Key Agreement over Wiretap Models with Non-Causal Side Information

by

Ali Zibaeenejad

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2012

AUTHOR'S DECLARATION

Abstract

The security of information is an indispensable element of a communication system when transmitted signals are vulnerable to eavesdropping. This issue is a challenging problem in a wireless network as propagated signals can be easily captured by unauthorized receivers, and so achieving a perfectly secure communication is a desire in such a wiretap channel. On the other hand, cryptographic algorithms usually lack to attain this goal due to the following restrictive assumptions made for their design. First, wiretappers basically have limited computational power and time. Second, each authorized party has often access to a reasonably large sequence of uniform random bits concealed from wiretappers.

To guarantee the security of information, Information Theory (IT) offers the following two approaches based on physical-layer security.

First, IT suggests using wiretap (block) codes to securely and reliably transmit messages over a noisy wiretap channel. No confidential common key is usually required for the wiretap codes. The secrecy problem investigates an optimum wiretap code that achieves the secrecy capacity of a given wiretap channel.

Second, IT introduces key agreement (block) codes to exchange keys between legitimate parties over a wiretap model. The agreed keys are to be reliable, secure, and (uniformly) random, at least in an asymptotic sense, such that they can be finally employed in symmetric key cryptography for data transmission. The key agreement problem investigates an optimum key agreement code that obtains the key capacity of a given wiretap model.

In this thesis, we study the key agreement problem for two wiretap models: a Discrete Memoryless (DM) model and a Gaussian model. Each model consists of a wiretap channel paralleled with an authenticated public channel. The wiretap channel is from a transmitter, called Alice, to an authorized receiver, called Bob, and to a wiretapper, called Eve. The Probability Transition Function (PTF) of the wiretap channel is controlled by a random sequence of Channel State Information (CSI), which is assumed to be non-causally available at Alice. The capacity of the public channel is $C_{P_1} \in [0, \infty)$ in the forward direction from Alice to Bob and $C_{P_2} \in [0, \infty)$ in the backward direction from Bob to Alice. For each model, the key capacity as a function of the pair (C_{P_1}, C_{P_2}) is denoted by $C_K(C_{P_1}, C_{P_2})$. We investigate the forward key capacity of each model, i.e., $C_K(C_{P_1}, 0)$ in this thesis. We also study the key generation over the Gaussian model when Eve's channel is less noisy than Bob's.

In the DM model, the wiretap channel is a Discrete Memoryless State-dependent Wiretap Channel (DM-SWC) in which Bob and Eve each may also have access to a sequence of Side Information (SI) dependent on the CSI. We establish a Lower Bound (LB) and an Upper Bound (UB) on the forward key capacity of the DM model. When the model is less noisy in Bob's favor, another UB on the forward key capacity is derived. The achievable key agreement code is asymptotically optimum as $C_{P_1} \to \infty$. For any given DM model, there also exists a finite capacity $C^*_{P_1}$, which is determined by the DM-SWC, such that the forward key capacity is achievable if $C_{P_1} \geq C^*_{P_1}$. Moreover, the key generation is saturated at capacity $C_{P_1} = C^*_{P_1}$, and thus increasing the public channel capacity beyond $C^*_{P_1}$ makes no improvement on the forward key capacity of the DM model. If the CSI is fully known at Bob in addition to Alice, $C^*_{P_1} = 0$, and so the public channel has no contribution in key generation when the public channel is in the forward direction.

The achievable key agreement code of the DM model exploits both a random generator and the CSI as resources for key generation at Alice. The randomness property of channel states can be employed for key generation, and so the agreed keys depend on the CSI in general. However, a message is independent of the CSI in a secrecy problem. Hence, we justify that the forward key capacity can exceed both the main channel capacity and the secrecy capacity of the DM-SWC.

In the Gaussian model, the wiretap channel is a Gaussian State-dependent Wiretap Channel (G-SWC) with Additive White Gaussian Interference (AWGI) having average power Λ. For simplicity, no side information is assumed at Bob and Eve. Bob's channel and Eve's channel suffer from Additive White Gaussian Noise (AWGN), where the correlation coefficient between noise of Bob's channel and that of Eve's channel is given by ϱ.

We prove that the forward key capacity of the Gaussian model is independent of ϱ. Moreover, we establish that the forward key capacity is positive unless Eve's channel is

less noisy than Bob's. We also prove that the key capacity of the Gaussian model vanishes if the G-SWC is physically degraded in Eve's favor. However, we justify that obtaining a positive key capacity is feasible even if Eve's channel is less noisy than Bob's according to our achieved LB on the key capacity for case $(C_{P_1}, C_{P_2}) \to (\infty, \infty)$. Hence, the key capacity of the Gaussian model is a function of ϱ.

In this thesis, an LB on the forward key capacity of the Gaussian model is achieved. For a fixed Λ, the achievable key agreement code is optimum for any $C_{P_1} \in [0, \infty)$ in both low Signal-to-Interference Ratio (SIR) and high SIR regimes. We show that the forward key capacity is asymptotically independent of C_{P_1} and Λ as the SIR goes to infinity, and thus the public channel and the interference have negligible contributions in key generation in the high SIR regime. On the other hand, the forward key capacity is a function of C_{P_1} and Λ in the low SIR regime. Contributions of the interference and the public channel in key generation are significant in the low SIR regime that will be illustrated by simulations.

The proposed key agreement code asymptotically achieves the forward key capacity of the Gaussian model for any SIR as $C_{P_1} \to \infty$. Hence, $C_K(\infty, 0)$ is calculated, and it is suggested as a UB on $C_K(C_{P_1}, 0)$. Using simulations, we also compute the minimum required C_{P_1} for which the forward key capacity is upper bounded within a given tolerance.

The achievable key agreement code is designed based on a generalized version of the Dirty Paper Coding (DPC) in which transmitted signals are correlated with the CSI. The correlation coefficient is to be determined by C_{P_1}. In contrast to the DM model, the LB on the forward key capacity of a Gaussian model is a strictly increasing function of C_{P_1} according to our simulations. This fact is an essential difference between this model and the DM model.

For $C_{P_1} = 0$ and a fixed Λ, the forward key capacity of the Gaussian model exceeds the main channel capacity of the G-SWC in the low SIR regime. By simulations, we show that the interference enhances key generation in the low SIR regime. In this regime, we also justify that the positive effect of the interference on the (forward) key capacity is generally more than its positive effect on the secrecy capacity of the G-SWC, while the interference has no influence on the main channel capacity of the G-SWC.

Acknowledgements

First and foremost, I would like to praise the Lord for all the blessings throughout my life, specially during the difficulties and challenges of my doctoral program.

Next, I deeply appreciate my supervisor, Professor Amir K. Khandani, for his consideration and support during the course of my PhD studies. Furthermore, I am grateful to all my advisory committee members, Professors Rei Safavi-Naini (the external examiner from the University of Calgary), Amir K. Khandani, Ian Goldberg, Guang Gong, and Patrick Mitran for their technical comments, suggestions, and the reviews of my thesis during their busy schedules.

I wish to express my gratitude to the administration of the Electrical and Computer Engineering (ECE) department in the University of Waterloo, particulary Professor Siva Sivoththaman (the associate chair) and Susan King (graduate records specialist) for their legislative assistance, time and advice.

I send my special thanks to my dear friends Dr. Hossein Bagheri, Dr. Vahid Pourahmadi, and Seyed Ali Ahmadzadeh as well as to my brother M. Hadi Zibaeenejad for their remarkable encouragement, help and advice. I also forward my regards and thanks to my collogues in the Coding and Signal Transmission (CST) Lab, specially Dr. Seyed Abolfazl Motahari, Akbar Ghasemi, M. Javad Abdoli for their valuable comments.

I sincerely appreciate my parents, Professor M. Javad Zibaeenejad and Khadijeh Mohammadi, for their endless love and unconditional support in my entire life for which my mere expression of thanks does not suffice.

Last but not least, I devote my heartfelt gratitude to my beloved wife, Fatemeh Jahanmiri, for her incredible support, patience, and understanding. The completion of my doctoral program would not be possible without her continual inspiration and dedication.

Dedication

To my darling,

Fatemeh

Table of Contents

List of Figures **xvii**

List of Abbreviations **xix**

List of Notations **xxiii**

1 Introduction **1**

1.1 Communication over Noisy Channels 2

1.2 One-Time Pad System 6

1.3 Information-Theoretic Security 10

1.4 The Key Agreement Problem 15

1.4.1 The Concept 15

1.4.2 The Source-Type Model vs. The Channel-Type Model 19

1.4.3 Privacy Amplification 25

1.5 Channels with Random States 28

1.6 State-Dependent Wiretap Models 31

1.7 Motivations 35

1.8 Contributions 40

1.8.1 The DM Model 40

1.8.2 The Gaussian Model 43

1.8.3 The Parallel Work 46

1.9 Organization of the Thesis 47

2 Fundamentals **49**

2.1 Mathematical Tools 50

2.1.1 Typicality 50

2.2 Wiretap Channels 53

2.2.1 Wiretap Channels without Side Information 53

2.2.2 Degraded Discrete Memoryless Wiretap Channel with Random States 59

2.2.3 Gaussian Wiretap Channel with Additive Interference 62

3 The Discrete Memoryless Model **71**

3.1 Problem Definitions 71

3.2 Statement of Main Results 76

3.3 The Proofs 79

3.3.1 The Proof of Theorem 3.1 80

3.3.2 The Proof of Corollary 3.1 95

3.3.3 The Proof of Theorem 3.2 96

3.3.4 The Proof of Theorem 3.3 104

3.3.5 The Proof of Corollary 3.2 106

3.3.6 The Proof of Theorem 3.4 106

3.3.7 The Proof of Corollary 3.3 106

3.3.8 The Proof of Corollary 3.4 107

4 The Gaussian Model 109

4.1 Problem Definitions 111

4.2 Statement of Main Results 116

4.2.1 Forward Public Channel 116

4.2.2 Two-Way Public Channel 118

4.3 Simulations 120

4.4 The Proofs 128

4.4.1 The Proof of Lemma 4.1 128

4.4.2 The Proof of Lemma 4.2 128

4.4.3 The Proof of Theorem 4.1 130

4.4.4 The Proof of Corollary 4.1 136

4.4.5 The Proof of Theorem 4.2 139

4.4.6 The Proof of Corollary 4.2 145

4.4.7 The Proof of Corollary 4.3 145

4.4.8 The Proof of Theorem 4.3 147

5 Conclusions 151

5.1 Future Work Directions 154

5.1.1 Bounds on the Forward Key Capacity of the Gaussian Model 155

5.1.2 Gaussian Models with Side Information at All Parties 155

5.1.3 Key Agreement over Wiretap Models with a Two-Way Public Channel156

5.1.4 Reliability-Exponent and Security-Exponent of the Key Agreement 156

APPENDICES 157

A Relation between Entropy Functions of Agreed Keys 159

B Existence of Maximum Values 161

C AEP for Gaussian Random Vectors 163

D Csiszár-Körner's Sum Identity 165

References 167

List of Figures

1.1 A communication system with a point-to-point noisy channel. 4

1.2 Shannon's model: Noise-free secrecy system. 6

1.3 Set-up of a one-time pad system in a (noisy) wiretap channel. 9

1.4 Wyner's model: Degraded noisy wiretap model. 11

1.5 The Gaussian wiretap channel. 14

1.6 Key agreement between Alice and Bob in presence of Eve. 17

1.7 Key agreement over a source-type model. 20

1.8 Key agreement over a channel-type model. 21

1.9 Key agreement over a channel-type model with a private DMC. 23

1.10 Costa's Model: Gaussian channel with known interference at the sender. . 31

1.11 A motivation of the discrete memoryless model. 36

1.12 A motivation of the Gaussian model. 37

2.1 A discrete memoryless wiretap channel (with no SI). 54

2.2 The capacity-equivocation region for a degraded symmetric DM-WC. . . . 59

2.3 A degraded DM-SWC with non-causal CSI at the transmitter. 60

2.4 Secrecy problem over a physically degraded G-SWC in Bob's favor. 63

2.5 Comparison of optimum values of α (in the DPC) versus SIR in the G-SWC. 67

2.6 Bounds on the secrecy capacity of the degraded G-SWC. 69

3.1 The key agreement problem over the DM model. 74

3.2 Structure of codebook $\mathbf{C}$ formed with a double random binning strategy. . 82

4.1 Key agreement over the Gaussian model. 112

4.2 Key agreement over a physically degraded Gaussian model in Bob's favor. . 114

4.3 Key agreement over a physically degraded Gaussian model in Eve's favor. . 114

4.4 The LB on key capacity of the Gaussian model with capacity C_{P_1}. 123

4.5 Coefficient γ^* for the Gaussian model with capacity C_{P_1}. 124

4.6 Correlation coefficient ρ^* for the Gaussian model with capacity C_{P_1}. 125

4.7 Comparison of key capacity, secrecy capacity and capacity of the G-SWC. . 126

4.8 The minimum required C_{P_1} to achieve $C_K(\infty, 0)$ within tolerance ζ. 127

4.9 Key generation over a Gaussian model with a backward public channel. . . 149

List of Abbreviations

AEP	Asymptotic equipartition property
AR	Asymptotic reliability
ARN	Asymptotic randomness
AS	Asymptotic security
AWGI	Additive white Gaussian interference
AWGN	Additive white Gaussian noise
BC	Broadcast channel
BSC	Binary symmetric channel
Ch.	Chapter
Corol.	Corollary
CSI	Channel state information
dB	Decibel
DM	Discrete memoryless
DM-SWC	Discrete memoryless state-dependent wiretap channel

DM-WC	Discrete memoryless wiretap channel
DMC	Discrete memoryless channel
DMMS	Discrete memoryless multiple source
DPC	Dirty paper coding
Eq.	Equation
G-SWC	Gaussian state-dependent wiretap channel
i.i.d.	independent and identically distributed
iff	if and only if
IT	Information theory
LB	Lower bound
Lem.	Lemma
ML	Maximum likelihood
PMF	Probability mass function
Prop.	Proposition
PTF	Probability transition function
Remk.	Remark
resp.	respectively
RV	Random variable
s. t.	such that
Sec.	Section

SI	Side information
SIR	signal-to-interference ratio
SNR	Signal-to-noise ratio
Thm.	Theorem
trans.	transmission
UB	Upper bound
vs.	versus

List of Notations

$\triangleq$	Defined as (e.g., $a \triangleq b$: a is defined as b)
$\hat{}$	The indicator for the legitimate receiver (Bob) (e.g., $\hat{M}$)
$\Leftrightarrow$	if and only if (iff)
$\square$	End of a proof
$\mathbb{N}$	Set of natural numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^{+}$	Set of positive members of $\mathbb{R}$
n	Coding block length
N, i, j, ℓ, τ	Reserved for natural numbers
R	Communication rate
C	(Ordinary) capacity
C_m	(Ordinary) capacity of the main channel
C_{mw}	(Ordinary) capacity of the overall main-wiretap channel
C_{P_1}	Forward public channel capacity
C_{P_2}	Backward public channel capacity
R_{E}	Equivocation rate
R_{L}	Leakage rate
C_{S}	Secrecy capacity
R_{K}	Key rate
C_{K}	Key capacity

Narrow, Latin capital letter	Random variable (e.g., A: random variable A)
Bold, Latin capital letter	Matrix (e.g., $\mathbf{A}_{N_1 \times N_2}$: matrix $\mathbf{A}$ with N_1 rows and N_2 columns)
Bold, Latin lower case letter	Vector of length n (e.g., vector $\mathbf{a} = (a_1, \ldots, a_n)$)
Blackboard bold, Capital letter	Set (e.g., $\mathbb{A}$: Set $\mathbb{A}$)
Calligraphy, Latin capital letter	Function (e.g., $\mathcal{F}(\alpha)$: function $\mathcal{F}$ of α)
Greek letter	Publicly known Parameter (e.g., Γ: average power)
$(x_1, \ldots, x_n)$	An n-tuple, it is called a pair/triple if $n = 2/n = 3$, resp.
x_i	For $i \leq n$, i^{th} element of vector $\mathbf{x} = (x_1, \ldots, x_i, \ldots, x_n)$
X_i^j	For $i \leq j \leq n$, truncated sequence $X_i^j \triangleq (X_i, \ldots, X_j)$ of random vector $\mathbf{x}$
$\mathbf{x}_i$	For $i \leq \ell$, i^{th} row of matrix $\mathbf{X}_{\ell \times n}$
X_{ij}	For $i \leq \ell$ and $j \leq n$, element (i, j) of random matrix $\mathbf{X}_{\ell \times n}$
$\inf(\mathbb{A})$	Infimum of set $\mathbb{A}$
$\sup(\mathbb{A})$	Supremum of set $\mathbb{A}$
$\lvert\mathbb{A}\rvert$	Cardinality of set $\mathbb{A}$
$\mathbb{A} \times \mathbb{B}$	The cartesian product of sets $\mathbb{A}$ and $\mathbb{B}$
$\mathbb{A}^N$	N times cartesian product of set $\mathbb{A}$, (i.e., $\mathbb{A}^N \triangleq \underbrace{\mathbb{A} \times \ldots \times \mathbb{A}}_{N \text{ times}}$)
$\mathbb{A}^c$	Complement of set $\mathbb{A}$
$\mathbf{0}$	All zero vector with length n
$\mathbf{x}^t$	Transpose of vector $\mathbf{x}$
$\log(\alpha)$	The log, base 2, of α, where $\alpha \in \mathbb{R}^+$
$\ln(\alpha)$	The natural logarithm of α, where $\alpha \in \mathbb{R}^+$
e	Euler's number
$\lvert\alpha\rvert$	For $\alpha \in \mathbb{R}$, the absolute value of α
$\lvert\mathbf{A}\rvert$	Determinant of matrix $\mathbf{A}$

$\mathscr{P}\{.\}$	Probability of an event
$\mathcal{P}_{error}(n)$	Probability of error as a function of block length n
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution function with mean μ and variance σ^2
$\mathcal{N}((\mu_1, \ldots, \mu_\ell), \mathbf{\Sigma}_{\ell\times\ell})$	Multivariate normal distribution function with mean vector $(\mu_1, \ldots, \mu_\ell)$ and covariance matrix $\mathbf{\Sigma}_{\ell\times\ell}$
$\mathcal{P}_X(x)$	Probability mass function of RV X at $X = x$
$X \sim \mathcal{P}_X(x)$	Random variable X has distribution $\mathcal{P}_X(x)$
$\mathcal{I}(X;Y)$	Mutual information function between RVs X and Y
$\mathcal{I}(X;Y\vert Z)$	Conditional mutual information between RVs X and Y given Z
$\mathcal{H}(X)$	(Shannon) Entropy function of RV X
$\mathcal{H}(X,Y)$	(Shannon) Joint entropy function of RVs X and Y
$\mathcal{H}(X\vert Y)$	Conditional entropy function of RV X given RV Y
$\hbar(X)$	Differential entropy function of continuous RV X
$\hbar(X,Y)$	Joint differential entropy function of continuous RVs X and Y
$\hbar(X\vert Y)$	Conditional differential entropy function of continuous RV X given continuous RV Y
$\mathcal{B}(\alpha)$	Binary entropy function (i.e., $\mathcal{B}(0) = \mathcal{B}(1) \triangleq 0$, and $\mathcal{B}(\alpha) \triangleq -\alpha\log(\alpha) - (1-\alpha)\log(1-\alpha)$ for $\alpha \in (0,1)$)
$\mathcal{E}(X)$	Expectation function of RV X
$\mathcal{E}(X\vert Y)$	Conditional expectation function of RV X given Y
$\mathcal{F} : \mathbb{X} \to \mathbb{Y}$	Function $\mathcal{F}$ is defined from domain $\mathbb{X}$ to codomain $\mathbb{Y}$
$\arg\max(\mathcal{F}(\alpha))$	Values of α that maximize function $\mathcal{F}(\alpha)$
$\arg\min(\mathcal{F}(\alpha))$	Values of α that minimize function $\mathcal{F}(\alpha)$
$n \to \infty$	as block length n goes to infinity
$a = \lim\limits_{x\to b} \mathcal{F}(x)$	a is equal to limit of $\mathcal{F}(x)$ as x goes to b
$X_1 \to X_2 \to X_3$	X_1, X_2, and X_3 form a Markov chain in the given order
$[x]^+$	$[x]^+ \triangleq \max\{x, 0\}$
$\lceil x \rceil$	The smallest integer number which is larger than x

Chapter 1

Introduction

The security of information is a challenging problem in communication systems where transmission signals are exposed to eavesdroppers. The secure communication plays a crucial role in a wireless network where propagated signals are easily accessible by (hidden) antennas of *unauthorized* receivers (*wiretappers*). Hence, a milestone in design of a wireless communication system is to guarantee that the transmitted information over a given *communication channel* [1] is *reliably* decoded [1,2] by authorized receivers such that the wiretappers remain ignorant about that information after interception of transmitted signals [3, Ch. 17].

In this thesis, we assume that the communication channels are *authenticated* which means wiretappers are not able to tamper with messages, communication signals, and labels of communication signals in any sense. We are also interested in scenarios with one authorized transmitter, called Alice, and two receivers: an intended receiver, called Bob, and a (passive) wiretapper, called Eve.

This chapter is organized as follows. We give an overview of the notion of channel coding in Section 1.1. We demonstrate the concept of security in cryptography and its advantages and drawbacks in Section 1.2. We introduce the physical layer security in Section 1.3 and Section 1.4: in the former, we define wiretap channels and the secrecy problem; in the latter, we define the concept of key agreement problems in information theory and privacy amplification, and we review the main related work in these fields. As this thesis is regarding wiretap channels with random states, we introduce the fundamental classes of state-dependent channels with the corresponding results and applications in Section 1.5. Next, we review seminal articles and methodology in the field of state-dependent wiretap models in Section 1.6; we then focus on wiretap models with random states. Motivations of this work are justified in Section 1.7. Also, the challenges of our models are given in this section. In Section 1.8, we briefly express the models together with the main contributions of this thesis. Finally, we state the organization of next chapters in Section 1.9.

1.1 Communication over Noisy Channels

In a real communication system, communication channels are often noisy. Shannon [4] represented a mathematical model for a noisy communication channel. According to this work, noisy channels are modeled by a set of *input alphabet(s)*, a set of *output alphabet(s)*, and a *probability transition function* (PTF) [1–3], which expresses the probability of observation of a channel output signal given a channel input signal. As an example, the mathematical model of a point-to-point communication channel is defined as follows.

Definition 1.1. The mathematical model of a point-to-point communication channel is determined by triple $(\mathbb{X}^n, \mathbb{Y}^n, \mathcal{P}_{\mathbf{y}|\mathbf{x}})$, where

- n is the number of transmissions over the channel, or equivalently, it is called the *block length* of an input vector (the input signal to the channel) and that of the output vector (the output signal from the channel);
- $\mathbb{X}^n$ is called an (n-tuple) input alphabet, i.e., $\mathbf{x} \in \mathbb{X}^n$, where $\mathbf{x}$ is the input random vector with length n;
- $\mathbb{Y}^n$ is called an (n-tuple) output alphabet, i.e., $\mathbf{y} \in \mathbb{Y}^n$, where $\mathbf{y}$ is the output random vector with length n;
- $\mathcal{P}_{\mathbf{y}|\mathbf{x}}$ is called a probability transition function, which is the conditional probability distribution function $\mathbf{y}$ given $\mathbf{x}$.

An important class of communication channels is the class of memoryless channels [1,2] in which channel output symbol at any time instant $i \in \{1, \ldots, n\}$ depends on only the channel input symbol of the time instant i and it is conditionally independent of other input symbols and output symbols at any time $\tau \neq i$, where $\tau \in \{1, \ldots, n\}$. In the following, a memoryless channel with finite input and output alphabets is defined.

Definition 1.2 (Discrete memoryless channel)**.** A point-to-point channel $(\mathbb{X}^n, \mathbb{Y}^n, \mathcal{P}_{\mathbf{y}|\mathbf{x}})$ is called a *discrete memoryless channel* (DMC) [1,2] if

$$\forall\, (\mathbf{x}, \mathbf{y}) \in \mathbb{X}^n \times \mathbb{Y}^n : \mathcal{P}_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} \mathcal{P}_{Y|X}(y_i|x_i)\,, \tag{1.1}$$

where $\mathcal{P}_{Y|X}$ is the conditional probability of a symbol $Y \in \mathbb{Y}$ given symbol $X \in \mathbb{X}$. $\mathcal{P}_{Y|X}$ is called the probability transition function (PTF) of the DMC. For simplicity, the DMC is characterized by triple $(\mathbb{X}, \mathbb{Y}, \mathcal{P}_{Y|X})$.

Now, assume Alice wishes to send message M through the DMC to Bob such that he can retrieve message $\hat{M}$ with a desired level of *reliability* determined by *average probability of error* $\mathscr{P}\{\hat{M} \neq M\}$. To achieve this goal, she maps M to an $n-$length vector $\mathbf{x}$ by an (encoding) function $\mathcal{F}$, which is called an *encoder.* Then, she sends the *codeword* $\mathbf{x}$ to Bob by n times transmissions over the communication channel. Finally, Bob maps the received signal $\mathbf{y}$ to message $\hat{M}$ by a (decoding) function $\mathcal{D}$, which is called a *decoder.* The pair of functions $(\mathcal{F}, \mathcal{D})$ is also referred to as a coding scheme. The mathematical model of this communication system is sketched in Figure (1.1) based on Shannon's study [4].

According to Shannon's work [4,5], a communication channel is said to be *perfectly reliable* (zero-error) if

$$\mathscr{P}\{\hat{M} \neq M\} = 0\,.$$

Shannon [5] proved that it is possible to send a positive rate over a point-to-point DMC with perfect reliability *if and only if* (iff) for every $y \in \mathbb{Y}$ there exists $(x, x') \in \mathbb{X}^2$, where $x \neq x'$, such that $P_{Y|X}(y|x')P_{Y|X}(y|x) = 0$. This result shows that achieving a perfectly reliable communication is not feasible in most channels. Hence, the asymptotic behavior

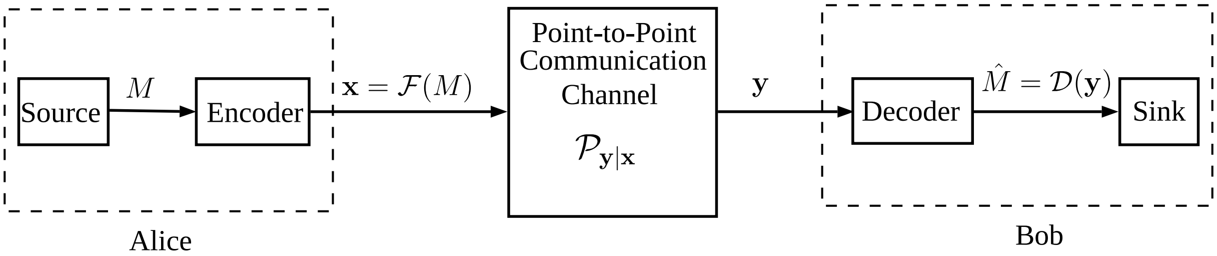

Source
M
Encoder
$\mathbf{x} = \mathcal{F}(M)$
Point-to-Point Communication Channel
$\mathcal{P}_{\mathbf{y}|\mathbf{x}}$
$\mathbf{y}$
Decoder
$\hat{M} = \mathcal{D}(\mathbf{y})$
Sink
Alice
Bob

Figure 1.1: A communication system with a point-to-point noisy channel.

of the average probability of error is considered in general to examine the reliability of a communication system as follows.

Definition 1.3 (*Asymptotic reliability (AR) condition*)**.** Let $\mathbb{M} = \{1, \ldots, \lceil 2^{nR} \rceil\}$, where $R \in \mathbb{R}^+ \cup \{0\}$ and $n \in \mathbb{N}$. Assume message $M \in \mathbb{M}$ is a uniformly distributed random variable and RV $\hat{M} \in \mathbb{M}$ is the decoded message at Bob. Define $\mathcal{P}_{error}(n) \triangleq \mathscr{P}\{\hat{M} \neq M\}$. The communication from Alice to Bob over the point-to-point channel $(\mathbb{X}^n, \mathbb{Y}^n, \mathcal{P}_{\mathbf{y}|\mathbf{x}})$ is said to be asymptotically reliable if there exists a pair of functions $(\mathcal{F}, \mathcal{D})$ with[1]

$$\mathcal{F} : \mathbb{M} \to \mathbb{X}^n \tag{1.2a}$$

$$\mathcal{D} : \mathbb{Y}^n \to \mathbb{M} \tag{1.2b}$$

such that

$$\lim_{n\to\infty} \mathcal{P}_{error}(n) = \lim_{n\to\infty} \frac{1}{\lceil 2^{nR} \rceil} \sum_{i=1}^{\lceil 2^{nR} \rceil} \mathscr{P}\{\hat{M} \neq i\} = 0\,, \tag{1.3}$$

where $\hat{M} = \mathcal{D}(\mathbf{y})$, $\mathbf{x} = \mathcal{F}(M)$, and conditional probability of $\mathbf{y}$ given $\mathbf{x}$ is $\mathcal{P}_{\mathbf{y}|\mathbf{x}}$. Further, the information rate R is called an *achievable rate* with respect to the asymptotic reliability (AR) condition (1.3).

In information theory, a fundamental objective of channel coding problem is to find the

[1] The encoder and the decoder are assumed to be deterministic functions. In fact, randomization at either the encoder or the decoder does not increase channel capacity of the point-to-point channel [3, 6].

channel capacity as follows.

Definition 1.4. (Channel capacity) The supremum of all achievable rates with respect to the asymptotic reliability (AR) condition (1.3) is defined as the channel capacity of the point-to-point channel $(\mathbb{X}^n, \mathbb{Y}^n, \mathcal{P}_{\mathbf{y}|\mathbf{x}})$.

Assuming rate R is an achievable rate, we sometimes need to examine how $\mathcal{P}_{error}(n) \to 0$ as $n \to \infty$. Specially, this issue has a practical importance when a negligible positive error is tolerable due to the application because large block length leads to delay in the decoding step. On the other hand, a trade-off between the finite block length and the average probability of error is an essential issue to design an encoder-decoder for a given channel. The *reliability-exponent* is utilized to measure how the average probability of error converges to zero.

Definition 1.5. The reliability condition 1.3 is said to be achievable with reliability-exponent ε_r if there exists a pair of functions $(\mathcal{F}, \mathcal{D})$ according to (1.2) such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log(\mathcal{P}_{error}(n)) \geq \varepsilon_r \ . \tag{1.4}$$

As an example, the capacity and reliability-exponent of a point-to-point DMC is given by [1,4]

$$C = \max_{\mathcal{P}_X} \mathcal{I}(X;Y)\,, \tag{1.5}$$

$$\varepsilon_r = \max_{\mathcal{P}_X} \max_{\lambda\in[0,1]} [\mathcal{T}(\lambda, \mathcal{P}_X) - \lambda R]\,, \tag{1.6}$$

where function $\mathcal{I}(X;Y) \triangleq \mathcal{E}_{P_{XY}(x,y)}(\log\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right))$ is the mutual information function between RVs X and Y, and

$$\mathcal{T}(\lambda, \mathcal{P}_X) = -\log\left(\sum_{y\in\mathbb{Y}}(\sum_{x\in\mathbb{X}} \mathcal{P}_X(\mathcal{P}_{Y|X})^{\frac{1}{1+\lambda}})^{1+\lambda})\right)\ . \tag{1.7}$$

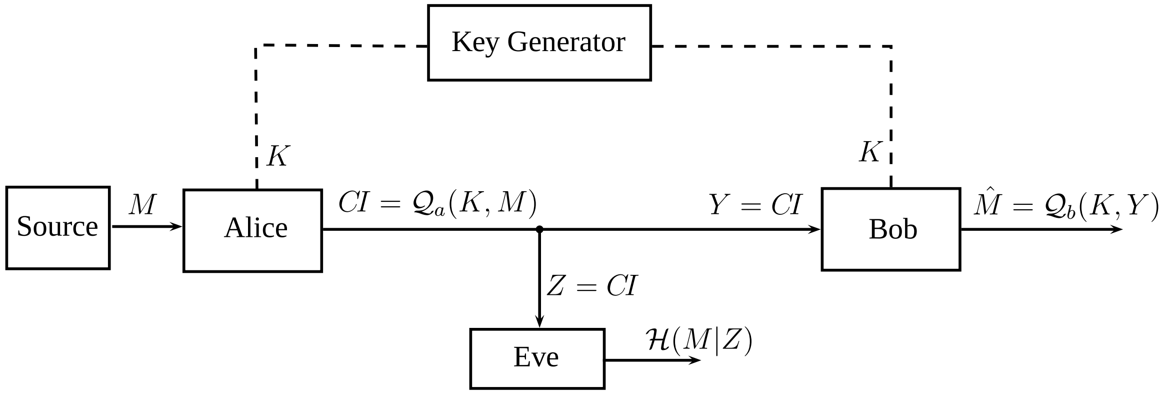


Figure 1.2: Shannon's model: Noise-free secrecy system.

1.2 One-Time Pad System

The problem of secure transmission of messages, which is known as a *secrecy problem*, was mathematically formulated by Shannon [7] in 1949. As shown in Figure 1.2, Shannon's model [7] consists of two noise-free communication channels: a *main channel* from Alice to Bob, and a *wiretap channel* from Alice to Eve. A third party, known as a *key generator*, distributes a common secret, called a *key*, to Alice and Bob before they launch communication such that Eve has no access to the key. In Shannon's model [7], Alice and Bob exploit the *symmetric-key cryptography* [8] to keep the data transmission secure [2].

Specifically, assume $\mathbb{M}$, $\mathbb{K}$, and $\mathbb{CI}$ are the *message set*, the *key set* and the *cipher set*, respectively, such that each set has a fixed finite size. Also, denote the message, the key, and the cipher, which are *random variables* (RVs), by M, K, and CI respectively. Define the function $\mathcal{Q}_a : \mathbb{K} \times \mathbb{M} \to \mathbb{CI}$ as the *encryption function* at Alice and the function $\mathcal{Q}_b : \mathbb{K} \times \mathbb{Y} \to \mathbb{M}$ as the *decryption function* at Bob, where $\mathbb{Y}$ is the output alphabet set of the main channel.

Alice wishes to send the message $M \in \mathbb{M}$ to Bob reliably ($\hat{M} = M$) in the presence of Eve such that Eve is unable to reduce her ambiguity about the message M having Z

[2] For definition of *asymmetric key (public key)* cryptography, see [9].

from the wiretap channel. To achieve this goal, Alice computes *cipher* $\mathit{CI} = \mathcal{Q}_a(K, M)$ and sends it to Bob. Bob, who receives $Y = \mathit{CI}$ from the noise-free main channel, retrieves the message $\hat{M} \in \mathbb{M}$ by $\hat{M} = \mathcal{Q}_b(K, Y)$. Eve obtains the same copy of the cipher as Bob does, i.e., $Z = Y = \mathit{CI}$. Then, she does her best to reduce her level of ambiguity about message M.

The *perfect security condition* guarantees that Eve is unable to reduce her ambiguity about message M with having intercepted signal Z, no matter how much time, memory and computational power she needs. Shannon [7] mathematically formulated this condition by the *(Shannon) entropy function* [4] (see also (2.1) and (2.2)) as follows.

Definition 1.6 (Perfect security (Shannon's criterion for security [7]))**.** Recalling Shannon's model in Figure 1.2, the communication system is said to be perfectly secure against Eve, if there exist functions $\mathcal{Q}_a$ and $\mathcal{Q}_b$ such that

$$\mathcal{H}(M) = \mathcal{H}(M|Z)\,, \tag{1.8}$$

where function

$$\mathcal{H}(M|Z) \triangleq \mathcal{H}(M, Z) - \mathcal{H}(Z)$$

is called the *equivocation function.*

Shannon [7] proved that condition

$$\mathcal{H}(K) \geq \mathcal{H}(M) \tag{1.9}$$

is a necessary condition for the perfect security of symmetric-key cryptography (Shannon's model in Figure 1.2). That is, the uncertainty of the key must be at least as large as the uncertainty of the message. Shannon [4] *proved* that a *one-time pad* system, which was originally proposed by Vernam [10], can implement a perfectly secure communication if a key-stream generator [8, Page 21] with $\mathcal{H}(K) = \mathcal{H}(M)$ is provided. As an example of the one-time pad system, the perfect security of Shannon's model with *Modulo-Additive* functions [8] is established in the following lemma.

Lemma 1.1 (Modulo-Additive functions for the one-time pad system [7]). *Let* $\mathbb{M} = \mathbb{K} = \{1, \ldots, N_0\}$*, where* $N_0 \in \mathbb{N}$*. Let* $\mathcal{Q}_a$ *and* $\mathcal{Q}_b$ *be addition functions modulo* N_0 [3]*, i.e.,*

$$CI{=}\mathcal{Q}_a(K, M){=}K + M \quad (mod\, N_0) \tag{1.10a}$$

$$\hat{M}{=}\mathcal{Q}_b(K, Y) = K + Y \quad (mod\, N_0)\,. \tag{1.10b}$$

Assume that the message $M \in \mathbb{M}$ *is picked up at random according to an arbitrary distribution. If* $K \in \mathbb{K}$ *is a uniformly distributed random variable whose realization is given by the key generator to both Alice and Bob, then Shannon's model is perfectly secure according to Definition 1.6.*

Although ease of implementation of the encryption and decryption functions in a one-time pad system, distribution of a confidential key with $\mathcal{H}(K) \geq \mathcal{H}(M)$, which demands private channels from the key generator to Alice and to Bob, remains a challenging issue in practice. Hence, the implementation of a one-time pad system is not usually practical.

The one-time pad system can also be exploited to provide security in a (noisy) wiretap channel. To illustrate this idea, let us first define the (noisy) wiretap channel with one sender (Alice) and two receivers (Bob and Eve), where Eve is a wiretapper, as follows.

Definition 1.7 (Wiretap channel). A wiretap channel with one sender (Alice) and one main receiver (Bob) and one wiretapper (Eve) is characterized by $(\mathbb{X}^n, \mathbb{Y}^n \times \mathbb{Z}^n, \mathcal{P}_{\mathbf{yz}|\mathbf{x}})$, where

- n is the block length;
- $\mathbf{x}$ is the channel input (from Alice), $\mathbf{y}$ is the first channel output (to Bob), $\mathbf{z}$ is the second channel output (to Eve);
- $\mathbb{X}^n$ is the input alphabet, i.e., $\mathbf{x} \in \mathbb{X}^n$;
- $\mathbb{Y}^n$ is the first output alphabet and $\mathbb{Z}^n$ is the second output alphabet, i.e., $(\mathbf{y}, \mathbf{z}) \in \mathbb{Y}^n \times \mathbb{Z}^n$;
- $\mathcal{P}_{\mathbf{yz}|\mathbf{x}}$ is the probability transition function (PTF) of the wiretap channel.

[3] When $N_0 = 2$, the addition function is called *Exclusive-OR* (XOR).

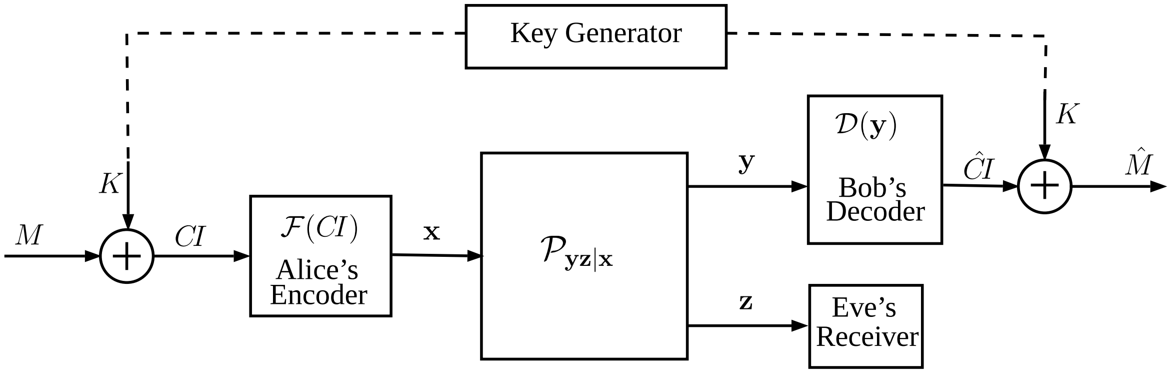


Figure 1.3: Set-up of a one-time pad system in a (noisy) wiretap channel.

The following example shows that concatenation of a one-time pad and an encoder-decoder can provide perfect security in a (noisy) wiretap channel if condition (1.9) is met.

Example 1.1. As shown in Figure 1.3, assume Alice is connected to Bob, the legitimate receiver, and to Eve, the wiretapper, through a (noisy) wiretap channel $(\mathbb{X}^n, \mathbb{Y}^n \times \mathbb{Z}^n, \mathcal{P}_{\mathbf{y},\mathbf{z}|\mathbf{x}})$. A uniformly distributed key $K \in \{1, \ldots, N_0\}$ is provided to both Alice and Bob by a key generator. Alice wishes to send a message $M \in \{1, \ldots, N_0\}$ over the channel to Bob with perfect security in presence of Eve. To do this, Alice, first enciphers the message M into a cipher CI by $CI = \mathcal{Q}_a(K, M)$ according to (1.10a); then, she encodes CI by an encoder $\mathcal{F}$ into $\mathbf{x}$ according to (1.2a) and sends it. Correspondingly, the receiver first decodes the received signal $\mathbf{y}$ by a decoder $\mathcal{D}$ according to (1.2b) so that he recovers his cipher $\hat{CI} \in \mathbb{M}$, and finally he calculates his message $\hat{M} = \mathcal{Q}_b(K, \hat{CI})$ according to (1.10b). In this case, the communication is perfectly secure according to Definition 1.6; however, the reliability of the communication depends on the choice of functions $\mathcal{F}$ and $\mathcal{D}$ as well as the communication rate $\frac{\mathcal{H}(M)}{n}$.

To design a practical symmetric-key cryptographic algorithm, a key with a reasonable size is applied which violates condition (1.9). That key must provide a given level of conditional security which depends on the application and on the proficiency of the (possible) wiretappers. That is, cryptographic algorithms are usually designed based on *computational security* and/or complexity of known attacks rather than the perfect security. In these cases, eavesdroppers are assumed to be equipped with limited processing power and time. A computationally secure algorithm is designed based on either the assumed yet unproven hardness of a certain problem such as discrete logarithm or a proof with certain computational restrictions. Generally, the accurate meaning of the computational security depends on the applications for which the cryptographic algorithm is to be designed and on the constraints assumed for wiretappers. Despite the practical benefit of such a cryptographic algorithm which demands a key with a reasonable size, it might be broken some day due to rapid growth of computational processors. Hence, these algorithms should be maintained with new cryptographic techniques such that they fulfill the requested level of computational security against eavesdroppers with up-to-date technologies.

1.3 Information-Theoretic Security

The perfect security in Figure 1.3 is acquired by concatenation of a one-time pad system and a channel encoder-decoder. That is, the enciphering-deciphering step is separated from the encoding-decoding step. In Example 1.1, the overall channel from input terminal of the encoder to the output terminal of the decoder is assumed to be noiseless. Hence, condition (1.9) is applied to the design of the one-time pad system in Figure 1.3.

Now, suppose a noisy wiretap channel in which the encoding and enciphering can be jointly implemented in one step as well as the decoding and deciphering. The question is if condition (1.9) is still required to provide perfect security in this new model. If yes, a perfectly secure communication system would be still practically infeasible because of the key-distribution issue. If no, violation of condition (1.9) might result in an implementation of a secure communication system with no key-distribution concerns.

In 1975, Wyner [11] *proved* that the security[4] is achievable in an asymptotic sense (see Definition 1.8) with no pre-shared key if Eve receives a noisy copy of Bob's received signal

[4] We reserve term *perfect security* only for Shannon's criterion for security given in Definition 1.6.

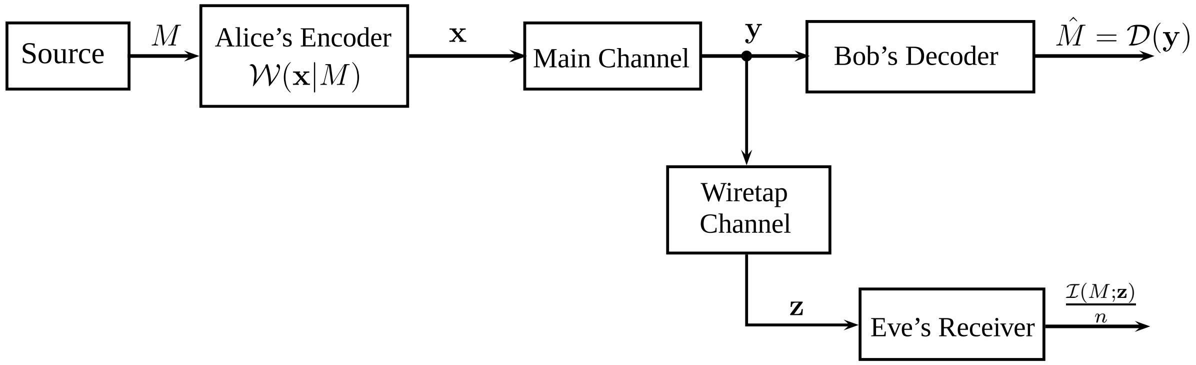


Figure 1.4: Wyner's model: Degraded noisy wiretap model.

from Alice. This work justified the idea that sending a positive secure rate is possible without using any key if Bob has an advantage over Eve. The achievable security, which is referred to as *information-theoretic security*, is guaranteed even if Eve has unlimited computational power and time. As shown in Figure 1.4, Wyner's model is based on the fact that the legitimate receiver has a physical advantage over the wiretapper in most real communication systems. In Wyner's model, Bob's advantage over Eve is that Eve receives a *degraded* version [2, 6] of Bob's signal, e.g., the eavesdropper is located further from the transmitter than the legitimate receiver. Wyner [11] exploited this advantage rather than a common key between Alice and Bob to securely transmit a message.

For his model in Figure 1.4, Wyner [11] assumed that the main channel and the wiretap channel are DMCs. He developed the idea of using a (block) wiretap coding scheme to achieve an (asymptotically) secure and reliable communication over the wiretap model. The wiretap coding scheme consists of three components: an *encoder* $\mathcal{W}$, a *decoder* $\mathcal{D}$, and a *wiretap codebook* [3, 6], where the block length is assumed to be n. In Figure 1.4, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are three random vectors which represent the emitted signal by Alice, the received signal by Bob, and the received signal by Eve, respectively. An objective of the wiretap coding scheme is to provide security in addition to reliability in an asymptotic sense.

When no security condition matters, in Section 1.1, we mentioned that the AR condition is met by using a pair of suitable deterministic encoder-decoder for a given DMC from Alice to Bob. When security is a concern in the wiretap channel, Wyner [11] suggested using a *stochastic encoder*[5], where a given message M maps randomly into a codeword $\mathbf{x}$ from a set of codewords according to stochastic function $\mathcal{W}$. In other words, a wiretap codebook consists of all mappings from any message to the corresponding *set*[6] of codewords such that the total sets partition the set of all codewords in that wiretap codebook. Hence, the randomized encoding function is represented by conditional distribution function $\mathcal{W}(\mathbf{x}|M)$, i.e.,

$$\forall M \in \mathbb{M} : \quad \sum_{\mathbf{x} \in \mathbb{X}^n} \mathcal{W}(\mathbf{x}|M) = 1 \tag{1.11}$$

where the elements of matrix $\mathcal{W}(\mathbf{x}|M)$ are non-negative. The role of the randomized encoder in Wyner's model is illustrated as follows.

The stochastic encoder encodes a message as if it adds (applies) an artificial random noise to the output of a deterministic encoder. Bob considers this noise as extra noise on top of noise of the main channel, and so he decodes the message from the noisy signal at expense of loosing some portion of his achievable rate with respect to the AR condition 1.3. However, this noise can exhaust the capacity of Eve's channel as she receives a degraded (noisier) version of Bob's signal, and so she can attain no information about the message.

Motivated by Wyner [11], the security level of a wiretap channel against Eve is measured by *equivocation rate* [13] (see Subsection 2.2.1). In this thesis, we examine the security of the communication in an asymptotic sense according to the following definition.

Definition 1.8 (Asymptotic security (AS) condition)**.** Let $\mathbb{M} = \{1, \ldots, \lceil 2^{nR} \rceil\}$, where $R \in \mathbb{R}^+ \cup \{0\}$ and $n \in \mathbb{N}$. Assume message $M \in \mathbb{M}$ is a uniformly distributed random variable. The communication from Alice to Bob over wiretap channel $(\mathbb{X}^n, \mathbb{Y}^n \times \mathbb{Z}^n, \mathcal{P}_{\mathbf{yz}|\mathbf{x}})$ is said to be asymptotically secure if there exists a stochastic function $\mathcal{W}$ according to (1.11) such that

$$\lim_{n \to \infty} \frac{1}{n} \mathcal{I}(M; \mathbf{z}) = 0 \, . \tag{1.12}$$

[5] It is also known as a randomized encoder [12].

[6] This set is also known as a *bin* in information theory.

where $\mathcal{I}(M;\mathbf{z})$ is obtained from joint distribution function

$$\mathcal{P}(M,\mathbf{z}) = \frac{1}{|\mathbb{M}|} \sum_{(\mathbf{x},\mathbf{y})\in\mathbb{X}^n\times\mathbb{Y}^n} \mathcal{P}_{\mathbf{yz}|\mathbf{x}}(\mathbf{y},\mathbf{z}|\mathbf{x})\mathcal{W}(\mathbf{x}|M)\ .$$

Further, rate R is said to satisfy the AS condition.

In the field of information-theoretic security, a fundamental objective of a secrecy problem is to find the *secrecy capacity* [11] of a given wiretap channel as introduced below[7].

Definition 1.9 (Secrecy capacity)**.** An achievable rate R with respect to the AR condition (1.3) in Definition 1.3 is called a *securely achievable rate* if it satisfies the AS condition according to Definition 1.8 as well. The supremum of all securely achievable rates is called the secrecy capacity.

Wyner [11] characterized the secrecy capacity of a discrete memoryless wiretap channel, i.e. the main channel and the wiretap channel are both DMCs in Figure 1.4 (see Subsection 2.2.1 for more details). Leung-Yan-Cheong [14] simplified the secrecy capacity of Wyner's model where both the main channel and wiretap channel are *symmetric* DMCs [1, Page 94]. This secrecy capacity, which is achieved at the uniform input distribution on the main channel, is simplified to

$$C_s = C_m - C_{mw}\ , \tag{1.13}$$

where C_m and C_{mw} are the capacity of the main channel, from Alice to Bob, and the capacity of the overall main-wiretap channel, from Alice to Eve, respectively.

As shown in Figure 1.5, Leung-Yan-Cheong and Hellman [13] extended Wyner's model to a (physically) degraded[8] Gaussian wiretap channel in which the main channel and the wiretap channel are two continuous[9] memoryless (and discrete time) channels [1, Ch. 7]

[7] In fact, the *capacity-equivocation region* of the wiretap channel is the main goal of the secrecy problem (see Subsection 2.2.1 for more details).

[8] See Subsection 2.2.1 for definitions of different types of degradedness.

[9] The input and output alphabets of a continuous channel each consist of a set of real numbers.

