,

$$C_s = \min\{n_1, \max\{n_1, m_1\} - 0\} = n_1.$$

Case 2: $n_1 \leq m_1$ and $n_2 \leq m_2$, but $m_1 - n_1 \neq m_2 - n_2$. In this case, the redundant nonzero rows of G are given by the redundant rows between the top $m_1 - n_1$ nonzero rows of A and the top $m_2 - n_2$ nonzero rows of B. Hence, $d_{AB} = \min\{m_1 - n_1, m_2 - n_2\}$. By Proposition 2,

$$\begin{aligned} C_s &= \min\{n_1, m_1 - \min\{m_1 - n_1, m_2 - n_2\}\} \\ &= \min\{n_1, \max\{n_1, m_1 - m_2 + n_2\}\} \\ &= n_1. \end{aligned}$$

Case 3: $n_1 > m_1$ and $n_2 > m_2$, but $m_1 - n_1 \neq m_2 - n_2$. In this case, the redundant nonzero rows of G are given by the redundant rows between the top $n_1 - m_1$ nonzero rows of A and the top $n_2 - m_2$ nonzero rows of B. Hence, $d_{AB} = \min\{n_1 - m_1, n_2 - m_2\}$. By Proposition 2,

$$\begin{aligned} C_s &= \min\{n_1, n_1 - \min\{n_1 - m_1, n_2 - m_2\}\} \\ &= n_1 - \min\{n_1 - m_1, n_2 - m_2\} \\ &= \max\{m_1, n_1 - n_2 + m_2\}. \end{aligned}$$

Case 4: $n_1 - m_1 = n_2 - m_2$ and $n_1 \geq m_1$. In this case, the redundant nonzero rows of G correspond to the redundant nonzero rows of

$$\begin{bmatrix} D^{q-n_1} \\ D^{q-n_2} \end{bmatrix}$$

so $d_{AB} = \min\{n_1, n_2\}$. By Proposition 2,

$$\begin{aligned} C_s &= \min\{n_1, n_1 - \min\{n_1, n_2\}\} \\ &= n_1 - \min\{n_1, n_2\} \\ &= (n_1 - n_2)^+. \end{aligned}$$

Case 5: $n_1 - m_1 = n_2 - m_2$ and $n_1 < m_1$. In this case, the redundant nonzero rows of G correspond to the redundant nonzero rows of

$$\begin{bmatrix} D^{q-m_1} \\ D^{q-m_2} \end{bmatrix}$$

so $d_{AB} = \min\{m_1, m_2\}$. By Proposition 2,

$$\begin{aligned} C_s &= \min\{n_1, m_1 - \min\{m_1, m_2\}\} \\ &= \min\{n_1, (m_1 - m_2)^+\} \\ &= \min\{n_1, (n_1 - n_2)^+\} \\ &= (n_1 - n_2)^+. \end{aligned}$$

Combining the results from the above five cases completes the proof of (3.12) and hence Theorem 8.