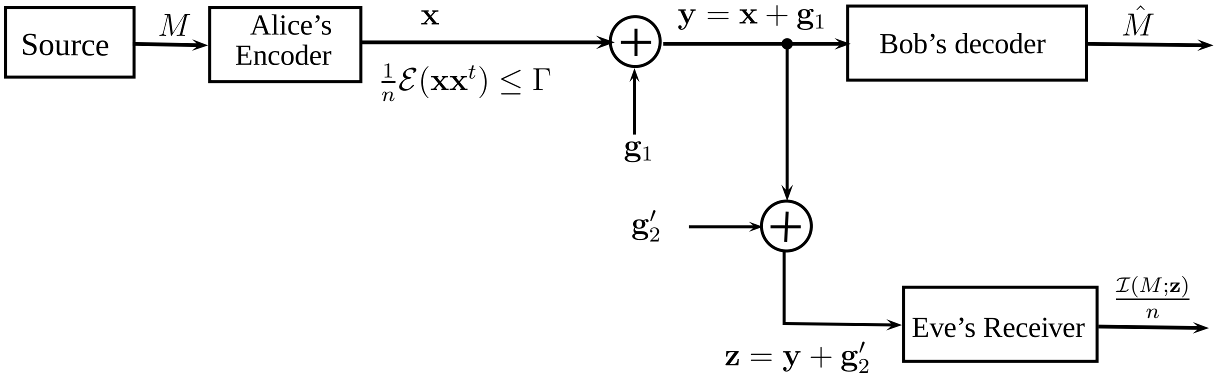

Source
M
Alice's Encoder
$\mathbf{x}$
$\frac{1}{n}\mathcal{E}(\mathbf{x}\mathbf{x}^t) \leq \Gamma$
$\mathbf{y} = \mathbf{x} + \mathbf{g}_1$
Bob's decoder
$\hat{M}$
$\mathbf{g}_1$
$\mathbf{g}_2'$
Eve's Receiver
$\frac{\mathcal{I}(M;\mathbf{z})}{n}$
$\mathbf{z} = \mathbf{y} + \mathbf{g}_2'$

Figure 1.5: The Gaussian wiretap channel.

with independent *additive white Gaussian noise* (AWGN) [1, Sec. 7.4]. Specifically, random vectors $\mathbf{g}_1$ and $\mathbf{g}_2'$ are independent Gaussian noise with *independent and identically distributed* (i.i.d.) components having distribution $\mathcal{N}(0, \sigma_1^2)$ and distribution $\mathcal{N}(0, \sigma_2^2 - \sigma_1^2)$, respectively, where $\sigma_2 \geq \sigma_1$. Alice's transmitter is subject to average power constraint $\frac{1}{n}\mathcal{E}(\mathbf{x}\mathbf{x}^t) \leq \Gamma$ as well [1, Sec. 7.4]. For the Gaussian wiretap channel, the authors [13] showed that (1.13) holds, i.e.,

$$C_S = \frac{1}{2}\log\left(\frac{1+\frac{\Gamma}{\sigma_1^2}}{1+\frac{\Gamma}{\sigma_2^2}}\right) . \tag{1.14}$$

Csiszár and Körner [12] generalized Wyner's result to a *broadcast channel* (BC) [2,15] with a transmitter (Alice) and two receivers (Bob and Eve). They [12] characterized the secrecy capacity of the BC, where both Bob and Eve decode a common message and Bob retrieves an extra private message from his received signal. For this model, they proved that a non-zero secrecy capacity always exists unless Eve's channel is *less noisy* [16] (see Definition 2.9) than Bob's channel.

1.4 The Key Agreement Problem

Another notable problem in the field of information-theoretic security is the *key agreement problem* which was introduced by Maurer [17] as well as Ahlswede and Csiszár [18, 19]. In this section, the concept of the key agreement problem is illustrated in Subsection 1.4.1. Main key agreement models and related work are reviewed in Subsection 1.4.2. Finally, the privacy amplification is introduced in Subsection 1.4.3 for security enhancement of agreed keys.

1.4.1 The Concept

In an information-theoretic key agreement problem, generally, some legitimate parties on a communication network wish to share a secret (key) reliably by using a (block) key agreement coding scheme such that wiretappers are unable to gain any information about the key. To do this, some signals are to be communicated between authorized parties (subject to constraints of the network) in presence of the wiretappers. This step, which is known as *key exchange*, is usually performed prior to the transmission of messages. At the end of the key exchange step, each authorized party retrieves his/her own key by decoding total available signals at his/her terminal. The agreed keys are finally to be exploited for symmetric-key cryptography [8] in the step of the transmission of messages [17, 18, 20]. Hence, the keys are selected from a same alphabet (key) set with a finite size, which is denoted by $\mathbb{K}$ in this thesis. Moreover, the agreed keys must satisfy the AR condition and the AS condition in the same way as messages $(M, \hat{M}) \in \mathbb{M}^2$ do in Definition 1.3 and Definition 1.8, respectively.

Further to these two conditions, any key is required to look like a uniformly distributed random variable as the third condition; thus, the key is often called a *secure common randomness* in the literature, e.g., [3, 18, 19, 21]. This condition is called the *asymptotic randomness condition* (ARN) as block length $n \to \infty$. Near uniformity of a RV with a finite size can be measured in various ways; one popular way in information theory, which is used in our work, is to compare the entropy rate of the RV with that of a uniformly distributed RV with the same size [3, Ch. 17].

These conditions are clarified as follows based on the key agreement model sketched in Figure 1.6. In Figure 1.6, assume that Alice and Bob want to share a key by key exchange over the wiretap channel $(\mathbb{X}^n, \mathbb{Y}^n \times \mathbb{Z}^n, \mathcal{P}_{\mathbf{yz}|\mathbf{x}})$ according to Definition 1.7 as well as over a *public channel* [17] (see Definition 3.1 for more details). Due to the nature of a public channel, any signal sent over this channel is accessible by Eve. Although the eavesdropper is able to obtain the same copy of transmitted signals over the public channel, she is not able to tamper those signals in any sense as the channel is assumed to be *authenticated*[10]. For generalization, assume that Alice, Bob, and Eve have also access to side information $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{e}$, respectively (see Chapter 3 for more details.). Let map the total signals transmitted over the public channel to a random variable denoted by P. After the key exchange step, Alice and Bob retrieve key $K \in \mathbb{K}$ and $\hat{K} \in \mathbb{K}$ from all available information at her/his terminal, respectively.

For the key exchange step an *admissible key agreement coding scheme*, including encoders, decoders, key generators, and a key agreement codebook, is utilized based on restrictions of a given model (see Chapter 3 and Chapter 4 for more details and examples). After the key exchange step, the agreed keys must satisfy the following conditions such that they are acceptable for symmetric-key cryptography.

Definition 1.10. Assume an admissible key agreement coding scheme with key set $\mathbb{K} = \{1, \ldots, \lceil 2^{nR_K} \rceil\}$, where $R_K \in \mathbb{R}^+ \cup \{0\}$ and $n \in \mathbb{N}$. Let $K \in \mathbb{K}$ and $\hat{K} \in \mathbb{K}$ be Alice's key and Bob's key at the end of the key exchange step. The efficiency of the admissible key agreement coding scheme is examined by the following functions.

1. The reliability of $(K, \hat{K})$ is measured by the average probability of error

$$\mathcal{P}_{error}(n) \triangleq \mathscr{P}\{\hat{K} \neq K\} , \tag{1.15}$$

which is a function of block length n.

[10] See [22–24] for key agreement problems over unauthenticated public channels.

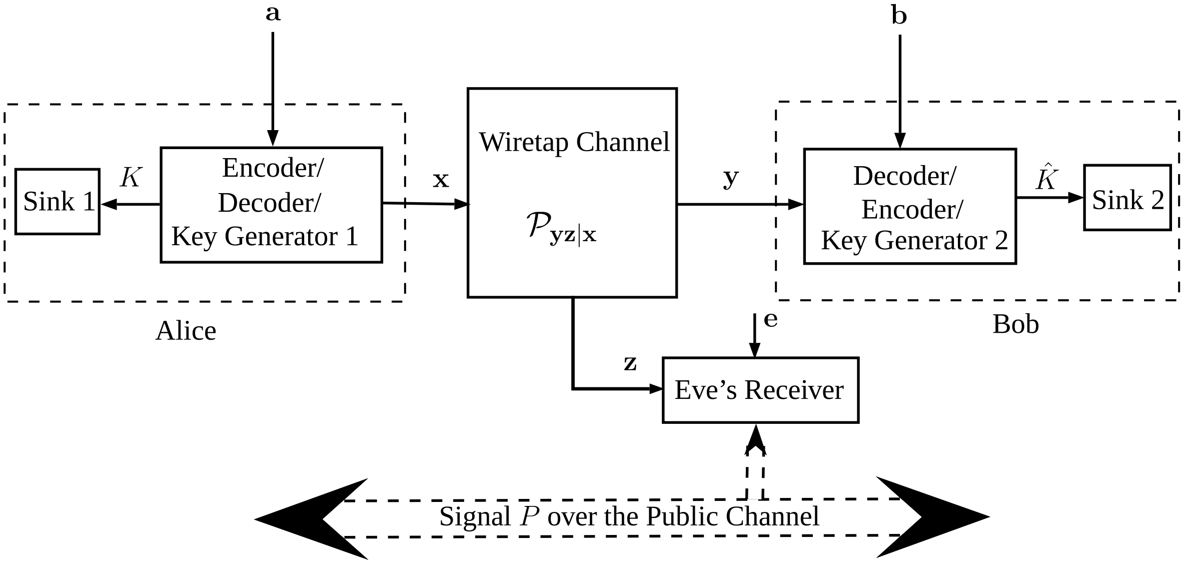


Figure 1.6: Key agreement between Alice and Bob in presence of Eve.

2. The security level of K is measured by the leakage rate

$$\mathcal{R}_{_L}(n) \triangleq \frac{1}{n}\mathcal{I}(K; \mathbf{z}, \mathbf{e}, P) \,, \tag{1.16}$$

   which is a function of block length n.

3. The randomness of K is measured by[11]

$$\mathcal{X}(n) \triangleq \frac{\log(|\mathbb{K}|) - \mathcal{H}(K)}{n} \,, \tag{1.17}$$

   which is a function of block length n.

A fundamental objective of a key agreement problem is to find the key capacity of a given model according to the following definition.

[11] $\log(|\mathbb{K}|) - \mathcal{H}(K) = \mathcal{D}_{KL}(\mathcal{P}_K||\mathcal{P}'_K)$, which is Kullback-Leibler divergence function [2,3] between distribution $\mathcal{P}_K$ of K and the uniform distribution $\mathcal{P}'_K(k) = \frac{1}{|\mathbb{K}|}$ on set $\mathbb{K}$.

Definition 1.11. A key rate $R_{_K}$ is said to be *achievable* if there exists an admissible key agreement code such that[12]

$$\begin{aligned}
\textit{The AR Condition}: &\quad \lim_{n\to\infty} \mathcal{P}_{error}(n) = 0\,, && (1.18a)\\
\textit{The AS Condition}: &\quad \lim_{n\to\infty} \mathcal{R}_{_L}(n) = 0\,, && (1.18b)\\
\textit{The ARN Condition}: &\quad \lim_{n\to\infty} \mathcal{X}(n) = 0\,, && (1.18c)\\
\textit{The Key Rate}: &\quad \liminf_{n\to\infty} \frac{\mathcal{H}(K)}{n} \geq R_{_K}\,. && (1.18d)
\end{aligned}$$

For a given key agreement problem, the supremum of all achievable key rates is called the key capacity of the given model, and it is denoted by $C_{_K}$.

As a special case, when the public channel is one-way from Alice to Bob (forward direction), it is referred to as the forward public channel and the corresponding key capacity is called *forward key capacity.* Similarly, a unilateral public channel from Bob to Alice (backward direction) is called a backward public channel, and the corresponding key capacity is called *backward key capacity.*

Now, suppose that Alice and Bob complete the key exchange step such that their agreed keys satisfy conditions (1.18). Then, Alice wishes to send a message securely to Bob by using a symmetric-key cryptographic algorithm. How much is the largest achievable secure rate by using the agreed keys for encryption-decryption? This question is addressed by the following lemma[13].

Lemma 1.2. *Let* $\mathbb{K} = \{1, \ldots, \lceil 2^{nR_K} \rceil\}$ *and* $N_0 = \lceil 2^{nR_K} \rceil$. *Assume that Alice and Bob agree on keys* $(K, \hat{K}) \in \mathbb{K}^2$, *which satisfy conditions* (1.18). *Having the agreed keys, Alice then encrypts and sends message* $M \in \{1, \ldots, N_0\}$, *independent of* K, *to Bob according to*

[12] In Appendix A, we prove that the AS condition for K together with the AR condition for $(K, \hat{K})$ implies the AS condition for $\hat{K}$.

[13] This lemma is originally given in Ahlswede-Csiszár [18, Lem. 2.1] for a noiseless wiretap channel. However, we state the lemma with a slight generalization for a noisy wiretap channel according to Example 1.1.

Figure 1.3, where

$$\begin{aligned} CI = & \mathcal{Q}_a(K, M) = K + M \quad (mod\, N_0) \\ \hat{M} = & \mathcal{Q}_b(\hat{K}, CI) = \hat{K} + \hat{C}I \quad (mod\, N_0)\,. \end{aligned}$$

according to (1.10)*. Then, rate* $R = \min\{R_{_K}, C_{_S}\}$ *is a securely achievable rate according to Definition 1.9, where* $C_{_S}$ *is the secrecy capacity of the wiretap channel.*

In general, any securely achievable rate sent over the wiretap channel $(\mathbb{X}^n, \mathbb{Y}^n \times \mathbb{Z}^n, \mathcal{P}_{\mathbf{yz}|\mathbf{x}})$ can be exploited as a secure common randomness (key) between Alice and Bob. Hence,

$$C_{_K} \geq C_{_S} \tag{1.19}$$

holds for any given wiretap model. In Chapter 3 and Chapter 4, we show that this inequality can be strict.

1.4.2 The Source-Type Model vs. The Channel-Type Model

For the key agreement between Alice and Bob in presence of Eve, Ahlswede and Csiszár [18] introduced two fundamental paradigms[14]: the source-type model and the channel-type model.

The source-type model [18] has a *discrete memoryless multiple source* (DMMS) with three-component vectors $(\mathbf{a}, \mathbf{b}, \mathbf{e})$ distributed according to *probability mass function* (PMF) $\mathcal{P}(\mathbf{a}, \mathbf{b}, \mathbf{e}) = \prod_{i=1}^{n} \mathcal{P}_{ABE}(a_i, b_i, e_i)$. Source vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{e}$ are available non-causally (prior to the key exchange step) at Alice, Bob, and Eve, respectively. Moreover, there exists a noiseless public channel with unlimited capacity for two-way communication between Alice and Bob. The source-type key agreement model is sketched in Figure 1.7.

Ahlswede and Csiszár [18] characterized the forward key capacity of the source-type

[14] These paradigms are still studied by many researchers in the field of information-theoretic key agreement, e.g., see publications [25–27]. On the other hand, papers [28, 29] studied the key agreement over two-way DMCs with no public channel.

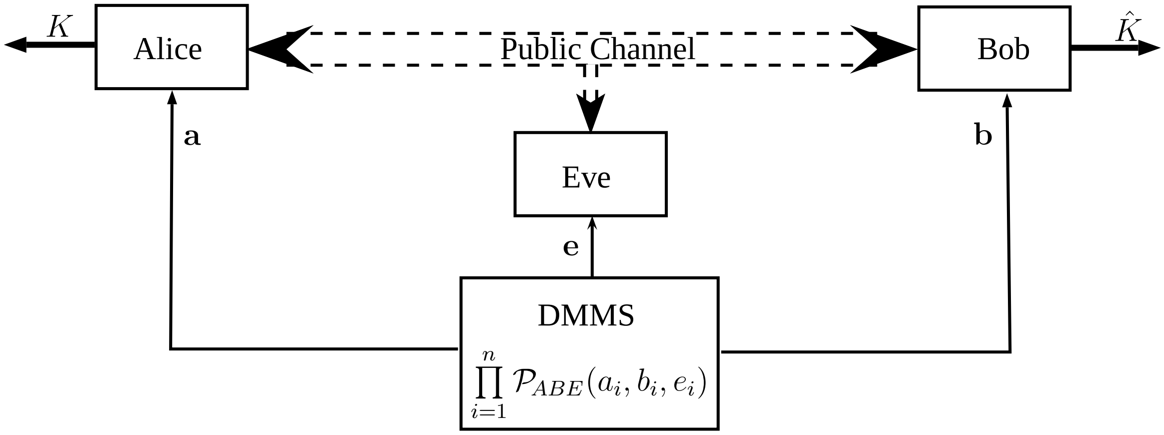


Figure 1.7: Key agreement over a source-type model.

model as

$$C_{_K} = \max_{\mathcal{P}_{U|A}\,\mathcal{P}_{W|U}} [\mathcal{I}(U; B|W) - \mathcal{I}(U; E|W)] \tag{1.20}$$

where U and W are auxiliary RVs such that $W \to U \to A \to (B, E)$ forms a Markov chain. They [18] also attained the following *upper bound* (UB) on the key capacity of the source-type model:

$$C_{_K} \leq \mathcal{I}(A; B|E) \,. \tag{1.21}$$

This UB is tight for the following special cases:

1. when $\mathbf{e}$ is independent of $(\mathbf{a}, \mathbf{b})$, e.g., $\mathbf{e} = \mathbf{0}$; in this case

$$C_{_K} = \mathcal{I}(A; B) \,, \tag{1.22}$$

2. when $A \to B \to E$ forms a Markov chain; in this case $C_{_K} = \mathcal{I}(A; B) - \mathcal{I}(A; E)$,

3. when $A \to E \to B$ forms a Markov chain, then $C_{_K} = 0$,

4. when either Alice or Bob has access to Eve's signal[15], then $C_{_K} = \mathcal{I}(A; B|E)$.

[15] This situation rarely occurs in real world as Eve has no motivation to share her intercepted signal with

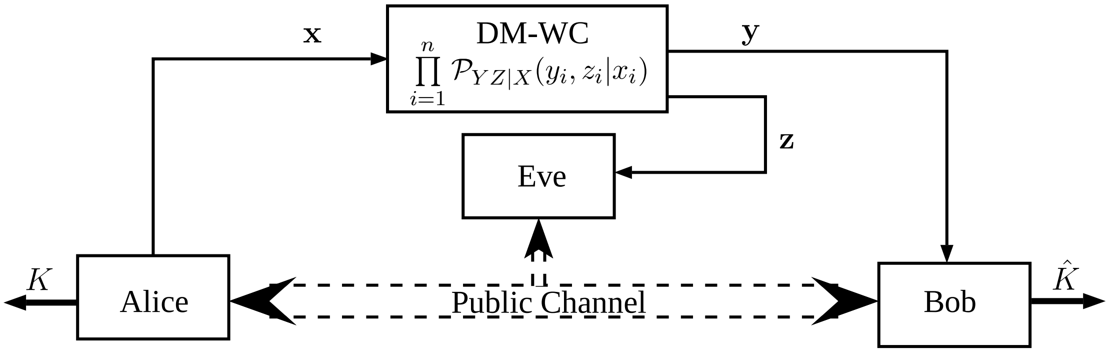


Figure 1.8: Key agreement over a channel-type model.

For the first case, the authors [18] constructed the key agreement coding scheme based on Slepian-Wolf source coding strategy [32]. In this case, the key capacity is achieved with only one single transmission over the public channel in either forward or backward direction. Hence, the key capacity, the forward key capacity, and the backward key capacity are all equal [18, Prop 1].

For the second case, the key capacity is achieved with only one single transmission over the public channel in the forward direction. Hence, the key capacity and the forward key capacity are equal [18, Thm. 1].

For the last case, the key capacity equals the backward or forward key capacity, respectively, according as Alice or Bob is informed [18, Thm. 3].

In a channel-type model, the parties have access to two channels: a (noisy) *discrete memoryless wiretap channel* (DM-WC) with PTF $\mathcal{P}(\mathbf{y}, \mathbf{z}|\mathbf{x}) = \prod_{i=1}^{n} \mathcal{P}_{YZ|X}(y_i, z_i|x_i)$, and a noiseless public channel with unlimited capacity in both directions. This model is depicted in Figure 1.8.

Ahlswede and Csiszár [18] obtained the forward key capacity of the channel-type model as

$$C_{\scriptscriptstyle K} = \max_{\mathcal{P}_U \mathcal{P}_{X|U}} \left[\mathcal{I}(U;Y) - \mathcal{I}(U;Z) \right], \tag{1.23}$$

legitimate parties [30, 31].

where U is an auxillary RV such that $U \to X \to (Y, Z)$. The forward key capacity equals the secrecy capacity of the corresponding DM-WC[16] [12], and no transmission over the public channel is required for key generation [18, Thm. 2]. Also, they [18] derived a UB on the key capacity of the channel-type model, i.e.,

$$C_{_K} \leq \max_{\mathcal{P}_X} \mathcal{I}(X; Y|Z) \,, \tag{1.24}$$

where $\mathcal{P}_{_X}$ is the input distribution on the DM-WC. This UB is tight for the following special cases:

1. when the DM-WC becomes a (private) DMC from Alice to Bob as sketched in Figure 1.9; in this case
$$C_{_K} = \max_{\mathcal{P}_X} \, \mathcal{I}(X; Y) \,, \tag{1.25}$$
2. when Eve's channel is a degraded version[17] of Bob's channel, i.e., $X \to Y \to Z$; in this case, $C_{_K} = \max\limits_{\mathcal{P}_X} [\mathcal{I}(X; Y) - \mathcal{I}(X; Z)]$,
3. when Bob's channel is a degraded version of Eve's, i.e., $X \to Z \to Y$, where the key capacity in this case is [18, Thm. 2]
$$C_{_K} = 0 \,, \tag{1.26}$$
4. when either Alice or Bob has full access to Eve's received signal $\mathbf{z}$; in this case $C_{_K} = \max\limits_{\mathcal{P}_X} \mathcal{I}(X; Y|Z)$,
5. when the outputs of the DM-WC are independent given its input, i.e., $Y \to X \to Z$; the key capacity in this case is [18, Thm. 2-Corol. 2]
$$C_{_K} = \max_{\mathcal{P}_X} [\mathcal{I}(X; Y) - \mathcal{I}(Y; Z)] \,. \tag{1.27}$$

[16] See Section 2.2.1 for the secrecy capacity of the DM-WC.

[17] See Section 2.2.1 for the definition of degradedness.

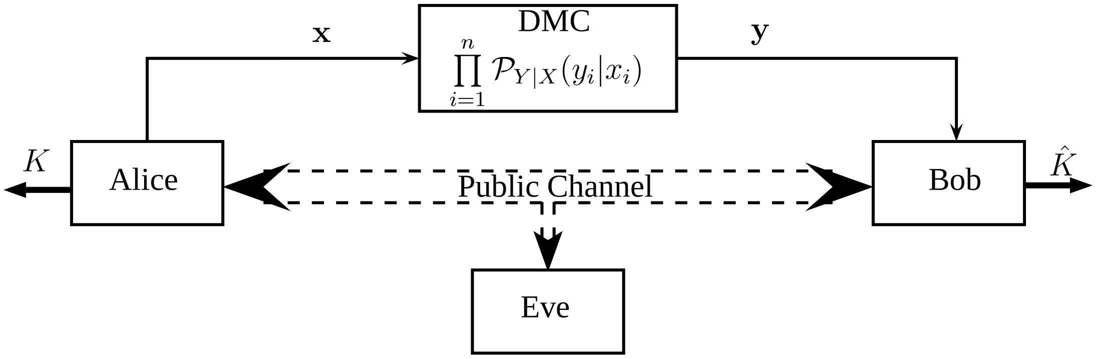


Figure 1.9: Key agreement over a channel-type model with a private DMC.

For the first case, the DMC is intrinsically secure against Eve, and the key capacity of the model equals the (ordinary) capacity of the DMC, which is achievable without using the public channel at all [18, Prop. 1]. For the second case, the key capacity equals the secrecy capacity of the wiretap channel, and the public channel is not used at all [18, Thm. 2]. This means that the public channel is not required for key generation in cases 1 and 2. In other words, for these cases, the key capacity, the forward key capacity, and the backward key capacity are all equal.

For the fourth case, the key capacity equals the backward or forward key capacity, respectively, according as Alice or Bob is informed [18, Thm. 3]. In this case, the known signal of the wiretapper at one of legitimate terminals may contribute in key generation more than the first case where the wiretapper has no access to the main channel. This is due to the fact that the positive effect of the information gained by authorized parties from Eve's signal can be more than the negative effect of her presence in the key agreement.

For the fifth case, the key capacity equals the backward key capacity, which is achieved with one single transmission over the public channel in the backward direction. Also, it is generally larger than forward key capacity of the model [18, Thm. 2-Corol. 2]. This case was originally studied by Maurer [17, 33]. He assumed a *binary symmetric channel* (BSC) [1] as the main channel (from Alice to Bob) and a BSC as the wiretap channel (from

Alice to Eve) such that the channels have *independent noise*. Using the public channel in *backward* direction for key generation, he proved that achieving a positive key rate is feasible even if the capacity of the wiretap channel (from Alice to Eve) exceeds that of the main channel (from Alice to Bob).

In similar and parallel publications, Khisti *et al.* [34] and Prabhakaran *et al.* [35, 36] merged the source-type model with the channel-type model, where the parties have access to a DMMS further to a DM-WC. This combined model has no public channel at all.

Khisti *et al.* [34] achieved a *lower bound* (LB) and a UB on the key capacity. The bounds coincide when the wiretap channel is a product of *reversely degraded* channels[18] [6, 37].

Prabhakaran *et al.* [35,36] investigated a trade-off between the key capacity and secrecy capacity of the model. They applied a *separation strategy* which converts the DM-WC into a public *bit pipe* and a private *bit pipe* [35,36]. According to this sub-optimal strategy, an achievable key rate is obtained by the use of the private pipe (from Alice to Bob). With the help of these pipes, another key rate is added to the last one from the correlated source components (see [35, Thm. 1] for more details). The strategy is shown to become optimum when both the DM-WC and the DMMS can be decomposed into product of two degraded BCs [37] and product of two degraded DMMS, respectively. The first sub-channel and sub-source components are degraded in Bob's favor, and the second ones are degraded in Eve's favor [35, 36].

Nitinawarat [38] as well as Watanabe and Oohama [39, 40] investigated the key capacity for a source-type model where the DMMS is replaced by a Gaussian multiple source. Specifically, Nitinawarat [38] assumed that Alice and Bob have correlated Gaussian SI and Eve has no SI. Applying a rate limited quantization on the Gaussian SI, the author calculated the key capacity given in (1.22) for the Gaussian RVs in his model [38]. Watanabe and Oohama [39, 40] considered three correlated Gaussian vectors generated i.i.d. according to a *fixed* covariance matrix. Each vector is given to one party. They investigated the forward key capacity over a rate limited one-way public channel with capacity C_{P_1}. The forward key capacity is obtained as a function of the public channel capacity for the

[18] Also, it is referred to as a degraded channel in Eve's favor [35].

following special case.

Theorem 1.1. *Consider a source-type model, which is sketched in Figure 1.7, with a Gaussian multiple source such that its public channel is one-way in the forward direction with capacity $C_{P_1} \in [0, \infty)$. Specifically, assume $(\mathbf{a}, \mathbf{b}, \mathbf{e})$ is i.i.d. according to $(A, B, E) \sim \mathcal{N}((0,0,0), \mathbf{\Sigma}_{3\times 3})$, where*

$$\mathbf{\Sigma}_{3\times 3} = \begin{pmatrix} \Sigma_a & \Sigma_{ab} & \Sigma_{ae} \\ \Sigma_{ba} & \Sigma_b & \Sigma_{be} \\ \Sigma_{ea} & \Sigma_{eb} & \Sigma_e \end{pmatrix} \tag{1.28}$$

is a positive definite covariance matrix with $\Sigma_{ab} \neq 0$. If Markov chain $A \to B \to E$ holds, then

$$C_K = \frac{1}{2} \log \left(\frac{\Sigma_{b|ae}\, 2^{-2C_{P_1}} + \Sigma_{b|e}(1 - 2^{-2C_{P_1}})}{\Sigma_{b|ae}} \right) , \tag{1.29}$$

where $\Sigma_{b|ae}$ and $\Sigma_{b|e}$ are the conditional variance of B given (A, E) and E, respectively.

1.4.3 Privacy Amplification

Comparing the security condition given in Definition 1.8 with Shannon's criterion of security (perfect security) given in Definition 1.6 implies that a secure communication is the sense of Definition 1.8 is *not* necessarily a secure communication in the sense of Definition 1.6. That is, Definition 1.8 requires only Eve's ratio of information about the message to be negligible. Hence, she can gain a possibly considerable amount of information about the message. Similarly, Condition 1.18b guarantees that only Eve's ratio of information about the key is negligible but not her information about any single bit of the key. On the other hand, a pair of agreed keys according to Definition 1.11 guarantees only a secure communication in the sense of Definition 1.8 according to Lemma 1.2.

Conclusively, Definition 1.8 and Definition 1.11 are too weak to provide privacy requirements [7,8] for the entire message and the entire key, respectively. Thus, the AS conditions in these definitions are referred to as the *weak sense* of security in the literature [3, 41]. Also, the achievable key rate and key capacity introduced in Definition 1.11 are known as the weak sense of achievable key rate and that of key capacity, respectively [19].

New definition for the achievable key rate is given below to assure that the entire key is secure. Following [3], we first combine the security and the randomness conditions of Definition 1.11 into a security index

$$S(n) \triangleq \log(|\mathbb{K}|) - \mathcal{H}(K|\mathbf{z}, \mathbf{e}, P) \,, \tag{1.30}$$

where K, $\mathbf{z}$, $\mathbf{e}$ and P are defined in Subsection 1.4.1 according to Figure 1.6.

The reliability-exponent for the average probability of error 1.15 in a key agreement problem can be defined in a similar way as it was given in Definition (1.5) for messages. In the following definition, the reliability-exponent and the security-exponent for the key agreement model sketched in Figure 1.6 are defined.

Definition 1.12. Assume an admissible key agreement coding scheme with key set $\mathbb{K} = \{1, \dots, \lceil 2^{nR_K} \rceil\}$, where $R_{_K} \in \mathbb{R}^+ \cup \{0\}$ and $n \in \mathbb{N}$. Let $K \in \mathbb{K}$ and $\hat{K} \in \mathbb{K}$ be Alice's key and Bob's key at the end of the key exchange step. The reliability condition (1.18a) and security condition 1.18b are said to be met with reliability-exponent ε_r and security-exponent ε_s, respectively, if there exists an admissible key agreement coding scheme such that

$$\liminf_{n\to\infty} -\frac{1}{n}\log(\mathcal{P}_{error}(n)) \geq \varepsilon_r \,, \tag{1.31a}$$

$$\liminf_{n\to\infty} -\frac{1}{n}\log(\mathcal{S}(n)) \geq \varepsilon_s \,. \tag{1.31b}$$

where $\mathcal{P}_{error}(n)$ and $\mathcal{S}(n)$ are given in (1.15) and (1.30), respectively.

In the following, a strong sense of key capacity is defined. This definition guarantees the privacy requirement [7, 8]of any single bit of the agreed keys.

Definition 1.13. A key rate $R_{_K}$ is said to be *strongly achievable* if conditions (1.18) are

[19] In this thesis, we treat the weak sense of security, achievable key rate, key capacity unless otherwise is stated.

satisfied with reliability-exponent $\varepsilon_r > 0$ and security-exponent $\varepsilon_s > 0$, respectively. The supremum of all strongly achievable key rates is called the key capacity in the strong sense.

Comparing Definition 1.13 with Definition 1.11, the key capacity in the strong sense is generally a lower bound on the key capacity in the weak sense. However, Maurer and Wolf [41] proved that the key capacity of a source-type model in the strong sense equals the key capacity of that model in the weak sense. They [41] also established this equality for a channel-type model with no public channel [20] (see [3, Page. 450-451] for more details). In the following, their technique is briefly explained.

At the end of the key exchange step, Alice and Bob retrieve key K and key $\hat{K}$, respectively, with a weakly achievable key rate $R_{_K}$ according to Definition 1.11; thus, Eve gains substantial information about the keys. Hence, the final step of a key agreement problem is to strengthen the security and reliability of the agreed keys. This step is called *privacy amplification* which gives Alice and Bob key K' and key $\hat{K}'$, respectively, with strongly achievable key rate $R'_{_K}$ according to Definition 1.13.

Privacy amplification was introduced by Bennett *et al.* [42] in the context of quantum cryptography as a method to extract secrecy from weakly random (partially secure) sequences. *Universal hashing* [43] and *extractors* [3, Sec. 17.1] are techniques for the privacy amplification [24, 41].

Maurer and Wolf [41] applied *extractors* to an achievable key rate in the weak sense to enhance its security and reliability specifications. For the privacy amplification, they [24,41] show that using extractors is a better technique than using hash functions in the sense that it requires shorter messages to be communicated between Alice and Bob.

[20] In fact, Maurer and Wolf [41] proved that the strong sense of secrecy capacity [41] and the weak sense of secrecy capacity (Definition 1.9) are equal for Csiszár-Körner's wiretap channel [12]. However, the secrecy capacity and key capacity of this model are equal when there is no public channel [18].

1.5 Channels with Random States

In this thesis, we focus on the class of *state-dependent* wiretap models. Before describing our model, we review the essential research in the field of state-dependent channels in this section.

Physical properties of a communication channel which control its PTF can be modeled by *channel state information* (CSI), and the corresponding channel is called a *state-dependent channel*. In wireless communications, the CSI includes scattering, fading, interference and power decay of a propagated signal. In wire-line communications, physical conditions of lines, like temperature or external forces (vibrations) are modeled as the CSI. Variations in the noise level due to interference or signal level due to fading can be also modeled as the CSI.

The output(s) of a state-dependent channel are a stochastic function of both the channel inputs and the CSI as a random vector. Specifically, a state-dependent channel has a collection of PTFs, where realization of the CSI at each transmission determines the actual PTF for that transmission. State-dependent channels model a large variety of practical channels, e.g., [6]:

- Compound channels [3]: The PTF is unknown for both sender and receiver, and it is only known over a set of PTFs;
- Arbitrary varying channels [44, 45]: The PTF alters per symbol transmission; e.g., channels suffered from Jamming signals;
- Host image in digital watermarking [46];
- Wireless fading channels [47];
- Writing on defected memories [48–50].

In this thesis, we are interested in the class of channels with random states where the CSI, which is represented by $\mathbf{s}$, is an i.i.d. random vector according to state distribution $\mathcal{P}(\mathbf{s}) = \prod_{i=1}^{n} \mathcal{P}_S(s_i)$. In this section, we review the capacity of memoryless channels with random states. We consider the channels with a transmitter (Alice), a receiver (Bob), and

PTF $\mathcal{P}(\mathbf{y}|\mathbf{x},\mathbf{s}) = \prod_{i=1}^{n} \mathcal{P}_{Y|XS}(y_i|x_i, s_i)$. No feedback is assumed from Bob to Alice. Alice and Bob each may have access to a random vector as *side information* (SI) . Let the SI at Alice be $\mathbf{a}$, which has the same length as that of $\mathbf{s}$. If the whole sequence $\mathbf{a}$ is available to Alice prior to each block transmission, the SI is said to be *non-causal.* For any $i \in \{1, \ldots, n\}$, if only sequence a_1^i is known at Alice before transmission of i^{th} symbol, the SI is said to be *causal*[21]. The SI is said to be *fully known* at the transmitter if $\mathbf{a} = \mathbf{s}$; otherwise, it is said to be *partially known.*

The significant publications which study the (ordinary) capacity of the channels with random states are as follows.

In 1958, Shannon [51] derived the ordinary capacity of a state-dependent DMC with casually known CSI at the transmitter. He proved that the capacity of this channel equals that of a DMC without any state, which has the same output alphabet and an extended input alphabet. The input alphabet of this equivalent channel consists of all mappings from the state alphabet to the input alphabet of the original channel. A particular input letter of the equivalent channel can be considered as a particular function (*strategy*) from the state alphabet to the input alphabet of the original channel. Shannon utilized strategies in which the input to the channel depends only on the current state of the channel but not on previous ones. He established that this type of strategies are sufficient for the equivalent channel to achieve capacity of the original channel [52].

Goldsmith and Varaiya [47] achieved the capacity of a fading channel in two cases: When the CSI is fully (whether causally or non-causally) known at both the transmitter and receiver, and when the CSI is known at the receiver alone. Optimal schemes are based on power adaption. For the former, they used a *water-pouring* in time domain as the optimal power adaption.

Kuznetsov and Tsybakov [48] studied coding for a memory with defective cells, which is an example of a DMC with non-causally known CSI at the transmitter. Although they offered some coding techniques for this channel, they did not attain its capacity.

[21] If there is no feedback from the receiver to the sender, it doesn't matter whether the SI at the receiver is available causally or non-causally in the sense of capacity. This is due to the fact that the receiver is assumed to decode after receiving the whole block code.

Gel ́fand and Pinsker [49] as well as Heegard and El Gamal [50] investigated capacity of a state-dependent DMC with random states, where the CSI is non-causally known at the transmitter but not at the receiver. The capacity is given by [49]

$$C = \max_{\mathcal{P}_{XU|S}} \left[\mathcal{I}(U;Y) - \mathcal{I}(U;S) \right] , \tag{1.32}$$

where U is an auxiliary RV such that $U \to (X, S) \to Y$ forms a Markov chain. Further, the capacity of this channel when i.i.d. random vector of SI dependent on the CSI is available at Bob was studied in [53, 54], and it is given by

$$C = \max_{\mathcal{P}_{XU|S}} \left[\mathcal{I}(U;Y,B) - \mathcal{I}(U;S) \right] , \tag{1.33}$$

where B represents the SI at Bob and U is an auxiliary RV such that $U \to (X, S) \to (B, Y)$ forms a Markov chain ($\mathcal{P}_{YB|XS}$ is given by the channel).

Costa [55] extended Gel ́fand-Pinsker's work to an AWGN channel with *additive white Gaussian interference* (AWGI), where the interference (state) is non-causally known at the transmitter. This model is sketched in Figure 1.10. In his model, Costa assumed that the interference, $\mathbf{s}$, and the noise, $\mathbf{g}$, are sequences of i.i.d. components distributed according to $S \sim \mathcal{N}(0, \Lambda)$ and $G \sim \mathcal{N}(0, \sigma^2)$, respectively. Also, the transmitter is subject to average power constraint $\frac{1}{n}(\mathbf{x}\mathbf{x}^t) \leq \Gamma$, where $\mathbf{x}$ is the transmission signal. In his celebrated paper, *writing on a dirty paper* [55], Costa showed that the capacity of this channel equals that of an AWGN channel with no interference, i.e.,

$$C = \frac{1}{2} \log(1 + \frac{\Gamma}{\sigma^2}) \, . \tag{1.34}$$

The optimum coding strategy, which is known as *dirty paper coding* (DPC), is to adapt transmitted signal to the state (dirts) such that the receiver obtains maximum possible information from contaminated signal. Costa [55] proved that equation (1.32) with $U = X + \gamma S$ can be applied to his model to compute the capacity. In this strategy, $X \sim \mathcal{N}(0, \Gamma)$ such that X and S are statistically independent. Then, he calculated the optimum value

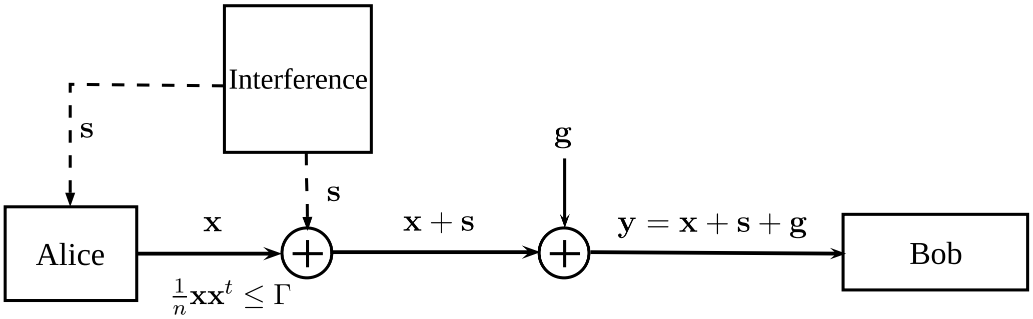

Interference
s
s
g
Alice
x
$\frac{1}{n}\mathbf{x}\mathbf{x}^t \le \Gamma$
$\mathbf{x}+\mathbf{s}$
$\mathbf{y}=\mathbf{x}+\mathbf{s}+\mathbf{g}$
Bob

Figure 1.10: Costa's Model: Gaussian channel with known interference at the sender.

of γ as

$$\gamma = \frac{\Gamma}{\Gamma + \sigma^2} \, , \tag{1.35}$$

which is interestingly independent of Λ. Based on DPC, the trivial idea of canceling the known interference at the transmitter is not generally optimal.

Consider the causal version of Costa's model, i.e., a power limited Gaussian channel with AWGN and AWGI where realizations of the interference are causally known only at the transmitter. In [56], the authors achieved the capacity of this channel in the high signal-to-noise ratio (SNR) regime by *lattice* strategy. Although an LB and a UB on the capacity are derived in [56], the capacity of this channel is not known for all SNRs.

1.6 State-Dependent Wiretap Models

As mentioned in the last section, the class of state-dependent channels models several important communication channels. Thus, the security issue of this class needs much consideration, specially in a state-dependent wireless network where unauthorized receivers inherently attend. To provide this demand, several research groups have studied various state-dependent wiretap models; among them, we highlight the following major studies in which two legitimate parties (Alice and Bob) and a wiretapper (Eve) have access to the channel:

The compound wiretap channel was investigated in [57–59]. The PTF of a compound channel is determined by the realization of the CSI, which is unknown to Alice, Bob, and Eve; however, the PTF (or the CSI) of the compound channel is known only within a set of candidates. This channel can be treated as a *multi-cast channel with multiple wiretappers.* In other words, the number of states available to Bob becomes the number of receivers such that each state corresponds to one receiver, and the number of states available to Eve becomes the number of eavesdroppers such that each state corresponds to one eavesdropper [57, 58]. An achievable coding scheme must be able to reliably transmit a message to all receivers such that it remains perfectly secure against all wiretappers. An LB and a UB on the secrecy capacity is reported in [57]. The LB achieves the secrecy capacity for a degraded compound channel [57]. When the CSI is known at the receiver, the secrecy capacity of the semi-deterministic compound channel and the parallel Gaussian compound channel are attained in [58] and [59], respectively.

Han Vinck's research group studied the secrecy problem in state-dependent channels with random states [60–62]. Mitrpant and Han Vinck [60] merged Costa's model (Figure 1.10) with Gaussian wiretap channel (Figure 1.5) to model a *Gaussian state-dependent wiretap channel* (G-SWC) with a physically degraded wiretapper. They [60] achieved an LB on the secrecy capacity which shows the secrecy capacity of the G-SWC with known CSI at Alice is generally larger than that of the corresponding Gaussian wiretap channel [13], which has no interference. Although known CSI at Alice gives no improvement on the ordinary capacity in Costa's model [55], it generally enlarges the secrecy capacity of a G-SWC. The achievable wiretap coding scheme is constructed based on DPC [55] with wiretap strategies [11, 31]. The scheme has not been proven yet to be optimum in general case. However, it is optimal in low *signal (power) to interference (power) ratio* (SIR) regime as well as high SIR regime; in low SIR, it equals the (ordinary) capacity of the main channel (from Alice to Bob) as well. A UB on the secrecy capacity is also derived in [60]. This bound is the secrecy capacity of an enhanced Gaussian wiretap channel in which Alice is generously permitted to control both her input signal and the interference.

In Chapter 2, this work will be reviewed in more details.

In [61,62], the authors considered a degraded *discrete memoryless state-dependent wiretap channel* (DM-SWC) with PTF $\mathcal{P}(\mathbf{y},\mathbf{z}|\mathbf{x},\mathbf{s}) = \prod_{i=1}^{n} \mathcal{P}_{Y|XS}(y_i|x_i,s_i)\mathcal{P}_{Z|Y}(z_i|y_i)$, where $\mathbf{s}$ is drawn i.i.d. according to $\prod_{i=1}^{n} \mathcal{P}_S(s_i)$ and it is non-causally known at Alice. Using *double random binning* [6], an LB on the secrecy capacity and an achievable *rate-equivocation region* (see Chapter 2) are established in this work. The LB is optimally equals the ordinary capacity of the main channel if sending a positive (information) rate from Alice to Eve is not possible at the capacity achieving distribution of the main channel. Moreover, two UBs are reported in [61]: the first one is the ordinary capacity of the main channel (from Alice to Bob), and the second one is the secrecy capacity of an enhanced channel in which $\mathbf{s}$ is governed by Alice as the second input further to $\mathbf{x}$. In Chapter 2, this work will be reviewed in more details.

Liu and Chen [63] extended the results of paper [61] to a DM-SWC with two sided non-causal SI at Alice and Bob. They [63] obtained an achievable rate-equivocation region for their model by combining the strategies of paper [61] with those of paper [53].

Khisti *et al.* [64] studied the key agreement problem over a DM-SWC with $\mathcal{P}(\mathbf{y},\mathbf{z}|\mathbf{x},\mathbf{s}) = \prod_{i=1}^{n} \mathcal{P}_{YZ|XS}(y_i,z_i|x_i,s_i)$. In their model, $\mathbf{s}$ is drawn i.i.d. according to $\prod_{i=1}^{n} \mathcal{P}_S(s_i)$ and it is *fully* known at both Alice and Bob in non-causal form. They [64] investigated the key capacity for the following two cases:

- When no public channel is available. In this case, the key capacity is given by

$$C_{K} = \max_{\mathcal{P}_{XU|S}} \left[\mathcal{I}(U;Y|S) - \mathcal{I}(U;Z|S) + \mathcal{H}(S|Z)\right], \tag{1.36}$$

  where U is an auxiliary RV such that $U \to (X,S) \to (Y,Z)$ form a Markov chain.

- When a public channel with unlimited capacity in both directions is available. In this case, if noise of Bob's channel and that of Eve's channel are mutually independent [64, Remk. 3], i.e., $Y \to (X, S) \to Z$, then the key capacity is given by

$$C_{_K} = \max_{\mathcal{P}_{X|S}} [\mathcal{I}(X; Y|Z, S) + \mathcal{H}(S|Z)] \, . \tag{1.37}$$

In [64], the key agreement coding scheme generates the achievable key rate based on two different resources: a random number generator, and the CSI. However, this result is not surprising as both Alice and Bob already have two exactly the same copies of the CSI at the beginning of each transmission. Hence, the CSI can be concluded as a secure common randomness when the portion which is revealed to Eve through the wiretap channel is removed.

Chia and El Gamal [65] investigated secrecy capacity of a DM-SWC when the CSI is *fully* known at both Alice and Bob. When the CSI is available causally, they achieved an LB on the secrecy capacity. When the CSI is available non-causally, they also derived a UB on the secrecy capacity. As the secrecy capacity of the non-causal case is not less than that of the causal case, they applied this UB to the causal case as well. From these bounds, they proved the secrecy capacity for the DM-SWC if Eve's channel is less noisy than Bob's [16] for every state vector $\mathbf{s}$. They [65, Thm. 3] showed that the secrecy capacity for that special channel is the same for both causal and non-causal cases. This means the optimal wiretap coding scheme for the non-causal case of a less noisy DM-SWC does not utilize the future instants of the given CSI for encoding at each symbol transmission. The achievable scheme is built on the following strategy:

For encoding, any message is split into two independent parts: the first part is encoded by wiretap coding strategy [11], and the second part is encrypted by a one-time pad system [10] and encoded with a deterministic encoder, where the key is taken from the CSI (The portion which is not revealed to Eve through her channel). The strategy exploits block Markov encoding [6] because the key used for encryption of the second part has been already agreed between Alice and Bob based on the released CSI at the last block transmission.

1.7 Motivations

In this thesis, we study the key agreement problem over a state-dependent wiretap channel paralleled with a public channel. The channel state is drawn i.i.d. according to a given distribution and it is non-causally known at the transmitter. In Chapter 3, we study the discrete memoryless (DM) model in which the wiretap channel is a DM-SWC (see Section 1.6). This model is an extension of Geĺfand-Pinsker's model demonstrated in Section 1.5. In Chapter 4, we study the Gaussian model in which the wiretap channel is a power-limited Gaussian wiretap channel with additive interference (channel state). This model is an extension of Costa's model sketched in Figure 1.10.

In this section, we illustrate the motivations for the key agreement problem definitions. We express a motivation of the DM model and that of the Gaussian model in Example 1.2 and Example 1.3, respectively.

Example 1.2. An essential part of cloud computing [66] is data storage and management. As shown in Figure 1.11, suppose that multiple users connect to a server through a wireless internet, which writes and stores data on multiple storage devices, e.g., hard disks. If a memory cell of a storage device is *proper*, the data read from it is the same as the data written on it. The written data is also referred to as the cell state. However, in practice, some cells of a storage device may be defected. The *defected cells* are those ones whose stored data can be read with random error [6, 50]. Hence, the relation between the read data and the written data is statistically determined by a PTF. In this scenario, the server knows the status of all memory cells at any time. On the other hand, the noisy channel from each user to the server is characterized by another PTF. Thus, an overall (noisy) channel, which is in fact memoryless, can be modeled from cell states to each receiver.

An important issue in cloud computing is privacy of the data received by each user. To apply symmetric-key cryptography for this issue, a key distribution step is required before a data transmission step begins. One solution is to model the scenario as a key agreement problem over a discrete memoryless state-dependent wiretap channel from the cells to the receivers in which the cell states are non-causally known at the transmitter. Investigation

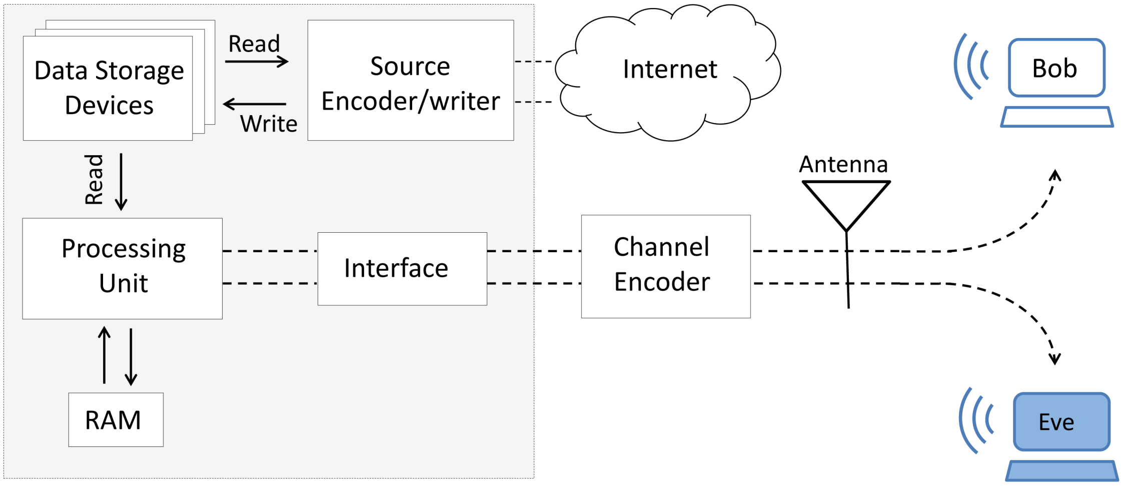


Figure 1.11: A motivation of the discrete memoryless model.

of a generalized model of this problem is the original motivation of the key agreement problem over the DM model in Chapter 3. Specifically, we wish to study the effects of the known states in key generation when a public channel, e.g., a web server with unrestricted accessibility, is available.

Example 1.3. Suppose a wireless network in which a transmitter is capable of serving multiple users at the same time, e.g., a base station in a cellular network [67] as sketched in Figure 1.12 for two users. In this case, the transmitter propagates an individual signal for each intended receiver. Each receiver's antenna collects a weighted sum of all propagated signals, where the coefficients of the summation are determined by the characteristic of the wireless channel. Each receiver is interested in its own signal and considers the weighted

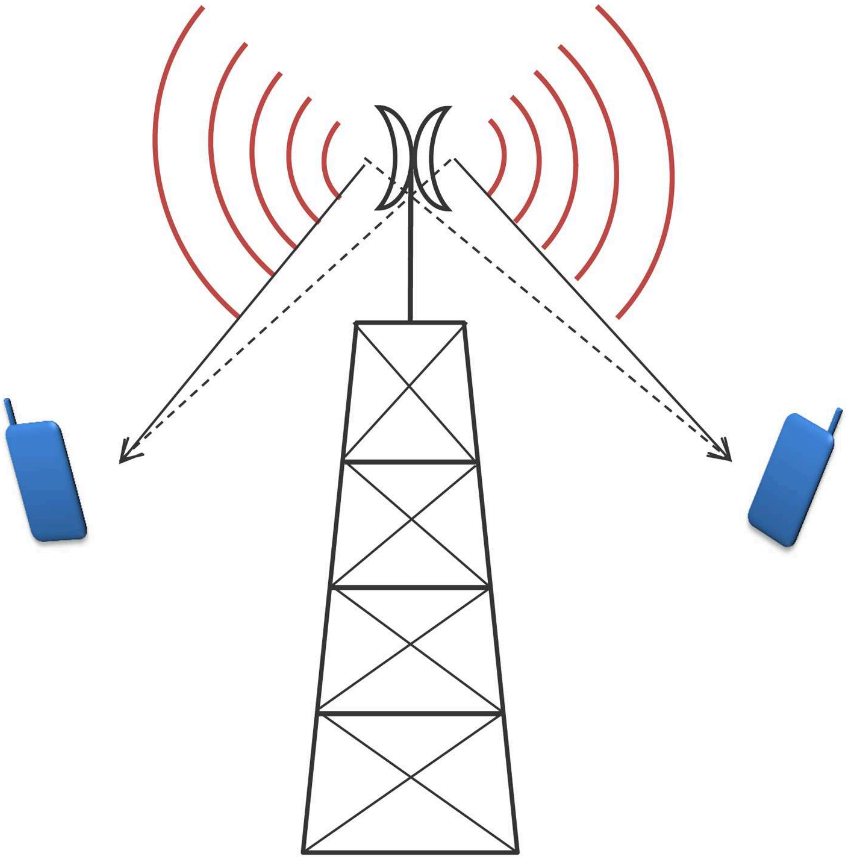


Figure 1.12: A motivation of the Gaussian model.

summation of the signals of other users as the interference signal. As a simple case, assume that the wireless channel coefficients are known at the sender and they do not change in the period of each block code transmission. In this scenario, the transmitter, who generates all the propagated signals, knows the interference of each user in advance of the transmission. This is due to the fact that the transmitted signal of each user can be calculated at the beginning of each block transmission in the transmitter as a function of the user's message. Although the interference is an impediment for data communication, it may be exploited for secure key generation between the transmitter and each receiver in presence of other users. Investigation of a simplified model of this problem is the original motivation of the key agreement problem over the Gaussian model in Chapter 4.

To explain why the key agreement of a state-dependent wiretap channel differs from that of a channel-type model as sketched in Figure 1.8, let us simplify the DM model as follows: Suppose the key agreement problem over a DM-SWC (with no public channel) where side information is available at neither Bob nor Eve. For any sent symbol over the state-dependent wiretap channel, the CSI leaks to Bob and to Eve, i.e., Eve and Bob obtain some information about the current channel state. The leakage of the CSI to Bob and that of the CSI to Eve each depends on the channel input distribution, the distribution of the CSI, and the PTF of the wiretap channel. The leakage of the CSI to Bob makes an advantage for key generation because he partially obtains some information about the CSI, which is fully known at Alice. This information can be exploited as a common randomness for the key agreement, and so it seems that Alice should select the channel input such that this leakage is improved. On the other hand, the leakage of the CSI to Eve makes a disadvantage for key generation because she obtains some information about the CSI and thus about the common information of Alice and Bob. It seems that Alice should choose the channel input such that this leakage is reduced. Hence, a trade-off between these two effects should be carefully considered in the proposed key agreement coding scheme. Further, Alice can govern these effects only by channel input distribution, where she can generate the transmission signal as a function of past, current, and future channel states. This is the main challenge of this problem which is still open in its general case. Note that the wiretap codebook for the Wyner model, which is optimum for the key agreement problem over a wiretap channel (channel-type model with no public channel) [18], is *not* optimum for the key agreement problem over a DM-SWC because it does not take into account the challenge of the leakage effects[22].

As mentioned in Subsection 1.4.2, the key capacity over a DM-WC in parallel with a

[22] Similarly, the secrecy capacity of a Gaussian wiretap channel with interference (G-SWC) is strictly larger than that of a Gaussian wiretap channel (with no interference) according to Figure 2.6 (see Subsection 2.2.3 for more details).

forward public channel (the channel-type model [18]) equals the secrecy capacity of the DM-WC as given in (2.11), and the public channel is useless in key generation. At first glance, similarly, the key capacity over a state-dependent wiretap channel paralleled with a forward public channel seems to be equal to the secrecy capacity of that wiretap channel and the public channel has no benefit in key generation (at least when Bob has no SI). Investigation of this hypothesis was an initial motivation of this research.

In this study, we prove that this hypothesis is not generally true. In fact, the public channel contributes in key generation even if Bob has no SI. Further, the following inspiring questions are addressed in this investigation.

- Do the i.i.d. random vectors of the CSI provide a helpful source of randomness at Alice for key generation?
- If the CSI is fully known at Bob as well, does this common information (between Alice and Bob) suffice to achieve key capacity of the model such that no key exchange over the wiretap channel or the public channel is required?
- If the public channel is involved in the key exchange, is the key capacity a strictly increasing function of the public channel capacity? What is the trade-off between the public channel capacity and the achieved key rate?
- How does the key capacity relate to the main channel capacity and the secrecy capacity of the wiretap channel?
- For a given wiretap channel, what is the maximum key capacity no matter how much the public channel capacity is? Is this key capacity achievable by using a finite capacity of the public channel? If no, is there any finite public channel capacity such that this key capacity can be achieved within a reasonable tolerance?
- Is it possible to obtain a positive key rate when Eve's channel is less noisy than Bob's channel in the Gaussian model? Does it depend on the direction of the public channel?
- What is the effect of the noise covariance matrix of the Gaussian wiretap channel in key generation?

1.8 Contributions

In this section, we briefly describe our models, contributions and methodology. The details will be given in Chapter 3 and Chapter 4.

We study the key agreement problem in two models: a *discrete memoryless* (DM) model and a Gaussian model. Each model has one sender (Alice), an authorized receiver (Bob), and an eavesdropper (Eve). Each model also consists of two parallel channels: a memoryless wiretap channel with random states, and an authenticated public channel between Alice and Bob. The capacity of the public channel is given by the pair (C_{P_1}, C_{P_2}), where $C_{P_1} \in [0, \infty)$ and $C_{P_2} \in [0, \infty)$ are the capacity of the public channel in the forward direction (from Alice to Bob) and in the backward direction (from Bob to Alice), respectively. The key capacity of each model is denoted by $C_K(C_{P_1}, C_{P_2})$ as a function of (C_{P_1}, C_{P_2}).

In Subsection 1.8.1, we express the DM model and the main results and methodology, which will be given in details in Chapter 3. In Subsection 1.8.2, we express the Gaussian model and the related contributions and methodology, which will be given in details in Chapter 4.

1.8.1 The DM Model

The wiretap channel in the DM model is a DM-SWC (see Section 1.6). The CSI is an i.i.d. random vector which controls the PTF of the wiretap channel. We assume that realizations of the CSI are non-causally known at the transmitter. We suppose that each receiver has access to an i.i.d. random vector dependent on the CSI according to a known joint PMF. In the DM model, we assume that the public channel is one-way in the forward direction, i.e., $C_{P_2} = 0$. The main objective of this work is to find (bound) the forward key capacity $C_K(C_{P_1}, 0)$ as a function of C_{P_1}.

The results of Chapter 3 are partially presented in paper [68]. The main contributions of this chapter are as follows.

(a). Lower bound on $C_{K}(C_{P_1},0)$. At the first look, the (forward) key capacity seems to be equal to the secrecy capacity of the DM-SWC. That is, a pair of messages $(M,\hat{M})$, where M is a uniform RV emitted from a source, in the corresponding secrecy problem is treated as a pair of agreed keys $(K,\hat{K})$ in the key agreement problem. However, this idea is not generally true, and the secrecy capacity should be just considered as the first achievable subkey rate of the key capacity. We will show that another achievable subkey rate can be generally generated on top of the secrecy capacity with the help of the known CSI at Alice (even if it is not known at Bob and the public channel capacity is zero). This key rate is determined by the DM-SWC. Our suggested key agreement code exploits both a random generator and the CSI as resources for the key generation. The i.i.d. channel state vector with a large enough length guarantees the randomness of the generated key according to the *asymptotic equipartition property* (AEP) [2, Ch. 3]. In other words, the CSI at Alice and the leaked CSI at Bob is a resource to provide common randomness for the key generation if the leaked CSI to Eve is removed from the common randomness. To do this, we apply a random quantization method to map a revealed channel state vector to a codeword from the key agreement codebook. On the other hand, the number of the codewords can be increased as the capacity of the public channel grows. As a result, this may lead to an increment of the achieved key rate as intuitively explained in the following: the key agreement codebook consists of multiple enumerated wiretap subcodebooks such that each one is constructed similar to Csiszár and Körner's wiretap codebook [12] (there is no CSI in their model though). At each block transmission, the released state sequence candidates a set of codewords from the whole key agreement codebook by using a selection rule. One codeword from that set is selected at random[23], by the encoder for key and signal generation. Similar to Csiszár and Körner's secrecy problem [12], the index of that codeword in its wiretap subcodebook determines Alice's key. Then, Alice sends the index of that wiretap subcodebook over the public channel to Bob. Hence, the number of wiretap subcodebooks is restricted by the capacity of the forward channel. Knowing this wiretap subcodebook, Bob can retrieve the codeword and achieve a secure key rate determined by

[23] We call it random quantization method.

that wiretap subcodebook. The encoder and decoder utilize *strong (letter) typicality* (see Subsection 2.1.1) to generate or to retrieve their keys, respectively. The random coding arguments [2, 6], Markov lemma (see Lemma 2.1), Fano's inequality [2, Thm. 2.10.1] and Wyner's wiretap strategy [11] are frequently used to establish the AR, AS, and ARN conditions for the agreed keys.

(b). Upper bounds on $C_K(C_{P_1}, 0)$. Two UBs on $C_K(C_{P_1}, 0)$ are derived in this thesis. Each UB is offered in the form of an optimization (maximization) problem subject to a constraint given by the capacity of the public channel. For each upper bound, we derive two inequality equations. The first one bounds the key capacity of the DM model, and the second one gives a condition based on the public channel capacity for the first one. The first UB is valid for any DM model in general, while the second UB is derived based on the assumption that the DM model is less noisy in Bob's favor. This UB can be also considered as a (loose) UB for the general model as well because the less noisy assumption makes an enhanced model in the sense of the key generation between Alice and Bob. We have frequently used Fano's inequality [2, Thm. 2.10.1], Csiszár-Körner's sum identity (see Appendix D), data processing inequality [2, Thm. 2.8.1], the chain rules for the entropy and mutual information functions, and the time-sharing strategy [2, 6, 12] in proofs of the UBs.

(c). Optimum Cases. Our achievable key agreement code is optimum for the following special cases:

- when the capacity of the public channel is unlimited in the forward direction, i.e., $C_{P_1} \to \infty$;
- when the capacity of the public channel exceeds a *finite* capacity $C^*_{P_1}$, which is determined by the DM-SWC. For any DM-SWC, we obtain $C^*_{P_1}$ and show that it is finite;
- when the wiretap channel does not exist. This special case coincides with the results of key agreement models with a common randomness given in publications [18, 21]

with the following difference. In Subsection 3.3.3, we will prove that the forward key capacity can be obtained by using one auxiliary RV and using two auxiliary RVs, as given in [18, 21], is not necessary to achieve the forward key capacity;

- when both Alice and Bob fully know the CSI. This special case was studied in paper [64], where the authors established the key capacity for special case $C_{P_1} = 0$. Extending their work, we prove that the achieved key capacity also equals the forward key capacity of the DM model for any $C_{P_1} > 0$. That is, the public channel has no benefit in this case for key generation.

1.8.2 The Gaussian Model

The wiretap channel in the Gaussian model is a G-SWC (see Section 1.6). The CSI is in the form of AWGI. We assume that realizations of the interference are non-causally known at the transmitter. In addition to the additive interference, Bob's channel and Eve's channel are affected by two individual AWGN, which are distributed i.i.d. according to $\mathcal{N}\left((0,0), \begin{bmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$, where $\varrho \in [-1, 1]$, σ_1^2 and σ_2^2 are the noise correlation coefficient, noise variance of Bob's channel and that of Eve's channel, respectively. In the Gaussian model, however, no SI at Bob and Eve is assumed for simplicity of calculations (see Subsection 5.1.2 for future work).

The main objective of this work is to find (bound) the forward key capacity $C_K(C_{P_1}, 0)$ as a function of C_{P_1}. Also, we want to determine if the key agreement is feasible if Eve's channel is less noisy than Bob's. Also, we are interested in the effect of the noise correlation coefficient ϱ on the key generation.

The results of Chapter 4 are partially presented in paper [69]. The main contributions of this chapter are as follows.

(a). Lower bound on $C_K(C_{P_1}, 0)$. The LB on $C_K(C_{P_1}, 0)$ of the DM model can not be directly extended to a continuous alphabet with the infinite size. We examine the possibility of this extension for the Gaussian model in this research. The LB on the forward key capacity is established as an extension of the achievable key agreement coding scheme of the DM model. We apply the *weakly (entropy) typicality* (see Subsection 2.1.1) for continuous RVs and the generalized Markov lemma[24] [71] (see Remark 2.3) for a Gaussian input distribution to justify the validity of the extension. Based on our achievable key agreement code, the input distribution on G-SWC is a Gaussian distribution which satisfies the input power constraint. To construct the transmission signal, we establish a generalized version of the DPC and we combine it with Wyner's wiretap strategies [11]. In the ordinary DPC strategy used by Costa [55] in his model (Figure 1.10), condition $\mathcal{E}(XS) = 0$ suffices to achieve the (ordinary) capacity as the message is independent of the CSI. However, in our work, X is to be correlated with S as the generated key is generally correlated with the CSI, and this correlation helps to increase the achievable key rate. In our achievable key agreement code, the transmitted signal conveys information about both the CSI and randomization to the receivers. The correlation coefficient between the transmitted sequence and the CSI sequence determines the weight of each key generation resource, and it is to be determined according to the transmitter power constraint, the interference average power, and the limited public channel capacity such that the maximum possible key rate achieves. We justify that the forward key capacity is positive as long as Bob's channel is less noisy than Eve's $\sigma_2 > \sigma_1$. In this thesis, we prove that the noise correlation coefficient has no effect on $C_{_K}(C_{P_1}, 0)$. Specifically, we establish that the forward key capacity of a given Gaussian model equals that of an equivalent Gaussian model with a physically degraded G-SWC.

(b). Upper bound on $C_{_K}(C_{P_1}, 0)$. The UBs on the forward key capacity derived for the DM model are not generally valid for the Gaussian model. This is due to the fact that the size of alphabets in the DM model is finite, and this fact is applied to establish those UBs. First, assume that Bob's channel is less noisy than Eve's, i.e., $\sigma_2 > \sigma_1$. To prove the UB on the forward key capacity of a Gaussian model, we prove the UB for its equivalent Gaussian model with a physically degraded channel. As the forward key capacity of both

[24] The author appreciates Prof. Mitran for this important point based on paper [70].

models are proved to be the same, the UB is also valid for the original Gaussian model with an arbitrary noise correlation coefficient ϱ. The UB is valid for any $C_{P_1} \geq 0$. We will prove this UB by use of Fano's inequality [2, Thm. 2.10.1], Jensen's inequality [2], *entropy power inequality* [2], and the fact that Eve's channel is assumed to be a physically degraded version of Alice's in the physically degraded G-SWC. Second, assume that Eve's channel is less noisy than Bob's, i.e., $\sigma_1 \geq \sigma_2$. We derive another UB on $C_K(C_{P_1}, 0)$ which shows the forward key capacity is always zero in this case.

(c). Special Cases. In this thesis, we prove that the suggested key agreement code is optimum in the sense that it achieves the forward key capacity in the following special cases:

- when the SIR goes to zero (and the interference average power is fixed) for any $C_{P_1} \geq 0$. This special case coincides with the results of paper [40] for special case *degraded* Gaussian multiple source (see Theorem 1.1 for more details);
- when the SIR goes to infinity, $C_K(C_{P_1}, 0) \to \frac{1}{2}\log\left(\frac{\sigma_2^2}{\sigma_1^2}\right)$ for any $C_{P_1} \geq 0$. In this case, the forward key capacity is independent of C_{P_1}, the interference average power, and the transmission power;
- when $C_{P_1} \to \infty$ for any SIR. In this case, we show that the transmission signal is asymptotically aligned with the interference to achieve the forward key capacity. That is, the transmitter *amplifies* the interference according to its maximum available power and *forwards* it to the wiretap channel.

Further, we study the key capacity $C_K(\infty, \infty)$ when Eve's channel is less noisy than Bob's, i.e., $\sigma_1 > \sigma_2$ to determine if key generation is possible in this case. We prove that $C_K(\infty, \infty) = 0$ if the G-SWC is physically degraded in Eve's favor, i.e., $\varrho = \frac{\sigma_2}{\sigma_1}$. If $\frac{\sigma_2}{2\sigma_1} \geq \varrho$, we construct a key agreement code for key generation by the extension of Maurer's method [17], which is originally given for BSCs, to the Gaussian model. Hence, we achieve a positive LB on $C_K(\infty, \infty)$ for case $\frac{\sigma_2}{2\sigma_1} > \varrho$. Consequently, the key capacity of the Gaussian model is a function of the noise correlation coefficient ϱ.

Moreover, we simulate the LB and the UB on $C_K(C_{P_1}, 0)$ for a Gaussian model by using

MATLAB®. With the simulations, the following facts are illustrated as well.

- An essential difference between the Gaussian model and the DM model is that the LB on $C_K(C_{P_1}, 0)$ is a strictly increasing function of C_{P_1}; however, it is bounded.
- The forward key capacity can be larger than both the secrecy capacity and main channel capacity (even) when no public channel is available.
- If the SIR is positive in dB, the difference between the LB and the calculated UB on $C_K(C_{P_1}, 0)$ is less than .15% for any $C_{P_1} \geq 0$.

1.8.3 The Parallel Work

In a parallel independent work to [72], we learned that Khisti *et al.* [73,74] have studied the key agreement problem over our investigated models for the following two special cases.

(a). No Public Channel. For the DM model with no public channel, the authors [74, Thm. 1] have proven an LB on the key capacity which is the same as Theorem 3.1 when $C_{P_1} = 0$ is relaxed. For this model, a UB on the key capacity is also given in [74, Thm. 2]. The UB is derived with generous assumptions to make an enhanced model (in Alice and Bob's favor) as follows. Bob is permitted to know Eve's received signal, and Alice is permitted to govern the CSI as the second channel input further to her transmission signal. With these assumptions, Eve's received signal is a degraded version of Bob's. Finally, the secrecy capacity of this model, which can be deduced from (2.16), is reported as a UB on the key capacity of the model of interest.

For the Gaussian model with a physically degraded G-SWC, Khisti [73, Prop. 2] achieved an LB on $C_K(0, 0)$ when the SNR at the main channel is not negative (in dB). In this special case, the LB of the Gaussian model is obtained from that of the DM model by direct calculation of the results for continuous Gaussian RVs. In the high SIR (with the high SNR) regime, this LB coincides our achieved LB (for special case $C_{P_1} = 0$), which results in Corollary 4.2 of this work for this special case. Khisti *et al.* [73,74] also derived a

UB on $C_{_K}(0,0)$ by direct calculation of the corresponding UB on $C_{_K}(0,0)$ of the DM model for Gaussian RVs. That UB equals $C_{_K}(\infty,0)$, which is achieved in [69] and in Theorem 4.2.

(b). Two-Way Public Channel with Unlimited Capacity in both Directions. In this case, Khisti [73] first studied the DM model. For this model, he achieved an LB on the key capacity as a maximization of two individual LBs on the key capacity [73, Thm. 3]. The first LB is the LB on the key capacity when $C_{_{P_1}} = 0$, which is achieved for the last case in part (a). The second LB is obtained by "a natural modification of Maurer's coding scheme [17, 18]" [73]. The second LB generally requires one forward transmission and one backward transmission over the public channel.

Using the strategy mentioned in the last case in part (a), a UB on the key capacity is derived for this case as well. The UB is tight when the outputs of the DM-SWC are independent given its inputs, i.e., $Y \to (X,S) \to Z$ forms a Markov chain. If this condition holds, the key capacity is characterized by [73, Thm. 5]

$$C_{_K}(\infty,\infty) = \max_{\mathcal{P}_{X|S}} \mathcal{I}(X,S;Y|Z)\,. \tag{1.38}$$

Khisti [73] extended the key capacity given in (1.38) to the Gaussian model in which Eve's channel is a physically degraded version of Bob's. For this special case, $C_{_K}(\infty,\infty)$ is reported [73, Prop. 5] by direct calculation of (1.38) for Gaussian RVs.

1.9 Organization of the Thesis

In this introduction, we have reviewed the essential points of information-theoretic security in wiretap communication channels with the focus on key agreement models. The organization of the rest of this thesis is as follows.

- **Chapter 2**: This chapter justifies the basic models and method that will be exploited to prove our results. The chapter consists of two main sections: in the first section, we will define mathematical tools required to establish our results. In the second section, we will present the seminal related work of secrecy problems, which will be served for comparison in Chapter 3 and Chapter 4.

- **Chapter 3**: The key agreement problem over the DM model is investigated in this chapter.
- **Chapter 4**: The key agreement problem over the Gaussian model is investigated in this chapter.
- **Chapter 5**: In this chapter, we will conclude our work. In the conclusion, we will mention the strategies used in our work. We will also highlight the ties of our research to previous work. Finally, we will offer the future work in this chapter.

Chapter 2

Fundamentals

In the this chapter, we introduce the mathematical methods and fundamental definitions and models which will be utilized in the next chapters. In this chapter, we survey the main related secrecy problems in the field of information-theoretic security. One purpose of this chapter is to review the essential points, results and wiretap strategies introduced in fundamental research [11, 12]. These strategies form the base of information-theoretic security. Accordingly, we study the secrecy problem over a discrete memoryless wiretap channel in Subsection 2.2.1.

Before we declare our key agreement problems, the main points, achievements, and strategies in significant publications [60,61] are surveyed in this chapter. These publications study the secrecy problem over wiretap channels with non-causal side information. Their strategies with novel modifications will be exploited in our problems. In the next chapter, the results of those papers will be further referred to compare our achieved LBs on the key capacity with the best known LBs/UBs on the secrecy capacity. To do this, the secrecy problem over a DM-SWC is reviewed in Subsection 2.2.2. The secrecy capacity over a G-SWC is surveyed in Subsection 2.2.3 as well.

2.1 Mathematical Tools

In this section, we present the fundamental definitions and basic lemmas, which will be used in the next chapters to establish our proofs.

2.1.1 Typicality

Shannon [4] introduced the notion of *typical sequences.* Two types of typicality are used in this thesis: Strong typicality [2, Sec. 10.6] and weak typicality [2, Ch. 3]. In this subsection, we define the strong typicality and the weak typicality. We also state the main useful lemmas of typical sequences, which are related to our work, without proofs. More details can be found in references [2,3,75].

The weak typicality, which is also known as *entropy-typicality* [75], is defined as follows.

Definition 2.1 (Weak typicality)**.** Assume $\epsilon \in (0,1)$. Let $\mathbb{X}$ be an alphabet set with a finite size. Let $\mathbf{x} \in \mathbb{X}^n$ be a sequence generated i.i.d. according to $\mathcal{P}(\mathbf{x}) = \prod_{i=1}^{n} \mathcal{P}_X(x_i)$. $\mathbf{x}$ is said to be an ϵ-*weakly typical* sequence with respect to the PMF $\mathcal{P}_X$ on $\mathbb{X}$ if

$$\left| -\frac{1}{n} \log(\mathcal{P}(\mathbf{x})) - \mathcal{H}(X) \right| < \epsilon \, ,$$

where function

$$\mathcal{H}(X) \triangleq -\sum_{x \in \mathbb{X}} \mathcal{P}_X(x) \log(\mathcal{P}_X(x)) \tag{2.1}$$

is (Shannon) entropy function. Moreover, $\mathbb{T}_\epsilon(\mathcal{P}_X)$ is called an ϵ-weakly typical set, which contains all $\mathbf{x} \in \mathbb{X}^n$ such that $\mathbf{x}$ is an ϵ-weakly typical sequence with respect to $\mathcal{P}_X$.

Definition 2.1 can be extended for multiple random vectors as follows.

Definition 2.2 (Weak joint typicality)**.** Assume $\epsilon \in (0,1)$ and $\ell \in \mathbb{N}$. Let $\mathbb{X}_i$ be an alphabet set with a finite size, where $i \in \{1, \ldots, \ell\}$. Assume an ℓ-tuple of sequences

$(\mathbf{x}_1, \ldots, \mathbf{x}_\ell) \in \mathbb{X}_1^n \times \ldots \times \mathbb{X}_\ell^n$ is drawn i.i.d. according to

$$\mathcal{P}(\mathbf{x}_1, \ldots, \mathbf{x}_\ell) = \prod_{i=1}^{n} \mathcal{P}_{X_1 \ldots X_\ell}(x_{1i}, \ldots, x_{\ell i}) \, .$$

The ℓ-tuple of sequences $(\mathbf{x}_1, \ldots, \mathbf{x}_\ell)$ is said to be ϵ-weakly (jointly) typical with respect to the distribution $\mathcal{P}_{X_1 \ldots X_\ell}$ on $\mathbb{X}_1 \times \ldots \times \mathbb{X}_\ell$ if

$$\left| -\frac{1}{n} \log(\mathcal{P}(\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_j})) - \mathcal{H}(X_{i_1}, \ldots, X_{i_j}) \right| < \epsilon$$

holds for *any* $j \in \{1, \ldots, \ell\}$ and *any* $1 \leq i_1 < i_2 < \ldots < i_j \leq \ell$, where function

$$\mathcal{H}(X_{i_1}, \ldots, X_{i_j}) \triangleq - \sum_{\upsilon_1 \in \mathbb{X}_{i_1}} \cdots \sum_{\upsilon_j \in \mathbb{X}_{i_j}} \mathcal{P}_{X_{i_1} \ldots X_{i_j}}(\upsilon_1, \ldots, \upsilon_j) \log(\mathcal{P}_{X_{i_1} \ldots X_{i_j}}(\upsilon_1, \ldots, \upsilon_j)) \quad (2.2)$$

is (Shannon) joint entropy function. Moreover, $\mathbb{T}_\epsilon(\mathcal{P}_{X_1 \ldots X_\ell})$ is called an ϵ-weakly (jointly) typical set, which contains all $(\mathbf{x}_1, \ldots, \mathbf{x}_\ell) \in \mathbb{X}_1 \times \ldots \times \mathbb{X}_\ell$ such that $(\mathbf{x}_1, \ldots, \mathbf{x}_\ell)$ is ϵ-weakly (jointly) typical with respect to the distribution $\mathcal{P}_{X_1 \ldots X_\ell}$.

Remark 2.1. Although weak typicality in Definition 2.1 and Definition 2.2 is introduced for discrete random sequences, it can be extended to continuous random sequences if the PMFs and the entropy functions are replaced by the corresponding probability density functions and the differential entropy functions, respectively [2, 75]. An *ϵ-weakly typical* set with respect to normal distribution $\mathcal{N}((\mu_1, \ldots, \mu_\ell), \mathbf{\Sigma}_{\ell \times \ell})$ on $\mathbb{R}^\ell$ is represented by $\mathbb{T}_\epsilon(\mathcal{N}((\mu_1, \ldots, \mu_\ell), \mathbf{\Sigma}_{\ell \times \ell}))$ in this thesis. This notation will be frequently used in Subsection 4.1.

The strong typicality, which is also known as *letter typicality* [75], is defined as follows.

Definition 2.3 (Strong typicality)**.** Assume $\epsilon \in (0, 1)$. Let $\mathbb{X}$ be an alphabet set with a finite size. A sequence $\mathbf{x} \in \mathbb{X}^n$ is said to be an *ϵ-strongly typical* sequence with respect to a distribution $\mathcal{P}_X$ on $\mathbb{X}$ if for every letter $\upsilon \in \mathbb{X}$

$$\left|\frac{1}{n}\eta(\upsilon|\mathbf{x}) - \mathcal{P}_X(\upsilon)\right| < \begin{cases} \frac{\epsilon}{|\mathbb{X}|}, & : if\ \mathcal{P}_X(\upsilon) > 0; \\ 0, & : if\ \mathcal{P}_X(\upsilon) = 0. \end{cases} \tag{2.3}$$

where $\eta(\upsilon|\mathbf{x})$ is the number of occurrences of the letter υ in vector $\mathbf{x}$. Moreover, $\mathbb{T}_\epsilon^*(\mathcal{P}_X)$ is called an ϵ-strongly typical set, which contains all $\mathbf{x} \in \mathbb{X}^n$ such that $\mathbf{x}$ is an ϵ-strongly typical sequence with respect to the distribution $\mathcal{P}_X$.

Definition 2.3 can be generalized to define strongly (jointly) typical sequences as follows.

Definition 2.4 (Strong joint typicality)**.** Assume $\epsilon \in (0,1)$ and $\ell \in \mathbb{N}$. Let $\mathbb{X}_i$ be an alphabet set with a finite size, where $i \in \{1,\ldots,\ell\}$. An ℓ-tuple of sequences $(\mathbf{x}_1,\ldots,\mathbf{x}_\ell) \in \mathbb{X}_1^n \times \ldots \times \mathbb{X}_\ell^n$ is said to be ϵ-strongly (jointly) typical with respect to distribution $\mathcal{P}_{X_1 \ldots X_\ell}$ on $\mathbb{X}_1 \times \ldots \times \mathbb{X}_\ell$ if for every $(\upsilon_1,\ldots,\upsilon_\ell) \in \mathbb{X}_1 \times \ldots \times \mathbb{X}_\ell$

$$\left|\frac{\eta(\upsilon_1,\ldots,\upsilon_\ell|\mathbf{x}_1,\ldots,\mathbf{x}_\ell)}{n} - \mathcal{P}_{X_1\ldots X_\ell}(\upsilon_1,\ldots,\upsilon_\ell)\right| < \begin{cases} \frac{\epsilon}{|\mathbb{X}_1|\ldots|\mathbb{X}_\ell|} & : if\ \mathcal{P}_{X_1\ldots X_\ell}(\upsilon_1,\ldots,\upsilon_\ell) > 0; \\ 0 & : if\ \mathcal{P}_{X_1\ldots X_\ell}(\upsilon_1,\ldots,\upsilon_\ell) = 0. \end{cases}$$

where $\eta(\upsilon_1,\ldots,\upsilon_\ell|\mathbf{x}_1,\ldots,\mathbf{x}_\ell)$ is the number of occurrences of the ℓ-tuple letters $(\upsilon_1,\ldots,\upsilon_\ell)$ of sequences $(\mathbf{x}_1,\ldots,\mathbf{x}_\ell)$. Moreover, $\mathbb{T}_\epsilon^*(\mathcal{P}_{X_1\ldots X_\ell})$ is called an ϵ-strongly (jointly) typical set, which contains all $(\mathbf{x}_1,\ldots,\mathbf{x}_\ell) \in \mathbb{X}_1^n \times \ldots \times \mathbb{X}_\ell^n$ such that $(\mathbf{x}_1,\ldots,\mathbf{x}_\ell)$ is ϵ-strongly typical sequences with respect to distribution $\mathcal{P}_{X_1\ldots X_\ell}$.

From Definition 2.4, if $(\mathbf{x}_1,\ldots,\mathbf{x}_\ell) \in \mathbb{T}_\epsilon^*(\mathcal{P}_{X_1\ldots X_\ell})$, then any sequences $(\mathbf{x}_{i_1},\mathbf{x}_{i_2},\ldots,\mathbf{x}_{i_j}) \in \mathbb{T}_\epsilon^*(\mathcal{P}_{X_{i_1}X_{i_2}\ldots X_{i_j}})$, where $j \in \{1,\ldots,\ell\}$ and $1 \leq i_1 < i_2 < \ldots < i_j \leq \ell$.

Remark 2.2. The strong typicality according to Definition 2.3 and Definition 2.4 can *not* be extended to continuous random vectors, because counting the occurrences of a given letter is meaningless in this case.

Markov lemma is a powerful tool for the proofs of Chapter 3. This lemma is stated in the following.

Lemma 2.1 (Markov lemma). *Assume discrete RVs* $(X, Y, Z) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$ *form Markov chain* $X \to Y \to Z$*, i.e.* $\mathcal{P}_{XYZ}(x, y, z) = \mathcal{P}_X(x)\mathcal{P}_{Y|X}(y|x)\mathcal{P}_{Z|Y}(z|y)$ *for any* $(x, y, z) \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$*. If for a* given $(\mathbf{y}, \mathbf{z}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{YZ})$, random vector $\mathbf{x}$ *is drawn according to* $\prod_{i=1}^n \mathcal{P}(x_i|y_i)$*, then*

$$\mathscr{P}\{(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{XYZ})|(\mathbf{y}, \mathbf{z}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{YZ})\} > 1 - \delta$$

where $\delta \to 0$ *as* $n \to \infty$*.*

In the following remark, we verify if the Markov lemma can be exploited for the Gaussian model in Chapter 4.

Remark 2.3. According to Remark 2.2, Markov lemma given in Lemma 2.1 does not apply to continuous RVs. However, Oohama [71] proved that Markov lemma still holds for Gaussian RVs if strongly typical sets in Lemma 2.1 are replaced by the corresponding weakly typical sets.

2.2 Wiretap Channels

In this section, we state the preliminary definitions of wiretap channels. These definitions are necessary to present our problems in the next chapters. We also review the secrecy problem in some wiretap models that will be recalled in Chapter 3 and Chapter 4 for comparison.

2.2.1 Wiretap Channels without Side Information

In this subsection, we define and review the secrecy problem over a memoryless wiretap channel (with no SI) based on publications [11], [12], [14], and [13].

Secrecy problem over a discrete memoryless wiretap channel (DM-WC) is presented in Figure 2.1. In the following, we declare the problem definitions.

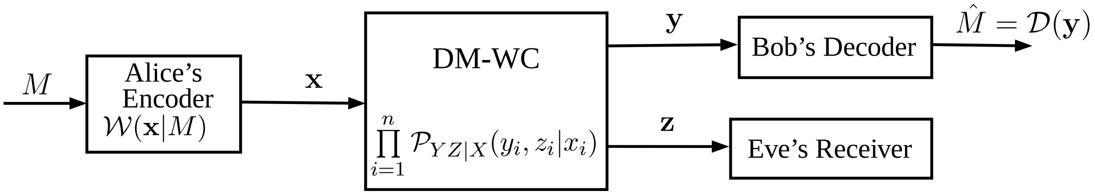

M
Alice's Encoder $\mathcal{W}(\mathbf{x}|M)$
$\mathbf{x}$
DM-WC
$\prod_{i=1}^{n} \mathcal{P}_{YZ|X}(y_i, z_i|x_i)$
$\mathbf{y}$
$\mathbf{z}$
Bob's Decoder
Eve's Receiver
$\hat{M} = \mathcal{D}(\mathbf{y})$

Figure 2.1: A discrete memoryless wiretap channel (with no SI).

Definition 2.5 (DM-WC)**.** Recalling Definition 1.7, assume the wiretap channel $(\mathbb{X}^n, \mathbb{Y}^n \times \mathbb{Z}^n, \mathcal{P}_{\mathbf{yz}|\mathbf{x}})$ with one sender (Alice) and two receivers (Bob and Eve). The wiretap channel is called a discrete memoryless wiretap channel (DM-WC) if

- $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ are finite sets;
- the PTF of the channel with input $\mathbf{x} \in \mathbb{X}^n$, and output $(\mathbf{y}, \mathbf{z}) \in \mathbb{Y}^n \times \mathbb{Z}^n$ is $\mathcal{P}(\mathbf{y}, \mathbf{z}|\mathbf{x}) = \prod_{i=1}^{n} \mathcal{P}_{YZ|X}(y_i, z_i|x_i)$.

For simplicity, a DM-WC is characterized by $(\mathbb{X}, \mathbb{Y} \times \mathbb{Z}, \mathcal{P}_{YZ|X})$.

Based on the discussions given in Section 1.3, an admissible wiretap coding scheme of a DM-WC for the secrecy problem is formulated as follows.

Definition 2.6. An admissible wiretap code $(\lceil 2^{nR} \rceil, n)$, where code block length $n \in \mathbb{N}$ and information rate $R \in \mathbb{R}^+ \cup \{0\}$, consists of the following components:

- a message set $\mathbb{M} = \{1, \ldots, \lceil 2^{nR} \rceil\}$. Message M is a RV which is uniformly distributed over $\mathbb{M}$;
- a (stochastic) encoder $\mathcal{W} : \mathbb{M} \to \mathbb{X}^n$ at Alice. This encoder maps each message $M \in \mathbb{M}$ to a codeword $\mathbf{x} \in \mathbb{X}^n$ according to PMF $\mathcal{W}(\mathbf{x}|M)$, where

$$\begin{cases} \mathcal{W}(\mathbf{x}|M) \geq 0 & : \forall \mathbf{x} \in \mathbb{X}^n, \forall M \in \mathbb{M}; \\ \sum\limits_{\mathbf{x} \in \mathbb{X}^n} \mathcal{W}(\mathbf{x}|M) = 1 & : \forall M \in \mathbb{M}; \end{cases} \tag{2.4}$$

- a decoder $\mathcal{D} : \mathbb{Y}^n \to \mathbb{M}$ at Bob. This decoder maps a received signal $\mathbf{y} \in \mathbb{Y}^n$ to a message $\hat{M} \in \mathbb{M}$ according to function $\hat{M} = \mathcal{D}(\mathbf{y})$;

As indicated in Sections 1.1 and 1.3, the reliability and security level of messages for a given admissible wiretap code are measured by the *average probability of error function* and (normalized) *equivocation rate function*, respectively [11, 30], as defined below.

Definition 2.7 (Measurements)**.** Assume an admissible wiretap code $(\lceil 2^{nR} \rceil, n)$ designed for the secrecy problem according to Definition 2.6.

1. The reliability of message M is measured by the average (block) probability of error
$$\mathcal{P}_{error}(n) \triangleq \mathscr{P}\{\hat{M} \neq M\} = \frac{1}{\lceil 2^{nR} \rceil} \sum_{i=1}^{\lceil 2^{nR} \rceil} \mathscr{P}\{\mathcal{D}(\mathbf{y}) \neq M | M = i\} \,, \tag{2.5}$$
which is a function of block length n.

2. The security level of message M at Eve is measured by the (normalized) equivocation rate
$$\mathcal{R}_{\scriptscriptstyle E}(n) \triangleq \frac{1}{n} \mathcal{H}(M|\mathbf{z}) \,, \tag{2.6}$$
which is a function of block length n.

The main objective of the secrecy problem is to find the *capacity-equivocation region* and *secrecy capacity* of a given model[1]. These terms are technically defined in the following.

Definition 2.8. A rate-equivocation pair $(R, R_{\scriptscriptstyle E})$ is said to be *achievable* if there exists an admissible wiretap code $(\lceil 2^{nR} \rceil, n)$ such that
$$\lim_{n \to \infty} \mathcal{P}_{error}(n) = 0 \,; \tag{2.7}$$
$$\liminf_{n \to \infty} \mathcal{R}_{\scriptscriptstyle E}(n) \geq R_{\scriptscriptstyle E} \,. \tag{2.8}$$

[1] Once the capacity-equivocation region of a given model is characterized, its secrecy capacity can be derived by an optimization problem. However, the secrecy capacity of a model is usually used as a criterion to compare it with other models in the field of information-theoretic security.

The capacity-equivocation region is the closure of the union of all achievable rate-equivocation pairs (R, R_E). Moreover, the *leakage rate* is defined as

$$R_L \triangleq R - R_E\ . \tag{2.9}$$

If the pair (R, R) is achievable, i.e., $R_L = 0$, then rate R is said to be (securely) achievable with respect to the AR condition 1.3 and the AS condition 1.12.

In Definition (1.9), the secrecy capacity is defined; equivalently, the supremum of all (securely) achievable rates R such that (R, R) is an achievable rate-equivocation pair is called the secrecy capacity.

The capacity-equivocation region and the secrecy capacity of the DM-WC is characterized by Csiszár and Körner [12] as follows.

Theorem 2.1. *The capacity-equivocation region of the DM-WC is*

$$\mathbb{C}_E = \bigcup_{\mathcal{P}_{VU}\mathcal{P}_{X|V}\mathcal{P}_{YZ|X}} \left\{ \begin{array}{l} (R, R_E) : \\ 0 \le R \le \mathcal{I}(V;Y), \\ 0 \le R_E \le R, \\ R_E \le \mathcal{I}(V;Y|U) - \mathcal{I}(V;Z|U) \end{array} \right\}, \tag{2.10}$$

where $U \in \mathbb{U}$ and $V \in \mathbb{V}$ are two auxiliary random variables such that $U \to V \to X \to (Y, Z)$, $|\mathbb{U}| \le |\mathbb{X}| + 3$, and $|\mathbb{V}| \le (|\mathbb{X}| + 1)(|\mathbb{X}| + 3)$. The secrecy capacity of the DM-WC is also given by

$$C_S = \max_{V \to X \to (Y,Z)} [\mathcal{I}(V;Y) - \mathcal{I}(V;Z)]\ . \tag{2.11}$$

The role of U and V in Theorem 2.1 can be explained as follows: V corresponds to the total message that can be decoded by Bob; however, U is a portion of that message which can be retrieved by both Bob and Eve. So, U is not secure by using the wiretap coding scheme. The other portion of V satisfies the AS condition 1.12 by using the wiretap coding scheme.

The secrecy capacity given in Theorem 2.1 can be simplified for some DM-WCs. To show this fact, let us first assume the following special cases of the DM-WCs[2].

- *Physically degraded wiretap channels*: a memoryless wiretap channel is said to be physically degraded in Bob's favor if the PTF of the channel can be factorized as

$$\forall x \in \mathbb{X}, \forall (y,z) \in \mathbb{Y} \times \mathbb{Z} : \quad \mathcal{P}_{YZ|X}(y,z|x) = \mathcal{P}_{Y|X}(y|x)\mathcal{P}_{Z|Y}(z|y) \, , \tag{2.12}$$

  As depicted in Figure 1.4, the Wyner's model [11] consists of a physically degraded wiretap channel in which $\mathcal{P}_{Y|X}$ and $\mathcal{P}_{Z|Y}$ are PTF of the main channel and that of the wiretap channel, respectively.

- *Stochastically degraded wiretap channels*: a memoryless wiretap channel is defined as a *stochastically degraded* wiretap channel in Bob's favor if its conditional marginal distributions are the same as those of a physically degraded wiretap channel; that is, if there exists a PTF $\mathcal{P}'_{Z|Y}$ such that

$$\forall \, x \in \mathbb{X}, z \in \mathbb{Z} : \quad \mathcal{P}_{Z|X}(z|x) = \sum_{y \, \in \mathbb{Y}} \mathcal{P}_{Y|X}(y|x)\mathcal{P}'_{Z|Y}(z|y) \, . \tag{2.13}$$

- *Less noisy wiretap channels*: This special case is a generalization of physically/ stochastically degraded wiretap channels. The less noisy ordering was introduced by Körner and Marton [16] to compare two noisy channels as follows.

  Definition 2.9 (Less noisy wiretap channel)**.** Bob's channel is said to be less noisy than Eve's channel (less noisy wiretap channel in Bob's favor) if for every auxiliary RV U with property $U \to X \to (Y,Z)$, we have

$$\mathcal{I}(U;Y) \geq \mathcal{I}(U;Z) \, . \tag{2.14}$$

[2] In fact, this classification is valid for a memoryless wiretap channel with continuous alphabets as well [2, 6].

According to this definition, a physically/stochastically degraded wiretap channel is a less noisy channel as well; however, the converse in not generally true [6]. Also, a physically degraded, stochastically degraded, and a less noisy wiretap channel in Eve's favor can be defined if Y is switched to Z in (2.12), (2.13), and (2.14), respectively.

With applying (2.14) to Theorem 2.1, the results given in this theorem can be simplified for less noisy wiretap channels (and so for degraded wiretap channels) if $V = X$ and $U = 0$ are relaxed [12, Thm. 3] as follows.

Theorem 2.2. *The capacity-equivocation region of a less noisy DM-WC is*

$$\mathbb{C}_{E} = \bigcup_{\mathcal{P}_X \mathcal{P}_{YZ|X}} \left\{ \begin{array}{l} (R, R_{E}) : \\ 0 \leq R \leq \mathcal{I}(X;Y), \\ 0 \leq R_{E} \leq R, \\ R_{E} \leq \mathcal{I}(X;Y) - \mathcal{I}(X;Z) \end{array} \right\} . \tag{2.15}$$

The secrecy capacity of a less noisy DM-WC is also given by

$$C_{S} = \max_{\mathcal{P}_X} [\mathcal{I}(X;Y) - \mathcal{I}(X;Z)] . \tag{2.16}$$

Theorem 2.2 gives the capacity-equivocation region and the secrecy capacity of Wyner's model [11] as well. According to [12, Corol. 3], the secrecy capacity of the DM-WC is always positive unless the DM-WC is less noisy in Eve's favor.

In a less noisy wiretap channel, if both $\mathcal{I}(X;Y)$ and $\mathcal{I}(X;Z)$ achieve their maximum values at one distribution $\mathcal{P}^*_X$, then the capacity-equivocation region given in Theorem 2.2 can be simplified as follows [14].

$$\mathbb{C}_{E} = \{(R, R_{E}) : 0 \leq R_{E} \leq R \leq C_m, R_{E} \leq C_m - C_{mw}\} , \tag{2.17}$$

where $C_m = \max\limits_{\mathcal{P}_X} \mathcal{I}(X;Y)$ and $C_{mw} = \max\limits_{\mathcal{P}_X} \mathcal{I}(X;Z)$ are the capacity of the main channel (from Alice to Bob) and the capacity of the overall wiretap channel (from Alice to Eve),

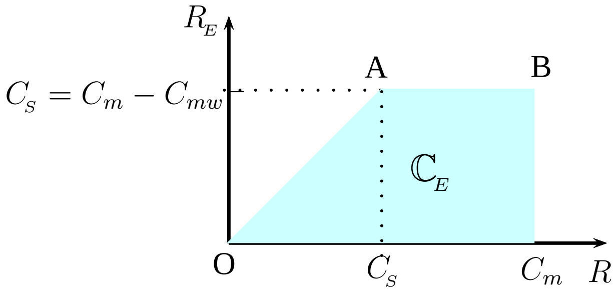


Figure 2.2: The capacity-equivocation region for a degraded symmetric DM-WC.

respectively. In this case, the secrecy capacity of the DM-WC is $C_S = C_m - C_{mw}$, which is given in (1.13) as well. Region $\mathbb{C}_E$, which is given in 2.17, is sketched in Figure 2.2. An example of this model is a physically degraded DM-WC in which both the main channel (from Alice to Bob) and the wiretap channel (from Bob to Eve) are symmetric DMCs [1,2].

According to Figure 2.2, line OA corresponds to the securely achievable rates, in which point A corresponds to the secrecy capacity. Increasing the information rate above C_S does not increase the (normarized) equivocation rate of the communication, and thus leakage rate R_L grows linearly as function of R. At point B, the leakage rate is C_{mw}; in other words, in average, portion $\frac{C_m - C_{mw}}{C_m}$ of the achievable rate satisfies the AS condition 1.12, and portion $\frac{C_{mw}}{C_m}$ of the rate is retrievable by Eve.

2.2.2 Degraded Discrete Memoryless Wiretap Channel with Random States

In this subsection, we define a (physically) degraded[3] discrete memoryless state-dependent wiretap channel (DM-SWC) with random states, which are non-causally known at the transmitter. The secrecy problem over this channel is reviewed in this subsection according to publications [61, 62]. The model is illustrated in Figure 2.3 and it is defined as follows.

[3] In this subsection, the wiretap channel is physically degraded in Bob's favor; however, the results are valid for a stochastically degraded channel as well.

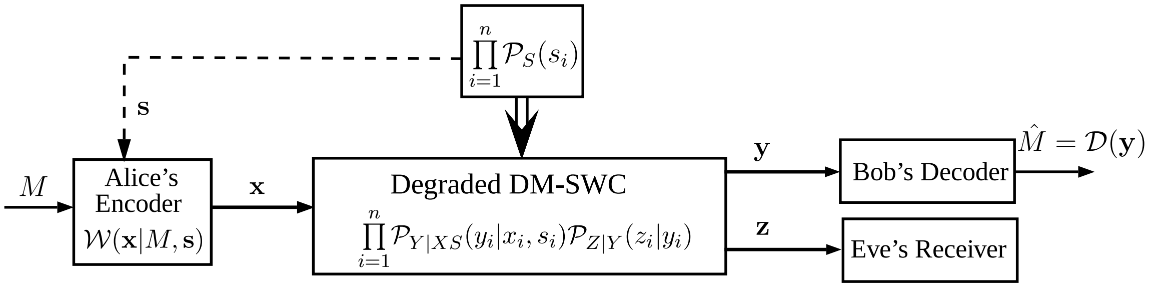

$\prod_{i=1}^{n}\mathcal{P}_S(s_i)$
$\mathbf{s}$
M
Alice's Encoder
$\mathcal{W}(\mathbf{x}|M,\mathbf{s})$
$\mathbf{x}$
Degraded DM-SWC
$\prod_{i=1}^{n}\mathcal{P}_{Y|XS}(y_i|x_i,s_i)\mathcal{P}_{Z|Y}(z_i|y_i)$
$\mathbf{y}$
$\mathbf{z}$
Bob's Decoder
Eve's Receiver
$\hat{M}=\mathcal{D}(\mathbf{y})$

Figure 2.3: A degraded DM-SWC with non-causal CSI at the transmitter.

Definition 2.10 (Degraded DM-SWC)**.** A (physically) degraded DM-SWC in Bob's favor with non-causal CSI at Alice is determined by $(\mathbb{S},\mathbb{X},\mathbb{Y}\times\mathbb{Z},\mathcal{P}_S,\mathcal{P}_{YZ|XS})$, where

- finite sets $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ are the channel input alphabet (from Alice), the first channel output alphabet (to Bob), and the second channel output alphabet (to Eve), respectively;
- $\mathbb{S}$ is the state alphabet with a finite size; also, random vector $\mathbf{s}\in\mathbb{S}^n$, which is called CSI, is drawn i.i.d. according to distribution $\mathcal{P}(\mathbf{s})=\prod_{i=1}^{n}\mathcal{P}_S(s_i)$;
- prior to each block transmission, Alice knows the realization of random vector $\mathbf{s}$;
- the PTF of the channel with the input $\mathbf{x}\in\mathbb{X}^n$, state $\mathbf{s}\in\mathbb{S}^n$, and output pair $(\mathbf{y},\mathbf{z})\in\mathbb{Y}^n\times\mathbb{Z}^n$ is $\mathcal{P}(\mathbf{y},\mathbf{z}|\mathbf{x},\mathbf{s})=\prod_{i=1}^{n}\mathcal{P}_{Y|XS}(y_i|x_i,s_i)\mathcal{P}_{Z|Y}(z_i|y_i)$, i.e., Eve receives a degraded version of Bob's signal through a channel with PTF $\mathcal{P}_{Z|Y}$.

An admissible wiretap coding scheme for the secrecy problem over the degraded DM-SWC is defined as follows.

Definition 2.11. An admissible wiretap code $(\lceil 2^{nR}\rceil,n)$, where code block length $n\in\mathbb{N}$ and information rate $R\in\mathbb{R}^+\cup\{0\}$, consists of the following components:

- a message set $\mathbb{M}=\{1,\ldots,\lceil 2^{nR}\rceil\}$. Message M is a RV which is uniformly distributed over $\mathbb{M}$;

- a (stochastic) encoder $\mathcal{W} : \mathbb{M} \times \mathbb{S}^n \to \mathbb{X}^n$ at Alice. Knowing side information $\mathbf{s}$, this encoder maps each message $M \in \mathbb{M}$ to a codeword $\mathbf{x} \in \mathbb{X}^n$ according to PMF $\mathcal{W}(\mathbf{x}|M, \mathbf{s})$, where

$$\begin{cases} \mathcal{W}(\mathbf{x}|M, \mathbf{s}) \geq 0 & : \forall \mathbf{x} \in \mathbb{X}^n, \forall M \in \mathbb{M}, \forall \mathbf{s} \in \mathbb{S}^n; \\ \sum\limits_{\mathbf{x} \in \mathbb{X}^n} \mathcal{W}(\mathbf{x}|M, \mathbf{s}) = 1 & : \forall M \in \mathbb{M}, \forall \mathbf{s} \in \mathbb{S}^n; \end{cases} \tag{2.18}$$

- a decoder $\mathcal{D} : \mathbb{Y}^n \to \mathbb{M}$ at Bob. This decoder maps a received signal $\mathbf{y} \in \mathbb{Y}^n$ to a message $\hat{M} \in \mathbb{M}$ according to function $\hat{M} = \mathcal{D}(\mathbf{y})$.

The measurements given in Definition 2.7 are also applied to the admissible wiretap codes defined in Definition 2.11. Bounds on the secrecy capacity of the degraded DM-SWC according to Definition 1.9 is given in the following theorem [61, 62].

Theorem 2.3. *The secrecy capacity of the degraded DM-SWC, which is introduced in Definition 2.10, is bounded by*

$$\max_{U \to (X,S) \to Y \to Z} \min\{\mathcal{I}(U;Y) - \mathcal{I}(U;S),\, \mathcal{I}(U;Y) - \mathcal{I}(U;Z)\} \leq C_{_S} \leq \min\{R_1, R_2\}\,, \tag{2.19}$$

where

$$R_1 = \max_{U \to (X,S) \to Y} [\mathcal{I}(U;Y) - \mathcal{I}(U;S)]\,, \tag{2.20}$$

$$R_2 = \max_{U \to (X,S) \to Y \to Z} [\mathcal{I}(U;Y) - \mathcal{I}(U;Z)]\,. \tag{2.21}$$

The structure of the achievable wiretap coding scheme is based on double random binning [6], which is generated by combination of Geĺfand-Pinsker's coding strategy [49] and wiretap coding strategies [11, 31]. Two upper bounds on the secrecy capacity are reported in Theorem 2.3: the first one (R_1) is the (ordinary) capacity of the main channel given in (1.32), and the second one (R_2) is the secrecy capacity of the DM-WC given in (2.11). The second UB is derived based on the fact that a (degraded) DM-WC with input (X, S) is an enhanced channel of the degraded DM-SWC in the sense that the secrecy capacity of the former is not less than the secrecy capacity of the latter.

It is still an open problem if any of the bounds in Theorem 2.3 is tight or not. However, the LB is proved to be optimal [61] in the following special case.

Corollary 2.1. *Assume the main channel capacity of the degraded DM-SWC, which is given in* (1.32)*, is achievable at input distribution* $\mathcal{P}^*_{X|SU}\mathcal{P}^*_{U|S}\mathcal{P}_S$*. At this distribution, if* $\mathcal{I}(U^*;S) \geq \mathcal{I}(U^*;Z^*)$*, then the secrecy capacity optimally equals the main channel capacity, i.e.,*

$$C_S = \max_{U \to (X,S) \to Y \to Z} [\mathcal{I}(U;Y) - \mathcal{I}(U;S)] = \mathcal{I}(U^*;Y^*) - \mathcal{I}(U^*;S) \, ,$$

where $\mathcal{I}(U^*;S)$*,* $\mathcal{I}(U^*;Y^*)$*, and* $\mathcal{I}(U^*;Z^*)$ *are calculated at distribution*

$$\mathcal{P}_{YZXUS} = \mathcal{P}_{Z|Y}\mathcal{P}_{Y|XS}\mathcal{P}^*_{X|SU}\mathcal{P}^*_{U|S}\mathcal{P}_S \, .$$

As implied by Corollary 2.1, the side information can assist Alice to (securely) achieve the main channel capacity with respect to the AS condition 1.12.

2.2.3 Gaussian Wiretap Channel with Additive Interference

As mentioned in Section 1.6, the secrecy problem of a physically degraded Gaussian wiretap channel with AWGI was studied by Mitrpant *et al.* [60, 76]. Their model is illustrated in Figure 2.4. In this subsection, we highlight the main results and ideas of that work.

A (physically) degraded Gaussian wiretap channel with AWGI is defined as follows.

Definition 2.12 (Physically degraded G-SWC)**.** A physically degraded Gaussian state-dependent wiretap channel (G-SWC) in Bob's favor (respectively, in Eve's favor) with non-causally known CSI at the transmitter is determined by 4-tuple $(\Gamma, \Lambda, \sigma_1^2, \sigma_2^2) \in \mathbb{R}^{+4}$, where

- random vector $\mathbf{x} \in \mathbb{R}^n$ is the channel input (from Alice), which is subject to average

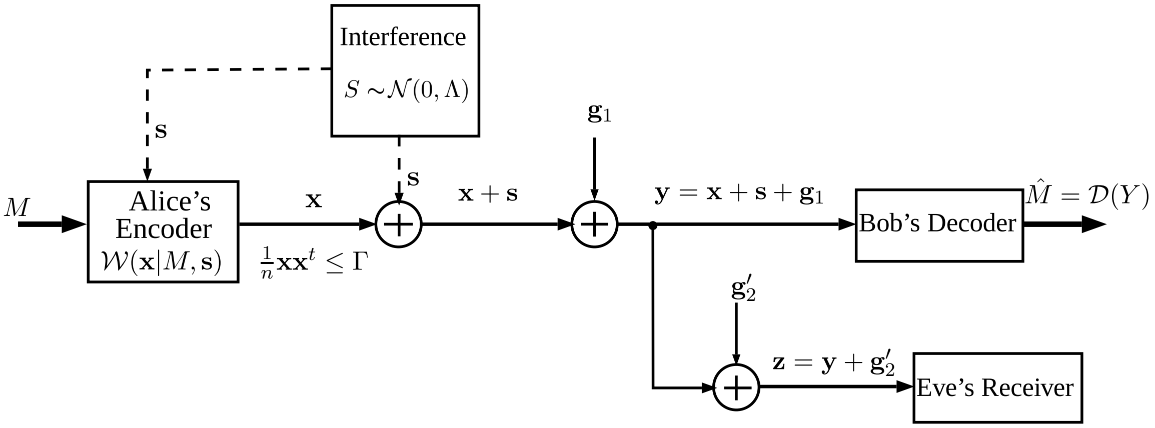

Interference
$S \sim \mathcal{N}(0,\Lambda)$
$\mathbf{s}$
$\mathbf{s}$
M
Alice's Encoder
$\mathcal{W}(\mathbf{x}|M,\mathbf{s})$
$\mathbf{x}$
$\frac{1}{n}\mathbf{x}\mathbf{x}^t \leq \Gamma$
$\mathbf{x}+\mathbf{s}$
$\mathbf{g}_1$
$\mathbf{y}=\mathbf{x}+\mathbf{s}+\mathbf{g}_1$
Bob's Decoder
$\hat{M}=\mathcal{D}(Y)$
$\mathbf{g}_2'$
$\mathbf{z}=\mathbf{y}+\mathbf{g}_2'$
Eve's Receiver

Figure 2.4: Secrecy problem over a physically degraded G-SWC in Bob's favor.

power constraint

$$\frac{1}{n}\mathbf{x}\mathbf{x}^t \leq \Gamma \,; \tag{2.22}$$

- random vector $\mathbf{s} \in \mathbb{R}^n$, which is known as *interference*, is drawn i.i.d. according to $S \sim \mathcal{N}(0,\Lambda)$;
- the realization of random vector $\mathbf{s}$ is known at Alice prior to each block transmission;
- if $\sigma_2 \geq \sigma_1$, the channel is physically degraded in Bob's favor, and then
  - random matrix $\mathbf{G} = \begin{bmatrix}\mathbf{g}_1\\ \mathbf{g}_2'\end{bmatrix}$ represents Gaussian noise of the channel, where $\mathbf{g}_1 \in \mathbb{R}^n$ and $\mathbf{g}_2' \in \mathbb{R}^n$ are independent random vectors such that components of $\mathbf{g}_1$ and those of $\mathbf{g}_2'$ are drawn i.i.d. according to $\mathcal{N}(0,\sigma_1^2)$ and $\mathcal{N}(0,\sigma_2^2-\sigma_1^2)$, respectively;
  - $\mathbf{y} \in \mathbb{R}^n$, and $\mathbf{z} \in \mathbb{R}^n$ are the first channel output (to Bob) and the second channel output (to Eve), respectively, such that

$$\mathbf{y} = \mathbf{x}+\mathbf{s}+\mathbf{g}_1 \,, \tag{2.23a}$$

$$\mathbf{z} = \mathbf{y}+\mathbf{g}_2' \,, \tag{2.23b}$$

which means that Eve receives a noisy version of Bob's signal;

- if $\sigma_1 > \sigma_2$, the channel is physically degraded in Eve's favor, and then
  - random matrix $\mathbf{G} = \begin{bmatrix} \mathbf{g}_1' \\ \mathbf{g}_2 \end{bmatrix}$ represents Gaussian noise of the channel, where $\mathbf{g}_1' \in \mathbb{R}^n$ and $\mathbf{g}_2 \in \mathbb{R}^n$ are independent random vectors such that components of $\mathbf{g}_1'$ and those of $\mathbf{g}_2$ are drawn i.i.d. according to $\mathcal{N}(0, \sigma_1^2 - \sigma_2^2)$ and $\mathcal{N}(0, \sigma_2^2)$, respectively;
  - $\mathbf{y} \in \mathbb{R}^n$, and $\mathbf{z} \in \mathbb{R}^n$ are the first channel output (to Bob) and the second channel output (to Eve), respectively, such that

$$\mathbf{z} = \mathbf{x} + \mathbf{s} + \mathbf{g}_2 \, , \tag{2.24a}$$

$$\mathbf{y} = \mathbf{z} + \mathbf{g}_1' \, , \tag{2.24b}$$

which means Bob receives a noisy version of Eve's signal.

An admissible wiretap coding scheme for the secrecy problem over the G-SWC is defined as follows.

Definition 2.13. An admissible wiretap code $(\lceil 2^{nR} \rceil, n)$, where code block length $n \in \mathbb{N}$ and information rate $R \in \mathbb{R}^+ \cup \{0\}$, consists of the following components:

- a message set $\mathbb{M} = \{1, \dots, \lceil 2^{nR} \rceil\}$. Message M is a RV which is uniformly distributed over $\mathbb{M}$;
- a (stochastic) encoder $\mathcal{W} : \mathbb{M} \times \mathbb{R}^n \to \mathbb{R}^n$ at Alice. Having state vector $\mathbf{s}$, the encoder maps each message $M \in \mathbb{M}$ to a codeword $\mathbf{x} \in \mathbb{R}^n$ according to conditional distribution $\mathcal{W}(\mathbf{x}|M, \mathbf{s})$ such that the average power constraint (2.22) is met, where

$$\begin{cases} \mathcal{W}(\mathbf{x}|M, \mathbf{s}) \geq 0 & : \forall \mathbf{x} \in \mathbb{R}^n, \forall M \in \mathbb{M}, \forall \mathbf{s} \in \mathbb{R}^n; \\ \sum\limits_{\mathbf{x} \in \mathbb{R}^n} \mathcal{W}(\mathbf{x}|M, \mathbf{s}) = 1 & : \forall M \in \mathbb{M}, \forall \mathbf{s} \in \mathbb{R}^n; \end{cases} \tag{2.25}$$

- a decoder $\mathcal{D} : \mathbb{R}^n \to \mathbb{M}$ at Bob. This decoder maps a received signal $\mathbf{y} \in \mathbb{R}^n$ to a message $\hat{M} \in \mathbb{M}$ according to function $\hat{M} = \mathcal{D}(\mathbf{y})$.

The measurements given in Definition 2.7 are also applied to the admissible wiretap codes defined in Definition 2.13. The secrecy capacity of the (physically) degraded G-SWC according to Definition (1.9) is bounded in the following theorem [60, 76].

Theorem 2.4. *Let C_s be the capacity of a (physically) degraded G-SWC with non-causally known interference at the transmitter. Let define*

$$\mathcal{R}(\alpha) \triangleq \frac{1}{2}\log\left(\frac{\Gamma(\Gamma+\Lambda+\sigma_1^2)}{(\Gamma+\alpha^2\Lambda)(\Gamma+\Lambda+\sigma_1^2)-(\Gamma+\alpha\Lambda)^2}\right), \tag{2.26}$$

$$\mathcal{R}_Z(\alpha) \triangleq \frac{1}{2}\log\left(\frac{(\Gamma+\Lambda+\sigma_1^2)[(1-\alpha)^2\Gamma\Lambda+\sigma_2^2(\Gamma+\alpha^2\Lambda)]}{(\Gamma+\Lambda+\sigma_2^2)[(1-\alpha)^2\Gamma\Lambda+\sigma_1^2(\Gamma+\alpha^2\Lambda)]}\right), \tag{2.27}$$

$$\alpha^* \triangleq \frac{\Gamma}{\Gamma+\sigma_1^2}, \tag{2.28}$$

$$\alpha_0 \triangleq \frac{\Gamma\Lambda+\Gamma\sqrt{\Lambda(\Gamma+\Lambda+\sigma_2^2)}}{\Lambda(\Gamma+\sigma_2^2)}. \tag{2.29}$$

$$\Gamma_l \triangleq \left[-\sigma_1^2-\frac{\Lambda}{2}+\frac{\sqrt{\Lambda^2+4\Lambda(\sigma_2^2-\sigma_1^2)}}{2}\right]^+, \tag{2.30}$$

$$\Gamma_h \triangleq -\frac{\Lambda}{2}+\frac{\sqrt{\Lambda^2+4\Lambda\sigma_2^2}}{2}. \tag{2.31}$$

Also, $C_m = \mathcal{R}(\alpha^*)$ is the capacity of the main channel [55]. Then, the secrecy capacity of the degraded G-SWC is lower bounded by

$$C_s \geq \begin{cases} C_m & : if\ 0 \leq \Gamma \leq \Gamma_l \\ \mathcal{R}(\alpha_0) & : if\ \Gamma_l \leq \Gamma \leq \Gamma_h \\ \mathcal{R}_Z(1) & : if\ \Gamma \geq \Gamma_h . \end{cases} \tag{2.32}$$

Mitrpant and Han Vinck [60] applied DPC strategy [55] to construct an achievable wiretap code based on double random binning. The structure of the double random binning for the degraded G-SWC is the same as that of the degraded DM-SWC in Theorem 2.3. To explain this idea, let $U = X + \alpha S$, where X and S are orthogonal RVs with Gaussian distributions according to $\mathcal{N}(0,\Gamma)$ and $\mathcal{N}(0,\Lambda)$, respectively; $\alpha \in \mathbb{R}^+$ is a constant as well. Hence, $\mathcal{R}(\alpha)$ and $\mathcal{R}_Z(\alpha)$, which is defined in (2.26) and (2.27) respectively, can be formulated as [60]

$$\begin{aligned} \mathcal{R}(\alpha) &= \mathcal{I}(U;Y) - \mathcal{I}(U;S) \ , \\ \mathcal{R}_Z(\alpha) &= \mathcal{I}(U;Y) - \mathcal{I}(U;Z) \ . \end{aligned} \tag{2.33}$$

Similar to Theorem 2.3, we have

$$C_s \geq \max_{\alpha} \min\{\mathcal{R}(\alpha), \mathcal{R}_Z(\alpha)\} \ . \tag{2.34}$$

Maximum value of $\mathcal{R}_Z(\alpha)$ and that of $\mathcal{R}(\alpha)$ are achieved at $\alpha = 1$ and $\alpha = \alpha^*$, respectively. Also, $\mathcal{R}(\alpha)$ and $\mathcal{R}_Z(\alpha)$ coincide at $\alpha = \alpha_0$. These values of α are given in Theorem 2.4. According to [55], $\alpha = \alpha^*$ is the optimum value for the DPC to achieve capacity $C_m = \mathcal{R}(\alpha^*)$. Value of α^* is independent of variance of the interference Λ. Now, consider the secrecy problem for the G-SWC, and denote the optimum value of α used in the corresponding DPC by α_{opt}. The following statements determine the value of α_{opt} as a function of power Γ (when other parameters are assumed to be fixed).

- $0 \leq \Gamma \leq \Gamma_l$: In this case $C_m = \mathcal{R}(\alpha^*) \leq \mathcal{R}_Z(\alpha^*)$, and so secrecy capacity $C_s = C_m$ is achievable with $\alpha_{opt} = \alpha^*$. At $\Gamma = \Gamma_l$, $\alpha_0 = \alpha^*$ and so $\mathcal{R}(\alpha^*) = \mathcal{R}_Z(\alpha^*)$.

- $\Gamma_l < \Gamma \leq \Gamma_h$: In this case

$$\mathcal{R}_Z(\alpha^*) \leq \mathcal{R}_Z(\alpha_0) = \mathcal{R}(\alpha_0) < \mathcal{R}(\alpha^*) = C_m \ , \tag{2.35}$$

  where $\alpha^* < \alpha_0 \leq 1$. As α goes from α^* to 1, $\mathcal{R}_Z(\alpha)$ increases and $\mathcal{R}(\alpha)$ decreases. In this case, $\alpha_{opt} = \alpha_0$ where these functions coincide, and so $C_s \geq \mathcal{R}(\alpha_0) = \mathcal{R}_Z(\alpha_0)$ according to (2.34). At $\Gamma = \Gamma_h$, $\alpha_0 = 1$, and so $\mathcal{R}_Z(\alpha) \leq \mathcal{R}(1) = \mathcal{R}_Z(1)$ for any α.

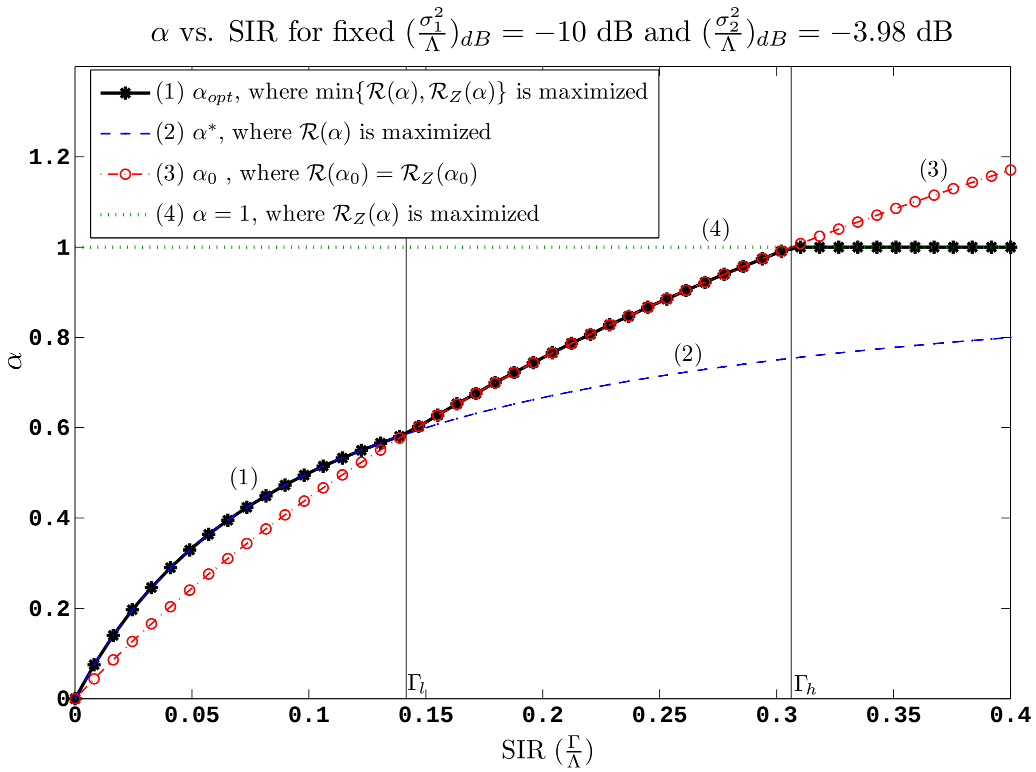


Figure 2.5: Comparison of optimum values of α (in the DPC) versus SIR in the G-SWC.

- $\Gamma > \Gamma_h$: In this case

$$\mathcal{R}(\alpha_0) = \mathcal{R}_Z(\alpha_0) \leq \mathcal{R}_Z(1) < \mathcal{R}(1)\,, \tag{2.36}$$

where $\alpha^* \leq 1 < \alpha_0$. $\mathcal{R}(\alpha)$ decreases from its maximum point $\mathcal{R}(\alpha^*)$ as α goes from α^* to α_0, and $R_Z(1) = \max_{\alpha} \mathcal{R}_Z(\alpha)$. Hence, in this case $\alpha_{opt} = 1$ and so

$$C_s \geq \mathcal{R}_Z(1) = \frac{1}{2}\log\left(\frac{(\Gamma+\Lambda+\sigma_1^2)\sigma_2^2}{(\Gamma+\Lambda+\sigma_2^2)\sigma_1^2}\right) \tag{2.37}$$

according to (2.34) and Theorem 2.4.

In Figure 2.5, we compare the optimum value α for which the capacity is achieved, i.e. α^*, with that for which the LB on the secrecy capacity, which is given in 2.34, is achieved, i.e. α_{opt}. The parameters in this figure are $\Lambda = 1$, $\sigma_1^2 = .1$, and $\sigma_2^2 = .4$. According to Figure 2.5, α_{opt} is the same as α^* for $0 \leq \Gamma \leq \Gamma_l$. For $\Gamma_l < \Gamma \leq \Gamma_h$, α_{opt} equals α_0 and it is larger than α^*. For $\Gamma > \Gamma_h$, $\alpha_{opt} = 1$ and $\alpha^* < \alpha_{opt} < \alpha_0$.

In Figure 2.6, the LB and UBs on the secrecy capacity of the (physically) degraded G-SWC is sketched [60], where $\Lambda = 1$, $\sigma_1^2 = .1$, and $\sigma_2^2 = .4$ are fixed. In this figure, curve (1) illustrates the LB on the secrecy capacity for the G-SWC according to Theorem 2.4; curve (2) is the main channel capacity (from Alice to Bob) of the G-SWC; curve (3) is an UB on the secrecy capacity claimed by Mitrpant and Han Vinck [60] (it will be discussed in the following); curve (4) is the secrecy capacity of the corresponding Gaussian wiretap channel sketched in Figure 1.5 (same parameters Γ, σ_1^2, σ_2^2, and no interference, i.e., $\Lambda = 0$) according to (1.14).

According to Theorem 2.4, the DPC is optimum for the secrecy problem when $\Gamma \leq \Gamma_l$ [60]. Mitrpant and Han Vinck [60] also claimed that the LB given in (2.37) is optimum for $\Gamma \geq \Gamma_h$ because of the following discussion. Assume Alice is able to govern the interference as its input. In other words, the interference is a part of Alice's transmitted signal which has a fixed power Λ. Thus, the authors [60] deduced that a Gaussian wiretap channel with transmission power $\Gamma + \Lambda$ is an enhanced wiretap channel of the given G-SWC. As the LB in (2.37) is actually the secrecy capacity of that enhanced channel (replace Γ by $\Gamma + \Lambda$ in (1.14)), the authors concluded that the LB (2.37) is tight.

In Figure 2.6, the LB on the secrecy capacity of the G-SWC is also compared with the secrecy capacity of a Gaussian wiretap channel (when there is no interference) [13]. This comparison shows that non-causal availability of the interference at Alice enhances the secrecy capacity of a Gaussian wiretap channel. This notable fact is in contrast with the fact that the channel capacity of Costa's model [55] is the same as that of the Gaussian channel (when there is no interference) [1, 2]. In other words, knowing the interference at Alice does not affect the main channel capacity of the G-SWC, but it enlarges its secrecy capacity (comparing with a corresponding Gaussian wiretap channel with no interference).

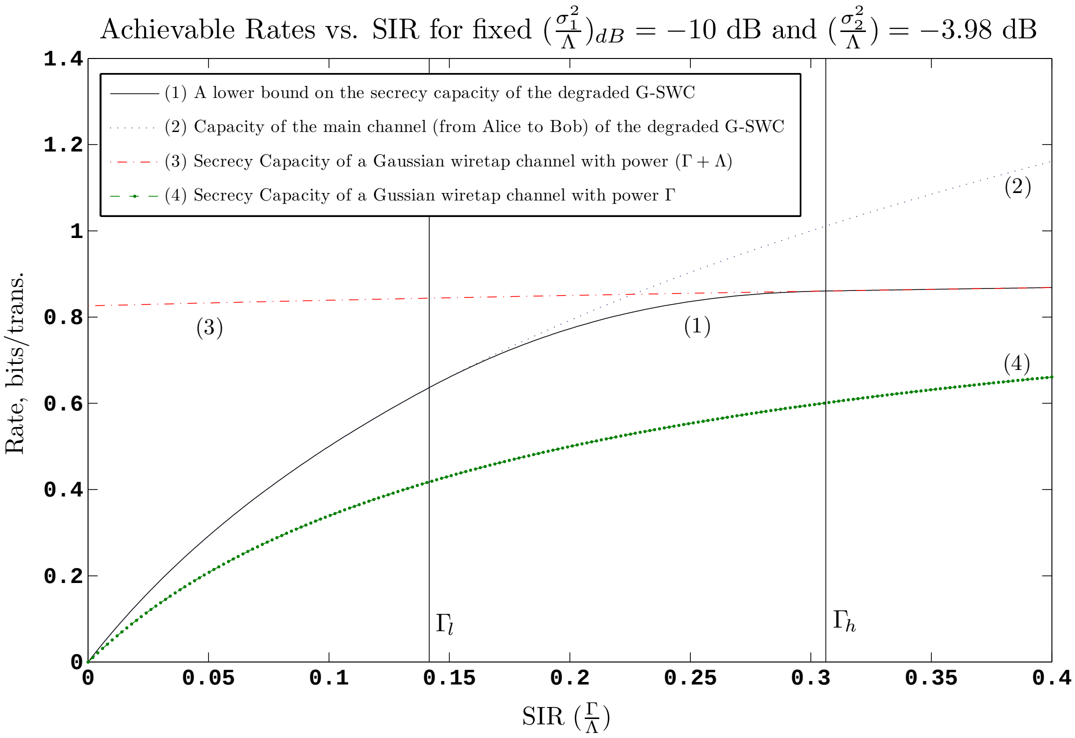


Figure 2.6: Bounds on the secrecy capacity of the degraded G-SWC.

Chapter 3

The Discrete Memoryless Model

In this chapter, we study the key agreement problem over a discrete memoryless (DM) model. The model consists of a wiretap channel with random states and a parallel one-way public channel in the forward direction with capacity $C_{P_1} \in [0, \infty)$. The CSI is drawn i.i.d., and its realizations are fully and non-causally known at Alice. The goal of this work is to characterize the (forward) key capacity of the DM model as a function of C_{P_1}.

The rest of this chapter is organized as follows. In Section 3.1, we will define the public channel in its general form, and we will introduce the key agreement problem over a DM model. In that section, we will technically define an admissible key agreement coding scheme for the key agreement problem. In Section 3.2, we will state the achieved bounds on the (forward) key capacity of the model. Finally, we will establish the proofs of the results in Section 3.3.

3.1 Problem Definitions

In this section, we define the key agreement problem over the DM model. Before we present the DM model, let us define a two-way (authenticated) public channel with given capacity in each direction. As a special case, a one-way public channel is also specified in the following definition.

Definition 3.1 (Public channel)**.** Assume $n \in \mathbb{N}$ is time duration, and side information $\mathbf{a}$ and $\mathbf{b}$ is available at Alice and Bob, respectively. Let $\mathbf{p}_1 \triangleq (P_{11}, P_{12}, \ldots, P_{1n}) \in \mathbb{P}_{11} \times \mathbb{P}_{12} \times \ldots \mathbb{P}_{1n}$ and $\mathbf{p}_2 \triangleq (P_{21}, P_{22}, \ldots, P_{2n}) \in \mathbb{P}_{21} \times \mathbb{P}_{22} \times \ldots \mathbb{P}_{2n}$ be the forward public message sequence and the backward public message sequence, respectively, where $\mathbb{P}_{1\tau}$ and $\mathbb{P}_{2\tau}$ are the alphabet set of the forward message and that of the backward message at time $\tau \in \{1, \ldots, n\}$, respectively. A public channel with a given capacity pair $(C_{_{P_1}}, C_{_{P_2}})$ is a noiseless channel between Alice and Bob which consists of a noiseless forward channel from Alice to Bob with capacity $C_{_{P_1}}$ and a noiseless backward channel from Bob to Alice with capacity $C_{_{P_2}}$ such that

- at time instant $\tau \in \{1, \ldots, n\}$, the forward message $P_{1\tau}$ is sent from Alice to Bob (in the forward direction), and the backward message $P_{2\tau}$ is sent from Bob to Alice (in the backward direction);
- $P_{1\tau}$ is a function of all available information at Alice up to time $\tau - 1$, i.e., it is a (stochastic) function of $\mathbf{a}$ and $(P_{21}, P_{22}, \ldots, P_{2(\tau-1)})$, where $P_{20} \triangleq 0$;
- $P_{2\tau}$ is a function of all available information at Bob up to time $\tau - 1$, i.e., it is a (stochastic) function of $\mathbf{b}$, $(P_{11}, P_{12}, \ldots, P_{1(\tau-1)})$, where $P_{10} \triangleq 0$, and $(Y_1, Y_2, \ldots, Y_{\tau-1})$, if there exists a wiretap channel connected from Alice to Bob with output symbol Y_τ at time τ;
- $\mathbf{p}_1$ and $\mathbf{p}_2$ are intercepted by Eve;
- the communication rate over the public channel is subject to

$$\text{forward capacity constraint:} \qquad \limsup_{n\to\infty} \frac{\log(|\mathbb{P}_1|)}{n} \leq C_{_{P_1}} , \tag{3.1a}$$

$$\text{backward capacity constraint:} \qquad \limsup_{n\to\infty} \frac{\log(|\mathbb{P}_2|)}{n} \leq C_{_{P_2}} , \tag{3.1b}$$

where $\mathbb{P}_1 \triangleq \mathbb{P}_{11} \times \ldots \times \mathbb{P}_{1n}$ and $\mathbb{P}_2 \triangleq \mathbb{P}_{21} \times \ldots \times \mathbb{P}_{2n}$.

If $C_{_{P_2}} = 0$, the public channel is called a (one-way) forward public channel; if $C_{_{P_1}} = 0$, the public channel is called a (one-way) backward public channel.

As sketched in Figure 3.1, the DM model consists of two parallel channels: a discrete memoryless state-dependent wiretap channel (DM-SWC) with non-causally known CSI at the transmitter and a forward public channel. In the following, the DM-SWC is defined.

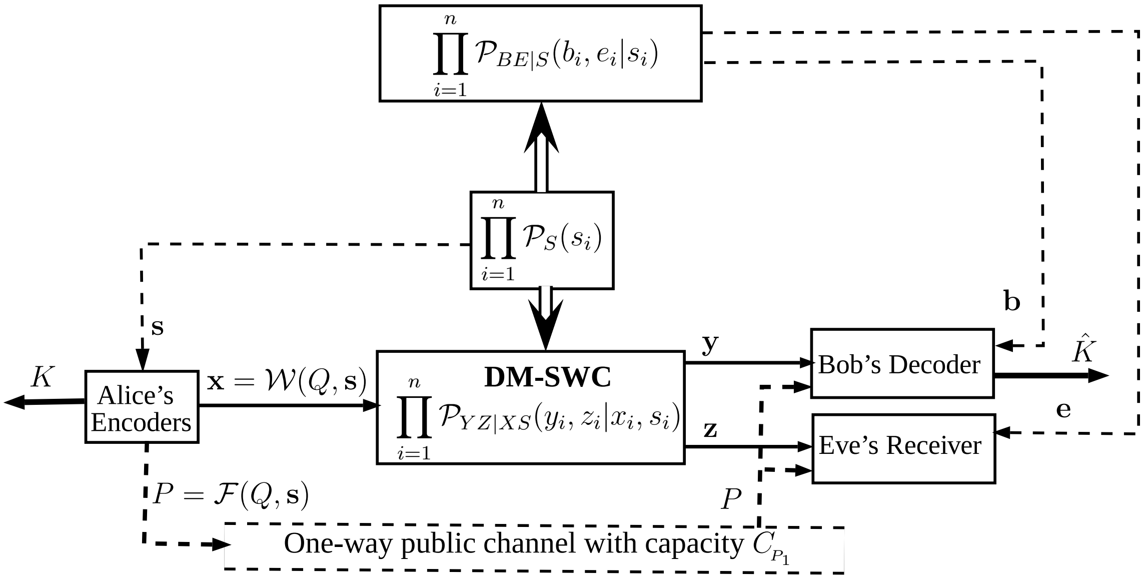
Definition 3.2 (DM-SWC)**.** A DM-SWC with known non-causal CSI at the transmitter is determined by 9-tuple $(\mathbb{S}, \mathbb{B}, \mathbb{E}, \mathbb{X}, \mathbb{Y}, \mathbb{Z}, \mathcal{P}_S(s), \mathcal{P}_{B\,E|S}(b, e|s), \mathcal{P}_{YZ|XS}(y, z|x, s))$, where

- finite sets $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ are the channel input alphabet (from Alice), the first channel output alphabet (to Bob), and the second channel output alphabet (to Eve), respectively;
- $\mathbb{S}$ is the state alphabet with a finite size; also, random vector $\mathbf{s} \in \mathbb{S}^n$, which is called CSI, is drawn i.i.d. according to distribution $\mathcal{P}(\mathbf{s}) = \prod_{i=1}^{n} \mathcal{P}_S(s_i)$;
- the PTF of the channel with the input $\mathbf{x} \in \mathbb{X}^n$, state $\mathbf{s} \in \mathbb{S}^n$, and output pair $(\mathbf{y}, \mathbf{z}) \in \mathbb{Y}^n \times \mathbb{Z}^n$ is $\mathcal{P}(\mathbf{y}, \mathbf{z}|\mathbf{x}, \mathbf{s}) = \prod_{i=1}^{n} \mathcal{P}_{YZ|XS}(y_i, z_i|x_i, s_i)$;
- random vectors $\mathbf{s} \in \mathbb{S}^n$, $\mathbf{b} \in \mathbb{B}^n$, and $\mathbf{e} \in \mathbb{E}^n$ are three-component SI available prior to each transmission at Alice, Bob, and Eve, respectively;
- the three-component SI is drawn i.i.d. according to

$$\mathcal{P}(\mathbf{b}, \mathbf{e}, \mathbf{s}) = \prod_{i=1}^{n} \mathcal{P}_{B\,E\,S}(b_i, e_i, s_i) = \prod_{i=1}^{n} \mathcal{P}_S(s_i) \prod_{i=1}^{n} \mathcal{P}_{B\,E|S}(b_i, e_i|s_i)\,. \tag{3.2}$$

In the DM model, Alice governs two encoders: the *wiretap channel encoder* $\mathcal{W}$ and the *public channel encoder* $\mathcal{F}$. Randomization at the encoders is permitted. Alice generates her key with a *key generator function* $\mathcal{K}_1$, and then she sends key agreement signals to Bob over both the wiretap channel and the noiseless public channel with capacity C_{P_1}.

The output of the wiretap channel encoder $\mathcal{W}$ is a block code with length n, which is sent in n successive transmissions over the wiretap channel. However, we assume, without any loss of generality, that the public channel encoder $\mathcal{F}$ transmits public message P at time instant n, i.e., $\mathbf{p}_1 = (\underbrace{0, \ldots, 0}_{(n-1)\text{ times}}, P)$. This is due to the following fact. There is no feedback from Bob to Alice, and Alice's SI is available non-causally prior to sending the public message. On the other hand, Bob is supposed to retrieve his key at the end of



$\prod_{i=1}^{n} \mathcal{P}_{BE|S}(b_i, e_i|s_i)$
$\prod_{i=1}^{n} \mathcal{P}_S(s_i)$
s
b
K
Alice's Encoders
$\mathbf{x} = \mathcal{W}(Q, \mathbf{s})$
DM-SWC
$\prod_{i=1}^{n} \mathcal{P}_{YZ|XS}(y_i, z_i|x_i, s_i)$
y
z
Bob's Decoder
$\hat{K}$
e
Eve's Receiver
$P = \mathcal{F}(Q, \mathbf{s})$
P
One-way public channel with capacity C_{P_1}

Figure 3.1: The key agreement problem over the DM model.

n^{th} transmission over the wiretap channel. Hence, transmission of the public message by successive uses of the public channel has the same effect on key generation as if the whole public message is sent in a single transmission parallel to n^{th} transmission of encoder $\mathcal{W}$.

Finally, Bob decodes his key at the end of n transmissions over the wiretap channel by applying a *key generator function*[1] $\mathcal{K}_2$ to his received signals $(\mathbf{y}, \mathbf{b}, P)$.

Eve has access to $\mathbf{e}$ and P from her SI and the public channel, respectively; she also receives $\mathbf{z}$ from the wiretap channel. Having $(\mathbf{z}, \mathbf{e}, P)$, Eve does her best to deduce any information about the keys.

The objective of the key agreement problem is to find the forward key capacity for the model by using an admissible key agreement code, which is defined as follows.

Definition 3.3. An admissible key agreement code $(\lceil 2^{nR_K} \rceil, n)$, where $R_{_K} \in \mathbb{R}^+ \cup \{0\}$ and code block length $n \in \mathbb{N}$, for the DM model consists of the following components:

[1] The key generator function at Bob is also called a *decoder* when the public channel is one-way in the forward direction.

- a key set $\mathbb{K} = \{1, \ldots, \lceil 2^{nR_K} \rceil\}$. This set is publicity known to all parties;
- a randomization RV Q with distribution $\mathcal{P}_Q$ over $\mathbb{Q}$, where $\mathbb{Q}$ is an arbitrary set with a finite size. For randomization, Alice generates a RV $Q \in \mathbb{Q}$ independent of $\mathbf{s}$;
- key generator function $\mathcal{K}_1 : \mathbb{Q} \times \mathbb{S}^n \to \mathbb{K}$. Observing $\mathbf{s}$, Alice computes her key $K = \mathcal{K}_1(Q, \mathbf{s})$;
- wiretap channel encoding function $\mathcal{W} : \mathbb{Q} \times \mathbb{S}^n \to \mathbb{X}^n$. Alice transmits $\mathbf{x} = \mathcal{W}(Q, \mathbf{s})$ over the wiretap channel in n transmissions;
- public channel encoding function $\mathcal{F} : \mathbb{Q} \times \mathbb{S}^n \to \mathbb{P}_{1n}$, where $\mathbb{P}_{1n}$ is assumed to be the range of function $\mathcal{F}$ which is subject to capacity constraint[2] (3.1a), where $|\mathbb{P}_1| = |\mathbb{P}_{1n}|$. Alice sends public message $P = \mathcal{F}(Q, \mathbf{s})$ over the public channel at time instant n;
- key generator (decoding) function $\mathcal{K}_2 : \mathbb{Y}^n \times \mathbb{B}^n \times \mathbb{P}_{1n} \to \mathbb{K}$. At the end of receiving $\mathbf{y}$, $\mathbf{b}$, and P, Bob decodes his key $\hat{K} = \mathcal{K}_2(\mathbf{y}, \mathbf{b}, P)$.

The efficiency of an admissible key agreement coding scheme is measured by the average probability of error, the leakage rate, and the randomness of $(K, \hat{K})$ as defined in Definition 1.10.

For a given public channel capacity pair $(C_{P_1}, 0)$, an achievable rate R_K is defined in Definition 1.11, where the admissible key agreement code is introduced in Definition 3.3. The (forward) key capacity of the DM model can be defined as follows.

Definition 3.4. Recall the admissible key agreement code in Definition 3.3. For a given public channel capacity $C_{P_1} \in [0, \infty)$, the supremum of all achievable key rates according to Definition 1.11 is called (forward) key capacity of that public channel. The forward key capacity is denoted by function $C_K(C_{P_1}, 0)$, where $(C_{P_1}, 0)$ is the pair of public channel capacity.

The ultimate objective of this chapter is to find (or bound) the forward key capacity of the DM model as a function of C_{P_1}.

Following Definition 2.9, we extend the definition of the less noisy property to the class

[2] In fact, we assumed $\mathbb{P}_{11} = \mathbb{P}_{12} = \ldots = \mathbb{P}_{1(n-1)} = \{0\}$.

of DM-SWCs. To do this, we assume a DM-SWC as a broadcast channel [2] with augmented input $(\mathbf{x}, \mathbf{s})$ and with augmented outputs $(\mathbf{y}, \mathbf{b})$ and $(\mathbf{z}, \mathbf{e})$. To simplify the notations, we represent augmented outputs with the new symbol ˇ on wiretap channel outputs, e.g.,

$$\check{Y}_i = (Y_i, B_i) \,, \tag{3.3a}$$

$$\check{Z}_i = (Z_i, E_i) \tag{3.3b}$$

for any $i \in \{1, \ldots, n\}$.

Definition 3.5. For a given channel state distribution $\mathcal{P}_S$, the state-dependent channel is said to be less noisy with respect to $\mathcal{P}_S$ in Bob's favor if

$$\mathcal{I}(U; \check{Y}) \geq \mathcal{I}(U; \check{Z}) \tag{3.4}$$

holds for *every* auxiliary random variable U such that $U \to (X, S) \to (\check{Y}, \check{Z})$ forms a Markov chain. A special class of less noisy DM-SWCs is the class of (*physically*) *degraded* DM-SWCs for which

$$\mathcal{P}_{\check{Y}\check{Z}|XS} = \mathcal{P}_{\check{Z}|\check{Y}} \mathcal{P}_{\check{Y}|XS} \,, \tag{3.5}$$

holds. A physically degraded DM-SWC is less noisy with respect to any channel state distribution $\mathcal{P}_S$.

3.2 Statement of Main Results

In this section, we declare the main results for the DM model defined in Section 3.1.

Theorem 3.1 (Lower bound on the forward key capacity)**.** *Assume the DM model with public channel capacity $C_{P_1} \in [0, \infty)$. Define the set*

$$\mathbb{O}(C_{P_1}) \triangleq \{\mathcal{P}_{XU|S} : C_{P_1} + \mathcal{I}(U; \check{Y}) \geq \mathcal{I}(U; S)\} \,, \tag{3.6}$$

where $U \in \mathbb{U}$ with cardinality $|\mathbb{U}| \leq |\mathbb{S}||\mathbb{X}| + 3$ is an auxiliary RV such that $U \to (X, S) \to$

$(\check{Y}, \check{Z})$. Then, the forward key capacity of the DM model with public channel capacity $C_{_{P_1}}$ is lower bounded by

$$C_{_K}(C_{_{P_1}}, 0) \geq \max_{\mathcal{P}_{XU|S} \in \mathbb{O}(C_{_{P_1}})} [\mathcal{I}(U; \check{Y}) - \mathcal{I}(U; \check{Z})] \,. \tag{3.7}$$

Comparing Theorem 3.1 with Corollary 2.1, we conclude that the key capacity exceeds both the main channel capacity and the secrecy capacity in some DM-SWCs, which are specified in the following corollary.

Corollary 3.1. *Let $C_{_{P_1}} = 0$. Assume a physically degraded DM-SWC according to* (3.5)*. Assume the main channel capacity of the DM-SWC, which is given in* (1.33)*, is achievable at input distribution $\mathcal{P}^*_{X|SU}\mathcal{P}^*_{U|S}\mathcal{P}_S$. At this distribution, if $\mathcal{I}(U^*; S) > \mathcal{I}(U^*; \check{Z}^*)$, then*

$$C_{_K}(0, 0) \geq \mathcal{I}(U^*; \check{Y}^*) - \mathcal{I}(U^*; \check{Z}^*) > C_{_S} = C_m = \mathcal{I}(U^*; \check{Y}^*) - \mathcal{I}(U^*; S) \,,$$

where $\mathcal{I}(U^*; S)$, $\mathcal{I}(U^*; \check{Y}^*)$, and $\mathcal{I}(U^*; \check{Z}^*)$ are calculated at distribution

$$\mathcal{P}_{\check{Z}|\check{Y}}\mathcal{P}_{\check{Y}XUS} = \mathcal{P}_{\check{Y}\check{Z}|XS}\mathcal{P}^*_{X|SU}\mathcal{P}^*_{U|S}\mathcal{P}_S \,.$$

Moreover, the UBs on the forward key capacity of the DM model are as follows.

Theorem 3.2 (Upper bound on the forward key capacity)**.** *Let U be an auxiliary RV such that $U \to (X, S) \to (\check{Y}, \check{Z})$. Then, the forward key capacity is upper bounded by*

$$C_{_K}(C_{_{P_1}}, 0) \leq \max_{\substack{U \to (X,S) \to (\check{Y}, \check{Z}) \\ s.t.\ C_{_{P_1}} + \mathcal{I}(U;\check{Y}) + \mathcal{H}(X|S,U) \geq \mathcal{I}(U;S)}} [\mathcal{I}(U; \check{Y}) - \mathcal{I}(U; \check{Z})] \,.$$

Theorem 3.3 (Upper bound on the forward key capacity of a less noisy DM-SWC)**.** *Let U and V be two auxiliary RVs such that $(U, V) \to (X, S) \to (\check{Y}, \check{Z})$. If the channel is less noisy with respect to the given distribution $\mathcal{P}_S$ in Bob's favor, then the forward key capacity* *is upper bounded by*

$$C_{_K}(C_{_{P_1}},0) \leq \max_{\substack{(U,V)\to(X,S)\to(\check{Y},\check{Z}) \\ s.t.\ C_{_{P_1}}+\mathcal{I}(U;\check{Y})\geq\mathcal{I}(U;S)}} [\mathcal{I}(U,V;\check{Y})-\mathcal{I}(U,V;\check{Z})] \,.$$

As a special case, the following corollary shows that the forward key capacity of a source-type model given in (1.20), which is originally established in [18, 21], is achievable by using a single auxiliary RV.

Corollary 3.2. *When no wiretap channel exists, i.e., $\mathbb{X}=\mathbb{Y}=\mathbb{Z}=\{0\}$, the forward key capacity of the model is given by*

$$C_{_K}(C_{_{P_1}},0) = \max_{\substack{U\to S\to(B,E) \\ s.t.\ C_{_{P_1}}+\mathcal{I}(U;B)\geq\mathcal{I}(U;S)}} [\mathcal{I}(U;B)-\mathcal{I}(U;E)] \,,$$

where $U\in\mathbb{U}$ is an auxiliary RV with $|\mathbb{U}|\leq|\mathbb{S}|+3$.

When the capacity of the public channel becomes unlimited, the forward key capacity is achieved as stated in the following theorem.

Theorem 3.4 (Public channel with unlimited capacity)**.** *If $C_{_{P_1}}\to\infty$, then the forward key capacity of the model is*

$$C_{_K}(\infty,0) = \max_{U\to(X,S)\to(\check{Y},\check{Z})} [\mathcal{I}(U;\check{Y})-\mathcal{I}(U;\check{Z})] \,. \tag{3.8}$$

Further, if the channel is less noisy with respect to the distribution $\mathcal{P}_S(s)$ in Bob's favor, then the forward key capacity can be simplified to

$$C_{_K}(\infty,0) = \max_{\mathcal{P}_{X|S}} [\mathcal{I}(X,S;\check{Y})-\mathcal{I}(X,S;\check{Z})] \,. \tag{3.9}$$

$C_{_K}(C_{_{P_1}},0)$ is a non-decreasing function of $C_{_{P_1}}$. Hence, $C_{_K}(\infty,0)$ is the maximum forward key capacity that can be attained by a DM model with a fixed DM-SWC. However, we do not demand an unlimited public channel capacity to achieve $C_{_K}(\infty, 0)$. Further, $C_{_K}(C_{_{P_1}}, 0)$ is *not* a strictly increasing function of $C_{_{P_1}}$. These facts are established in the following corollary.

Corollary 3.3. *For a given DM-SWC, let distribution*

$$\mathcal{P}^*_{XU|S} \in \arg\max_{\mathcal{P}_{XU|S}} [\mathcal{I}(U; \check{Y}) - \mathcal{I}(U; \check{Z})] \,,$$

where $U \to (X, S) \to (\check{Y}, \check{Z})$*. Define* $C^*_{_P} \triangleq [\mathcal{I}(U^*; S) - \mathcal{I}(U^*; \check{Y}^*)]^+$*, which is calculated according to distribution* $\mathcal{P}_{\check{Y}|XS}\mathcal{P}^*_{XU|S}\mathcal{P}_S$*. Then,* $C^*_{_P}$ *is finite and for any* $C_{_{P_1}} \geq C^*_{_P}$*,* $C_{_K}(C_{_{P_1}}, 0) = C_{_K}(\infty, 0)$*, where* $C_{_K}(\infty, 0)$ *is given by Theorem 3.4.*

Now, consider a DM-SWC in which the CSI is fully known at both Alice and Bob (no matter if it is fully known at Eve or not). In [64], the (forward) key capacity of this model is obtained for the special case no public channel as given in (1.36). Based on Corollary 3.3, we show that the public channel does not help the forward key capacity in this case. That is, the result of [64] is optimum for any $C_{_{P_1}} \geq 0$ as well. In this case, the forward key capacity is obtained from Theorem 3.1 and Corollary 3.3 as follows.

Corollary 3.4. *If Bob has access to the CSI, i.e.,* $\mathbf{b} = \mathbf{s}$*, then the forward key capacity of the model for any public channel capacity* $C_{_{P_1}} \geq 0$ *is*

$$C_{_K}(C_{_{P_1}}, 0) = \max_{U \to (X,S) \to (Y,Z,E)} [\mathcal{I}(U; Y|S) - \mathcal{I}(U; Z|S) + \mathcal{H}(S|Z, E)] \,.$$

3.3 The Proofs

In this section, we prove the results given in Section 3.2 as follows.

3.3.1 The Proof of Theorem 3.1

The strategy of the proof is as follows. First, we prove that $R_{_K} = \mathcal{I}(U;\check{Y}) - \mathcal{I}(U;\check{Z})$ is an achievable key rate if $\mathcal{P}_{XU|S} \in \mathbb{O}(C_{_{P_1}})$. Proving this fact, the LB is obtained by taking the supremum on all achievable key rates. On the other hand, $|\mathbb{U}| \leq |\mathbb{S}||\mathbb{X}| + 3$ follows from support lemma [6] as proved in [12, Appendix]. Hence, the size of set $\mathbb{U}$ is finite and we can switch the supremum with the maximum according to Appendix B.

In the following, we prove that $R_{_K}$ is an achievable key rate with the use of the strong typicality introduced in Subsection 2.1.1. In Subsubsection 3.3.1.1, the preliminaries of the proof is expressed where the distribution on RVs for different cases are established. In Subsubsection 3.3.1.2, we create an admissible key agreement code for the DM model satisfying the LB given in Theorem 3.1. In Subsubsection 3.3.1.3, we analyze the efficiency of the admissible key agreement code according to Definition 1.10.

3.3.1.1 Preliminaries

Select set $\mathbb{U}$ such that $|\mathbb{U}| \leq |\mathbb{S}||\mathbb{X}| + 3$. Given $C_{_{P_1}}$, $\mathcal{P}_{BE|S}$, and $\mathcal{P}_{YZ|XS}$ from the DM model, arbitrarily select conditional PMFs $\mathcal{P}_{U|S}$ and $\mathcal{P}_{X|US}$ to fix

$$\mathcal{P}_{YZXUBE|S} = \mathcal{P}_{BE|S}\mathcal{P}_{U|S}\mathcal{P}_{X|US}\mathcal{P}_{YZ|XS} \tag{3.10}$$

such that $C_{_{P_1}} \geq \mathcal{I}(U;S) - \mathcal{I}(U;Y,B)$. From (3.10), Markov chains $(B,E) \to S \to U$ and $(B,E) \to (U,S) \to X$ and $(U,B,E) \to (X,S) \to (Y,Z)$ are justified. Also, $\mathcal{P}_U(u) = \sum\limits_{s\in\mathbb{S}} \mathcal{P}_{U|S}(u|s)\mathcal{P}_S(s)$ is fixed by (3.10), where $u \in \mathbb{U}$.

If $\mathcal{I}(U;\check{Z}) \geq \mathcal{I}(U;\check{Y})$, then no key rate can be agreed between Alice and Bob, and so $\mathbb{K} = \{1\}$; if not, the admissible key agreement code is constructed in the next subsubsection.

3.3.1.2 Key Agreement Code Generation

Sequences $\mathbf{u}$ are generated i.i.d. according to $\mathcal{P}_U$. For any $\epsilon \in (0,1)$, randomly pick up $N_0 \leq 2^{n(H(U)-\epsilon)}$ sequences $\mathbf{u}$ for the codebook generation. Each sequence $\mathbf{u}$ is called a codeword. The structure of the key agreement codebook is based on double random binning as explained below.

As shown in Figure 3.2, the (key agreement) codebook is constructed with N_1 enumerated bins such that each bin consists of N_2 enumerated sub-bins. N_0 codewords are randomly partitioned in all $N_1 N_2$ sub-bins such that each sub-bin is filled with the same number of codewords. Let consider each sub-bin as a *set* of codewords. Denote sub-bin j located in bin i by the locator set $\mathbb{L}_{ij}$, where $i \in \{1, \ldots, N_1\}$ and $j \in \{1, \ldots, N_2\}$; each locator set has also the same number $|\mathbb{L}| \triangleq |\mathbb{L}_{ij}| = \frac{N_0}{N_1 N_2}$ codewords. Define

$$\mathbb{I}_i \triangleq \bigcup_{j=1}^{N_2} \mathbb{L}_{ij} \tag{3.11}$$

as the set of all codewords placed in bin i, where $i \in \{1, \ldots, N_1\}$; each bin has the same $|\mathbb{I}_i| = \frac{N_0}{N_1}$ codewords. For $i \in \{1, \ldots, N_1\}$ and $j \in \{1, \ldots, N_2\}$, (i, j) is called the *label* of codeword $\mathbf{u}$, if $\mathbf{u} \in \mathbb{L}_{ij}$; i is also called its *bin-label* as $\mathbf{u} \in \mathbb{I}_i$. Finally, the *matrix* of all labeled codewords is publicly released as the (key agreement) codebook, i.e.,

$$\mathbf{C} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{N_0} \end{bmatrix}. \tag{3.12}$$

At each block code transmission, the encoder selects a codeword, which is specified by $\tilde{\mathbf{u}}$, from the codebook $\mathbf{C}$. Having $\mathbf{s}$ and codeword $\tilde{\mathbf{u}}$, transmitted signal $\mathbf{x}$ is generated according to $\mathcal{P}(\mathbf{x}|\mathbf{s}, \tilde{\mathbf{u}}) = \prod_{i=1}^{n} \mathcal{P}_{X|SU}(x_i|s_i, \tilde{u}_i)$. At the end of the transmission of the whole block code, Bob decodes a codeword $\hat{\mathbf{u}}$ from the codebook $\mathbf{C}$. Specifications of the codebook, the encoding and the decoding functions are individually determined for the following two cases.

(a). Case $\mathcal{I}(U; \check{Y}) \geq \mathcal{I}(U; S)$.

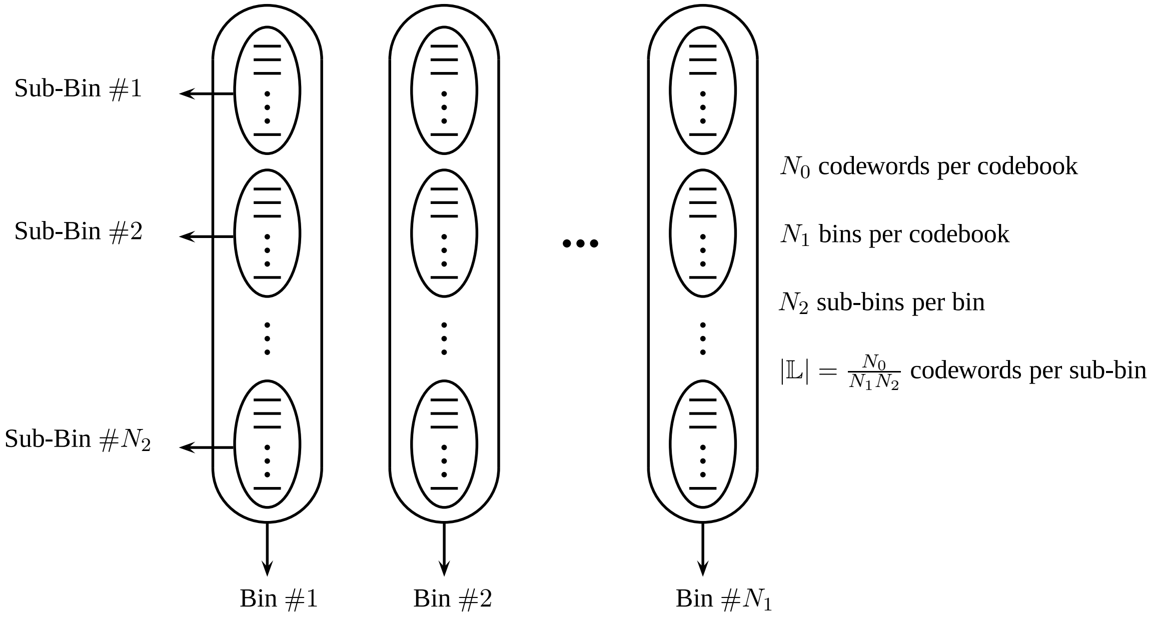


Figure 3.2: Structure of codebook $\mathbf{C}$ formed with a double random binning strategy.

- *Codebook Specifications.*

$$N_0 = 2^{n(\mathcal{I}(U;\check{Y})-\epsilon_1)}, \tag{3.13}$$

$$N_1 = 2^{n[\mathcal{I}(U;\check{Y})-\epsilon_1-\max\{\mathcal{I}(U;S)+\epsilon_0,\, \mathcal{I}(U;\check{Z})-\epsilon_2\}]}, \tag{3.14}$$

$$N_2 = 2^{n[\max\{\mathcal{I}(U;S)+\epsilon_0,\, \mathcal{I}(U;\check{Z})-\epsilon_2\}-(\mathcal{I}(U;\check{Z})-\epsilon_3)]}, \tag{3.15}$$

$$|\mathbb{L}| = 2^{n[\mathcal{I}(U;\check{Z})-\epsilon_3]}, \tag{3.16}$$

where $1 > \epsilon_0 > \epsilon_1 > \epsilon > \epsilon_3 > \epsilon_2 > 0$ are fixed real values.

- *Encoding.* Let define $\mathbb{K}_1 \triangleq \{1,\dots,N_1\}$, $\mathbb{K}_2 \triangleq \{1,\dots,N_2\}$, and assume key set $\mathbb{K} = \{1,\dots,N_1.N_2\}$. Alice generates a uniformly distributed RV $Q \in \mathbb{K}_1$. Given $\mathbf{s}$ and Q, she picks up codeword $\tilde{\mathbf{u}}$ at uniformly random from the set $\{\mathbf{u} : \mathbf{u} \in \mathbb{I}_Q, (\mathbf{u},\mathbf{s}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{US})\}$. If the set is void, the encoder randomly sends a codeword from set $\mathbb{I}_Q$, but an error will be declared at the encoder. Assume $\tilde{\mathbf{u}} \in \mathbb{L}_{ij}$, where $i \in \{1, \ldots, N_1\}$ and $j \in \{1, \ldots, N_2\}$. Fix a one-to-one function $\tilde{\mathcal{K}}_1 : \mathbb{K}_1 \times \mathbb{K}_2 \to \mathbb{K}$. Then, $K = \tilde{\mathcal{K}}_1(i, j)$. In this case the encoder sends no public message over the public channel, i.e., $P = 0$. From equations (3.14) and (3.15), size of the key set is

$$|\mathbb{K}| = N_1 N_2 = 2^{n[\mathcal{I}(U;\check{Y}) - \mathcal{I}(U;\check{Z}) - \epsilon_1 + \epsilon_3]} \,. \tag{3.17}$$

- *Decoding.* Observing $\mathbf{b}$ and receiving $\mathbf{y}$ from the DM-SWC, the legitimate decoder builds the decoding set

$$\mathbb{D}_1 = \{\mathbf{u} : \mathbf{u} \in \mathbf{C}, (\mathbf{u}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{UYB})\} \,.$$

If $|\mathbb{D}_1| \neq 1$ or $\tilde{\mathbf{u}} \notin \mathbb{D}_1$, the decoder declares an error; if not, $\hat{K} = \tilde{\mathcal{K}}_1(i, j)$, where $\tilde{\mathbf{u}} \in \mathbb{D}_1$, and (i, j) is the label of codeword $\hat{\mathbf{u}} = \tilde{\mathbf{u}}$.

(b). Case $\mathcal{I}(U;\check{Y}) < \mathcal{I}(U;S)$.

- *Codebook Specifications.*

$$N_0 = 2^{n(\mathcal{I}(U;S) + \epsilon_0)} \,, \tag{3.18}$$

$$N_1 = 2^{n[\mathcal{I}(U;S) - \mathcal{I}(U;\check{Y}) + \epsilon_0 + \epsilon_1]} \,, \tag{3.19}$$

$$N_2 = 2^{n[\mathcal{I}(U;\check{Y}) - \mathcal{I}(U;\check{Z}) - \epsilon_1 + \epsilon_2]} \,, \tag{3.20}$$

$$|\mathbb{L}| = 2^{n(\mathcal{I}(U;\check{Z}) - \epsilon_2)} \,, \tag{3.21}$$

where $1 > \epsilon_0 > \epsilon_1 > \epsilon > \epsilon_2 > 0$ are fixed real values.

- *Encoding.* Let key set $\mathbb{K} = \{1, \ldots, N_2\}$. Observing $\mathbf{s}$, Alice picks up codeword $\tilde{\mathbf{u}}$ at uniformly random from set $\{\mathbf{u} : \mathbf{u} \in \mathbf{C}, (\mathbf{u}, \mathbf{s}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{US})\}$. The encoder declares an error if this set is void, and then codeword $\tilde{\mathbf{u}}$ is randomly chosen from codebook $\mathbf{C}$. Assume that the label of $\tilde{\mathbf{u}}$ is (I, J), where $I \in \{1, \ldots, N_1\}$ and $J \in \{1, \ldots, N_2\}$.

Then, $P = I$ and $K = J$. Therefore, $P \in \{1, \ldots, N_1\}$ and $K \in \mathbb{K}$. In this case

$$|\mathbb{P}_1| = 2^{n[\mathcal{I}(U;S) - \mathcal{I}(U;\check{Y}) + \epsilon_0 + \epsilon_1]} , \tag{3.22}$$

$$|\mathbb{K}| = 2^{n[\mathcal{I}(U;\check{Y}) - \mathcal{I}(U;\check{Z}) - \epsilon_1 + \epsilon_2]} , \tag{3.23}$$

due to equations (3.19) and (3.20).

- *Decoding.* Bob's decoder knows the bin-label of the sent codeword from public message P. Having P, $\mathbf{y}$ and $\mathbf{b}$, it constructs the set

$$\mathbb{D}_2 = \{\mathbf{u} : \mathbf{u} \in \mathbb{I}_P, (\mathbf{u}, \mathbf{y}, \mathbf{b}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{UYB})\} .$$

An error will be declared by the decoder if $|\mathbb{D}_2| \neq 1$ or $\tilde{\mathbf{u}} \notin \mathbb{D}_2$; if not, then $\tilde{\mathbf{u}} = \hat{\mathbf{u}}$ and assume that the label of codeword $\tilde{\mathbf{u}} \in \mathbb{D}_2$ is $(P, \hat{K})$, where $\hat{K} \in \mathbb{K}$. Finally, $\hat{K}$ is chosen by Bob as his key.

3.3.1.3 Analysis

Any codebook, as a random matrix, has codewords $\mathbf{u}_1, \ldots, \mathbf{u}_{N_0}$ such that each codeword is drawn i.i.d. according to $\mathcal{P}_U(u)$. Hence, a random codebook is generated according to

$$\mathcal{P}(\mathbf{C}) = \prod_{j=1}^{n} \prod_{i=1}^{N_0} \mathcal{P}_U(u_{ij}) . \tag{3.24}$$

We prove that the average probability of error and the average leakage rate over all randomly generated key agreement codebooks vanishes as $n \to \infty$. Then, this turns out that there exists at least one codebook with small enough probability of error and small enough leakage rate as $n \to \infty$ due to the random coding argument [2,6]. Specifically, assume that for any $\epsilon \in (0, 1)$ there exists $n \geq N$ such that

$$\mathcal{E}(\mathcal{P}_{error}(n)) \leq \epsilon , \tag{3.25a}$$

$$\frac{1}{n}\mathcal{I}(K; \check{\mathbf{z}}, P, \mathbf{C}) \leq \epsilon , \tag{3.25b}$$

where $\mathbf{C}$ is a random codebook with distribution (3.24) and $\mathcal{E}$ indicates an average over the random codebook ensemble. Thus, from equations (3.25) we have

$$\begin{aligned}2\epsilon &\geq \mathcal{E}(\mathcal{P}_{error}(n)) + \frac{1}{n}\mathcal{I}(K;\check{\mathbf{z}},P,\mathbf{C})\ , \qquad (3.26)\\ &= \mathcal{E}(\mathcal{P}_{error}(n)) - \frac{1}{n}\mathcal{H}(K|\check{\mathbf{z}},P,\mathbf{C}) + \frac{1}{n}\mathcal{H}(K)\\ &= \mathcal{E}(\mathcal{P}_{error}(n)) - \frac{1}{n}\sum_{\mathbf{C}_0}\mathcal{H}(K|\check{\mathbf{z}},P,\mathbf{C}=\mathbf{C}_0)\mathscr{P}(\mathbf{C}=\mathbf{C}_0) + \frac{1}{n}\mathcal{H}(K)\end{aligned}$$

where (3.26) follows from equations (3.25a) and (3.25b). From (3.26), it can be concluded [2] that there exists a key agreement codebook $\mathbf{C}_0$ such that

$$\mathcal{P}_{error}(n) - \frac{1}{n}\mathcal{H}(K|\check{\mathbf{z}},P,\mathbf{C}_0) + \frac{1}{n}\mathcal{H}(K) \leq 2\epsilon$$

or equivalently

$$\mathcal{P}_{error}(n) + \frac{1}{n}\mathcal{I}(K;\check{\mathbf{z}},P,\mathbf{C}_0) \leq 2\epsilon \qquad (3.27)$$

Therefore, from (3.27) we conclude

$$\mathcal{P}_{error}(n) \leq 2\epsilon\ , \qquad (3.28a)$$

$$\frac{1}{n}\mathcal{I}(K;\check{\mathbf{z}},P,\mathbf{C}_0) \leq 2\epsilon\ , \qquad (3.28b)$$

that leads to the approval of the AR and AS conditions according to Definition 1.11.

In the sequel, we prove equations (3.25a) and (3.25b) in part (a) and part (b), respectively to justify the AR and AS conditions. Moreover, we will prove that $\mathcal{H}(K|\check{\mathbf{z}},P,\mathbf{C}) \geq \log(|\mathbb{K}|) - n\epsilon$ for any $\epsilon \in (0,1)$ in part (b), and thus the ARN condition, given in Definition 1.11, is established according to the fact that $\mathcal{H}(K) \geq \mathcal{H}(K|\check{\mathbf{z}},P,\mathbf{C})$.

To simplify notations, let define $\mathbb{I}_0 \triangleq \{\mathbf{u}_1,\ldots,\mathbf{u}_{N_0}\}$ as the set of all codewords in this part. According to Definition 1.10, we examine the following conditions for the proposed key agreement code.

(a). The Average Probability of Error. Assume that $\mathbb{EV}_1$ and $\mathbb{EV}_2$ are error events at the encoding step and the decoding step, respectively. We have

$$\begin{aligned}
\mathcal{E}(\mathcal{P}_{error}(n)) &\leq \sum_{\mathbf{s}\,\in\,\mathbb{S}^n} \mathcal{P}(\mathbf{s})[\mathscr{P}(\mathbb{EV}_1 \cup \mathbb{EV}_2|\mathbf{s})]\;, && (3.29)\\
&\leq \sum_{\mathbf{s}\,\in\,\mathbb{T}^*_\epsilon(\mathcal{P}_S)} \mathcal{P}(\mathbf{s})[\mathscr{P}(\mathbb{EV}_1 \cup \mathbb{EV}_2|\mathbf{s})] + \sum_{\mathbf{s}\,\notin\,\mathbb{T}^*_\epsilon(\mathcal{P}_S)} \mathcal{P}(\mathbf{s})\;, && (3.30)\\
&\leq \sum_{\mathbf{s}\,\in\,\mathbb{T}^*_\epsilon(\mathcal{P}_S)} \mathcal{P}(\mathbf{s})[\mathscr{P}(\mathbb{EV}_1 \cup \mathbb{EV}_2|\mathbf{s})] + \frac{\epsilon}{4}\;, && (3.31)\\
&= \sum_{\mathbf{s}\,\in\,\mathbb{T}^*_\epsilon(\mathcal{P}_S)} \mathcal{P}(\mathbf{s})[\mathscr{P}(\mathbb{EV}_1|\mathbf{s}) + \mathscr{P}(\mathbb{EV}_2|\mathbb{EV}_1^c, \mathbf{s})] + \frac{\epsilon}{4}\;, && (3.32)
\end{aligned}$$

where the expectation in (3.29) is over all code-books $\mathbf{C}$ with distribution given in (3.24); equation (3.31) follows from the fact that

$$\sum_{\mathbf{s}\,\notin\,\mathbb{T}^*_\epsilon(\mathcal{P}_S)} \mathcal{P}(\mathbf{s}) \leq \frac{\epsilon}{4}$$

for large enough $n \geq n_{11}(\epsilon)$ [2, Sec. 11.2]. Hence, we focus our attention to $\mathbf{s} \in \mathbb{T}^*_\epsilon(\mathcal{P}_S)$. With this assumption, the terms in (3.32) are evaluated in the following.

Error at the Encoder. Error event at the encoder is

$$\mathbb{EV}_1 \triangleq \{\nexists \mathbf{u} \in \mathbb{I}_Q : (\mathbf{s}, \mathbf{u}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{SU})\}\;. \qquad (3.33)$$

Now, the first term in (3.32) is bounded as follows:

$$\begin{aligned}
\mathscr{P}(\mathbb{EV}_1|\mathbf{s}) &\leq (1 - 2^{-n(\mathcal{I}(U;S)+\epsilon)})^{|\mathbb{I}_Q|} && (3.34)\\
&\leq (1 - 2^{-n(\mathcal{I}(U;S)+\epsilon)})^{2^{n(\mathcal{I}(U;S)+\epsilon_0)}} && (3.35)\\
&\leq exp(-2^{-n(\mathcal{I}(U;S)+\epsilon)})^{2^{n(\mathcal{I}(U;S)+\epsilon_0)}} && (3.36)\\
&= exp(-2^{-n(\mathcal{I}(U;S)+\epsilon)}.2^{n(\mathcal{I}(U;S)+\epsilon_0)})\\
&= exp(-2^{n(\epsilon_0-\epsilon)})
\end{aligned}$$

$$\leq \frac{\epsilon}{4} \tag{3.37}$$

where

- (3.34) follows from the fact that the codewords have been generated independently and from [2, Lem. (10.6.2)] for large enough $n \geq n_{12}(\epsilon)$ such that

$$\epsilon_0 \geq \epsilon + \frac{1}{n}\log(-\ln(\frac{\epsilon}{4})) > \epsilon \tag{3.38}$$

- (3.35) holds because $0 < 1 - 2^{-n(\mathcal{I}(U;S)+\epsilon)} < 1$ and $|\mathbb{I}_Q| = \frac{N_0}{N_1} \geq 2^{n(\mathcal{I}(U;S)+\epsilon_0)}$ due to equations (3.13) and (3.14) as well as equations (3.18) and (3.19);
- (3.36) holds from inequality $1 + a \leq exp(a)$ for any real value a;
- (3.37) is valid due to (3.38);

The encoder finally sends codeword $\tilde{\mathbf{u}}$ based on observation of $\mathbf{s}$. If no error declared at the encoder (conditioned on $\mathbb{EV}_1^c$ and given vector $\mathbf{s}$), we have $(\tilde{\mathbf{u}}, \mathbf{s}) \in \mathbb{T}_\epsilon^*(\mathcal{P}_{US})$.

Error at the Decoder. Error event at the decoder is defined as

$$\mathbb{EV}_2 \triangleq \mathbb{EV}_{21} \cup \mathbb{EV}_{22} \;, \tag{3.39}$$

where

$$\mathbb{EV}_{21} = \{(\mathbf{y}, \mathbf{b}, \tilde{\mathbf{u}}) \notin \mathbb{T}_\epsilon^*(\mathcal{P}_{YBU})\} \;, \tag{3.40}$$

$$\mathbb{EV}_{22} = \{\exists \mathbf{u} \neq \tilde{\mathbf{u}} \in \mathbb{I}_P : (\mathbf{y}, \mathbf{b}, \mathbf{u}) \in \mathbb{T}_\epsilon^*(\mathcal{P}_{YBU})\} \;. \tag{3.41}$$

Therefore, we have

$$\mathscr{P}(\mathbb{EV}_2|\mathbb{EV}_1^c, \mathbf{s}) = \mathscr{P}(\mathbb{EV}_{21}|\mathbb{EV}_1^c, \mathbf{s}) + \mathscr{P}(\mathbb{EV}_{22}|\mathbb{EV}_{21}^c, \mathbb{EV}_1^c, \mathbf{s}) \;, \tag{3.42}$$

according to (3.39). The first term in the right-hand-side of (3.42) is bounded by the following lemma.

Lemma 3.1. *For any $\epsilon \in (0, 1)$, there exists $n \geq n_{21}(\epsilon)$ such that $\mathscr{P}(\mathbb{EV}_{21}|\mathbb{EV}_1^c, \mathbf{s}) \leq \frac{\epsilon}{4}$.*

Proof.

$$\mathscr{P}(\mathbb{EV}_{21}|\mathbb{EV}_1^c, \mathbf{s}) \leq \mathscr{P}(\mathbb{EV}_{211}|\mathbb{EV}_1^c, \mathbf{s}) + \mathscr{P}(\mathbb{EV}_{212}|\mathbb{EV}_{211}^c, \mathbb{EV}_1^c, \mathbf{s}) \tag{3.43}$$

where

$$\mathbb{EV}_{211} \triangleq \{(\mathbf{b}, \mathbf{s}, \tilde{\mathbf{u}}) \notin \mathbb{T}^*_\epsilon(\mathcal{P}_{BSU})\}\,, \tag{3.44}$$

$$\mathbb{E}_{212} \triangleq \{(\mathbf{y}, \mathbf{b}, \mathbf{s}, \tilde{\mathbf{u}}) \notin \mathbb{T}^*_\epsilon(\mathcal{P}_{YBSU})\}\,. \tag{3.45}$$

From (3.10), we have $U \rightarrow S \rightarrow B$; from condition $\mathbb{EV}_1^c$, we conclude $(\tilde{\mathbf{u}}, \mathbf{s}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{US})$ for given $\mathbf{s}$; hence, there exists $n \geq n_{211}(\epsilon)$ such that $\mathscr{P}(\mathbb{E}_{211}) < \frac{\epsilon}{8}$ due to Markov lemma [2, Lem. 15.8.1] and the fact that $\mathbf{b}$ is drawn according to $\prod_{i=1}^{n} \mathcal{P}_{B|S}(b_i|s_i)$, where $\mathcal{P}_{B|S}(b|s) = \sum_{e\in\mathbb{E}} \mathcal{P}_{BE|S}(b, e|s)$.

From (3.10), we have $B \rightarrow (U, S) \rightarrow Y$; from condition $\mathbb{E}_{211}^c(\mathbf{s})$, we conclude $(\tilde{\mathbf{u}}, \mathbf{b}, \mathbf{s}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{U\Upsilon BS})$ for given $\mathbf{s}$; thus, there exists $n \geq n_{212}(\epsilon)$ such that $\mathscr{P}(\mathbb{E}_{212}) \leq \frac{\epsilon}{8}$ due to Markov lemma [2, Lem. 15.8.1] and the fact that $\mathbf{y}$ is drawn according to $\prod_{i=1}^{n} \mathcal{P}_{Y|US}(y_i|\tilde{u}_i, s_i)$, where $\mathcal{P}_{Y|US}(y|u, s) = \sum_{x\in\mathbb{X}} \mathcal{P}_{Y|XS}(y|x, s)\mathcal{P}_{X|US}(x|u, s)$ due to equation (3.10).
Thus, the proof of the lemma is completed if $n_{21}(\epsilon) > \max\{n_{211}(\epsilon), n_{212}(\epsilon)\}$. □

The second term in the right-hand-side of (3.42) is bounded as follows:

$$\begin{aligned}
\mathscr{P}(\mathbb{EV}_{22}|\mathbb{EV}_{21}^c, \mathbb{EV}_1^c, \mathbf{s}) &= \mathscr{P}(\mathbb{EV}_{22}|\mathbb{EV}_{21}^c, \mathbf{s}) && (3.46)\\
&\leq \sum_{\substack{\mathbf{u} \neq \tilde{\mathbf{u}} \\ \mathbf{u} \in \mathbb{I}_P}} \mathscr{P}((\mathbf{y}, \mathbf{b}, \mathbf{u}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{YBU})|\mathbb{EV}_{21}^c, \mathbf{s}) && (3.47)\\
&= \sum_{\mathbf{u}\in\mathbb{I}_P} \sum_{\substack{\mathbf{u} \neq \tilde{\mathbf{u}} \\ \mathbf{u} \in \mathbb{I}_P}} \mathscr{P}(\tilde{\mathbf{u}} = \mathbf{u}|\mathbb{EV}_{21}^c, \mathbf{s})\mathscr{P}((\mathbf{y}, \mathbf{b}, \mathbf{u}) \in \mathbb{T}^*_\epsilon(\mathcal{P}_{YBU})|\mathbb{EV}_{21}^c, \mathbf{s}, \tilde{\mathbf{u}})
\end{aligned}$$

$$\leq \sum_{\mathbf{u}\in\mathbb{I}_P} \sum_{\substack{\mathbf{u} \neq \tilde{\mathbf{u}} \\ \mathbf{u} \in \mathbb{I}_P}} \mathscr{P}(\tilde{\mathbf{u}} = \mathbf{u}|\mathbb{EV}_{21}^c, \mathbf{s}) 2^{-n(\mathcal{I}(U;Y,B)-\epsilon)} \tag{3.48}$$

$$\leq \sum_{\mathbf{u}\in\mathbb{I}_P} \mathscr{P}(\tilde{\mathbf{u}} = \mathbf{u}|\mathbb{EV}_{21}^c, \mathbf{s})[2^{-n(\mathcal{I}(U;Y,B)-\epsilon)}(|\mathbb{I}_P| - 1)]$$

$$< [2^{-n(\mathcal{I}(U;Y,B)-\epsilon)} 2^{n(\mathcal{I}(U;Y,B)-\epsilon_1)}] \sum_{\mathbf{u}\in\mathbb{I}_P} \mathscr{P}(\tilde{\mathbf{u}} = \mathbf{u}|\mathbb{EV}_{21}^c, \mathbf{s}) \tag{3.49}$$

$$= 2^{-n(\epsilon_1-\epsilon)} \tag{3.50}$$

$$\leq \frac{\epsilon}{4} \tag{3.51}$$

where

- (3.46) follows from the fact that $\mathbb{EV}_1^c \subseteq \mathbb{EV}_{12}^c$;
- (3.47) is a union bound on (3.46);
- (3.48) follows from [2, Lem. (10.6.2)] for large enough $n \geq n_{22}(\epsilon)$;
- (3.49) follows from $|\mathbb{I}_0| = N_0$ for the first case according to (3.13), also from $|\mathbb{I}_P| = \frac{N_0}{N_1}$ for $P \in \{1, \ldots, N_1\}$ (the second case) according to equations (3.18) and (3.19);
- (3.50) follows from $\sum_{\mathbf{u}\in\mathbb{I}_P} \mathscr{P}(\tilde{\mathbf{u}} = \mathbf{u}|\mathbb{EV}_{21}^c, \mathbf{s}, P) = 1$;
- (3.51) follows if

$$\epsilon_1 \geq \epsilon - \frac{1}{n}\log(\frac{\epsilon}{4}) > \epsilon\ , \tag{3.52}$$

As a result, if ϵ_1 satisfies (3.52), from equations (3.42), (3.51), and Lemma 3.1, we conclude

$$\mathscr{P}(\mathbb{EV}_2|\mathbb{EV}_1^c, \mathbf{s}) \leq \frac{\epsilon}{2} \tag{3.53}$$

for $n \geq n_2(\epsilon)$, where $n_2(\epsilon) \triangleq \max\{n_{21}(\epsilon), n_{22}(\epsilon)\}$.

Conditions (3.38) and (3.52) have been already satisfied due to the specifications of the key agreement codebook. From equations (3.32), (3.37), and (3.53), we conclude that there exists $n \geq \max\{n_1(\epsilon), n_2(\epsilon)\}$ such that

$$\mathcal{E}(\mathscr{P}_{error}(n)) \leq \epsilon \tag{3.54}$$

for any $\epsilon \in (0, 1)$. This equation approves the validity of (3.25a).

(b). The Average Leakage Rate.

Lemma 3.2. *For any $\epsilon_5 > 0$, there exists $n \geq n_5(\epsilon_5)$ such that $\mathcal{H}(\mathbf{u}|K, \check{\mathbf{z}}, P, \mathbf{C}) < n\epsilon_5$, provided $\epsilon_0 > \epsilon_2 > \epsilon$.*

Proof. Suppose an arbitrary decoder Δ wishing to retrieve sent codeword $\tilde{\mathbf{u}}$ from K, $\check{\mathbf{z}}$, and P as its input information. The output of this decoder is denoted by $\hat{\mathbf{u}}$. Having K and P, the label of codeword $\tilde{\mathbf{u}}$ is uniquely determined with no error according to the encoding functions, which are publicly known. Hence, decoder Δ is to find $\hat{\mathbf{u}}$ from the sub-bin in which $\tilde{\mathbf{u}}$ is located. No matter what the decoding function of Δ is, from Fano's inequality [2, Thm. 2.10.1], we have

$$\begin{aligned} \mathcal{H}(\mathbf{u}|K, \check{\mathbf{z}}, P, \mathbf{C}) &\leq \mathcal{H}(\mathbf{u}|K, \check{\mathbf{z}}, P) && (3.55) \\ &\leq \mathcal{B}(\vartheta) + \vartheta \log(|\mathbb{L}|)\ , \\ &\leq \mathcal{B}(\vartheta) + n(\mathcal{I}(U; \check{Z}) - \epsilon_2)\vartheta\ , && (3.56) \end{aligned}$$

where

$$\begin{aligned} \mathcal{B}(x) &\triangleq -x\log(x) - (1-x)\log(1-x), && (3.57) \\ \vartheta &\triangleq \mathscr{P}(\hat{\mathbf{u}} \neq \tilde{\mathbf{u}}|K, \check{\mathbf{z}}, P). && (3.58) \end{aligned}$$

and (3.56) follows from equations (3.16) and (3.21) as $\epsilon_3 > \epsilon_2 > 0$.

The right-hand-side of (3.56) is valid for any decoder and it is determined by the decoder's probability of error ϑ. However, the value of left-hand-side is independent of decoder Δ, and it is determined by the joint PMF of $(\mathbf{u}, K, \check{\mathbf{z}}, P)$. So, any decoder Δ with probability of error ϑ gives a corresponding UB, as a function of ϑ, on $\mathcal{H}(\mathbf{u}|K, \check{\mathbf{z}}, P)$.

On the other hand, we can claim that there exists a (strongly) jointly typical decoder Δ for which $\vartheta \leq \epsilon$ for any $\epsilon \in (0,1)$. This decoder is called Δ^*. The proof of this claim is similar to that of the AR condition in the last part with the following modifications.

- B,Y, $\check{Y}$, and their corresponding vectors are replaced by E,Z, $\check{Z}$, and their corresponding vectors, respectively;
- $\mathbb{I}_P$ is changed to $\mathbb{L}_{IJ}$, where the random pair (I, J) is the label of codeword $\mathbf{u}$
- Markov chain $U \rightarrow (X, S) \rightarrow Z$ holds due to (3.10).

Finally, from (3.56) we conclude the lemma for decoder Δ^* as for any $\epsilon_5 \in (0,1)$ there exists a decoder Δ^* with probability of error $\vartheta \leq \epsilon$ such that $\mathcal{B}(\vartheta) + n(\mathcal{I}(U; \check{Z}) - \epsilon_2)\vartheta \leq \epsilon_5$ holds. □

Lemma 3.3. *$\mathcal{H}(\mathbf{u}|P, \mathbf{C}) \geq n[\mathcal{I}(U; \check{Y}) - \epsilon_1]$, where $\epsilon_1 \in (0,1)$ was fixed in the codebook generation.*

Proof. Consider $\mathbf{C}$ as a random matrix of independently drawn codewords. Let RV I denote the label of codeword $\mathbf{u}$ in codebook $\mathbf{C}$. In other words, I is the row number of random matrix C in which $\mathbf{u}$ is located.

$$\begin{aligned}
\mathcal{H}(\mathbf{u}|P, \mathbf{C}) &= \mathcal{H}(\mathbf{u}, P|\mathbf{C}) - \mathcal{H}(P|\mathbf{C}) \\
&= \mathcal{H}(\mathbf{u}|\mathbf{C}) + \mathcal{H}(P|\mathbf{u}, \mathbf{C}) - \mathcal{H}(P|\mathbf{C}) \\
&= \mathcal{H}(\mathbf{u}) + \mathcal{H}(\mathbf{C}|\mathbf{u}) - \mathcal{H}(\mathbf{C}) + \mathcal{H}(P|\mathbf{u}, \mathbf{C}) - \mathcal{H}(P|\mathbf{C}) \\
&= \mathcal{H}(\mathbf{u}) + \mathcal{H}(\mathbf{C}|\mathbf{u}) - \mathcal{H}(\mathbf{C}) - \mathcal{H}(P|\mathbf{C}) \qquad (3.59) \\
&= \mathcal{H}(\mathbf{u}) + \mathcal{H}(\mathbf{C}|\mathbf{u}, I) + \mathcal{H}(I|\mathbf{u}) - \mathcal{H}(I|\mathbf{u}, \mathbf{C}) - \mathcal{H}(\mathbf{C}) - \mathcal{H}(P|\mathbf{C}) \\
&= \mathcal{H}(\mathbf{u}) + \mathcal{H}(\mathbf{C}|\mathbf{u}, I) + \mathcal{H}(I|\mathbf{u}) - \mathcal{H}(\mathbf{C}) - \mathcal{H}(P|\mathbf{C}) \qquad (3.60) \\
&\geq \mathcal{H}(\mathbf{u}) + \sum_{\substack{\ell=1 \\ \ell \neq I}}^{N_0} \mathcal{H}(\mathbf{u}_\ell) + \mathcal{H}(I|\mathbf{u}) - \sum_{\ell=1}^{N_0} \mathcal{H}(\mathbf{u}_\ell) - \log(|\mathbb{P}_1|) \qquad (3.61) \\
&= \mathcal{H}(I|\mathbf{u}) - \log(|\mathbb{P}_1|) \qquad (3.62)
\end{aligned}$$

$$= \log(N_0) - \log(|\mathbb{P}_1|) \tag{3.63}$$

$$= n[\mathcal{I}(U;\check{Y}) - \epsilon_1] \tag{3.64}$$

where

- (3.59) holds because P, if it is not zero as in the first case of code-books, is determined by bin-index of $\mathbf{u}$ in $\mathbf{C}$;
- (3.60) holds because I is uniquely determined by $\mathbf{u}$ and $\mathbf{C}$;
- (3.61) follows from the following facts:

  a. Codebook $\mathbf{C}$ is filled with codewords $\mathbf{u}$ at uniformly random;

  b. Codewords $\mathbf{u}$ are drawn independently;

  c. $\mathcal{H}(P|\mathbf{C}) \leq \log(|\mathbb{P}_1|)$, where

  $$\log(|\mathbb{P}_1|) = \begin{cases} 0 & : \text{if } \mathcal{I}(U;\check{Y}) \geq \mathcal{I}(U;S), \\ n[\mathcal{I}(U;S) - \mathcal{I}(U;\check{Y}) + \epsilon_0 + \epsilon_1] & : \text{if } \mathcal{I}(U;\check{Y}) < \mathcal{I}(U;S) \end{cases} \tag{3.65}$$

  due to public message encoding functions.

- (3.62) follows from $\mathcal{H}(\mathbf{u}) = \mathcal{H}(\mathbf{u}_I)$;
- (3.63) follows from the fact that when $\mathbf{C}$ is not given $\mathscr{P}\{I = i|\mathbf{u}\} = \frac{1}{N_0}$ for any $i \in \{1, \ldots, N_0\}$, because a permutation of a codebook containing $\mathbf{u}$ results in another codebook with the same PMF in which the label of codeword $\mathbf{u}$ is shifted correspondingly. Hence, $\sum\limits_{\mathbf{C}} \mathscr{P}(I = i|\mathbf{C}, \mathbf{u})\mathcal{P}(\mathbf{C}|\mathbf{u})$ is the same for any $i \in \{1, \ldots, N_0\}$ ($\mathscr{P}(I = i|\mathbf{C}, \mathbf{u}) = 1$ if label of $\mathbf{u}$ in $\mathbf{C}$ is i; otherwise, it is zero.);
- (3.64) follows from equations (3.13), (3.18), and (3.65).

□

Lemma 3.4. *For any $\epsilon_6 \in (0,1)$, there exists $n \geq n_6(\epsilon_6)$ such that $\frac{1}{n}\mathcal{H}(\check{\mathbf{z}}|\mathbf{u}) \geq \mathcal{H}(\check{Z}|U) - \epsilon_6$.*

Proof.

$$
\begin{aligned}
\mathcal{H}(\check{\mathbf{z}}|\mathbf{u}) &= \mathcal{H}(\mathbf{z},\mathbf{e}|\mathbf{u}) && (3.66)\\
&= -\sum_{(\mathbf{u},\mathbf{z},\mathbf{e})\in(\mathbb{U}^n,\mathbb{Z}^n,\mathbb{E}^n)} \mathcal{P}(\mathbf{z},\mathbf{e}|\mathbf{u})\mathcal{P}(\mathbf{u})\log(\mathcal{P}(\mathbf{z},\mathbf{e}|\mathbf{u})) && (3.67)\\
&\geq -\sum_{(\mathbf{u},\mathbf{z},\mathbf{e})\in\mathbb{T}^*_\epsilon(\mathcal{P}_{UZE})} \mathcal{P}(\mathbf{z},\mathbf{e}|\mathbf{u})\mathcal{P}(\mathbf{u})\log(\mathcal{P}(\mathbf{z},\mathbf{e}|\mathbf{u})) && (3.68)\\
&\geq -n\sum_{\mathbf{u}\in\mathbb{T}^*_\epsilon(\mathcal{P}_U)} \mathcal{P}(\mathbf{u})\times && (3.69)\\
&\qquad [\sum_{\upsilon\in\mathbb{U}}\frac{\eta(\upsilon|\mathbf{u})}{n}\sum_{(z,e)\in(\mathbb{Z},\mathbb{E})}\mathcal{P}_{Z,E|U}(z,e|u)\log(\mathcal{P}_{Z,E|U}(z,e|u))] && (3.70)\\
&\geq n\sum_{\mathbf{u}\in\mathbb{T}^*_\epsilon(P_U)} \mathcal{P}(\mathbf{u})\times && (3.71)\\
&\qquad [\sum_{u\in\mathbb{U}}(\mathcal{P}_U(U=u)-\epsilon')\sum_{(z,e)\in(\mathbb{Z},\mathbb{E})}-\mathcal{P}_{Z,E|U}(z,e|u)\log(\mathcal{P}_{Z,E|U}(z,e|u))] && (3.72)\\
&= n\sum_{\mathbf{u}\in\mathbb{T}^*_\epsilon(\mathcal{P}_U)} \mathcal{P}(\mathbf{u})\mathcal{H}(Z,E|U)(1-\epsilon'') && (3.73)\\
&\geq n(1-\epsilon)\mathcal{H}(Z,E|U)(1-\epsilon'') && (3.74)\\
&= n(\mathcal{H}(\check{Z}|U)-\epsilon_6) && (3.75)\\
& && (3.76)
\end{aligned}
$$

where

- (3.68) distribution $\mathcal{P}_{UZE}$ is calculated from (3.10);
- (3.70) $\eta(\upsilon|\mathbf{u})$ represents the number of indices $i \in \{1,\ldots,n\}$ such that $\upsilon = u_i$ (see Section 2.1.1 for more details); also,

$$\mathcal{P}(\mathbf{s},\mathbf{x}|\mathbf{u})\mathcal{P}(\mathbf{z},\mathbf{e}|\mathbf{x},\mathbf{s}) = \prod_{i=1}^{n}\mathcal{P}_{SX|U}(s_i,x_i|u_i)\mathcal{P}_{ZE|XS}(z_i,e_i|x_i,s_i)\,,$$

  where $\mathcal{P}_{SX|U}$ and $\mathcal{P}_{ZE|XS}$ are calculated according to (3.10) (see also [53, Page 1634]);
- (3.72) follows from the fact that sequence $\mathbf{u}$ is ϵ-strongly typical (see [3, Sec. 2]);

- (3.74) follows from $\sum\limits_{\mathbf{u}\in\mathbb{T}^*_\epsilon(\mathcal{P}_U)} \mathcal{P}(\mathbf{u}) \geq 1-\epsilon$;
- (3.75) holds for any $\epsilon_6 \in (0,1)$ if ϵ', ϵ'', and ϵ were selected small enough such that previous inequalities hold.

□

Finally, for any $\epsilon \in (0,1)$, we complete the proof as follows:

$$\begin{aligned}
\mathcal{H}(K|\check{\mathbf{z}},P,\mathbf{C}) &= \mathcal{H}(K|\check{\mathbf{z}},P,\mathbf{u},\mathbf{C}) + \mathcal{I}(K;\mathbf{u}|\check{\mathbf{z}},P,\mathbf{C}) \\
&= \mathcal{H}(\mathbf{u}|\check{\mathbf{z}},P,\mathbf{C}) - \mathcal{H}(\mathbf{u}|K,\check{\mathbf{z}},P,\mathbf{C}) && (3.77) \\
&= \mathcal{H}(\mathbf{u}|P,\mathbf{C}) - \mathcal{H}(\check{\mathbf{z}}|P,\mathbf{C}) + \mathcal{H}(\check{\mathbf{z}}|\mathbf{u},P,\mathbf{C}) - n\epsilon_5 && (3.78) \\
&\geq \mathcal{H}(\mathbf{u}|P,\mathbf{C}) - \mathcal{H}(\check{\mathbf{z}}) + \mathcal{H}(\check{\mathbf{z}}|\mathbf{u}) - n\epsilon_5 && (3.79) \\
&\geq n[\mathcal{I}(U;\check{Y}) - \epsilon_1] - \sum_{i=1}^{n}\mathcal{H}(\check{Z}_i) + \sum_{i=1}^{n}\mathcal{H}(\check{Z}_i|U_i) - n(\epsilon_5+\epsilon_6) && (3.80) \\
&= n[\mathcal{I}(\check{Y};U) - \mathcal{I}(\check{Z};U)] - n(\epsilon_1+\epsilon_5+\epsilon_6) && (3.81) \\
&\geq \log(|\mathbb{K}|) - n(\epsilon_2+\epsilon_5+\epsilon_6) && (3.82) \\
&\geq \mathcal{H}(K) - n\epsilon && (3.83)
\end{aligned}$$

where

- (3.77) follows from the fact that K is a function of $\mathbf{u}$ for given codebook $\mathbf{C}$;
- (3.78) follows from Lemma 3.2 for $n \geq n_5(\epsilon_5)$;
- (3.79) holds as $\mathcal{H}(\check{\mathbf{z}}|\mathbf{u},\mathbf{C}) = \mathcal{H}(\check{\mathbf{z}}|\mathbf{u})$ follows from Markov chain $(\mathbf{C},P) \to \mathbf{u} \to (\mathbf{e},\mathbf{z})$; Also, $\mathcal{H}(\check{\mathbf{z}}) \geq \mathcal{H}(\check{\mathbf{z}}|P,\mathbf{C})$;
- (3.81) holds because of the following facts:

  a. The first term follows from Lemma 3.3;

b. The second term follows from

$$
\begin{aligned}
\mathcal{H}(\check{\mathbf{z}}) &\leq \sum_{i=1}^{n} \mathcal{H}(\check{Z}_i) && (3.84)\\
&= n \sum_{i=1}^{n} \frac{1}{n} \mathcal{H}(\check{Z}_\tau | \tau = i) && (3.85)\\
&= n\mathcal{H}(\check{Z}_\tau | \tau) && (3.86)\\
&\leq n\mathcal{H}(\check{Z}_\tau) && (3.87)\\
&= n\mathcal{H}(\check{Z})
\end{aligned}
$$

for a uniformly distributed RV $\tau \in \{1, \ldots, n\}$. Finally, let $Z \triangleq Z_\tau$;

c. The third term follows from Lemma 3.4;

- (3.81) follows from equations (3.13), (3.21);
- (3.82) holds from equations (3.17) and (3.23);
- (3.83) is valid from $\log(|\mathbb{K}|) \geq \mathcal{H}(K)$ and $\epsilon > \epsilon_1 + \epsilon_5 + \epsilon_6$.

Hence, (3.83) approves the validity of (3.25b).

3.3.2 The Proof of Corollary 3.1

According to Corollary 2.1, we have

$$C_{S} = C_m = \mathcal{I}(U^*; \check{Y}^*) - \mathcal{I}(U^*; S)$$

which is also valid if side information is available at Bob or/and Eve[3]. From Theorem 3.1, we also understand that $R_{K} = \mathcal{I}(U^*; \check{Y}^*) - \mathcal{I}(U^*; \check{Z}^*)$ is an achievable key rate if the DM-SWC is less noisy in Bob's favor. The fact that $\mathcal{I}(U^*; S) > \mathcal{I}(U^*; \check{Z}^*)$ completes the proof of this corollary.

[3] If side information is available at any receiver, an augmented output can be considered at that receiver, and thus the secrecy capacity can be extended to this case.

3.3.3 The Proof of Theorem 3.2

The proof of the UB is established in two steps. In the first step, we obtain a UB on the forward key capacity of the DM model due to the effect of the DM-SWC on the achievable key rate. In the second step, we derive a constraint imposed by C_{P_1} due to the public channel.

First, fix an arbitrarily small $\epsilon \geq 0$. According to Definition 1.11, a key rate R_K is achievable if there exists an admissible key agreement code $(\lceil 2^{nR_K} \rceil, n)$, which returns $(K, \hat{K})$, for the given ϵ such that

$$\mathcal{P}_{error}(n) \leq \frac{\epsilon}{3}\,, \tag{3.88a}$$

$$\mathcal{R}_L(n) \leq \frac{\epsilon}{3}\,, \tag{3.88b}$$

$$\mathcal{X}(n) \leq \frac{\epsilon}{3}\,, \tag{3.88c}$$

$$\frac{\mathcal{H}(K)}{n} + \frac{\epsilon}{3} \geq R_K\,. \tag{3.88d}$$

where $\mathcal{P}_{error}(n)$, $\mathcal{R}_L(n)$, $\mathcal{X}(n)$ are defined in Definition (1.10). Then, we can derive the UB in the following two steps.

(a). The effect of the wiretap channel on R_K. First, we establish the following lemma.

Lemma 3.5. *Let $\mathcal{P}_{WUBESXYZ}$ be a joint distribution on $\mathbb{W} \times \mathbb{U} \times \mathbb{B} \times \mathbb{E} \times \mathbb{S} \times \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$ such that Markov chains $W \to (U, B, E) \to (X, S) \to (Y, Z)$ and $W \to (U, B, E, S) \to X$ hold, where the marginal distribution $\mathcal{P}_{BES}$ is fixed. Recalling notations* (3.3)*, define the function*

$$\mathcal{L}(\mathcal{P}_{USX\check{Y}\check{Z}|W=w_m}) \triangleq \mathcal{I}(U; \check{Y}|W = w_m) - \mathcal{I}(U; \check{Z}|W = w_m)\,,$$

where $w_m \in \mathbb{W}$ is fixed. Then, there exists a distribution $\mathcal{P}^*_{USX\check{Y}\check{Z}}$ with marginal distribution $\mathcal{P}_{BES}$ such that

$$\mathcal{I}(U^*; \check{Y}^*) - \mathcal{I}(U^*; \check{Z}^*) = \mathcal{L}(\mathcal{P}_{USX\check{Y}\check{Z}|W=w_m})\,, \tag{3.89}$$

where the left-hand-side is calculated at distribution $\mathcal{P}^*_{USX\check{Y}\check{Z}}$.

Proof. Let

$$\mathcal{P}^*_{WUSX\check{Y}\check{Z}} \triangleq \mathcal{P}_{BES}\mathcal{P}^*_{W|BES}\mathcal{P}^*_{U|WBES}\mathcal{P}_{X|UBES}\mathcal{P}_{YZ|XS} \tag{3.90}$$

where $\mathcal{P}_{BES}$, $\mathcal{P}_{X|UBES}$, and $\mathcal{P}_{YZ|XS}$ are marginal distributions of $\mathcal{P}_{WUSXBYEZ}$ and

$$\begin{aligned}&\forall (b,e,s) \in \mathbb{B}\times\mathbb{E}\times\mathbb{S} :\\ &\qquad \mathcal{P}^*_{W|BES}(w|b,e,s) \triangleq \begin{cases} 1, & \text{:If } w = w_m;\\ 0, & \text{:Otherwise.}\end{cases}\end{aligned} \tag{3.91a}$$

$$\begin{aligned}&\forall (u,b,e,s) \in \mathbb{U}\times\mathbb{B}\times\mathbb{E}\times\mathbb{S} :\\ &\qquad \mathcal{P}^*_{U|WBES}(u|w,b,e,s) \triangleq \begin{cases} \frac{\mathcal{P}_{W|BES}(w_m|b,e,s)\mathcal{P}_{U|BESW}(u|b,e,s,w_m)}{\mathcal{P}_W(w_m)}, & \text{:If } w = w_m;\\ 0, & \text{:Otherwise.}\end{cases}\end{aligned} \tag{3.91b}$$

From (3.90) and (3.91a), we have

$$\mathcal{P}^*_W(w) = \sum_{s\in\,\mathbb{S}} \mathcal{P}^*_{W|S}(w|s)\mathcal{P}_S(s) = \begin{cases} 1, & \text{:If } w = w_m;\\ 0, & \text{:Otherwise.}\end{cases} \tag{3.92}$$

Moreover, from (3.90) and (3.91b) and (3.92), we have

$$\forall u \in \mathbb{U} : \mathcal{P}^*_U(u) = \mathcal{P}_{U|W}(u|w_m)\,. \tag{3.93}$$

From (3.90) and (3.92), for any $(w_m, u, b, e, s) \in \mathbb{W}\times\mathbb{U}\times\mathbb{B}\times\mathbb{E}\times\mathbb{S}$ (with fixed given w_m) we conclude that

$$\begin{aligned}\mathcal{P}^*_{UBES|W}(u,b,e,s|w_m) &= \mathcal{P}^*_{BES|W}(b,e,s|w_m)\mathcal{P}^*_{U|BESW}(u|b,e,s,w)\\ &= \frac{\mathcal{P}_{BES}(b,e,s)\mathcal{P}^*_{W|BES}(w_m|b,e,s)}{\mathcal{P}^*_W(w_m)}\mathcal{P}^*_{U|BESW}(u|b,e,s,w_m)\end{aligned}$$

$$= \mathcal{P}_{BES}(b,e,s)\left(\frac{\mathcal{P}_{W|BES}(w_m|b,e,s)}{\mathcal{P}_W(w_m)}\mathcal{P}_{U|BESW}(u|b,e,s,w_m)\right)$$
$$= \mathcal{P}_{UBES|W}(u,b,e,s|w_m)$$

From this equation, for any $(u,b,e,s,x,y,z) \in \mathbb{U}\times\mathbb{B}\times\mathbb{E}\times\mathbb{S}\times\mathbb{X}\times\mathbb{Y}\times\mathbb{Z}$ and fixed $w_m \in \mathbb{W}$, we have

$$\mathcal{P}^*_{UBESXYZ|W}(u,b,e,s,x,y,z|w_m) = \mathcal{P}_{USXYZ|W}(u,b,e,s,x,y,z|w_m)$$

due to Markov chains $W \to U \to (X,S) \to (Y,Z)$ and $W \to (U,B,E,S) \to X$ and (3.90). As a result,

$$\mathcal{L}(\mathcal{P}_{USX\check{Y}\check{Z}|W=w_m}) = \mathcal{L}(\mathcal{P}^*_{USX\check{Y}\check{Z}|W=w_m})\ .$$

On the other hand, from (3.92), which is valid for distribution $\mathcal{P}^*_{WUSX\check{Y}\check{Z}}$, we can redefine $\mathbb{W} = \{w_m\}$ for this distribution, and so $\mathcal{I}(U;\check{Y}|W=w_m) - \mathcal{I}(U;\check{Z}|W=w_m) = \mathcal{I}(U;\check{Y}) - \mathcal{I}(U;\check{Z})$ at $\mathcal{P}^*_{UBESXYZ}(u,b,e,s,x,y,z) = \mathcal{P}^*_{UBESXYZ|W}(u,b,e,s,x,y,z|w_m)$. Hence, the lemma is established. □

Now, we prove the UB as follows.

$$nR_{_K} \le \mathcal{H}(K) + n\frac{\epsilon}{3} \tag{3.94}$$
$$\le \mathcal{I}(K;\hat{K}) + n\frac{2\epsilon}{3} \tag{3.95}$$
$$\le \mathcal{I}(K;\mathbf{y},\mathbf{b},P) + n\frac{2\epsilon}{3} \tag{3.96}$$
$$\le \mathcal{I}(K;\mathbf{y},\mathbf{b},P) - \mathcal{I}(K;\mathbf{z},\mathbf{e},P) + n\epsilon \tag{3.97}$$
$$= \mathcal{I}(K;\check{Y}_1^n|P) - \mathcal{I}(K;\check{Z}_1^n|P) + n\epsilon \tag{3.98}$$
$$= \sum_{i=1}^n [\mathcal{I}(K;\check{Y}_i|\check{Y}_1^{i-1},P) - \mathcal{I}(K;\check{Z}_i|\check{Z}_{i+1}^n,P)] + n\epsilon \tag{3.99}$$
$$= \sum_{i=1}^n [\mathcal{I}(K;\check{Y}_i|\check{Y}_1^{i-1},\check{Z}_{i+1}^n,P) - \mathcal{I}(K;\check{Z}_i|\check{Y}_1^{i-1},\check{Z}_{i+1}^n,P)] + n\epsilon \tag{3.100}$$

$$= \sum_{i=1}^{n} [\mathcal{I}(U_i; \check{Y}_i | W_i) - \mathcal{I}(U_i; \check{Z}_i | W_i)] + n\epsilon \tag{3.101}$$

$$= n[\frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(U_\tau; \check{Y}_\tau | W_\tau, \tau = i) - \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}(U_\tau; \check{Z}_\tau | W_\tau, \tau = i)] + n\epsilon$$

$$= n[\mathcal{I}(U_\tau; \check{Y}_\tau | W_\tau, \tau) - \mathcal{I}(U_\tau; \check{Z}_\tau | W_\tau, \tau)] + n\epsilon \tag{3.102}$$

$$= n[\mathcal{I}(U; \check{Y} | W) - \mathcal{I}(U; \check{Z} | W)] + n\epsilon \tag{3.103}$$

$$= n \sum_{w \in \mathbb{W}} [\mathcal{I}(U; \check{Y} | W = w) - \mathcal{I}(U; \check{Z} | W = w)] \mathcal{P}_W(w) + n\epsilon$$

$$\leq n[\mathcal{I}(U; \check{Y} | W = w_m) - \mathcal{I}(U; \check{Z} | W = w_m)] + n\epsilon \tag{3.104}$$

$$= n[\mathcal{I}(U^*; \check{Y}^*) - \mathcal{I}(U^*; \check{Z}^*)] + n\epsilon \tag{3.105}$$

$$\leq n \sup_{U \to (X,S) \to (Y,Z)} [\mathcal{I}(U; \check{Y}) - \mathcal{I}(U; \check{Z})] + n\epsilon$$

$$= n \max_{U \to (X,S) \to (Y,Z)} [\mathcal{I}(U; \check{Y}) - \mathcal{I}(U; \check{Z})] + n\epsilon \tag{3.106}$$

where

- (3.94) follows from (3.88d);
- (3.95) follows from (3.88a) and $\mathcal{H}(K|\hat{K}) \leq n\frac{\epsilon}{3}$ due to Fano's inequality [2, Thm. 2.10.1];
- (3.96) holds due to data processing inequality [2, Thm. 2.8.1] as $\hat{K} = \mathcal{K}_2(\mathbf{y}, \mathbf{b}, P)$ and so $K \to (\mathbf{y}, \mathbf{b}, P) \to \hat{K}$;
- (3.97) follows from (3.88b);
- (3.98) follows from (3.3).
- (3.99) is valid because of the chain rule for mutual information [2, Thm. 2.5.2];
- (3.100) holds from Csiszár-Körner's sum identity (see Appendix D);
- (3.101) holds by definitions $W_i \triangleq (P\check{Y}_1^{i-1}\check{Z}_{i+1}^n)$ and $U_i \triangleq (K, W_i)$, i.e.,

$$U_i \triangleq (KP\check{Y}_1^{i-1}\check{Z}_{i+1}^n) ; \tag{3.107}$$

- (3.102) holds as a time-sharing RV τ with a uniform distribution over $\{1, \ldots, n\}$, which is independent of $(K, P, S_1^n, X_1^n, \check{Y}_1^n, \check{Z}_1^n)$, is applied;

- (3.103) follows from definitions

$$U \triangleq (U_\tau, \tau)\ , \tag{3.108a}$$

$$W \triangleq (W_\tau, \tau)\ , \tag{3.108b}$$

$$X \triangleq X_\tau\ , \tag{3.108c}$$

$$\check{Y} \triangleq \check{Y}_\tau\ , \tag{3.108d}$$

$$\check{Z} \triangleq \check{Z}_\tau\ ; \tag{3.108e}$$

  Also, note that Markov chain $W \to U \to (X, S) \to (\check{Y}, \check{Z})$ is valid; $\mathbb{U}$ and $\mathbb{W}$ are also reserved for alphabet set of U and that of W, respectively;

- (3.104) follows from $\sum\limits_{w \in \mathbb{W}} \mathcal{P}_{W}(w) = 1$ and selection

$$w_m \in \arg\max_{w \in \mathbb{W}} [\mathcal{I}(U; \check{Y}|W = w) - \mathcal{I}(U; \check{Z}|W = w)]\ ,$$

  and so the equality in (3.104) holds if

$$\forall w \in \mathbb{W}: \quad w \in \arg\max_{w \in \mathbb{W}} [\mathcal{I}(U; \check{Y}|W = w) - \mathcal{I}(U; \check{Z}|W = w)]\ ; \tag{3.109}$$

- (3.105) follows from Lemma 3.5 according to distribution $\mathcal{P}^*_{USX\check{Y}\check{Z}}$ for which Markov chains $W \to U \to (X, S) \to (Y, Z)$ and $(U, W, X, Y, Z) \to S \to (B, E)$ hold;
- (3.106) follows from Appendix B and the fact that $\mathbb{U}$ is a finite set according to (3.107) and (3.108a) for a given n;

(b) The effect of the public channel on $R_{_K}$.

$$\begin{aligned} nC_{_P} &\geq \log(|\mathbb{P}_1|) \qquad (3.110)\\ &\geq \mathcal{H}(P) \\ &= \mathcal{H}(KP) - \mathcal{H}(K|P) \\ &= \mathcal{H}(KP) - \mathcal{I}(K; Y_1^n B_1^n | P) - \mathcal{H}(K|PY_1^n B_1^n) \end{aligned}$$

$$\geq \mathcal{H}(KP) - \mathcal{I}(K; Y_1^n B_1^n | P) - n\epsilon \tag{3.111}$$

$$\geq \mathcal{H}(KP) - \mathcal{I}(KP; \check{Y}_1^n) - n\epsilon \tag{3.112}$$

$$\geq \mathcal{H}(KP) - \mathcal{H}(KP | X_1^n S_1^n) - \mathcal{I}(KP; \check{Y}_1^n) - n\epsilon \tag{3.113}$$

$$= \mathcal{I}(KP; X_1^n S_1^n) - \mathcal{I}(KP; \check{Y}_1^n) - n\epsilon$$

$$= \sum_{i=1}^{n} \mathcal{I}(KP; X_i S_i | X_{i+1}^n S_{i+1}^n) - \sum_{i=1}^{n} \mathcal{I}(KP; \check{Y}_i | \check{Y}_1^{i-1}) - n\epsilon \tag{3.114}$$

$$\begin{aligned}
&= \sum_{i=1}^{n} [\mathcal{I}(KP\check{Y}_1^{i-1}; X_i S_i | X_{i+1}^n S_{i+1}^n) - \mathcal{I}(Y_1^{i-1}; X_i S_i | KP X_{i+1}^n S_{i+1}^n)] \\
&\quad + \sum_{i=1}^{n} [\mathcal{I}(KP X_{i+1}^n S_{i+1}^n; \check{Y}_i | \check{Y}_1^{i-1}) - \mathcal{I}(X_{i+1}^n S_{i+1}^n; \check{Y}_i | KP\check{Y}_1^{i-1})] - n\epsilon
\end{aligned}$$

$$= \sum_{i=1}^{n} [\mathcal{I}(KP\check{Y}_1^{i-1}; X_i S_i | X_{i+1}^n S_{i+1}^n) - \mathcal{I}(KP X_{i+1}^n S_{i+1}^n; \check{Y}_i | \check{Y}_1^{i-1})] - n\epsilon \tag{3.115}$$

$$\begin{aligned}
&= \sum_{i=1}^{n} [\mathcal{I}(KP X_{i+1}^n S_{i+1}^n \check{Y}_1^{i-1}; X_i S_i) - \mathcal{I}(KP X_{i+1}^n S_{i+1}^n \check{Y}_1^{i-1}; \check{Y}_i)] \\
&\quad - \sum_{i=1}^{n} [\mathcal{I}(X_{i+1}^n S_{i+1}^n; X_i S_i) - \mathcal{I}(\check{Y}_1^{i-1}; \check{Y}_i)] - n\epsilon
\end{aligned}$$

$$\begin{aligned}
&\geq \sum_{i=1}^{n} [\mathcal{I}(KP X_{i+1}^n S_{i+1}^n \check{Y}_1^{i-1}; X_i S_i) - \mathcal{I}(KP X_{i+1}^n S_{i+1}^n \check{Y}_1^{i-1}; \check{Y}_i)] \\
&\quad - \sum_{i=1}^{n} \mathcal{I}(X_{i+1}^n S_{i+1}^n; X_i S_i) - n\epsilon
\end{aligned} \tag{3.116}$$

$$\begin{aligned}
&= \sum_{i=1}^{n} [\mathcal{I}(KP X_{i+1}^n S_{i+1}^n \check{Z}_{i+1}^n \check{Y}_1^{i-1}; X_i S_i) - \mathcal{I}(KP X_{i+1}^n S_{i+1}^n \check{Z}_{i+1}^n \check{Y}_1^{i-1}; \check{Y}_i)] \\
&\quad - \sum_{i=1}^{n} [\mathcal{I}(\check{Z}_{i+1}^n; X_i S_i | KP X_{i+1}^n S_{i+1}^n \check{Y}_1^{i-1}) - \mathcal{I}(\check{Z}_{i+1}^n; \check{Y}_i | KP X_{i+1}^n S_{i+1}^n \check{Y}_1^{i-1})] \\
&\quad - \sum_{i=1}^{n} \mathcal{I}(X_{i+1}^n S_{i+1}^n; X_i S_i) - n\epsilon
\end{aligned}$$

$$= \sum_{i=1}^{n} [\mathcal{I}(KP X_{i+1}^n S_{i+1}^n \check{Z}_{i+1}^n \check{Y}_1^{i-1}; X_i S_i) - \mathcal{I}(KP X_{i+1}^n S_{i+1}^n \check{Z}_{i+1}^n \check{Y}_1^{i-1}; \check{Y}_i)]$$

$$-\sum_{i=1}^{n}\mathcal{I}(X_{i+1}^{n}S_{i+1}^{n};X_iS_i)-n\epsilon \tag{3.117}$$

$$=\sum_{i=1}^{n}[\mathcal{I}(KP\check{Z}_{i+1}^{n}\check{Y}_1^{i-1};X_iS_i)-\mathcal{I}(KP\check{Z}_{i+1}^{n}\check{Y}_1^{i-1};\check{Y}_i)]$$

$$+\sum_{i=1}^{n}[\mathcal{I}(X_{i+1}^{n}S_{i+1}^{n};X_iS_i|KP\check{Z}_{i+1}^{n}\check{Y}_1^{i-1})-\mathcal{I}(X_{i+1}^{n}S_{i+1}^{n};\check{Y}_i|KP\check{Z}_{i+1}^{n}\check{Y}_1^{i-1})]$$

$$-\sum_{i=1}^{n}\mathcal{I}(X_{i+1}^{n}S_{i+1}^{n};X_iS_i)-n\epsilon$$

$$\geq\sum_{i=1}^{n}[\mathcal{I}(KP\check{Z}_{i+1}^{n}\check{Y}_1^{i-1};S_i)+\mathcal{I}(X_{i+1}^{n}S_{i+1}^{n};X_i|S_i)-\mathcal{I}(KP\check{Z}_{i+1}^{n}\check{Y}_1^{i-1};\check{Y}_i)]$$

$$-\sum_{i=1}^{n}[\mathcal{H}(X_iS_i)-\mathcal{H}(X_iS_i|X_{i+1}^{n}S_{i+1}^{n})]-n\epsilon \tag{3.118}$$

$$=\sum_{i=1}^{n}[\mathcal{I}(KP\check{Z}_{i+1}^{n}\check{Y}_1^{i-1};S_i)+\mathcal{H}(X_i|S_i)-\mathcal{H}(X_i|KP\check{Z}_{i+1}^{n}\check{Y}_1^{i-1}S_i)-\mathcal{I}(KP\check{Z}_{i+1}^{n}\check{Y}_1^{i-1};\check{Y}_i)]$$

$$-\sum_{i=1}^{n}\mathcal{H}(X_iS_i)+\mathcal{H}(X_1^nS_1^n)-n\epsilon \tag{3.119}$$

$$=\sum_{i=1}^{n}[\mathcal{I}(U_i;S_i)-\mathcal{H}(X_i|U_iS_i)-\mathcal{I}(U_i;\check{Y}_i)]$$

$$+\sum_{i=1}^{n}\mathcal{H}(X_i|S_i)-\sum_{i=1}^{n}\mathcal{H}(X_i|S_i)+\mathcal{H}(X_1^n|S_1^n)+[\sum_{i=1}^{n}\mathcal{H}(S_i)-\mathcal{H}(S_1^n)]-n\epsilon \tag{3.120}$$

$$=\sum_{i=1}^{n}[\mathcal{I}(U_i;S_i)-\mathcal{H}(X_i|U_iS_i)-\mathcal{I}(U_i;\check{Y}_i)]+\mathcal{H}(X_1^n|S_1^n)-n\epsilon \tag{3.121}$$

$$\geq\sum_{i=1}^{n}[\mathcal{I}(U_i;S_i)-\mathcal{H}(X_i|U_iS_i)-\mathcal{I}(U_i;\check{Y}_i)]-n\epsilon \tag{3.122}$$

$$=n[\sum_{i=1}^{n}\frac{1}{n}\mathcal{I}(U_i;S_i|\tau=i)-\sum_{i=1}^{n}\frac{1}{n}\mathcal{I}(U_i;\check{Y}_i|\tau=i)$$

$$-\sum_{i=1}^{n}\frac{1}{n}\mathcal{H}(X_i|U_iS_i,\tau=i)]-n\epsilon$$

$$=n[\mathcal{I}(U_\tau;S_\tau|\tau)-\mathcal{H}(X_\tau|U_\tau S_\tau,\tau)-\mathcal{I}(U_\tau;\check{Y}_\tau|\tau)]-n\epsilon \tag{3.123}$$

$$\geq n[\mathcal{I}(U_\tau, \tau; S_\tau) - \mathcal{I}(U_\tau, \tau; \check{Y}_\tau) - \mathcal{H}(X_\tau | U_\tau, \tau, S_\tau)] - n\epsilon \tag{3.124}$$

$$= n[\mathcal{I}(U; S) - \mathcal{I}(U; \check{Y}) - \mathcal{H}(X|U, S)] - n\epsilon \tag{3.125}$$

where

- (3.110) follows from public channel capacity constraint given in (3.1a);
- (3.111) holds due to Fano's inequality [2, Thm. 2.10.1] as well as satisfaction of the AR condition (1.18a) by the code, because $\hat{K} = \mathcal{K}_2(Y_1^n, B_1^n, P)$;
- (3.112) follows from substitution $\check{Y}_1^n = (Y_1^n, B_1^n)$; also, let $\check{Y}_i = (Y_i B_i)$;
- (3.113) holds due to $\mathcal{H}(KP|X_1^n S_1^n) \geq 0$;
- (3.114) follows from the mutual information chain rule [2, Ch. 2];
- (3.115) follows from Csiszár-Körner's sum identity (see Appendix D);
- (3.116) follows from $\mathcal{I}(\check{Y}_1^{i-1}; \check{Y}_i) \geq 0$;
- (3.117) holds due to Markov chains $X_i S_i \to X_{i+1}^n S_{i+1}^n \to \check{Z}_{i+1}^n$ and $\check{Y}_i \to X_{i+1}^n S_{i+1}^n \to \check{Z}_{i+1}^n$ as the wiretap channel is memoryless;
- (3.118) follows from data processing inequality [2, Thm. 2.8.1] because Markov chain $KP\check{Y}_1^{i-1}\check{Z}_{i+1}^n X_{i+1}^n S_{i+1}^n \to X_i S_i \to Y_i$ holds as the wiretap channel is memoryless;
- (3.119) follows from $\sum_{i=1}^{n} \mathcal{H}(X_i S_i | X_{i+1}^n S_{i+1}^n) = H(X_1^n S_1^n)$ due to entropy chain rule [2];
- (3.120) is valid due to Definition 3.107;
- (3.121) follows from the fact that S_1^n is drawn i.i.d., and so $\sum_{i=1}^{n} \mathcal{H}(S_i) = \mathcal{H}(S_1^n)$;
- (3.122) follows from $\mathcal{H}(X_1^n | S_1^n) \geq 0$;
- (3.123) follows by assuming τ is a uniform RV over $\{1, \ldots, n\}$ such that it is independent of $(KPS_1^n X_1^n \check{Y}_1^n \check{Z}_1^n)$;
- (3.124) holds as distribution of S_τ does not depend on τ (S_1^n is i.i.d.); also, $\mathcal{I}(U_\tau, \tau; \check{Y}_\tau) \geq \mathcal{I}(U_\tau; \check{Y}_\tau | \tau)$, where the equality holds if $\check{Y}_1^n$ is i.i.d.;
- (3.125) follows from definitions (3.108), also S_τ can be replaced by S as distribution of S_τ does not depend on τ (S_1^n is i.i.d.);

As equations (3.106) and (3.125) are valid for any achievable rate R_{κ}, and any $\epsilon \in (0, 1)$

and a correspondingly large enough n, the theorem is concluded.

3.3.4 The Proof of Theorem 3.3

For any $i \in \{1, \ldots, n\}$, let $U_i = (P\check{Y}_1^{i-1}S_{i+1}^n)$ and $V_i = (KX_{i+1}^n)$; thus, the Markov chain $(U_i, V_i) \to (X_i, S_i) \to (\check{Y}_i, \check{Z}_i)$ is valid for every i. Further, we can write

$$\begin{aligned}
nC_{\scriptscriptstyle P} &\geq \log(|\mathbb{P}_1|) && (3.126)\\
&\geq \mathcal{H}(P)\\
&\geq \mathcal{I}(P; S_1^n)\\
&= \sum_{i=1}^{n} \mathcal{I}(P; S_i | S_{i+1}^n)\\
&= \sum_{i=1}^{n} [\mathcal{I}(P\check{Y}_1^{i-1}; S_i | S_{i+1}^n) - \mathcal{I}(\check{Y}_1^{i-1}; S_i | S_{i+1}^n, P)]\\
&= \sum_{i=1}^{n} [\mathcal{I}(P\check{Y}_1^{i-1}S_{i+1}^n; S_i) - \mathcal{I}(S_{i+1}^n; \check{Y}_i | \check{Y}_1^{i-1}, P)] && (3.127)\\
&\geq \sum_{i=1}^{n} [\mathcal{I}(U_i; S_i) - \mathcal{I}(U_i; \check{Y}_i)]\\
&= n[\frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(U_\tau; S_\tau | \tau = i) - \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}(U_\tau; \check{Y}_\tau | \tau = i)]\ ,\\
&= n[\mathcal{I}(U_\tau; S_\tau | \tau) - \mathcal{I}(U_\tau; \check{Y}_\tau | \tau)] && (3.128)\\
&\geq n[\mathcal{I}(U_\tau, \tau; S_\tau) - \mathcal{I}(U_\tau, \tau; \check{Y}_\tau)] && (3.129)\\
&= n[\mathcal{I}(U; S) - \mathcal{I}(U; \check{Y})]\ , && (3.130)
\end{aligned}$$

where

- (3.126) follows from public channel capacity constraint given in (3.1a);
- (3.127) follows from the fact that $\mathbf{s}$ is i.i.d. and from Csiszár-Körner's sum identity (see Appendix D);
- (3.128) holds as a time-sharing RV τ with a uniform distribution over $\{1, \ldots, n\}$, which is independent of $(K, P, S_1^n, X_1^n, \check{Y}_1^n, \check{Z}_1^n)$, is applied;

- (3.129) holds because τ is independent of S_τ ($\mathbf{s}$ is i.i.d.);
- (3.130) follows from definitions $U \triangleq (U_\tau, \tau)$, $\check{Y} \triangleq \check{Y}_\tau$, and $S \triangleq S_\tau$.

Restarting from (3.100), we also have

$$
\begin{aligned}
nR_{_K}(C_{_{P_1}}) &\leq \sum_{i=1}^{n}[\mathcal{I}(K, \check{Y}_1^{i-1}, \check{Z}_{i+1}^n, P; \check{Y}_i) - \mathcal{I}(K, \check{Y}_1^{i-1}, \check{Z}_{i+1}^n, P; \check{Z}_i)] + n\epsilon && (3.131)\\
&= \sum_{i=1}^{n}[\mathcal{I}(K, \check{Y}_1^{i-1}, \check{Z}_{i+1}^n, X_{i+1}^n, S_{i+1}^n, P; \check{Y}_i) - \mathcal{I}(K, \check{Y}_1^{i-1}, \check{Z}_{i+1}^n, X_{i+1}^n, S_{i+1}^n, P; \check{Z}_i)]\\
&\quad - \sum_{i=1}^{n}[\mathcal{I}(X_{i+1}^n, S_{i+1}^n; \check{Y}_i | K, \check{Y}_1^{i-1}, \check{Z}_{i+1}^n, P) - \mathcal{I}(X_{i+1}^n, S_{i+1}^n; \check{Z}_i | K, \check{Y}_1^{i-1}, \check{Z}_{i+1}^n, P)] + n\epsilon\\
&\leq \sum_{i=1}^{n}[\mathcal{I}(U_i, V_i; \check{Y}_i) - \mathcal{I}(U_i, V_i; \check{Z}_i)] + n\epsilon && (3.132)\\
&= n[\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}(U_\tau, V_\tau; \check{Y}_\tau | \tau = i) - \frac{1}{n}\sum_{i=1}^{n}\mathcal{I}(U_\tau, V_\tau; \check{Z}_\tau | \tau = i)] + n\epsilon && (3.133)\\
&= n[\mathcal{I}(U_\tau, V_\tau; \check{Y}_\tau | \tau) - \mathcal{I}(U_\tau, V_\tau; \check{Z}_\tau | \tau)] + n\epsilon\\
&\leq n[\mathcal{I}(U_\tau, V_\tau, \tau; \check{Y}_\tau) - \mathcal{I}(U_\tau, V_\tau, \tau; \check{Z}_\tau)] + n\epsilon && (3.134)\\
&= n[\mathcal{I}(U, V; \check{Y}) - \mathcal{I}(U, V; \check{Z})] + n\epsilon\,, && (3.135)
\end{aligned}
$$

where

- (3.131) holds due to less noisy property given in Definition 3.5;
- (3.132) follows from Markov chain $(K, \check{Y}_1^{i-1}, P) \rightarrow (X_{i+1}^n, S_{i+1}^n) \rightarrow \check{Z}_{i+1}^n$ (The DM-SWC is memoryless) and from less noisy property in Definition 3.5 (The second bracket in (3.132) is always non-negative);
- (3.133) holds because a time-sharing RV τ with a uniform distribution over $\{1, \ldots, n\}$, which is independent of $(K, P, S_1^n, X_1^n, \check{Y}_1^n, \check{Z}_1^n)$, is introduced;
- (3.134) holds due to less noisy property in Definition 3.5, i.e., $\mathcal{I}(\tau; \check{Z}) - \mathcal{I}(\tau; \check{Y}) \leq 0$;
- (3.135) follows from definitions $U \triangleq (U_\tau, \tau)$, $V \triangleq (V_\tau, \tau)$, $\check{Y} \triangleq \check{Y}_\tau$, and $\check{Z} \triangleq \check{Z}_\tau$.

3.3.5 The Proof of Corollary 3.2

The proof directly follows from Theorems 3.1 and 3.2 when $\mathbb{X} = \mathbb{Y} = \mathbb{Z} = \{0\}$ is relaxed.

3.3.6 The Proof of Theorem 3.4

When $C_{P_1} \to \infty$, $\mathbb{O}(C_{P_1})$ defined in Theorem 3.1 contains any arbitrary distribution $\mathcal{P}_{XU|S}(x, u|s)$ without any restriction as $\mathcal{I}(U; S) \leq \mathcal{H}(S) \leq \log(|\mathbb{S}|) < \infty$ always holds. This confirms the direct part from Theorem 3.1. The converse part is also concluded from Theorem 3.2 as the constraint on the maximization is relaxed as $C_{P_1} \to \infty$.

When the DM model is less noisy in Bob's favor, the upper bound (3.106) is simplified as follows:

$$
\begin{aligned}
nR_K &\leq \max_{U \to (X,S) \to (Y,Z)} [\mathcal{I}(U; \check{Y}) - \mathcal{I}(U; \check{Z})] + n\epsilon \\
&= \max_{U \to (X,S) \to (Y,Z)} [\mathcal{I}(U, X, S; \check{Y}) - \mathcal{I}(U, X, S; \check{Z})] - [\mathcal{I}(X, S; \check{Y}|U) - \mathcal{I}(X, S; \check{Z}|U)] + n\epsilon \\
& \hspace{30em} (3.136) \\
&\leq \max_{\mathcal{P}_{X|S}} [\mathcal{I}(X, S; \check{Y}) - \mathcal{I}(X, S; \check{Z})] + n\epsilon \,, \qquad (3.137)
\end{aligned}
$$

where (3.137) holds from Markov chain $U \to (X, S) \to (\check{Y}, \check{Z})$ and from less noisy property stated in Definition 3.5 (The second bracket in (3.136) is always non-negative as a conditional mutual information is expected value of unconditional ones [12].). On the other hand, upper bound (3.137) is achievable according to Theorem 3.1 for $U = (X, S)$ as $C_{P_1} \to \infty$.

3.3.7 The Proof of Corollary 3.3

The proof is concluded directly from Theorems 3.1 and 3.4 based on the following facts:

(a) $P^*_{XU|S}(x, u|s) \in \mathbb{O}(C_{P_1})$ for any $C_{P_1} \geq C^*_P$;

(b) $\mathbb{O}(C_{P_1}) \subseteq \mathbb{O}(\infty)$ for any $C_{P_1} \geq 0$.

Also, the existence of the maximum value is given by Appendix B. Also, $C^*_{_P}$ is finite because

$$C^*_{_P} = [\mathcal{I}(U^*;S) - \mathcal{I}(U^*;\check{Y}^*)]^+ \leq \mathcal{H}(S) \leq \log(|\mathbb{S}|) < \infty \,.$$

3.3.8 The Proof of Corollary 3.4

Let substitute $\check{Y} = (Y,B)$ and $\check{Z} = (Z,E)$ in Corollary 3.3. For any distribution $\mathcal{P}_{XU|S}$, we have $\mathcal{I}(U;S) - \mathcal{I}(U;Y,S) \leq 0$; thus, $C^*_{_P} = 0$, and so $C_{_K}(C_{_{P_1}},0)$ is independent of $C_{_{P_1}}$. Hence, $C_{_K}(C_{_{P_1}},0) = C_{_K}(0,0)$. The proof of the corollary is finalized from [64, Thm. 1] and the fact that $U \to S \to E$ according to (3.10).

Chapter 4

The Gaussian Model

In this chapter, we study the key agreement problem over a Gaussian model. The model consists of a Gaussian wiretap channel with AWGI, which is non-causally known at the transmitter, and a parallel public channel. This model can be considered as an extension of the DM model studied in Chapter 3. In Definition 2.12, a physically degraded Gaussian wiretap channel with AWGI is introduced. In this chapter, we assume a generalized Gaussian wiretap channel with two receivers, Bob and Eve, such that the channel is not necessarily physically degraded in either Bob's or Eve's favor. In fact, noise at Bob's channel and noise at Eve's channel are assumed to be correlated, which is determined by noise covariance matrix.

In this chapter, we carefully examine the possibility of extending the results of Chapter 3 to the Gaussian model. Specifically, we justify the extension of Theorem 3.1 by using generalized Markov lemma [71] for Gaussian RVs (see Remark 2.3). The LB on the forward key capacity is obtained by using a Gaussian auxiliary random variable. However, the existence of the *maximum* value of all achievable key rates, as claimed in Theorem 3.1 and proved in Appendix B, can not be extended to the Gaussian model because the size of the channel state alphabet is not finite. Hence, the *supremum* of all achievable rates will be established in Theorem 4.1 as the forward key capacity.

On the other hand, the UBs given in Theorem 3.2 and Theorem 3.3 do not apply to the Gaussian model because the corresponding proofs rely on the finiteness of the alphabets of the RVs in the DM model. For the same reason, we can not claim that the forward key capacity is a constant (equals $C_K(\infty, 0)$) when the public channel capacity is beyond a finite value based on Corollary 3.3. In fact, our simulations show that the forward key capacity is a strictly increasing function of the public channel capacity in the Gaussian model.

In lemma 4.2, we justify that the Gaussian model is equivalent to a Gaussian model with a physically degraded Gaussian wiretap channel as long as the forward key capacity is concerned. Based on this fact, the forward key capacity is not a function of the correlation coefficient between noise of Bob's channel and that of Eve's channel. Using this equivalence, we prove the UB on the forward key capacity of the corresponding Gaussian model with a physically degraded Gaussian wiretap channel, and then we apply that UB to the forward key capacity of the original Gaussian model. This strategy is similar to obtaining the UB on the capacity of a Gaussian BC [2,6].

On the other hand, we show that this equivalence does not hold when transmissions over the public channel are permitted in the backward direction. Specifically, we establish that the backward key capacity vanishes if the Gaussian wiretap channel is physically degraded in Eve's favor, but the backward key capacity can be positive even if the Gaussian wiretap channel is less noisy in Eve's favor. In this case, we use a strategy similar to Maurer's method [17], which was originally offered for a wiretap model where the main channel, from Alice to Bob, and the wiretap channel, from Alice to Eve, are two independent BSCs. Motivated by this method, we show that the backward key capacity and thus the key capacity are functions of the correlation coefficient between noise of Bob's channel and that of Eve's channel.

After the UB on the forward key capacity is obtained, we establish the optimality of the achievable scheme in Theorem 4.2 for the special case of unlimited capacity of the public channel. We also calculate the forward key capacity for any public channel capacity in low SIR and high SIR regimes.

The rest of this chapter is organized as follows. In Section 4.1, we will present the Gaussian model. In this section, we will technically define an admissible key agreement coding scheme for the model and the objective of the problem. In Section 4.2, we will declare our main results. We will simulate our results for a Gaussian model with given parameters in Section 4.3. Finally, we will establish the proofs of our results in Section 4.4.

4.1 Problem Definitions

We consider a Gaussian model with three parties: a sender (Alice), a legitimate receiver (Bob), and an eavesdropper (Eve). As depicted in Figure 4.1, the model consists of a public channel between Alice and Bob according to Definition 3.1 and a Gaussian state-dependent wiretap channel (G-SWC) defined as follows.

Definition 4.1 (G-SWC)**.** A G-SWC with non-causal SI at the transmitter and no SI at the receivers is determined by $(\Gamma, \Lambda, \sigma_1^2, \sigma_2^2, \varrho) \in \mathbb{R}^{+4} \times [-1, 1]$, where

- random vector $\mathbf{x} \in \mathbb{R}^n$, which is the channel input (from Alice), is subject to average power constraint
$$\frac{1}{n}\mathbf{x}\mathbf{x}^t \leq \Gamma \; ; \tag{4.1}$$
- random vector $\mathbf{s} \in \mathbb{R}^n$, which is known as *interference*, is drawn i.i.d. according to $\mathcal{N}(0, \Lambda)$;
- random matrix $\mathbf{G}_{2\times n} = \begin{bmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \end{bmatrix}$ represents Gaussian noise of the channel, where $\mathbf{g}_1 \in \mathbb{R}^n$ and $\mathbf{g}_2 \in \mathbb{R}^n$ are Gaussian noise of Bob's channel and that of Eve's channel, respectively. The components of $(\mathbf{g}_1, \mathbf{g}_2)$ are drawn i.i.d. according to $(G_1, G_2) \sim \mathcal{N}\left((0,0), \begin{bmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$, where ϱ is called the noise correlation coefficient;
- $\mathbf{y}$ and $\mathbf{z}$ are the first channel output (to Bob), and the second channel output (to Eve), respectively, such that
$$\mathbf{y} = \mathbf{x} + \mathbf{s} + \mathbf{g}_1 \; , \tag{4.2a}$$
$$\mathbf{z} = \mathbf{x} + \mathbf{s} + \mathbf{g}_2 \; ; \tag{4.2b}$$
- the realization of random vector $\mathbf{s}$ is known at Alice prior to each block transmission[1].

Alice and Bob wish to agree on a common secret key by transmission(s) over the Gaussian model in presence of Eve. To do this, Alice and Bob exploit an admissible key agreement code defined as follows.

Definition 4.2. Let $i \in \{1, \ldots, n\}$ be the time index. Recall the characteristics of a public

[1] No SI at Bob and Eve is assumed for simplicity of calculations, i.e., $\mathbf{b} = \mathbf{e} = \mathbf{0}$.

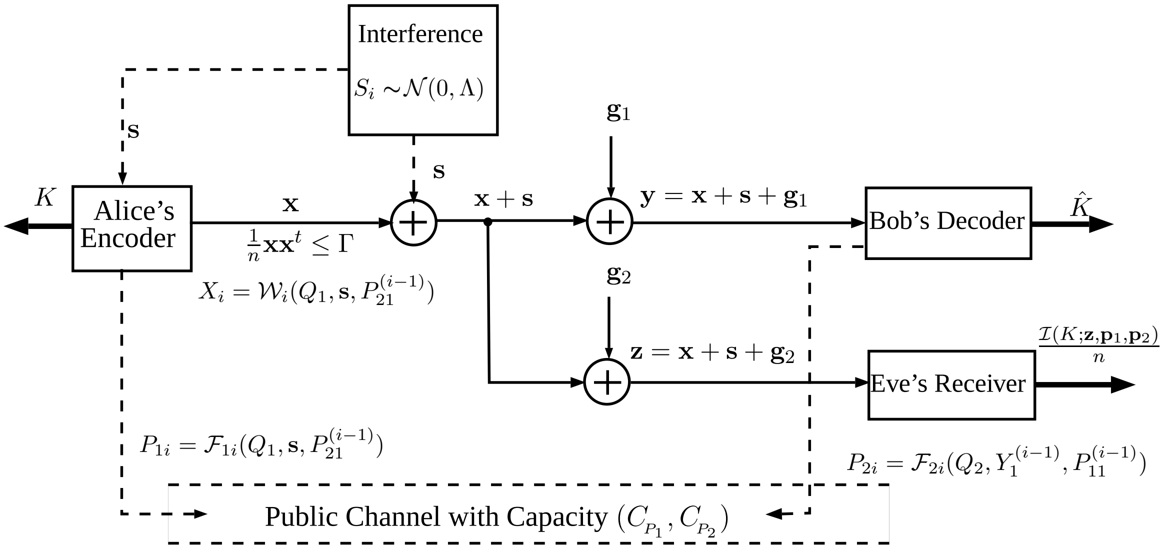

Interference
$S_i \sim \mathcal{N}(0,\Lambda)$
$\mathbf{s}$
$\mathbf{s}$
$\mathbf{g}_1$
K
Alice's Encoder
$\mathbf{x}$
$\frac{1}{n}\mathbf{x}\mathbf{x}^t \leq \Gamma$
$X_i = \mathcal{W}_i(Q_1, \mathbf{s}, P_{21}^{(i-1)})$
$\mathbf{x}+\mathbf{s}$
$\mathbf{y}=\mathbf{x}+\mathbf{s}+\mathbf{g}_1$
Bob's Decoder
$\hat{K}$
$\mathbf{g}_2$
$\mathbf{z}=\mathbf{x}+\mathbf{s}+\mathbf{g}_2$
Eve's Receiver
$\frac{\mathcal{I}(K;\mathbf{z},\mathbf{p}_1,\mathbf{p}_2)}{n}$
$P_{1i} = \mathcal{F}_{1i}(Q_1, \mathbf{s}, P_{21}^{(i-1)})$
$P_{2i} = \mathcal{F}_{2i}(Q_2, Y_1^{(i-1)}, P_{11}^{(i-1)})$
Public Channel with Capacity (C_{P_1}, C_{P_2})

Figure 4.1: Key agreement over the Gaussian model.

channel from Definition 3.1. An admissible key agreement code $(\lceil 2^{nR_K} \rceil, n)$, where $R_K \in \mathbb{R}^+ \cup \{0\}$ and code block length $n \in \mathbb{N}$, for the Gaussian model consists of the following components:

- a *key set* $\mathbb{K} = \{1, \ldots, \lceil 2^{nR_K} \rceil\}$. This set is publicity known to all parties;
- two *randomization* RVs with distributions $\mathcal{P}_{Q_1}$ and $\mathcal{P}_{Q_2}$ over $\mathbb{Q}_1$ and $\mathbb{Q}_2$, respectively, where $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are arbitrarily finite sets. For randomization, Alice and Bob generate RV $Q_1 \in \mathbb{Q}_1$ and RV $Q_2 \in \mathbb{Q}_2$, respectively, such that Q_1, Q_2, and $\mathbf{s}$ are mutually independent. Also, Q_2 is independent of $\mathbf{y}$;
- *public channel encoding functions*

$$\text{(at Alice)} \qquad \mathcal{F}_{1i} : \mathbb{Q}_1 \times \mathbb{R}^n \times (\mathbb{P}_{21} \times \ldots \times \mathbb{P}_{2(i-1)}) \to \mathbb{P}_{1i}\,, \tag{4.3a}$$

$$\text{(at Bob)} \qquad \mathcal{F}_{2i} : \mathbb{Q}_2 \times \mathbb{R}^{(i-1)} \times (\mathbb{P}_{11} \times \ldots \times \mathbb{P}_{1(i-1)}) \to \mathbb{P}_{2i}\,. \tag{4.3b}$$

Alice and Bob transmit forward public message $P_{1i} = \mathcal{F}_{1i}(Q_1, \mathbf{s}, P_{21}^{(i-1)})$ and backward public message $P_{2i} = \mathcal{F}_{2i}(Q_2, Y_1^{(i-1)}, P_{11}^{(i-1)})$ over the public channel at time instant i

subject to *capacity constraints*

$$\text{forward capacity constraint:} \qquad \limsup_{n\to\infty} \frac{1}{n}\log(|\mathbb{P}_1|) \leq C_{P_1} \,, \tag{4.4a}$$

$$\text{backward capacity constraint:} \qquad \limsup_{n\to\infty} \frac{1}{n}\log(|\mathbb{P}_2|) \leq C_{P_2} \,. \tag{4.4b}$$

- *wiretap channel encoding function* $\mathcal{W}_i : \mathbb{Q}_1 \times \mathbb{R}^n \times (\mathbb{P}_{21} \times \ldots \times \mathbb{P}_{2(i-1)}) \to \mathbb{R}$. Alice generates $X_i = \mathcal{W}_i(Q_1, \mathbf{s}, P_{21}^{(i-1)})$ and transmits it at time instant i over the wiretap channel such that average power constraint (4.1) is met;
- *key generator functions*[2]

$$\text{(at Alice):} \qquad \mathcal{K}_1 : \mathbb{Q}_1 \times \mathbb{R}^n \times \mathbb{P}_2 \to \mathbb{K} \,, \tag{4.5a}$$

$$\text{(at Bob):} \qquad \mathcal{K}_2 : \mathbb{Q}_2 \times \mathbb{R}^n \times \mathbb{P}_1 \to \mathbb{K} \,. \tag{4.5b}$$

At the end of all transmissions, Alice and Bob compute $K = \mathcal{K}_1(Q_1, \mathbf{s}, \mathbf{p}_2)$ and $\hat{K} = \mathcal{K}_2(Q_2, \mathbf{y}, \mathbf{p}_1)$, respectively.

As illustrated in the following remark, a physically degraded G-SWC as defined in Definition 2.12 is a special case of the G-SWC given in Definition 4.1.

Remark 4.1. In Definition 4.1, if $\varrho = \frac{\sigma_1}{\sigma_2}$ then $\mathcal{I}(X + S; Z|Y) = 0$ and the G-SWC is a physically degraded in Bob's favor according to Definition 2.12. Similarly, if $\varrho = \frac{\sigma_2}{\sigma_1}$ then $\mathcal{I}(X + S; Y|Z) = 0$ and the G-SWC is a physically degraded in Eve's favor according to Definition 2.12. See equations 4.108 for more details.

The Gaussian model with a physically degraded G-SWC in Bob's favor (respectively, in Eve's favor) is sketched in Figure 4.2 (respectively, in Figure 4.3).

When the public channel is one-way in the forward direction, the functions introduced in Definition 4.2 can be simplified as follows with an abuse of the notions.

Definition 4.3. Assume no public channel is available from Bob to Alice, i.e., $C_{P_2} = 0$. Let

[2] The key generator function at Bob is also called a decoder when $C_{P_2} = 0$.

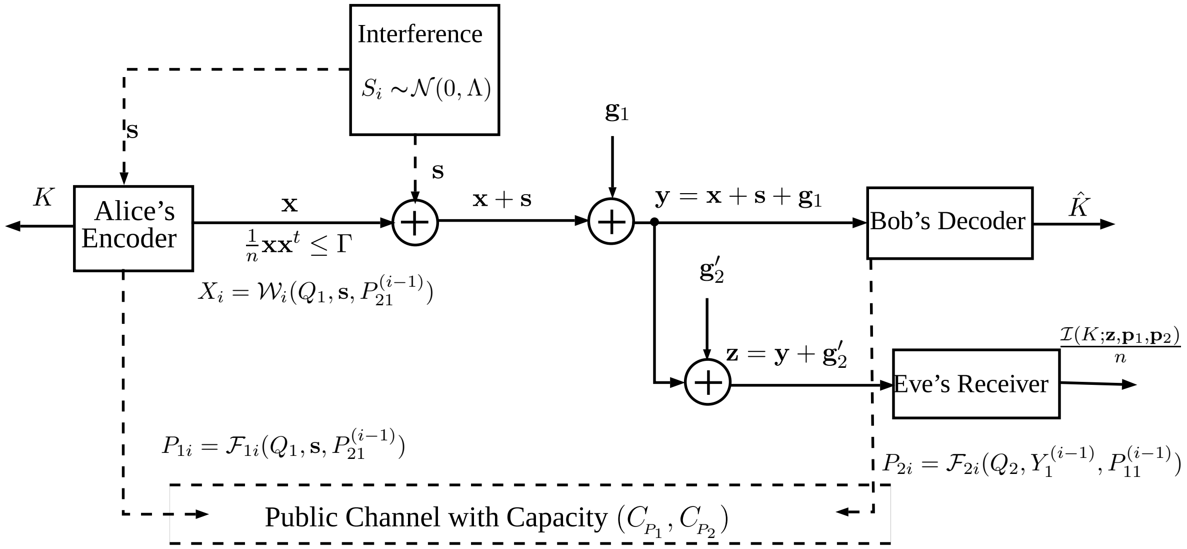

Interference
$S_i \sim \mathcal{N}(0,\Lambda)$
$\mathbf{s}$
$\mathbf{s}$
$\mathbf{g}_1$
K
Alice's Encoder
$\mathbf{x}$
$\frac{1}{n}\mathbf{x}\mathbf{x}^t \leq \Gamma$
$\mathbf{x}+\mathbf{s}$
$\mathbf{y}=\mathbf{x}+\mathbf{s}+\mathbf{g}_1$
Bob's Decoder
$\hat{K}$
$X_i = \mathcal{W}_i(Q_1, \mathbf{s}, P_{21}^{(i-1)})$
$\mathbf{g}_2'$
$\frac{\mathcal{I}(K;\mathbf{z},\mathbf{p}_1,\mathbf{p}_2)}{n}$
$\mathbf{z}=\mathbf{y}+\mathbf{g}_2'$
Eve's Receiver
$P_{1i} = \mathcal{F}_{1i}(Q_1, \mathbf{s}, P_{21}^{(i-1)})$
$P_{2i} = \mathcal{F}_{2i}(Q_2, Y_1^{(i-1)}, P_{11}^{(i-1)})$
Public Channel with Capacity (C_{P_1}, C_{P_2})

Figure 4.2: Key agreement over a physically degraded Gaussian model in Bob's favor.

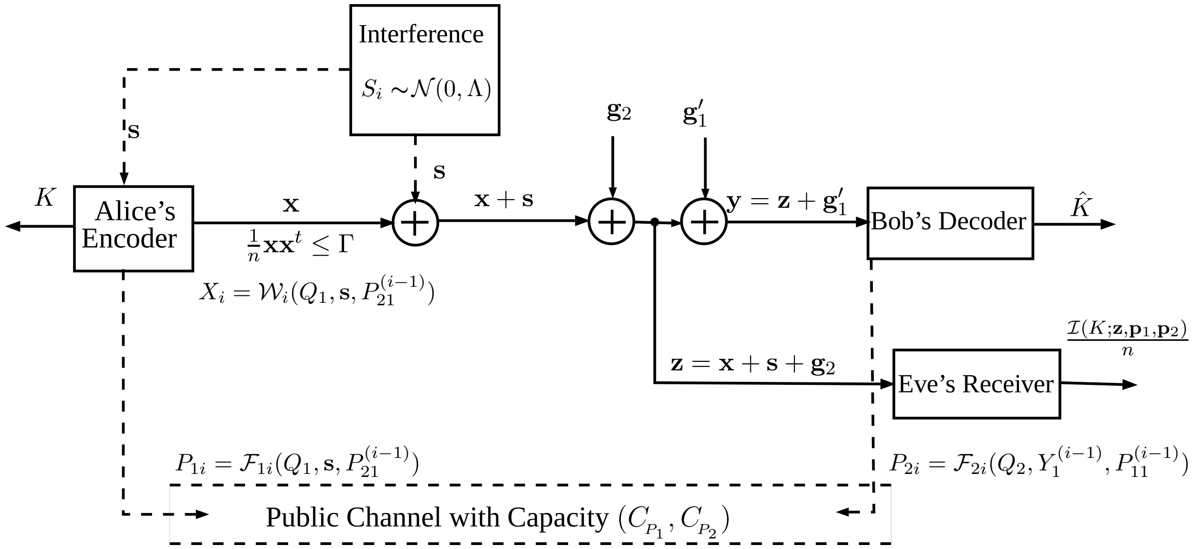

Interference
$S_i \sim \mathcal{N}(0,\Lambda)$
$\mathbf{s}$
$\mathbf{s}$
$\mathbf{g}_2$
$\mathbf{g}_1'$
K
Alice's Encoder
$\mathbf{x}$
$\frac{1}{n}\mathbf{x}\mathbf{x}^t \leq \Gamma$
$\mathbf{x}+\mathbf{s}$
$\mathbf{y}=\mathbf{z}+\mathbf{g}_1'$
Bob's Decoder
$\hat{K}$
$X_i = \mathcal{W}_i(Q_1, \mathbf{s}, P_{21}^{(i-1)})$
$\frac{\mathcal{I}(K;\mathbf{z},\mathbf{p}_1,\mathbf{p}_2)}{n}$
$\mathbf{z}=\mathbf{x}+\mathbf{s}+\mathbf{g}_2$
Eve's Receiver
$P_{1i} = \mathcal{F}_{1i}(Q_1, \mathbf{s}, P_{21}^{(i-1)})$
$P_{2i} = \mathcal{F}_{2i}(Q_2, Y_1^{(i-1)}, P_{11}^{(i-1)})$
Public Channel with Capacity (C_{P_1}, C_{P_2})

Figure 4.3: Key agreement over a physically degraded Gaussian model in Eve's favor.

$Q_2 = 0$ and $\mathbf{p}_2 = \mathbf{0}$. In this case, the key generator functions are specified by

$$\text{(at Alice):} \qquad \mathcal{K}_1 : \mathbb{Q}_1 \times \mathbb{R}^n \rightarrow \mathbb{K} \tag{4.6a}$$

$$\text{(at Bob):} \qquad \mathcal{K}_2 : \mathbb{R}^n \times \mathbb{P}_1 \rightarrow \mathbb{K}\,, \tag{4.6b}$$

where Alice and Bob generate the keys according to $K = \mathcal{K}_1(Q_1, \mathbf{s})$ and $\hat{K} = \mathcal{K}_2(\mathbf{y}, \mathbf{p}_1)$, respectively. In this case, the public channel encoder at Alice is also denoted by function

$$\mathcal{F} : \mathbb{Q}_1 \times \mathbb{R}^n \rightarrow \mathbb{P}_1\,, \tag{4.7}$$

where she sends public message $\mathbf{p}_1 = \mathcal{F}(Q_1, \mathbf{s})$ over the public channel in n transmissions. The wiretap channel encoder is also represented by function

$$\mathcal{W} : \mathbb{Q}_1 \times \mathbb{R}^n \rightarrow \mathbb{R}^n\,, \tag{4.8}$$

where Alice transmits signal $\mathbf{x} = \mathcal{W}(Q_1, \mathbf{s})$ over the wiretap channel in n transmissions.

When $C_{P_2} = 0$, the assumptions $Q_2 = 0$ and $\mathbf{p}_2 = \mathbf{0}$ in Definition 4.3 impose no loss on the forward key capacity due to Lemma 4.1 and Definition 3.1, respectively.

An admissible key agreement code for the Gaussian model is defined in Definition 4.2. For the Gaussian model, the efficiency of an admissible key agreement code is measured by the average probability of error, the leakage rate, and the randomness of $(K, \hat{K})$ as introduced in Definition 1.10 when $\mathbf{b} = \mathbf{e} = \mathbf{0}$ is relaxed and $P \triangleq (\mathbf{p}_1, \mathbf{p}_2)$[3].

For a given public channel capacity pair (C_{P_1}, C_{P_2}), an achievable rate R_K is defined in Definition 1.11, where the admissible key agreement code is introduced in Definition 4.2. The key capacity of the Gaussian model can be also defined as follows.

Definition 4.4. Recall the admissible key agreement code in Definition 4.2. For a given public channel capacity $C_{P_1} \in [0, \infty)$ and $C_{P_2} \in [0, \infty)$, the supremum of all achievable key rates according to Definition 1.11 is called the key capacity of that public channel. The key capacity is denoted by function $C_K(C_{P_1}, C_{P_2})$, where (C_{P_1}, C_{P_2}) is the pair of public channel capacity.

[3] As mentioned in Subsection 1.4.1, P is a one-to-one function of all transmitted signals over the public channel during n-time slot transmissions.

In this chapter, we are interested in $C_K(C_{P_1}, 0)$ according to Definition 4.4. Also, we investigate $C_K(\infty, \infty)$ when Eve's channel is less noisy than Bob's. We also study the effect of ϱ on the key capacity.

4.2 Statement of Main Results

In this section, we focus on a Gaussian model with a G-SWC $(\Gamma, \Lambda, \sigma_1^2, \sigma_2^2, \varrho)$ as defined in Section 4.1. This section consists of two subsections: in Subsection 4.2.1, we investigate the forward key capacity of the model; and in Subsection 4.2.2, we study effects of the public channel feedback on the key capacity when Eve's channel is less noisy than Bob's.

4.2.1 Forward Public Channel

In this subsection, we assume the public channel is one-way in the forward direction, i.e., $C_{P_2} = 0$. We are interested in bounds on $C_K(C_{P_1}, 0)$.

The following lemma allows to relax randomization $Q_2 = 0$ at Bob when $C_{P_2} = 0$.

Lemma 4.1. *Randomization at Bob does not increase the forward key capacity of the Gaussian model.*

When the public channel is in the forward direction, the following lemma justifies that the key capacity of a G-SWC as defined in Definition 4.1 equals to the key capacity of its corresponding physically degraded G-SWC.

Lemma 4.2. *Recall Definition 2.12. The forward key capacity of the Gaussian model equals the forward key capacity of the Gaussian model with a physically degraded G-SWC* $(\Gamma, \Lambda, \sigma_1^2, \sigma_2^2)$ *in either Bob's favor (if* $\sigma_2^2 > \sigma_1^2$*), as sketched in Figure 4.2, or Eve's favor (if* $\sigma_1 \geq \sigma_2$*), as sketched in Figure 4.3.*

In the following theorems, we establish the LB and the UB on the forward key capacity of the Gaussian model. To do this, we first prove the bounds on the forward key capacity of the equivalent physically degraded model and then we apply the results to the actual Gaussian model according to Lemma 4.2.

Theorem 4.1 (Lower bound on the forward key capacity)**.** *Let $\sigma_2^2 > \sigma_1^2$. Assume the Gaussian model with public channel capacity $C_{P_1} \in [0, \infty)$. Define the set*

$$\mathbb{O}(C_{P_1}) \triangleq \left\{ \begin{array}{ll} (\gamma, \rho): & \gamma \in [0,1], \rho \in (-1,1), \\ & C_{P_1} \geq \frac{1}{2} \log \left(\frac{(1-\rho^2)(1-\gamma)^2 \Gamma\Lambda + \sigma_1^2 (\Gamma + \gamma^2 \Lambda + 2\rho\, \gamma \sqrt{\Gamma\Lambda}\,)}{(1-\rho^2)\Gamma(\Gamma + 2\rho\sqrt{\Gamma\Lambda} + \Lambda + \sigma_1^2)} \right) \end{array} \right\} . \tag{4.9}$$

Then, the key rate

$$\mathcal{R}_K(\gamma, \rho, C_{P_1}) = \frac{1}{2} \log \left(\frac{[(1-\rho^2)(1-\gamma)^2 \Gamma\Lambda + \sigma_2^2 (\Gamma + \gamma^2 \Lambda + 2\rho\, \gamma \sqrt{\Gamma\Lambda}\,)](\Gamma + 2\rho\sqrt{\Gamma\Lambda} + \Lambda + \sigma_1^2)}{[(1-\rho^2)(1-\gamma)^2 \Gamma\Lambda + \sigma_1^2 (\Gamma + \gamma^2 \Lambda + 2\rho\, \gamma \sqrt{\Gamma\Lambda}\,)](\Gamma + 2\rho\sqrt{\Gamma\Lambda} + \Lambda + \sigma_2^2)} \right) \tag{4.10}$$

is an achievable key rate of the Gaussian model for any $(\gamma, \rho) \in \mathbb{O}(C_{P_1})$. Also, let

$$R_K(C_{P_1}, 0) \triangleq \sup_{(\gamma,\rho) \in\, \mathbb{O}(C_{P_1})} \mathcal{R}_K(\gamma, \rho, C_{P_1}) \tag{4.11}$$

then, the forward key capacity of the model is lower bounded by

$$C_K(C_{P_1}, 0) \geq R_K(C_{P_1}, 0)\,. \tag{4.12}$$

As a special case, an achievable key rate on the forward key capacity of the Gaussian model in low SIR regime can be calculated from Theorem 4.1. Using Lemma 4.2 together with Theorem 1.1, which is originally proved by Watanabe and Oohama [40], the optimality of the achievable key agreement code is illustrated. Hence, the forward key capacity in low SIR is given by the following corollary.

Corollary 4.1. *Assume σ_1^2, σ_2^2, and Λ in the Gaussian model are fixed, where $\sigma_2^2 > \sigma_1^2$. In low SIR regime, i.e., $\frac{\Gamma}{\Lambda} \to 0$, the forward key capacity of the Gaussian model is given by*

$$\lim_{\Gamma \to 0} C_K(C_{P_1}, 0) = \frac{1}{2} \log \left((1 - 2^{2C_K(\infty,0)}) 2^{-2C_{P_1}} + 2^{2C_K(\infty,0)} \right) , \tag{4.13}$$

where $C_K(\infty, 0) = \frac{1}{2} \log \left(\frac{\sigma_2^2(\Lambda+\sigma_1^2)}{\sigma_1^2(\Lambda+\sigma_2^2)} \right)$ is the forward key capacity of the Gaussian model when $C_{P_1} \to \infty$ and $\Gamma \to 0$.

When the public channel has unlimited capacity, the following important theorem proves that $C_K(\infty, 0)$ is optimally achievable. Further, $C_K(\infty, 0)$ gives a UB on the forward key capacity, which is a non-decreasing function of C_{P_1}.

Theorem 4.2 (Key capacity for unlimited public channel capacity)**.** *Let $\sigma_2^2 > \sigma_1^2$. If $C_{P_1} \to \infty$, then the key capacity of the model is*

$$C_K(\infty, 0) = \frac{1}{2} \log \left(\frac{(\Gamma + \Lambda + 2\sqrt{\Gamma\Lambda} + \sigma_1^2)\sigma_2^2}{(\Gamma + \Lambda + 2\sqrt{\Gamma\Lambda} + \sigma_2^2)\sigma_1^2} \right) . \tag{4.14}$$

Moreover, for any $C_{P_1} \geq 0$, $C_K(C_{P_1}, 0)$ is upper bounded by

$$C_K(\infty, 0) \geq C_K(C_{P_1}, 0) . \tag{4.15}$$

When (power) signal-to-interference ratio (SIR) is high enough, $C_K(\infty, 0)$ is asymptotically achievable by the following corollary.

Corollary 4.2. *Let $\sigma_2^2 > \sigma_1^2$. In high SIR regime, i.e., $\frac{\Gamma}{\Lambda} \to \infty$, the forward key capacity of the Gaussian model is*

$$\lim_{\frac{\Gamma}{\Lambda} \to \infty} C_K(C_{P_1}, 0) = \frac{1}{2} \log \left(\frac{\sigma_2^2}{\sigma_1^2} \right) \tag{4.16}$$

for any $C_{P_1} \in [0, \infty)$.

A natural question regarding forward key capacity of the Gaussian model is that if Bob's channel is required to be less noisy than Eve's channel such that the key generation is possible. The following corollary answers this question.

Corollary 4.3. *The forward key capacity is non-zero iff Bob's channel is less noisy than Eve's channel, i.e.,*

$$\sigma_2^2 > \sigma_1^2 \quad \Leftrightarrow \quad C_K(C_{P_1}, 0) > 0 . \tag{4.17}$$

4.2.2 Two-Way Public Channel

In this subsection, we assume the public channel is two-way, but Eve's channel is less noisy than Bob's channel, i.e. $\sigma_1^2 \geq \sigma_2^2$. We are interested in bounds on $C_K(\infty, \infty)$.

Although condition $\sigma_2^2 > \sigma_1^2$ is required to have a positive forward key capacity according to Corollary 4.3, this condition might be unnecessary for key generation when the public channel is two-way, i.e., $C_{P_2} > 0$. In the following theorem, we focus on the key capacity of a Gaussian model when Eve's channel is less noisy than Bob's channel. This theorem proves that Lemma 4.1 can *not* be extended to general case when a feedback from Bob to Alice exists, i.e., $C_{P_2} > 0$.

Theorem 4.3. *Let Eve's channel be less noisy than Bob's channel, i.e., $\sigma_1^2 \geq \sigma_2^2$, in the Gaussian model.*

(a) If the G-SWC is a physically degraded Gaussian wiretap channel in Eve's favor as sketched in Figure 4.3, i.e., $\varrho = \frac{\sigma_2}{\sigma_1}$, then

$$C_K(\infty, \infty) = 0\,.$$

(b) If correlation coefficient of noise satisfies $\frac{\sigma_2}{2\sigma_1} \geq \varrho$, then

$$C_K(\infty, \infty) \geq \frac{1}{2} \log \left(1 + \frac{\sigma_2^2 - 2\rho\sigma_1\sigma_2}{\sigma_1^2} \right)\,.$$

Part (a) of Theorem 4.3 is an extension of the key capacity of a degraded DM model, as given in (1.26) by Ahlswede and Csiszár [18], to the Gaussian model with a physically degraded G-SWC in Eve's favor. Part (b) of Theorem 4.3 is an extension of Maurer's scheme [17] to the Gaussian model with the backward public channel.

Comparing Theorem 4.2 with Theorem 4.3, we understand that the key capacity is a function of the correlation coefficient of noise (ϱ) when the backward public channel exists. However, the forward public channel is independent of ϱ. The characterization of the key capacity for any $\varrho \in [-1, 1]$ remains as an open problem for future work.

4.3 Simulations

In this section, we illustrate simulation results for a Gaussian model with a forward public channel. Assume a Gaussian model with parameters

$$\begin{cases} \frac{\sigma_1^2}{\Lambda} = 0.1 \ , \\ \frac{\sigma_2^2}{\Lambda} = 0.4 \ . \end{cases} \tag{4.18}$$

In our simulations σ_1^2, σ_2^2, Λ are fixed, and the desired plots are sketched as a function of SIR, i.e., $(\frac{\Gamma}{\Lambda})_{dB}$. In the following, we consider the concepts behind each figure:

- Figure 4.4: Recalling Theorem 4.1, the LB on $C_K(C_{P_1}, 0)$ is simulated in this figure for 5 different values of C_{P_1}. The UB on the key capacity is also sketched in this figure according to Theorem 4.2. This UB is the maximum achievable key capacity over the Gaussian model with any $C_{P_1} \geq 0$. for a given G-SWC. According to this simulation, inequality

  $$R_K(0,0) < R_K(.25) < R_K(.5) < R_K(1) < R_K(2) < C_K(\infty, 0) \tag{4.19}$$

  holds for any SIR. When $\frac{\Gamma}{\Lambda} = 1$, we have

  $$\begin{cases} R_K(0,0) = .9478 \ , \\ C_K(\infty, 0) = .9491 \ . \end{cases} \tag{4.20}$$

  Hence, we observe that $R_K(0,0)$ is .137% less than $C_K(\infty, 0)$, which is the maximum achievable key capacity for the given G-SWC. This difference is even less for other values of C_{P_1} according to (4.19). Moreover, $|R_K(0,0) - C_K(\infty, 0)| \to 0$ as $\frac{\Gamma}{\Lambda} \to \infty$. In other words, the LB on $C_K(C_{P_1}, 0)$ for any $C_{P_1} \geq 0$ asymptotically achieves

  $$\lim_{\frac{\Gamma}{\Lambda} \to \infty} C_K(C_{P_1}, 0) = \frac{1}{2} \log \left(\frac{\sigma_2^2}{\sigma_1^2} \right) = 1 \, bit/trans. \, .$$

  This fact was established before in Corollary 4.2. Hence, the public channel has negligible contribution in key generation in high SIR regime.

- Figure 4.5 and Figure 4.6: Assume (γ^*, ρ^*) is the optimum[4] value of the pair (γ, ρ) in the sense that the supremum of (4.11) is obtained. Figure 4.5 and Figure 4.6 exhibit γ^* and ρ^* versus SIR, respectively, for 6 values of C_{P_1}. According to these figures, both γ^* and ρ^* are increasing functions of C_{P_1} for any SIR. Specially,

$$\forall\, (\frac{\Gamma}{\Lambda}) \in \mathbb{R}^+ : \quad \lim_{C_{P_1} \to \infty} (\gamma^*, \rho^*) = (1, 1)\ .$$

  Moreover, both γ^* and ρ^* are increasing functions of SIR $(\frac{\Gamma}{\Lambda})$ for any $C_{P_1} \geq 0$. Specially

$$\forall\, C_{P_1} \geq 0 : \quad \lim_{\frac{\Gamma}{\Lambda} \to \infty} (\gamma^*, \rho^*) = (1, 1).$$

- Figure 4.7: Assume the Gaussian model with $C_{P_1} = 0$ (the G-SWC alone). In this figure, we compare the (ordinary) capacity of the main channel [55], the known LB on the secrecy capacity [60], and the LB on the key capacity given in Theorem 4.1. Obviously, the secrecy capacity is upper bounded by the main channel capacity. However, as it is illustrated in this figure, $R_K(0,0)$ exceeds the ordinary capacity of the G-SWC when $(\frac{\Gamma}{\Lambda})_{dB} < -8.5$ dB. In this region, this means that

$$C_S \leq C_m < R_K(0,0) \leq C_K(0)\ ,$$

  where C_S and C_m are the secrecy capacity and the main channel capacity of the G-SWC, respectively. In other words, the key capacity in low SIR regime is generally greater than both the main channel capacity and the secrecy capacity as it can be generated with assistance of the interference.

- Figure 4.8: According to the achievable scheme stated in Theorem 4.1, the Gaussian model demands an unlimited public channel capacity to achieve $C_K(\infty, 0)$, which is the maximum achievable key capacity for a given G-SWC. This fact was shown in (4.19) as well. This is in contrast with the DM model in which $C_K(\infty, 0)$ is achievable by using a finite public channel capacity. However, with a finite public channel capacity in the Gaussian model,

[4] In this section, we seek the optimum values based on simulations which is subject to computational approximations.

the achieved key rate can be significantly close to $C_K(\infty, 0)$. Specifically, let ζ (in percent) be the relative difference between $C_K(\infty, 0)$ and $R_K(C_{P_1}, 0)$, which is the achieved LB on $C_K(C_{P_1}, 0)$ according to Theorem 4.1. In Figure 4.8, we investigate the minimum public channel capacity $C^*_{P_1}$ such that

$$\left| \frac{R_K(C^*_{P_1}, 0) - C_K(\infty, 0)}{C_K(\infty, 0)} \right| \times 100 < \zeta \, . \tag{4.21}$$

According to this figure, the minimum required $C^*_{P_1}$ versus SIR is sketched for three values of ζ. Based on this figure, we show that there exists a *finite* public channel capacity $C^*_{P_1} < \infty$ for the Gaussian model such that $R_K(C^*_{P_1}, 0)$ *approximately* equals $C_K(\infty, 0)$ in the sense of (4.21). In other words, for any $C_{P_1} \geq C^*_{P_1}$, $R_K(C_{P_1}, 0)$ achieves $C_K(\infty, 0)$, which is a UB on $C_K(C_{P_1}, 0)$, in approximate sense

$$\left| \frac{R_K(C_{P_1}, 0) - C_K(\infty, 0)}{C_K(\infty, 0)} \right| \times 100 < \zeta \, , \tag{4.22}$$

which is valid due to $R_K(C^*_{P_1}, 0) \leq R_K(C_{P_1}, 0) \leq C_K(C_{P_1}, 0) \leq C_K(\infty, 0)$.

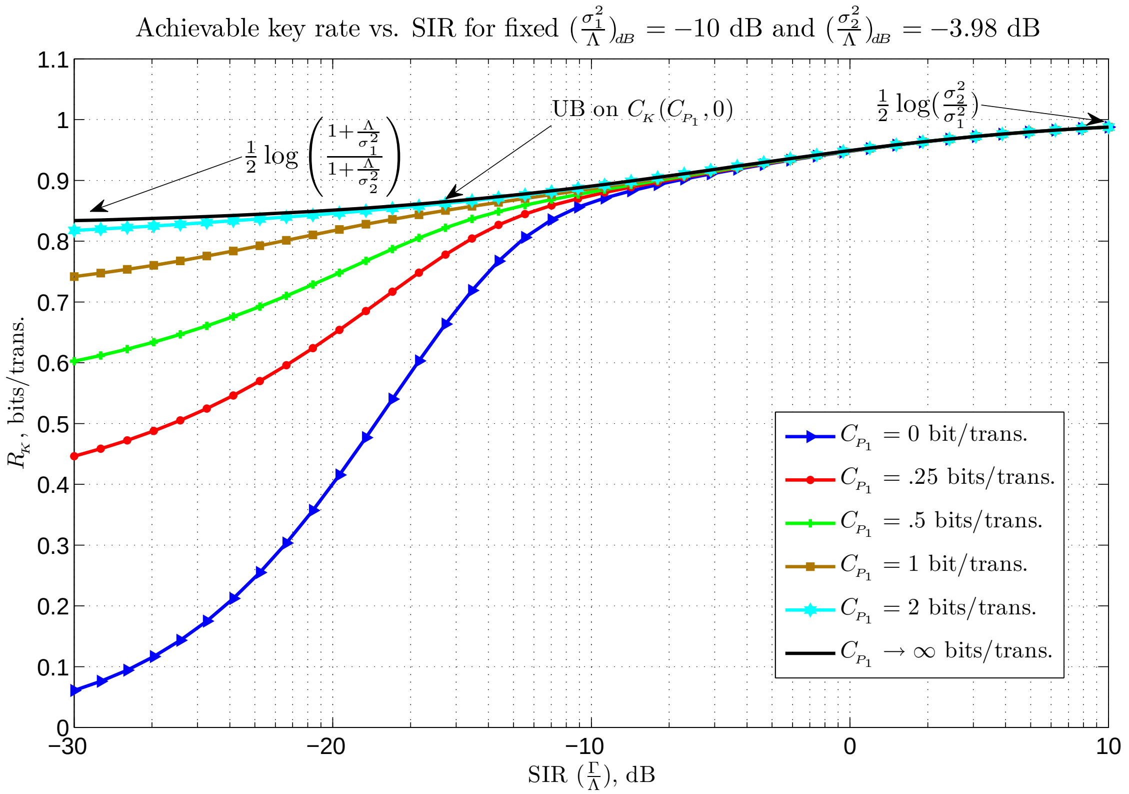


Figure 4.4: The LB on key capacity of the Gaussian model with capacity C_{P_1}.

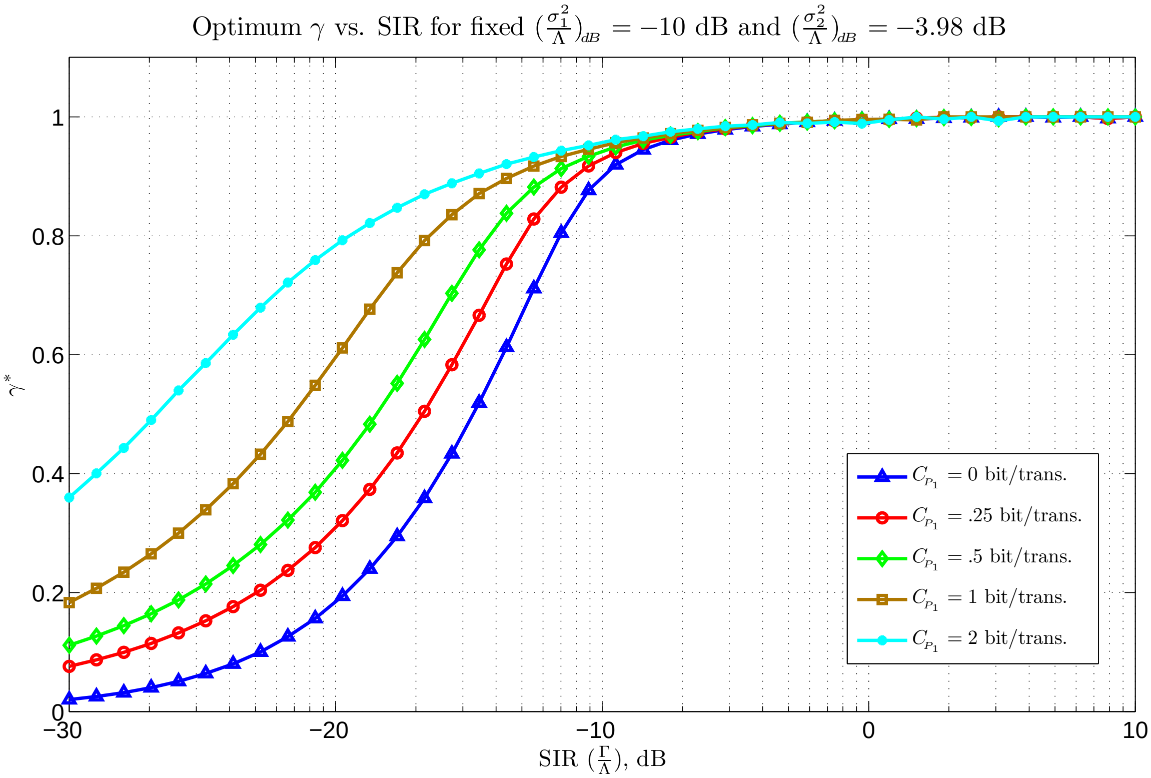


Figure 4.5: Coefficient γ^* for the Gaussian model with capacity C_{P_1}.

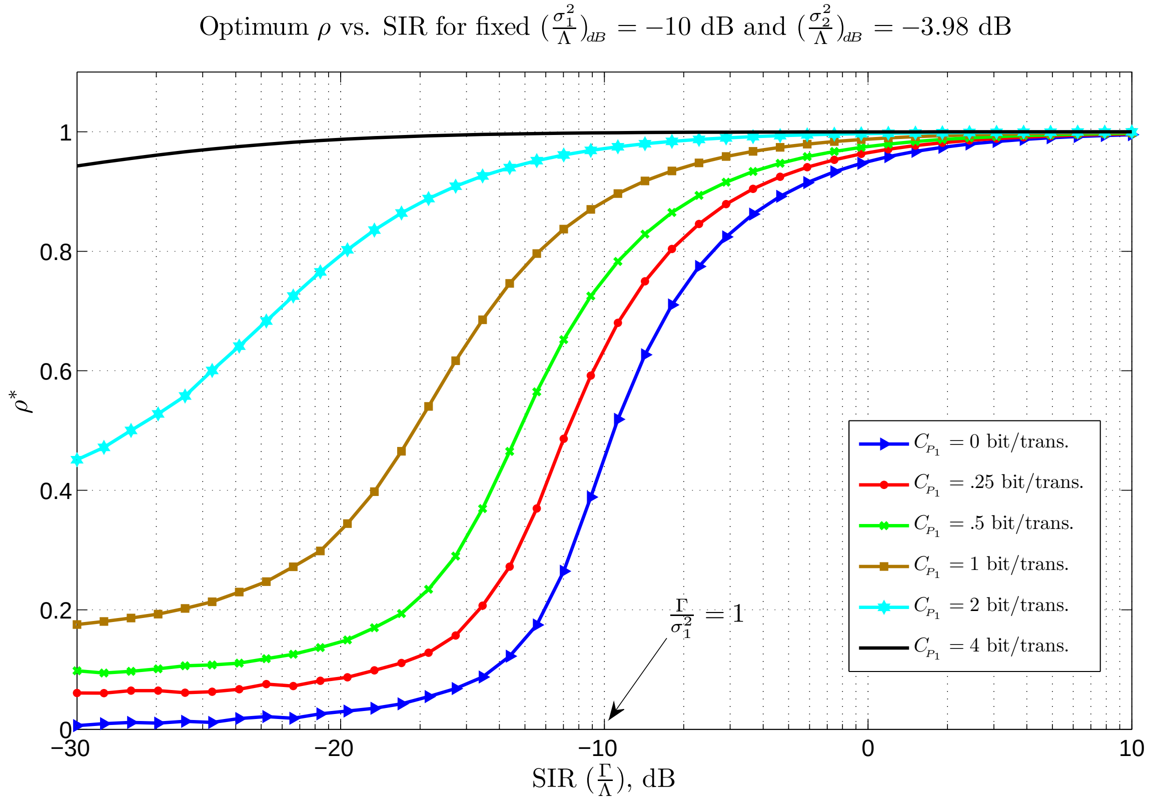


Figure 4.6: Correlation coefficient ρ^* for the Gaussian model with capacity C_{P_1}.

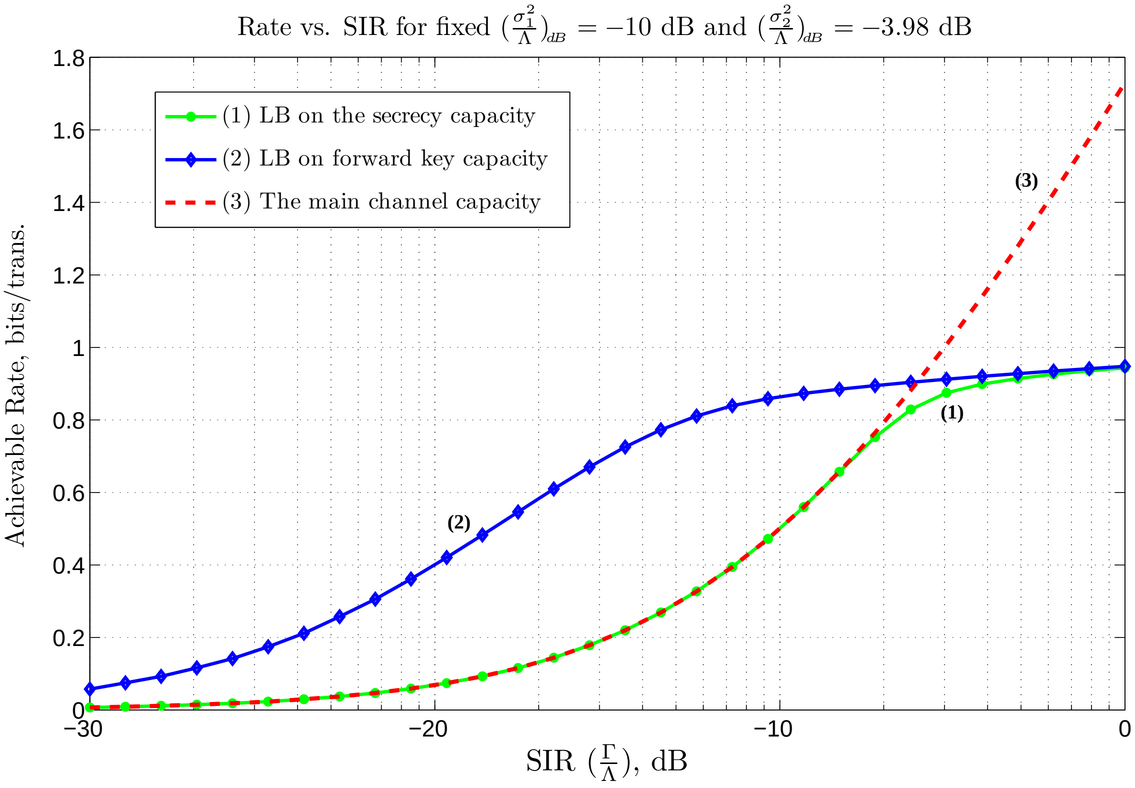


Figure 4.7: Comparison of key capacity, secrecy capacity and capacity of the G-SWC.

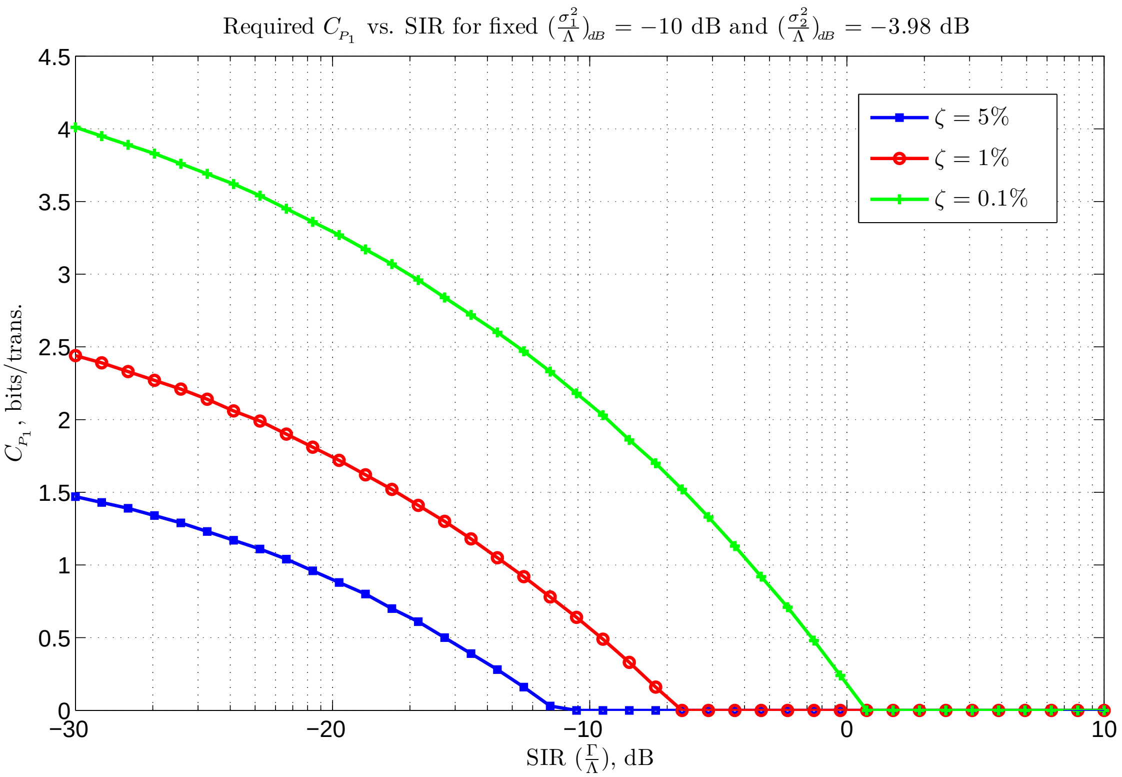


Figure 4.8: The minimum required C_{P_1} to achieve $C_K(\infty, 0)$ within tolerance ζ.

4.4 The Proofs

The detailed proofs of the results given in Section 4.2 are established in this section as follows.

4.4.1 The Proof of Lemma 4.1

Let $C_{P_2} = 0$ due to the assumption of this lemma. Having $(\mathbf{z}, \mathbf{e}, P)$, the security and randomness of K does not depend on $\hat{K}$ according to Definition 1.10. Also, K does not depend on Q_2. Hence, using a randomization at Bob has no effect on the security and randomness of K.

On the other hand, $K = \mathcal{K}_1(Q_1, \mathbf{s})$, $\mathbf{x} = \mathcal{W}(Q_1, \mathbf{s})$, and $\mathbf{p}_1 = \mathcal{F}(Q_1, \mathbf{s})$. Hence, Markov chain $K \to (Q_1, \mathbf{s}) \to (\mathbf{p}_1, \mathbf{y}) \to (Q_2, \mathbf{p}_1, \mathbf{y}) \to \hat{K}$ holds as Q_2 is independent of $(Q_1, \mathbf{s}, \mathbf{p}_1, \mathbf{y})$ according to Definition 4.2. Hence, $(\mathbf{p}_1, \mathbf{y})$ is a *sufficient statistic* [2, Sec. 2.9] of $(Q_2, \mathbf{p}_1, \mathbf{y})$ for $\hat{K}$. Hence, we conclude that randomization at Bob (Q_2) does not decrease the average probability of error $\mathscr{P}\{\hat{K} \neq K\}$.

As a result, the randomization at Bob does not enhance the reliability, the security level, and the randomness of $(K, \hat{K})$ as defined in Definition 1.10. Hence, $Q_2 = 0$ is relaxed when forward key capacity is investigated.

4.4.2 The Proof of Lemma 4.2

Recalling Definition (4.3), let first justify the following lemma.

Lemma 4.3. *$\mathcal{P}(K, \hat{K})$ and $\mathcal{P}(K, \mathbf{z}, \mathbf{p}_1)$ are uniquely determined by probability density functions*[5] *$\mathcal{P}(\mathbf{y}|\mathbf{x}+\mathbf{s})$, $\mathcal{P}(\mathbf{z}|\mathbf{x}+\mathbf{s})$, $\mathcal{P}(\mathbf{s})$, and probability mass function $\mathcal{P}(Q_1)$.*

[5] With an abuse of the notations in this section, $\mathcal{P}$ is also used as a probability density function for continuous RVs.

Proof. $\mathcal{P}(K,\hat{K})$ is given by

$$\begin{aligned}\mathcal{P}(K,\hat{K}) &= \int_{\mathbf{s}\in\,\mathbb{R}^n}\int_{\mathbf{y}\in\,\mathbb{R}^n}\sum_{Q_1\in\,\mathbb{Q}_1}\mathcal{P}(K,\hat{K},Q_1,\mathbf{s},\mathbf{y})\,d\mathbf{y}\,d\mathbf{s}\\ &= \int_{\mathbf{s}\in\,\mathbb{R}^n}\int_{\mathbf{y}\in\,\mathbb{R}^n}\sum_{Q_1\in\,\mathbb{Q}_1}\mathcal{P}(\hat{K}|Q_1,\mathbf{s},K,\mathbf{y})\mathcal{P}(\mathbf{y}|Q_1,\mathbf{s},K)\mathcal{P}(K|Q_1,\mathbf{s})\mathcal{P}(Q_1)\mathcal{P}(\mathbf{s})\,d\mathbf{y}\,d\mathbf{s} \qquad (4.23)\\ &= \int_{\mathbf{s}\in\,\mathbb{R}^n}\int_{\mathbf{y}\in\,\mathbb{R}^n}\sum_{\substack{Q_1\in\,\mathbb{Q}_1\\ K=\mathcal{K}_1(Q_1,\mathbf{s})\\ \mathbf{x}=\mathcal{W}(Q_1,\mathbf{s})}}\mathcal{P}(\hat{K}|Q_1,\mathbf{s},K,\mathbf{x},\mathbf{p}_1,\mathbf{y})\mathcal{P}(\mathbf{y}|Q_1,\mathbf{s},\mathbf{x},K)\mathcal{P}(Q_1)\mathcal{P}(\mathbf{s})\,d\mathbf{y}\,d\mathbf{s} \qquad (4.24)\\ &= \int_{\mathbf{s}\in\,\mathbb{R}^n}\int_{\mathbf{y}\in\,\mathbb{R}^n}\sum_{\substack{Q_1\in\,\mathbb{Q}_1\\ K=\mathcal{K}_1(Q_1,\mathbf{s})\\ \mathbf{x}=\mathcal{W}(Q_1,\mathbf{s})\\ \hat{K}=\mathcal{K}_2(\mathbf{y},\mathcal{F}(Q_1,\mathbf{s}))}}\mathcal{P}(\mathbf{y}|\mathbf{x}+\mathbf{s})\mathcal{P}(Q_1)\mathcal{P}(\mathbf{s})\,d\mathbf{y}\,d\mathbf{s} \qquad (4.25)\end{aligned}$$

where

- (4.23) is valid because Q_1 and $\mathbf{s}$ are independent;
- (4.24) holds as $\mathbf{x} = \mathcal{W}(Q_1,\mathbf{s})$ and $\mathbf{p}_1 = \mathcal{F}(Q_1,\mathbf{s})$ are deterministic functions of $(Q_1,\mathbf{s})$;
- (4.25) follows from the fact that $\hat{K} = \mathcal{K}_2(\mathbf{y},\mathbf{p}_1) = \mathcal{K}_2(\mathbf{y},\mathcal{F}(Q_1,\mathbf{s}))$ and Markov chain $(K,Q_1,\mathbf{s},\mathbf{x}) \to (\mathbf{x}+\mathbf{s}) \to \mathbf{y}$.

Further, $\mathcal{P}(K,\mathbf{z},\mathbf{p}_1)$ is given by

$$\begin{aligned}\mathcal{P}(K,\mathbf{z},\mathbf{p}_1) &= \int_{\mathbf{s}\in\,\mathbb{R}^n}\sum_{Q_1\in\,\mathbb{Q}_1}\mathcal{P}(K,\mathbf{z},\mathbf{p}_1,\mathbb{Q}_1,\mathbf{s})\,d\mathbf{s}\\ &= \int_{\mathbf{s}\in\,\mathbb{R}^n}\sum_{Q_1\in\mathbb{Q}_1}\mathcal{P}(\mathbf{z}|K,Q_1,\mathbf{s},\mathbf{p}_1)\mathcal{P}(K|Q_1,\mathbf{s},\mathbf{p}_1)\mathcal{P}(\mathbf{p}_1|Q_1,\mathbf{s})\mathcal{P}(Q_1)\mathcal{P}(\mathbf{s})\,d\mathbf{s} \qquad (4.26)\\ &= \int_{\mathbf{s}\in\,\mathbb{R}^n}\sum_{\substack{Q_1\in\,\mathbb{Q}_1\\ \mathbf{x}=\mathcal{W}(Q_1,\mathbf{s})}}\mathcal{P}(\mathbf{z}|K,Q_1,\mathbf{s},\mathbf{x},\mathbf{p}_1)\mathcal{P}(K|Q_1,\mathbf{s})\mathcal{P}(\mathbf{p}_1|Q_1,\mathbf{s})\mathcal{P}(Q_1)\mathcal{P}(\mathbf{s})\,d\mathbf{s} \qquad (4.27)\end{aligned}$$

$$= \int_{\mathbf{s}\in\, \mathbb{R}^n} \sum_{\substack{Q_1 \in\, \mathbb{Q}_1 \\ K=\mathcal{K}_1(Q_1,\mathbf{s}) \\ \mathbf{p}_1=\mathcal{F}(Q_1,\mathbf{s}) \\ \mathbf{x}=\mathcal{W}(Q_1,\mathbf{s})}} \mathcal{P}(\mathbf{z}|\mathbf{x}+\mathbf{s})\mathcal{P}(Q_1)\mathcal{P}(\mathbf{s})\, d\mathbf{s} \tag{4.28}$$

where

- (4.26) holds because Q_1 and $\mathbf{s}$ are independent;
- (4.27) follows from Markov chain $\mathbf{p}_1 \to (Q_1, \mathbf{s}) \to K$ and from the fact that $\mathbf{x}$ is a deterministic function of $(Q_1, \mathbf{s})$;
- (4.28) holds due to Markov chain $(K, Q_1, \mathbf{s}, \mathbf{x}, \mathbf{p}_1) \to (\mathbf{x}+\mathbf{s}) \to \mathbf{z}$.

Finally, equations (4.25) and (4.28) establish the lemma. □

According to Definition 1.10, the reliability of the pair $(K, \hat{K})$ is determined by $\mathcal{P}(K, \hat{K})$, and its security level and randomness are uniquely characterized by $\mathcal{P}(K, \mathbf{z}, \mathbf{p}_1)$.

According to Lemma 4.3, the reliability, security, and randomness of $(K, \hat{K})$, as defined in Definition 1.10, depend on joint distribution $\mathcal{P}(\mathbf{p}_1, \mathbf{y}, \mathbf{z}, \mathbf{x}, \mathbf{s}, Q_1)$ but only through the conditional marginal distributions $\mathcal{P}(\mathbf{y}|\mathbf{x}+\mathbf{s})$, $\mathcal{P}(\mathbf{z}|\mathbf{x}+\mathbf{s})$, $\mathcal{P}(\mathbf{s})$, and $\mathcal{P}(Q_1)$. Hence, we conclude the following statement.

First, let $\sigma_2 > \sigma_1$. Based on Lemma 4.3, if $\mathbf{z}$ is replaced by $\mathbf{z}' = \mathbf{y} + \mathbf{g}_2'$, where $\mathbf{g}_2'$ is drawn i.i.d. according to $\mathcal{N}(0, \sigma_2^2 - \sigma_1^2)$ independent of $\mathbf{g}_1$, the efficiency of an admissible key agreement code does not change. Second, let $\sigma_1 \geq \sigma_2$. Based on Lemma 4.3, similarly, if $\mathbf{y}$ is replaced by $\mathbf{y}' = \mathbf{z} + \mathbf{g}_1'$, where $\mathbf{g}_1'$ is drawn i.i.d. according to $\mathcal{N}(0, \sigma_1^2 - \sigma_2^2)$ independent of $\mathbf{g}_2$, the efficiency of an admissible key agreement code does not change. As a result, the forward key capacity of a Gaussian model equals that of its corresponding Gaussian model with a physically degraded wiretap channel.

4.4.3 The Proof of Theorem 4.1

We first examine if we can extend the proof of Theorem 3.1, which is given in Subsection 3.3.1, to that of Theorem 4.1 (subject to possible modifications).

The mutual information function can be defined for continuous RVs by using the concept of quantization with asymptotically small error (see [77, Lem. 5.5.1], [2, Ch. 8], and [6, Page 23] for more details). Assume that $(X, Y) \in \mathbb{R}^2$ is a pair of continuous RVs which is drawn according to probability density function $\mathcal{P}(x, y)$. Then, the mutual information function between $X \sim \mathcal{P}(x)$ and $Y \sim \mathcal{P}(y)$ is defined as

$$\mathcal{I}(X; Y) \triangleq \int_{y \in \mathbb{R}} \int_{x \in \mathbb{R}} \mathcal{P}(x, y) \log \left(\frac{\mathcal{P}(x, y)}{\mathcal{P}(x)\mathcal{P}(y)} \right) dx\, dy\,. \tag{4.29}$$

Recall the definitions of typicality in Subsection 2.1.1. According to Remark 2.2, the strong typicality used in the proof of Theorem 3.1 can *not* be defined for continuous RVs. However, the definition of weak (entropy) typicality can be applied to both discrete and continuous RVs according to Remark 2.1.

On the other hand, equations (3.37) and (3.51) still hold if the statements of strong typicality are replaced by those of weak typicality. However, the validity of Lemma 3.1 and Lemma 3.2 relies on the the validity of Markov lemma, which is not generally true for weak typicality. However, according to Remark 2.3, the Markov lemma can be generalized for Gaussian input distributions based on weak typicality [70].

With applying weak typicality and generalized Markov lemma [71], the proof of Lemma 3.1 and that of Lemma 3.2 can be extended to the Gaussian model subject to finding a Gaussian input distribution satisfying average power constraint (4.1). Once these lemmas are established, the AR, AS, and ARN conditions will be approved by following the lines of Subsection 3.3.1.

To complete the proof by using this approach, we seek an admissible key agreement code to generate the claimed achievable key rate as well as to produce signal $\mathbf{x}$ according to a Gaussian distribution satisfying constraint (4.1).

To do this, for any $\epsilon \in (0, 1)$, select real numbers $\Gamma_0 = \Gamma - \epsilon$, $\gamma \in [0, 1]$, and $\rho \in (-1, 1)$. Define

$$\beta \triangleq \rho \sqrt{\frac{\Gamma_0}{\Lambda}}\,, \tag{4.30a}$$

$$\alpha \triangleq \gamma + \beta\,. \tag{4.30b}$$

Also, let

$$X = T + \beta S\,, \tag{4.31a}$$

$$U = T + \alpha S\,, \tag{4.31b}$$

where T is a Gaussian RV generated according to $T \sim \mathcal{N}(0, \Gamma_0 - \beta^2\Lambda)$ such that T and S are independent, i.e., $\mathcal{E}(TS) = 0$. Hence, from (4.31a), X is a Gaussian RV drawn according to $X \sim \mathcal{N}(0, \Gamma_0)$ such that $\mathcal{E}(XS) = \beta\Lambda = \rho\sqrt{\Gamma_0\Lambda}$. Also, from (4.31b), U is a Gaussian RV drawn according to $U \sim \mathcal{N}(0, \Gamma_0 + (\alpha^2 - \beta^2)\Lambda)$.

Equations (4.31) will be later used for generation of $\mathbf{x}$ by using generalized DPC. By applying (4.30b), equations (4.31) can be merged into

$$U = X + \gamma S\,, \tag{4.32}$$

where the correlation coefficient between X and S is $\rho \in (-1, 1)$.

If $\mathbf{x}$ is generated i.i.d. according to $X \sim \mathcal{N}(0, \Gamma_0)$ (as we demonstrate it in the sequel), the channel outputs $\mathbf{y}$ and $\mathbf{z}$ are also i.i.d. Gaussian random vectors because of equations (4.2) and the fact that $\mathbf{s}$, $\mathbf{g}_1$, and $\mathbf{g}_2$ are i.i.d. Gaussian vectors. Hence, distributions of X_i, S_i, Y_i, Z_i, G_{1i}, and G_{2i} do not depend on time instant $i \in \{1, \ldots, n\}$. Thus, we have

$$Y = X + S + G_1 \tag{4.33a}$$

$$Z = X + S + G_2 \tag{4.33b}$$

where $G_1 \sim \mathcal{N}(0, \sigma_1^2)$ and $G_2 \sim \mathcal{N}(0, \sigma_2^2)$ are AWGN.

$\mathcal{I}(U; S)$, $\mathcal{I}(U; Y)$, and $\mathcal{I}(U; Z)$ are computable from equations (4.32) and (4.33). To do this, we first need the following expressions.

$$\begin{aligned} \mathcal{E}(U^2) &= \mathcal{E}(X^2) + \gamma^2\mathcal{E}(S^2) + 2\gamma\mathcal{E}(XS) \\ &= \Gamma_0 + \gamma^2\Lambda + 2\rho\,\gamma\sqrt{\Gamma_0\Lambda}\,, \end{aligned} \tag{4.34a}$$

$$\begin{aligned} \mathcal{E}(US) &= \mathcal{E}(XS) + \gamma\mathcal{E}(S^2) \\ &= \rho\sqrt{\Gamma_0\Lambda} + \gamma\Lambda\,, \end{aligned} \tag{4.34b}$$

$$\begin{aligned} \mathcal{E}(Y^2) &= \mathcal{E}(X^2) + 2\mathcal{E}(XS) + \mathcal{E}(S^2) + \mathcal{E}(G_1^2) + 2\mathcal{E}((X+S)G_1) \\ &= \Gamma_0 + 2\rho\sqrt{\Gamma_0\Lambda} + \Lambda + \sigma_1^2\,, \end{aligned} \tag{4.34c}$$

$$\begin{aligned}\mathcal{E}(UY) &= \mathcal{E}(X^2) + (1+\gamma)\mathcal{E}(XS) + \gamma\mathcal{E}(S^2) + \mathcal{E}(G_1X) + \gamma\mathcal{E}(G_1S) \\ &= \Gamma_0 + \rho(1+\gamma)\sqrt{\Gamma_0\Lambda} + \gamma\Lambda\,, \qquad (4.34\text{d})\end{aligned}$$

$$\begin{aligned}\mathcal{E}(Z^2) &= \mathcal{E}(X^2) + 2\mathcal{E}(XS) + \mathcal{E}(S^2) + \mathcal{E}(G_2^2) + 2\mathcal{E}((X+S)G_2) \\ &= \Gamma_0 + 2\rho\sqrt{\Gamma_0\Lambda} + \Lambda + \sigma_2^2\,, \qquad (4.34\text{e})\end{aligned}$$

$$\begin{aligned}\mathcal{E}(UZ) &= \mathcal{E}(X^2) + (1+\gamma)\mathcal{E}(XS) + \gamma\mathcal{E}(S^2) + \mathcal{E}(G_2X) + \gamma\mathcal{E}(G_2S) \\ &= \Gamma_0 + \rho(1+\gamma)\sqrt{\Gamma_0\Lambda} + \gamma\Lambda\,. \qquad (4.34\text{f})\end{aligned}$$

where equations (4.34c), (4.34d), (4.34e) and (4.34f) hold from the fact that (G_1, G_2) is independent of (X, S). We use [2, Thm. 8.4.1] to calculated entropy of a (multivariate) normal distribution. Thus, from equations (4.34), we have

$$\begin{aligned}\mathcal{I}(U;S) &= \hbar(U) + \hbar(S) - \hbar(U,S) \\ &= \frac{1}{2}\log\left(\frac{\mathcal{E}(U^2)\mathcal{E}(S^2)}{\mathcal{E}(U^2)\mathcal{E}(S^2) - (\mathcal{E}(US))^2}\right) \\ &= \frac{1}{2}\log\left(\frac{\Gamma_0 + \gamma^2\Lambda + 2\rho\,\gamma\sqrt{\Gamma_0\Lambda}}{(1-\rho^2)\Gamma_0}\right)\,, \qquad (4.35)\end{aligned}$$

$$\begin{aligned}\mathcal{I}(U;Y) &= \hbar(U) + \hbar(Y) - \hbar(U,Y) \\ &= \frac{1}{2}\log\left(\frac{\mathcal{E}(U^2)\mathcal{E}(Y^2)}{\mathcal{E}(U^2)\mathcal{E}(Y^2) - (\mathcal{E}(UY))^2}\right) \\ &= \frac{1}{2}\log\left(\frac{(\Gamma_0 + \gamma^2\Lambda + 2\rho\,\gamma\sqrt{\Gamma_0\Lambda}\,)(\Gamma_0 + 2\rho\sqrt{\Gamma_0\Lambda} + \Lambda + \sigma_1^2)}{(1-\rho^2)(1-\gamma)^2\Gamma_0\Lambda + \sigma_1^2(\Gamma_0 + \gamma^2\Lambda + 2\rho\,\gamma\sqrt{\Gamma_0\Lambda}\,)}\right)\,, \qquad (4.36)\end{aligned}$$

$$\begin{aligned}\mathcal{I}(U;Z) &= \hbar(U) + \hbar(Z) - \hbar(U,Z) \\ &= \frac{1}{2}\log\left(\frac{\mathcal{E}(U^2)\mathcal{E}(Z^2)}{\mathcal{E}(U^2)\mathcal{E}(Z^2) - (\mathcal{E}(UZ))^2}\right) \\ &= \frac{1}{2}\log\left(\frac{(\Gamma_0 + \gamma^2\Lambda + 2\rho\,\gamma\sqrt{\Gamma_0\Lambda}\,)(\Gamma_0 + 2\rho\sqrt{\Gamma_0\Lambda} + \Lambda + \sigma_2^2)}{(1-\rho^2)(1-\gamma)^2\Gamma_0\Lambda + \sigma_2^2(\Gamma_0 + \gamma^2\Lambda + 2\rho\,\gamma\sqrt{\Gamma_0\Lambda}\,)}\right)\,. \qquad (4.37)\end{aligned}$$

Now, we revise the admissible key agreement code given for the DM model to obtain an admissible key agreement code for the Gaussian model. In the Gaussian model, codewords $\mathbf{u}$ are generated i.i.d. according to $U \sim \mathcal{N}(0, \Gamma_0 + (\alpha^2 - \beta^2)\Lambda)$. Having $\mathcal{I}(U;S)$, $\mathcal{I}(U;Y)$, and $\mathcal{I}(U;Z)$, the key agreement codebook of the Gaussian model is constructed with the same specifications as those of the DM model (see Subsection 3.3.1). The key generator, encoding and decoding functions of the Gaussian model are similar to those of the DM model with the difference that the strong typicality rule used in those functions is to be replaced by the weak typicality rule for the Gaussian model.

A revealed random vector $\mathbf{s}$ is weakly typical due to AEP [2, Thm. 8.2.2] with high probability as $n \to \infty$, i.e., $\mathscr{P}\{\mathbf{s} \in \mathbb{T}_{\epsilon_0}(\mathcal{N}(0,\Lambda))\} > 1 - \epsilon_0$ for any $\epsilon_0 \in (0,1)$ and block length $n \geq n_0(\epsilon_0)$. This is equivalent to

$$|\frac{1}{n}\mathbf{s}\mathbf{s}^t - \Lambda| \leq \epsilon_0' \,, \tag{4.38}$$

where $\epsilon_0' = 2\ln(2)\Lambda\,\epsilon_0$ according to Lemma C.1.

Having $\mathbf{s}$, Alice selects codeword $\tilde{\mathbf{u}}$ from the key agreement codebook such that the pair $(\tilde{\mathbf{u}}, \mathbf{s}) \in \mathbb{T}_{\epsilon_1}(\mathcal{N}\left((0,0), \begin{bmatrix} \mathcal{E}(U^2) & \mathcal{E}(US) \\ \mathcal{E}(US) & \mathcal{E}(S^2) \end{bmatrix}\right))$ for $0 < \epsilon_0 \leq \epsilon_1 < 1$ and $n \geq n_1(\epsilon_1)$. If no error declares at the encoder[6] such codeword $\tilde{\mathbf{u}}$ exists in the codebook. According to Lemma C.2 and (4.31b), this joint typicality leads to

$$|\frac{1}{n}\mathbf{t}\mathbf{s}^t| \leq \epsilon_1' \tag{4.39}$$

where $\epsilon_1' = \frac{\epsilon_1 \ln(2)}{\alpha}[3(\Gamma_0 - \beta^2\Lambda) + 2\alpha^2\Lambda]$ and

$$\mathbf{t} = \tilde{\mathbf{u}} - \alpha\mathbf{s} \,. \tag{4.40}$$

Once codeword $\tilde{\mathbf{u}}$ is selected, signal $\mathbf{x}$ is generated by

$$\mathbf{x} = \mathbf{t} + \beta\mathbf{s} \tag{4.41}$$

$$= \tilde{\mathbf{u}} - \gamma\mathbf{s} \tag{4.42}$$

[6] As mentioned at the beginning of the proof, equations (3.37) and (3.51) are still valid if weak typicality is applied for their proofs. So, the probability of error at the encoder vanishes as $n \to \infty$.

where (4.42) follows from equations (4.30b) and (4.40). From (4.42), we conclude that $\mathbf{x}$ is a Gaussian vector as $\tilde{\mathbf{u}}$ and $\mathbf{s}$ are i.i.d. Gaussian random vectors. $\mathbf{x}$ is also i.i.d. according to $X \sim \mathcal{N}(0, \Gamma_0)$ due to equations (4.31) and (4.42). Moreover, ϵ_1-weak typicality of $(\tilde{\mathbf{u}}, \mathbf{s})$ leads to ϵ_2-weak typicality of $(\mathbf{x}, \tilde{\mathbf{u}}, \mathbf{s})$ for $0 < \epsilon_1 \leq \epsilon_2 < 1$ and $n \geq n_2(\epsilon_2)$ due to (4.42) [2, Thm. 15.2.1]. Hence $\mathbf{x} \in \mathbb{T}_{\epsilon_2}(\mathcal{N}(0, \Gamma_0))$, which is resulted in

$$|\frac{1}{n}\mathbf{x}\mathbf{x}^t - \Gamma_0| \leq \epsilon_2' \tag{4.43}$$

for $\epsilon_2' = 2\ln(2)\Gamma_0\,\epsilon_2$ due to Lemma C.1. If

$$\epsilon \geq \epsilon_2' \tag{4.44}$$

holds, power constraint (4.1) is met from (4.43) as $\Gamma_0 = \Gamma - \epsilon$. On the other hand, (4.44) is achievable for any $\epsilon \in (0, 1)$ and $n \geq \max\{n_0(\epsilon_0), n_1(\epsilon_1), n_2(\epsilon_2)\}$ if ϵ_0, ϵ_1, and ϵ_2 are initially selected small enough.

We also have

$$\begin{align}
|\frac{1}{n}\mathbf{x}\mathbf{s}^t - \rho\sqrt{\Gamma_0\Lambda}| &= |\frac{1}{n}(\mathbf{t} + \beta\mathbf{s})\mathbf{s}^t - \beta\Lambda| \tag{4.45}\\
&\leq |\frac{1}{n}\mathbf{t}\mathbf{s}^t| + \beta|\frac{1}{n}\mathbf{s}\mathbf{s}^t - \Lambda| \tag{4.46}\\
&\leq \epsilon_1' + \beta\epsilon_0' \tag{4.47}\\
&< \epsilon_1' + \sqrt{\frac{\Gamma_0}{\Lambda}}\epsilon_0' \tag{4.48}\\
&\leq \epsilon_3\,, \tag{4.49}
\end{align}$$

where

- (4.45) holds due to equations (4.41) and (4.30a);
- (4.46) holds due to triangle inequality [78];
- (4.47) follows from equations (4.38) and (4.39);
- (4.48) follows from $\beta < \sqrt{\frac{\Gamma_0}{\Lambda}}$ due to (4.30a) and the fact that $\rho \in (-1, 1)$;
- (4.49) holds for any $\epsilon_3 \in (0, 1)$ if ϵ_1 and ϵ_0 are initially selected such that $1 > \epsilon_3 \geq \epsilon_1' + \sqrt{\frac{\Gamma_0}{\Lambda}}\,\epsilon_0'$.

Equation (4.49) shows the essential difference between our strategy with that of ordinary DPC [55]. Let θ be the angle between vectors $\mathbf{x}$ and $\mathbf{s}$. In DPC [55], $\mathbf{x}$ and $\mathbf{s}$ are asymptotically orthogonal, i.e., $\cos(\theta) \to 0$ as $n \to \infty$. However, in our strategy, $\cos(\theta) \to \rho$ as $n \to \infty$.

Eventually, we provide a Gaussian distribution for the channel input satisfying the power constraint (4.1). As mentioned before, the proof can be completed with the lines of Subsection 3.3.1 when generalized Markov lemma [70, 71] for the Gaussian distribution is applied in place of the ordinary Markov lemma according to Remark 2.3. Hence, rate

$$\begin{aligned}\mathcal{R}_K(\gamma,\rho,C_{P_1}) &= \mathcal{I}(U;Y) - \mathcal{I}(U;Z)\\ &= \frac{1}{2}\log\left(\frac{[(1-\rho^2)(1-\gamma)^2\Gamma_0\Lambda + \sigma_2^2(\Gamma_0+\gamma^2\Lambda+2\rho\,\gamma\sqrt{\Gamma_0\Lambda}\,)](\Gamma_0+2\rho\sqrt{\Gamma_0\Lambda}+\Lambda+\sigma_1^2)}{[(1-\rho^2)(1-\gamma)^2\Gamma_0\Lambda + \sigma_1^2(\Gamma_0+\gamma^2\Lambda+2\rho\,\gamma\sqrt{\Gamma_0\Lambda}\,)](\Gamma_0+2\rho\sqrt{\Gamma_0\Lambda}+\Lambda+\sigma_2^2)}\right)\end{aligned}$$

is achievable subject to

$$\begin{aligned}C_{P_1} &\geq \mathcal{I}(U;S) - \mathcal{I}(U;Y) \\ &= \frac{1}{2}\log\left(\frac{(1-\rho^2)(1-\gamma)^2\Gamma_0\Lambda + \sigma_1^2(\Gamma_0+\gamma^2\Lambda+2\rho\,\gamma\sqrt{\Gamma_0\Lambda}\,)}{(1-\rho^2)\Gamma_0(\Gamma_0+2\rho\sqrt{\Gamma_0\Lambda}+\Lambda+\sigma_1^2)}\right)\end{aligned} \tag{4.50}$$

where $\Gamma_0 = \Gamma - \epsilon$ for a given $\epsilon \in (0,1)$. The LB on the forward key capacity is also obtained by taking the supremum of all achievable rates $\mathcal{R}_K(\gamma,\rho,C_{P_1})$.

4.4.4 The Proof of Corollary 4.1

The proof consists of two parts. In the first part, we prove that the key rate given in (4.13) is achievable. In the second part, we show that the key rate coincides the forward key capacity given in Theorem 1.1.

(a). The Direct Part. Let

$$\rho = 1 + \frac{2\sigma_1^2}{\Lambda - (\Lambda+\sigma_1^2)2^{2C_{P_1}}}\,, \tag{4.51}$$

$$\gamma = \sqrt{\frac{\Gamma}{\Lambda}}\,. \tag{4.52}$$

Recalling Theorem 4.1, we prove that $(\gamma, \rho) \in \mathbb{O}(C_{P_1})$ when $\Gamma \to 0$ (and Λ is fixed) in the following.

$$\begin{aligned}
C_{P_1} &= \frac{1}{2}\log\left(\frac{\Lambda}{\Lambda+\sigma_1^2} + \frac{2\sigma_1^2}{(1-\rho)(\Lambda+\sigma_1^2)}\right) && (4.53)\\
&= \frac{1}{2}\log\left(\frac{(1-\rho^2)\Lambda + 2\sigma_1^2(1+\rho)}{(1-\rho^2)(\Lambda+\sigma_1^2)}\right)\\
&= \lim_{\frac{\Gamma}{\Lambda}\to 0}\frac{1}{2}\log\left(\frac{(1-\rho^2)(1-\sqrt{\frac{\Gamma}{\Lambda}})^2 + 2\frac{\sigma_1^2}{\Lambda}(1+\rho)}{(1-\rho^2)(\frac{\Gamma}{\Lambda} + 2\rho\sqrt{\frac{\Gamma}{\Lambda}} + 1 + \frac{\sigma_1^2}{\Lambda})}\right)\\
&= \lim_{\frac{\Gamma}{\Lambda}\to 0}\frac{1}{2}\log\left(\frac{(1-\rho^2)(1-\gamma)^2\Gamma\Lambda + \sigma_1^2(\Gamma + \gamma^2\Lambda + 2\rho\,\gamma\sqrt{\Gamma\Lambda}\,)}{(1-\rho^2)\Gamma(\Gamma + 2\rho\sqrt{\Gamma\Lambda} + \Lambda + \sigma_1^2)}\right) && (4.54)
\end{aligned}$$

where

- (4.53) follows from (4.51);
- (4.54) follows from (4.52);

From (4.54), we conclude $(\gamma, \rho) \in \mathbb{O}(C_{P_1})$ as $\Gamma \to 0$.

$$\begin{aligned}
\lim_{\Gamma\to 0}\mathcal{R}_K(\gamma,\rho,C_{P_1}) &= \lim_{\Gamma\to 0}\frac{1}{2}\log\left(\frac{[(1-\rho^2)\Gamma\Lambda + 2\sigma_2^2\Gamma(1+\rho)](\Lambda+\sigma_1^2)}{[(1-\rho^2)\Gamma\Lambda + 2\sigma_1^2\Gamma(1+\rho)(\Lambda+\sigma_2^2)}\right) && (4.55)\\
&= \frac{1}{2}\log\left(\frac{[(1-\rho)\Lambda + 2\sigma_2^2](\Lambda+\sigma_1^2)}{[(1-\rho)\Lambda + 2\sigma_1^2](\Lambda+\sigma_2^2)}\right)\\
&= \frac{1}{2}\log\left(\frac{\Lambda\sigma_1^2 + \sigma_2^2[(\Lambda+\sigma_1^2)2^{2C_{P_1}} - \Lambda]}{\sigma_1^2(\Lambda+\sigma_1^2)2^{2C_{P_1}}}\right) && (4.56)\\
&= \frac{1}{2}\log\left(\frac{\Lambda(\sigma_1^2-\sigma_2^2)2^{-2C_{P_1}} + \sigma_2^2(\Lambda+\sigma_1^2)}{\sigma_1^2(\Lambda+\sigma_2^2)}\right) && (4.57)
\end{aligned}$$

where

- (4.55) follows from (4.10) and (4.52);
- (4.56) follows from (4.51).

From (4.57), we conclude that the key rate (4.13) is achievable, where $C_K(\infty, 0) = \frac{1}{2}\log\left(\frac{\sigma_2^2(\Lambda+\sigma_1^2)}{\sigma_1^2(\Lambda+\sigma_2^2)}\right)$.

(b). The Converse Part. The converse part follows from Theorem 1.1 and Lemma 4.2 as follows.

First, assume that the G-SWC in the Gaussian model is physically degraded in Bob's favor, i.e., $\varrho = \frac{\sigma_1}{\sigma_2}$. Hence,

$$\mathcal{E}(G_1 G_2) = \sigma_1^2 \tag{4.58}$$

When $\Gamma \to 0$, Alice, Bob, and Eve have access to S, $S{+}G_1$, and $S{+}G_2$, respectively. Consequently, multivariate Gaussian RV $(S, S+G_1, S+G2) \sim \mathcal{N}((0,0,0), \mathbf{\Sigma}_{3\times 3})$, where

$$\mathbf{\Sigma}_{3\times 3} = \begin{pmatrix} \Lambda & \Lambda & \Lambda \\ \Lambda & \Lambda+\sigma_1^2 & \Lambda+\sigma_1^2 \\ \Lambda & \Lambda+\sigma_1^2 & \Lambda+\sigma_2^2 \end{pmatrix} \tag{4.59}$$

due to the fact that S is independent of (G_1, G_2) as well as (4.58). Recalling Theorem 1.1, we have

$$\begin{aligned} \Sigma_{b|e} &= \Sigma_b - \Sigma_{be}\Sigma_e^{-1}\Sigma_{eb} \\ &= \frac{(\Lambda+\sigma_1^2)(\sigma_2^2-\sigma_1^2)}{\Lambda+\sigma_2^2} \end{aligned} \tag{4.60a}$$

$$\begin{aligned} \Sigma_{b|ae} &= \Sigma_b - \begin{pmatrix} \Sigma_{be} & \Sigma_{ba} \end{pmatrix} \begin{pmatrix} \Sigma_e & \Sigma_{ea} \\ \Sigma_{ae} & \Sigma_a \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{eb} \\ \Sigma_{ab} \end{pmatrix} \\ &= \sigma_1^2(1 - \frac{\sigma_1^2}{\sigma_2^2}) \end{aligned} \tag{4.60b}$$

where (4.60a) and (4.60b) follow from (4.59).

From Theorem 1.1 and equations (4.60), we conclude that the achievable key rate (4.57) is the forward key capacity of the Gaussian model when $\varrho = \frac{\sigma_1}{\sigma_2}$ and $\Gamma \to 0$. On the other hand, from Lemma 4.2, we conclude that the forward key capacity (4.57) is valid for a Gaussian model with $\Gamma \to 0$ and any $\varrho \in [-1, 1]$.

4.4.5 The Proof of Theorem 4.2

The proof consists of two parts: in the first part, we prove the direct part (achievability) of the theorem; in the second one, we derive its converse part (optimality).

(a). The Direct Part. Let $\gamma = 1$ and $\rho = 1 - \delta$, where $\delta \in (0, 1)$. Assume $\delta \to 0$; thus, $(1, 1-\delta) \in \mathbb{O}(C_{P_1})$ if condition

$$C_{P_1} \geq \frac{1}{2} \log \left(\frac{\sigma_1^2(\Gamma + \Lambda + 2\,\sqrt{\Gamma\Lambda}\,)}{\delta\,\Gamma(\Gamma + 2\sqrt{\Gamma\Lambda} + \Lambda + \sigma_1^2)} \right) . \tag{4.61}$$

is met. To satisfy this condition, select C_{P_1} (as a function of δ) large enough such that condition (4.61) is met for the given $\delta \in (0, 1)$. In other words, $(1, 1-\delta) \in \mathbb{O}(C_{P_1})$ with $\delta \to 0$ as $C_{P_1} \to \infty$. Hence,

$$\mathcal{R}_K(1, 1-\delta, C_{P_1}) = \frac{1}{2} \log \left(\frac{\sigma_2^2(\Gamma + \Lambda + 2(1-\delta)\sqrt{\Gamma\Lambda} + \sigma_1^2)}{\sigma_1^2(\Gamma + \Lambda + 2(1-\delta)\sqrt{\Gamma\Lambda} + \sigma_2^2)} \right) \tag{4.62}$$

is an achievable key rate according to (4.10), where $\delta \to 0$, $\delta > 0$, as $C_{P_1} \to \infty$. As a result, the following LB on $C_K(\infty, 0)$ is asymptotically achievable according to (4.62) as $\delta \to 0$, $\delta > 0$, and $C_{P_1} \to \infty$:

$$C_K(\infty, 0) \geq \frac{1}{2} \log \left(\frac{(\Gamma + \Lambda + 2\sqrt{\Gamma\Lambda} + \sigma_1^2)\sigma_2^2}{(\Gamma + \Lambda + 2\sqrt{\Gamma\Lambda} + \sigma_2^2)\sigma_1^2} \right) . \tag{4.63}$$

(b). The Converse Part. We prove a UB on the forward key capacity of a Gaussian model with a physically degraded G-SWC $(\Gamma, \Lambda, \sigma_1^2, \sigma_2^2)$ in Bob's favor. According to Lemma 4.2, the UB on the forward key capacity is valid for its equivalent Gaussian model with the G-SWC $(\Gamma, \Lambda, \sigma_1^2, \sigma_2^2, \varrho)$ as defined in Section 4.1.

Let $i \in \{1, \ldots, n\}$ be the time instant. According to (2.23), for a physically degraded G-SWC in Bob's favor ($\sigma_2^2 > \sigma_1^2$), we have

$$Y_i = X_i + S_i + G_{1i} \tag{4.64a}$$

$$Z_i = Y_i + G'_{2i} \,, \tag{4.64b}$$

where $G_{1i} \sim \mathcal{N}(0, \sigma_1^2)$ and $G'_{2i} \sim \mathcal{N}(0, \sigma_2^2 - \sigma_1^2)$ are independent. We first establish the following lemma.

Lemma 4.4. *Let $i \in \{1, \ldots, n\}$ be the time instant and define*

$$\Upsilon_i \triangleq \mathcal{E}((X_i + S_i)^2) \,. \tag{4.65}$$

Assume a G-SWC with average power constraint (4.1) *and the interference average power Λ. Then, there exists a real number $\rho \in [-1, 1]$ such that*

$$\frac{1}{n}\sum_{i=1}^{n} \Upsilon_i \leq \Gamma + \Lambda + 2\rho\sqrt{\Gamma\Lambda} \,, \tag{4.66}$$

where the equality holds if $X_i \sim \mathcal{N}(0, \Gamma)$ for any $i \in \{1, \ldots, n\}$.

Proof. The proof is similar to [79, Eq. 90-92] as follows.

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \Upsilon_i &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}(S_i^2) + \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}(X_i^2) + \frac{2}{n}\sum_{i=1}^{n} \mathcal{E}(X_i S_i) \\ &\leq \Lambda + \Gamma + \frac{2}{n}\sum_{i=1}^{n} \mathcal{E}(X_i S_i) \,, \end{aligned} \tag{4.67}$$

where (4.67) follows from

- $\mathcal{E}(S_i^2) = \Lambda$ as $S_i \sim \mathcal{N}(0, \Lambda)$ for any $i \in \{1, \ldots, n\}$,
- $\frac{1}{n}\sum\limits_{i=1}^{n} \mathcal{E}(X_i^2) = \mathcal{E}(\frac{1}{n}\sum\limits_{i=1}^{n} X_i^2) \leq \Gamma$ due to average power constraint (4.1).

Also, the equality in (4.67) holds if $X_i \sim \mathcal{N}(0, \Gamma)$ for any $i \in \{1, \ldots, n\}$.

On the other hand,

$$
\begin{aligned}
|\frac{2}{n}\sum_{i=1}^{n}\mathcal{E}(X_iS_i)| &\le \frac{2}{n}\sum_{i=1}^{n}\sqrt{\mathcal{E}(X_i^2)\mathcal{E}(S_i^2)} && (4.68)\\
&\le 2\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{E}(X_i^2)\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{E}(S_i^2)\right)} && (4.69)\\
&= 2\sqrt{\mathcal{E}(\frac{1}{n}\sum_{i=1}^{n}X_i^2)\Lambda} && (4.70)\\
&\le 2\sqrt{\Gamma\Lambda}\,, && (4.71)
\end{aligned}
$$

where

- (4.68) holds due to Cauchy-Schwarz inequality [80] $(\mathcal{E}(X_iS_i))^2 \le \mathcal{E}(X_i^2)\mathcal{E}(S_i^2)$;
- (4.69) follows from Cauchy-Schwarz inequality

$$
\left(\sum_{i=1}^{n}\sqrt{\mathcal{E}(X_i^2)}\sqrt{\mathcal{E}(S_i^2)}\right)^2 \le \left(\sum_{i=1}^{n}\mathcal{E}(X_i^2)\right)\left(\sum_{i=1}^{n}\mathcal{E}(S_i^2)\right)\ ;
$$

- (4.70) holds as $S_i \sim \mathcal{N}(0,\Lambda)$ for any $i \in \{1,\ldots,n\}$;
- (4.71) holds due to average power constraint (4.1).

Thus, from (4.70), we conclude that there exists a real number $-1 \le \rho \le 1$ such that

$$
\frac{2}{n}\sum_{i=1}^{n}\mathcal{E}(X_iS_i) = 2\rho\sqrt{\Gamma\Lambda}\,. \tag{4.72}
$$

Applying (4.72) to (4.67), the lemma is approved. □

Fix an arbitrarily small $\epsilon \ge 0$. According to Definition 1.11, a key rate R_K is achievable if there exists an admissible key agreement code $(\lceil 2^{nR_K}\rceil, n)$, which returns $(K,\hat{K})$, for the given ϵ such that equations (3.88) are held. Assuming $\sigma_1^2 \ge \sigma_2^2$, we derive a UB on the achievable key rate R_K, when $C_{P_2} = 0$, as follows.

$$nR_{K} \leq \mathcal{H}(K) + n\frac{\epsilon}{3} \tag{4.73}$$

$$\leq \mathcal{I}(K;\hat{K}) + n\frac{2\epsilon}{3} \tag{4.74}$$

$$\leq \mathcal{I}(K;\mathbf{y},\mathbf{p}_1,Q_2) + n\frac{2\epsilon}{3} \tag{4.75}$$

$$\leq \mathcal{I}(K;\mathbf{y},\mathbf{p}_1) + n\frac{2\epsilon}{3} \tag{4.76}$$

$$\leq \mathcal{I}(K;\mathbf{y},\mathbf{p}_1) - \mathcal{I}(K;\mathbf{z},\mathbf{p}_1) + n\epsilon \tag{4.77}$$

$$= \mathcal{I}(K;\mathbf{y}|\mathbf{p}_1) - \mathcal{I}(K;\mathbf{z}|\mathbf{p}_1) + n\epsilon \tag{4.78}$$

$$= \sum_{i=1}^{n}[\mathcal{I}(U_i;Y_i|W_i) - \mathcal{I}(U_i;Z_i|W_i)] + n\epsilon \tag{4.79}$$

$$= \sum_{i=1}^{n}\left([\mathcal{I}(U_i;Y_i) - \mathcal{I}(U_i;Z_i)] - [\mathcal{I}(W_i;Y_i) - \mathcal{I}(W_i;Z_i)]\right) + n\epsilon \tag{4.80}$$

$$\leq \sum_{i=1}^{n}[\mathcal{I}(U_i,X_i,S_i;Y_i) - \mathcal{I}(U_i,X_i,S_i;Z_i) - \mathcal{I}(X_i,S_i;Y_i|U_i) + \mathcal{I}(X_i,S_i;Z_i|U_i)] + n\epsilon \tag{4.81}$$

$$= \sum_{i=1}^{n}\left([\mathcal{I}(X_i,S_i;Y_i) - \mathcal{I}(X_i,S_i;Z_i)] - [\mathcal{I}(X_i,S_i;Y_i|U_i) - \mathcal{I}(X_i,S_i;Z_i|U_i)]\right) + n\epsilon \tag{4.82}$$

$$\leq \sum_{i=1}^{n}[\mathcal{I}(X_i,S_i;Y_i) - \mathcal{I}(X_i,S_i;Z_i)] + n\epsilon \tag{4.83}$$

$$= \sum_{i=1}^{n}[\hbar(Y_i) - \hbar(Y_i|X_i,S_i) - \hbar(Z_i) + \hbar(Z_i|X_i,S_i)] + n\epsilon$$

$$= \sum_{i=1}^{n}[\hbar(Y_i) - \hbar(X_i+S_i+G_{1i}|X_i,S_i) - \hbar(Z_i) + \hbar(X_i+S_i+G_{2i}|X_i,S_i)] + n\epsilon \tag{4.84}$$

$$= \sum_{i=1}^{n}[\hbar(Y_i) - \hbar(G_{1i}) - \hbar(Z_i) + \hbar(G_{2i})] + n\epsilon \tag{4.85}$$

$$= \frac{n}{2}\log(2\pi e\sigma_2^2) - \frac{n}{2}\log(2\pi e\sigma_1^2) + \sum_{i=1}^{n}[\hbar(Y_i) - \hbar(Y_i+G'_{2i})] + n\epsilon \tag{4.86}$$

$$\leq \frac{n}{2}\log(\frac{\sigma_2^2}{\sigma_1^2}) + \sum_{i=1}^{n}[\hbar(Y_i) - \frac{1}{2}\log(2^{2\hbar(G'_{2i})} + 2^{2\hbar(Y_i)})] + n\epsilon \tag{4.87}$$

$$\leq \frac{n}{2}\log(\frac{\sigma_2^2}{\sigma_1^2}) + \sum_{i=1}^{n}[\frac{1}{2}\log(2\pi e\mathcal{E}(Y_i^2)) - \frac{1}{2}\log(2^{2\hbar(G'_{2i})} + 2^{\log(2\pi e\mathcal{E}(Y_i^2))})] + n\epsilon \tag{4.88}$$

$$= \frac{n}{2}\log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) + \sum_{i=1}^{n}[\frac{1}{2}\log(2\pi e(\Upsilon_i + \sigma_1^2))$$

$$-\frac{1}{2}\log(2\pi e(\sigma_2^2 - \sigma_1^2) + 2\pi e(\Upsilon_i + \sigma_1^2))] + n\epsilon \tag{4.89}$$

$$= \frac{n}{2}\log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) + \frac{1}{2}\sum_{i=1}^{n}\log\left(\frac{\Upsilon_i + \sigma_1^2}{\Upsilon_i + \sigma_2^2}\right) + n\epsilon$$

$$\leq \frac{n}{2}\log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) + \frac{n}{2}\log\left(\frac{\frac{1}{n}\sum_{i=1}^{n}\Upsilon_i + \sigma_1^2}{\frac{1}{n}\sum_{i=1}^{n}\Upsilon_i + \sigma_2^2}\right) + n\epsilon \tag{4.90}$$

$$\leq \frac{n}{2}\log\left(\frac{\sigma_2^2(\Gamma + \Lambda + 2\rho\sqrt{\Gamma\Lambda} + \sigma_1^2)}{\sigma_1^2(\Gamma + \Lambda + 2\rho\sqrt{\Gamma\Lambda} + \sigma_2^2)}\right) + n\epsilon \tag{4.91}$$

$$\leq \frac{n}{2}\log\left(\frac{\sigma_2^2(\Gamma + \Lambda + 2\sqrt{\Gamma\Lambda} + \sigma_1^2)}{\sigma_1^2(\Gamma + \Lambda + 2\sqrt{\Gamma\Lambda} + \sigma_2^2)}\right) + n\epsilon \tag{4.92}$$

where

- (4.73) and (4.74) follow from the lines of (3.94) and (3.95), respectively, which are valid in the Gaussian model as well;
- (4.75) follows from data processing inequality [2, Thm. 2.8.1] due to $K \to (Q_2, \mathbf{y}, \mathbf{p}_1) \to \hat{K}$;
- (4.76) follows from the fact that Q_2 is independent of $(K, \mathbf{y}, \mathbf{p}_1)$ when $C_{P_2} = 0$;
- (4.77) follows from (3.88b), where P is a one-to-one function of $\mathbf{p}_1$ and $\mathbf{e} = \mathbf{0}$ in Definition 1.10;
- (4.79) follows from the proof of (3.101), which is valid for the Gaussian model as well, where P is a one-to-one function of $\mathbf{p}_1$ as well as $\check{Y}_i = Y_i$ and $\check{Z}_i = Z_i$ for any $i \in \{1, \ldots, n\}$ because neither Bob nor Eve has SI in the Gaussian model. Also, $W_i \triangleq (Y_1^{i-1}Z_{i+1}^{n}P)$ and $U_i \triangleq (K, W_i)$;
- (4.80) follows from Markov chain

$$W_i \to U_i \to (X_i, S_i) \to Y_i \to Z_i \tag{4.93}$$

which is valid due to definition $U_i \triangleq (K, W_i)$ and equations (4.64) for any $i \in \{1, \dots, n\}$;

- (4.81) and (4.82) follow from data processing inequality [2, Thm. 2.8.1] due to (4.93);
- (4.83) follows from data processing inequality [2, Thm. 2.8.1] due to $(X_i, S_i) \to (U_i, Y_i) \to Z_i$, which is valid because of (4.64) and (4.93), i.e.,

$$\begin{aligned}\mathcal{I}(X_i, S_i; Y_i|U_i) - \mathcal{I}(X_i, S_i; Z_i|U_i) &= \mathcal{I}(X_i, S_i; Y_i, Z_i|U_i) - \mathcal{I}(X_i, S_i; Z_i|U_i, Y_i) - \mathcal{I}(X_i, S_i; Z_i|U_i) \\ &= \mathcal{I}(X_i, S_i; Y_i|U_i, Z_i) \\ &\geq 0\,;\end{aligned}$$

- (4.84) follows from equations (4.64), where $G_{2i} = G_{1i} + G'_{2i}$;
- (4.85) follows from the fact that G_{1i} and $G_{2i} = G_{1i} + G'_{2i}$ are independent of (X_i, S_i);
- (4.86) follows from entropy of normal distributions [2, Thm. 8.4.1] and (4.64b);
- (4.87) follows from $2^{2\hbar(Y_i+G'_{2i})} \geq 2^{2\hbar(Y_i)} + 2^{2\hbar(G'_{2i})}$ due to *entropy power inequality* [2, Thm. 17.7.3] and the fact that G'_{2i} is independent of Y_i in (4.64b);
- (4.88) follows from the fact that $\hbar(Y_i) - \frac{1}{2}\log(2^{2\hbar(G'_{2i})} + 2^{2\hbar(Y_i)})$ is an increasing function of $\hbar(Y_i)$ as $\hbar(G'_{2i}) = \frac{1}{2}\log(2\pi e(\sigma_2^2 - \sigma_1^2))$ [2, Thm. 8.4.1] is fixed; it also follows from $\hbar(Y_i) \leq \frac{1}{2}\log(2\pi e\mathcal{E}(Y_i^2))$ due to [2, Thm. 8.6.5] as $\mathcal{E}(Y_i) = 0$ with the equality when $Y_i \sim \mathcal{N}(0, \mathcal{E}(Y_i^2))$;
- (4.89) holds due to $\mathcal{E}(Y_i^2) = \Upsilon_i + \sigma_1^2$ according to (4.65) and the fact that G_{1i} is independent of (X_i, S_i);
- (4.90) holds by applying Jensen's inequality [2, Thm. 2.6.2] to $\log\left(\frac{\Upsilon_i+\sigma_1^2}{\Upsilon_i+\sigma_2^2}\right)$ as a concave function of Υ_i due to $\sigma_2^2 > \sigma_1^2$;
- (4.91) follows from the fact that $\log\left(\frac{\xi+\sigma_1^2}{\xi+\sigma_2^2}\right)$ is an increasing function of $\xi \in \mathbb{R}^+ \cup \{0\}$ due to $\sigma_2^2 > \sigma_1^2$; then, it is valid because of Lemma 4.4 for some $-1 \leq \rho \leq 1$;
- (4.92) holds because function $\frac{1}{2}\log\left(\frac{\sigma_2^2(\Gamma+\Lambda+2\rho\sqrt{\Gamma\Lambda}+\sigma_1^2)}{\sigma_1^2(\Gamma+\Lambda+2\rho\sqrt{\Gamma\Lambda}+\sigma_2^2)}\right)$ is an increasing function of ρ provided $\sigma_2^2 > \sigma_1^2$.

4.4.6 The Proof of Corollary 4.2

First, as $\frac{\Gamma}{\Lambda} \to \infty$ the UB on the key capacity is

$$\begin{aligned}\lim_{\frac{\Gamma}{\Lambda}\to\infty} C_K(C_{P_1}, 0) &\leq \lim_{\frac{\Gamma}{\Lambda}\to\infty} C_K(\infty, 0) \\ &= \frac{1}{2}\log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) \qquad (4.94)\end{aligned}$$

according to Theorem 4.2.

Second, for any fixed $C_{P_1} \in [0,\infty)$, $\gamma \in [0,1]$, and $\rho \in (-1,1)$ we have

$$\lim_{\frac{\Gamma}{\Lambda}\to\infty} \frac{1}{2}\log\left(\frac{(1-\rho^2)(1-\gamma)^2\Gamma\Lambda + \sigma_1^2(\Gamma + \gamma^2\Lambda + 2\rho\,\gamma\sqrt{\Gamma\Lambda}\,)}{(1-\rho^2)\Gamma(\Gamma + 2\rho\sqrt{\Gamma\Lambda} + \Lambda + \sigma_1^2)}\right) = -\infty\,;$$

hence, $(\gamma,\rho) \in \mathbb{O}(C_{P_1})$ according to (4.9). Therefore, we obtain the following LB on the key capacity from Theorem 4.1:

$$\lim_{\frac{\Gamma}{\Lambda}\to\infty} C_K(C_{P_1}, 0) \geq \frac{1}{2}\log\left(\frac{\sigma_2^2}{\sigma_1^2}\right)\,, \qquad (4.95)$$

for $\gamma = 1$ and $\rho \to 1$. Eventually, equations (4.94) and (4.95) confirm the corollary.

4.4.7 The Proof of Corollary 4.3

The proof consists of two parts: in the direct part, we prove that if $\sigma_2^2 > \sigma_1^2$ then $C_K(C_{P_1}, 0) > 0$; in the converse part, we prove that if $\sigma_1^2 \geq \sigma_2^2$ then $R_K \leq n\epsilon$, for any achievable forward key rate R_K and arbitrarily small $\epsilon \geq 0$.

(a). The Direct Part. We show that for any $C_{P_1} > 0$ there exists $(\gamma,\rho) \in \mathbb{O}(C_{P_1})$ such that $\mathcal{R}_K(\gamma,\rho,C_{P_1}) > 0$. Recalling Theorem 4.1, let $(\gamma,\rho) = (0,0)$; then, $(\gamma,\rho) \in \mathbb{O}(C_{P_1})$ because

$$\begin{aligned}\frac{1}{2}\log\left(\frac{(1-\rho^2)(1-\gamma)^2\Gamma\Lambda + \sigma_1^2(\Gamma + \gamma^2\Lambda + 2\rho\,\gamma\sqrt{\Gamma\Lambda}\,)}{(1-\rho^2)\Gamma(\Gamma + 2\rho\sqrt{\Gamma\Lambda} + \Lambda + \sigma_1^2)}\right) &= \frac{1}{2}\log\left(\frac{\Gamma\Lambda + \sigma_1^2\Gamma}{\Gamma(\Gamma + \Lambda + \sigma_1^2)}\right) \\ &< 0\end{aligned}$$

$$\leq C_{P_1} \, .$$

The achievable rate with $(\gamma, \rho) = (0, 0)$ is

$$\begin{aligned}
\mathcal{R}_K(0, 0, C_{P_1}) &= \frac{1}{2} \log \left(\frac{(\Gamma\Lambda + \sigma_2^2\Gamma)(\Gamma + \Lambda + \sigma_1^2)}{(\Gamma\Lambda + \sigma_1^2\Gamma)(\Gamma + \Lambda + \sigma_2^2)} \right) \\
&= \frac{1}{2} \log \left(\frac{1 + \frac{\Gamma}{(\Lambda+\sigma_1^2)}}{1 + \frac{\Gamma}{(\Lambda+\sigma_2^2)}} \right) \, ,
\end{aligned}$$

which is positive when $\sigma_2^2 > \sigma_1^2$. Hence, the forward key capacity is positive when $\sigma_2^2 > \sigma_1^2$.

(b). The Converse Part. We prove a UB on the forward key capacity of a Gaussian model with a physically degraded G-SWC $(\Gamma, \Lambda, \sigma_1^2, \sigma_2^2)$ in Eve's favor. According to Lemma 4.2, the UB on the forward key capacity is valid for its equivalent Gaussian model with the G-SWC $(\Gamma, \Lambda, \sigma_1^2, \sigma_2^2, \varrho)$ as defined in Section 4.1.

According to (2.24) and Definition 4.1, for a physically degraded G-SWC in Eve's favor $(\sigma_1^2 \geq \sigma_2^2)$, we have

$$\mathbf{y} = \mathbf{z} + \mathbf{g}_1' \tag{4.96}$$

where $\mathbf{g}_1'$ is independent of $(Q_1, \mathbf{s}, \mathbf{g}_2)$, and it is distributed i.i.d. according to $\mathcal{N}(0, \sigma_1^2 - \sigma_2^2)$.

Fix an arbitrarily small $\epsilon \geq 0$. According to Definition 1.11, a key rate R_K is achievable if there exists an admissible key agreement code $(\lceil 2^{nR_K} \rceil, n)$, which returns $(K, \hat{K})$, for the given ϵ such that equations (3.88) are met.

Assuming $\sigma_1^2 \geq \sigma_2^2$, we derive a UB on the achievable key rate R_K, when $C_{P_2} = 0$, in the following. We can restart from (4.78) because this step is valid when $\sigma_1^2 \geq \sigma_2^2$ as well.

$$\begin{aligned}
nR_K &\leq \mathcal{I}(K; \mathbf{y}|\mathbf{p}_1) - \mathcal{I}(K; \mathbf{z}|\mathbf{p}_1) + n\epsilon \\
&= \mathcal{I}(K; \mathbf{y}|\mathbf{p}_1) - \mathcal{I}(K; \mathbf{y}, \mathbf{z}|\mathbf{p}_1) + \mathcal{I}(K; \mathbf{y}|\mathbf{p}_1, \mathbf{z}) + n\epsilon \\
&= -\mathcal{I}(K; \mathbf{z}|\mathbf{y}, \mathbf{p}_1) + \mathcal{I}(K; \mathbf{y}|\mathbf{p}_1, \mathbf{z}) + n\epsilon \\
&= -\mathcal{I}(K; \mathbf{z}|\mathbf{y}, \mathbf{p}_1) + \mathcal{I}(K; \mathbf{g}_1'|\mathbf{p}_1, \mathbf{z}) + n\epsilon && (4.97) \\
&= -\mathcal{I}(K; \mathbf{z}|\mathbf{y}, \mathbf{p}_1) + n\epsilon && (4.98) \\
&\leq n\epsilon && (4.99)
\end{aligned}$$

where

- (4.97) follows from (4.96);
- (4.98) is valid because $\mathbf{g}_1'$ is independent of $(Q_1, \mathbf{s}, \mathbf{g}_2)$, and so it is independent of $K = \mathcal{K}_1(Q_1, \mathbf{s})$, $\mathbf{x} = \mathcal{W}(Q_1, \mathbf{s})$, $\mathbf{p}_1 = \mathcal{F}(Q_1, \mathbf{s})$ and $\mathbf{z} = \mathbf{x} + \mathbf{s} + \mathbf{g}_2$ according to Definition 4.3.

Hence, if $\epsilon \to 0$, equation (4.99) proves that the forward key capacity of a Gaussian model with a physically degraded G-SWC in Eve's favor, i.e., $\sigma_1^2 \geq \sigma_2^2$, vanishes. Using Lemma 4.2, we can extend this result to the Gaussian model with a G-SWC having parameters $(\Gamma, \Lambda, \sigma_1^2, \sigma_2^2, \varrho)$, where $\sigma_1^2 \geq \sigma_2^2$ and $\varrho \in [-1, 1]$, as specified in Definition 4.1.

4.4.8 The Proof of Theorem 4.3

Recalling Definition 4.2 for the notations, we prove this theorem in the following two parts.

(a). Part (a) of Theorem 4.3. To prove this part, we need the following lemma, which is proved in [3, Lem. 17.18].

Lemma 4.5. *For any $i \in \{1, \ldots, n\}$, assume that P_{1i} is a function of A, E, and $P_{21}^{i-1} \triangleq (P_{21}, P_{22}, \ldots, P_{2(i-1)})$, while P_{2i} is a function of B, E, and $P_{11}^{i} \triangleq (P_{11}, P_{12}, \ldots, P_{1i})$. Then*

$$\mathcal{I}(A; B|E) \geq \mathcal{I}(A; B|E, \mathbf{p}_1, \mathbf{p}_2) \, ,$$

where $\mathbf{p}_1 \triangleq P_{11}^n$ and $\mathbf{p}_2 \triangleq P_{21}^n$.

Fix an arbitrarily small $\epsilon \geq 0$. According to Definition 1.11, a key rate R_K is achievable if there exists an admissible key agreement code $(\lceil 2^{nR_K} \rceil, n)$, which returns $(K, \hat{K})$, for the given ϵ such that equations (3.88) are held.

As the channel is assumed to be physically degraded in Eve's favor, (4.96) holds. Using this equation, we derive a UB on the achievable key rate R_K in the following.

$$
\begin{aligned}
nR_{K} &\leq \mathcal{H}(K) + n\frac{\epsilon}{3} && \text{(4.100)}\\
&\leq \mathcal{I}(K;\hat{K}) + n\frac{2\epsilon}{3} && \text{(4.101)}\\
&\leq \mathcal{I}(K;\mathbf{y},\mathbf{p}_1,Q_2) + n\frac{2\epsilon}{3} && \text{(4.102)}\\
&\leq \mathcal{I}(K;\mathbf{y},\mathbf{p}_1,\mathbf{p}_2,Q_2) + n\frac{2\epsilon}{3} && \\
&\leq \mathcal{I}(K;\mathbf{y},\mathbf{p}_1,\mathbf{p}_2,Q_2) - \mathcal{I}(K;\mathbf{z},\mathbf{p}_1,\mathbf{p}_2) + n\epsilon && \text{(4.103)}\\
&= \mathcal{I}(K;\mathbf{y},\mathbf{z},\mathbf{p}_1,\mathbf{p}_2,Q_2) - \mathcal{I}(K;\mathbf{z}|\mathbf{y},\mathbf{p}_1,\mathbf{p}_2,Q_2) - \mathcal{I}(K;\mathbf{z},\mathbf{p}_1,\mathbf{p}_2) + n\epsilon && \\
&= \mathcal{I}(K;\mathbf{y},Q_2|\mathbf{z},\mathbf{p}_1,\mathbf{p}_2) - \mathcal{I}(K;\mathbf{z}|\mathbf{y},\mathbf{p}_1,\mathbf{p}_2,Q_2) + n\epsilon && \\
&\leq \mathcal{I}(K;\mathbf{y},Q_2|\mathbf{z},\mathbf{p}_1,\mathbf{p}_2) + n\epsilon && \\
&\leq \mathcal{I}(K,Q_1,\mathbf{s};\mathbf{y},Q_2|\mathbf{z},\mathbf{p}_1,\mathbf{p}_2) + n\epsilon && \\
&= \mathcal{I}(Q_1,\mathbf{s};\mathbf{y},Q_2|\mathbf{z},\mathbf{p}_1,\mathbf{p}_2) + n\epsilon && \text{(4.104)}\\
&\leq \mathcal{I}(Q_1,\mathbf{s};\mathbf{y},Q_2|\mathbf{z}) + n\epsilon && \text{(4.105)}\\
&= \mathcal{I}(Q_1,\mathbf{s};\mathbf{y}|\mathbf{z}) + n\epsilon && \text{(4.106)}\\
&= \sum_{i=1}^{n} \mathcal{I}(Q_1,\mathbf{s};Y_i|\mathbf{z},Y_1^{(i-1)}) + n\epsilon && \\
&\leq n\epsilon && \text{(4.107)}
\end{aligned}
$$

where

- (4.100) follows from (3.88d);
- (4.101) follows from (3.88a) and thus $\mathcal{H}(K|\hat{K}) \leq n\frac{\epsilon}{3}$ due to Fano's inequality [2, Thm. 2.10.1];
- (4.102) holds due to data processing inequality [2, Thm. 2.8.1] as $\hat{K} = \mathcal{K}_2(Q_2,\mathbf{y},\mathbf{p}_1)$ and so $K \to (Q_2,\mathbf{y},\mathbf{p}_1) \to \hat{K}$;
- (4.103) follows from (3.88b), where P is a one-to-one function of $(\mathbf{p}_1,\mathbf{p}_2)$ and $\mathbf{e} = \mathbf{0}$ in Definition 1.10;
- (4.104) follows from the fact that K is a deterministic function of $(Q_1,\mathbf{s},\mathbf{p}_2)$;
- (4.105) follows from Lemma 4.5, where A, B, and E in this lemma must be substituted by $(Q_1,\mathbf{s})$, $(Q_2,\mathbf{y})$, and $\mathbf{z}$, respectively;
- (4.106) follows from the fact that Q_2 is independent of $(Q_1,\mathbf{s},\mathbf{y})$ according to Definition 4.2;

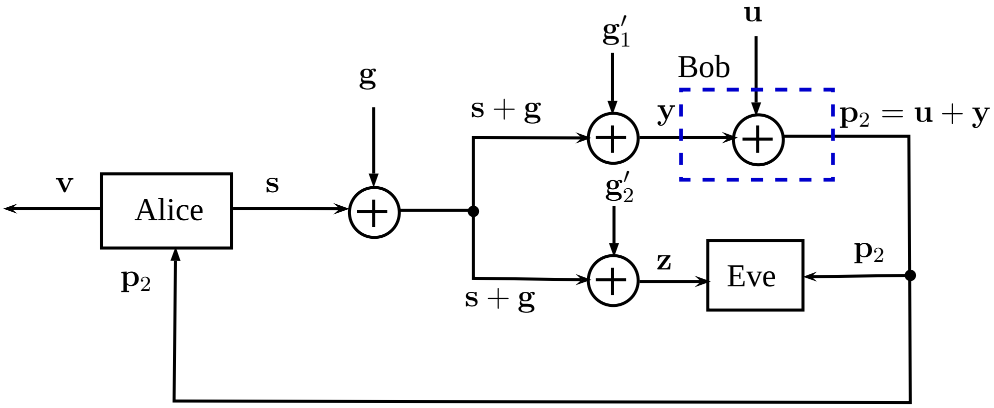


Figure 4.9: Key generation over a Gaussian model with a backward public channel.

- (4.107) follows from the fact that for any given $i \in \{1, \ldots, n\}$ and Z_i, $Y_i = Z_i + G'_{1i}$ is independent of $(Q_1, \mathbf{s}, (Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n), Y_1^{(i-1)})$ because the G-SWC is memoryless and physically degraded.

Hence, if $\epsilon \to 0$, equation (4.107) proves that the key capacity $C_K(\infty, \infty)$ of a Gaussian model with a physically degraded G-SWC in Eve's favor, i.e., $\sigma_1^2 \geq \sigma_2^2$, vanishes.

(b). Part (b) of Theorem 4.3. First, let define independent Gaussian vectors $\mathbf{g}$, $\mathbf{g}_1'$, and $\mathbf{g}_2'$, which are drawn i.i.d. according to $\mathcal{N}(0, \varrho\sigma_1\sigma_2)$, $\mathcal{N}(0, \sigma_1^2 - \rho\sigma_1\sigma_2)$, and $\mathcal{N}(0, \sigma_2^2 - \rho\sigma_1\sigma_2)$, respectively. Then, the Gaussian noise vectors can be written as

$$\mathbf{g}_1 = \mathbf{g} + \mathbf{g}_1' \,, \tag{4.108a}$$

$$\mathbf{g}_2 = \mathbf{g} + \mathbf{g}_2' \,. \tag{4.108b}$$

This is due to the fact that the distribution of a pair of jointly Gaussian RVs is uniquely determined by a covariance matrix. From (4.108), if $\varrho = \frac{\sigma_1}{\sigma_2}$ then $\mathbf{g}_1' = \mathbf{0}$ (in probability) and the channel is physically degraded in Bob's favor. Similarly, if $\varrho = \frac{\sigma_2}{\sigma_1}$ then $\mathbf{g}_2' = \mathbf{0}$ (in probability) and the channel is physically degraded in Eve's favor (see Remark 4.1).

The proof is justified as follows according to Figure 4.9. According to Corollary 4.3, no key rate can be agreed by using the G-SWC alone because Eve has an advantage over Bob on the G-SWC. Hence, Alice sends nothing in our key agreement scheme, i.e. $\mathbf{x} = \mathbf{0}$. Bob generates a Gaussian wiretap codebook [13]. He generates the key according to randomization RV Q_2, which is uniformly distributed. Bob treats Q_2 as a message and encodes it into codeword $\mathbf{u}$ by using the Gaussian wiretap codebook. Observing $\mathbf{y} = \mathbf{s} + \mathbf{g}_1$ from the wiretap channel, he sends $\mathbf{p}_2 = \mathbf{u} + \mathbf{y} = \mathbf{u} + \mathbf{s} + \mathbf{g}_1$ over the public channel in the backward direction. Alice, who knows $\mathbf{s}$, calculates $\mathbf{v} = \mathbf{p}_2 - \mathbf{s} = \mathbf{u} + \mathbf{g}_1$. On the other hand, Eve, who receives $\mathbf{z} = \mathbf{s} + \mathbf{g}_2$ from the wiretap channel, is able to obtain $\mathbf{p}_2 - \mathbf{z} = \mathbf{u} + \mathbf{g}_1 - \mathbf{g}_2 = \mathbf{u} + \mathbf{g}'_1 - \mathbf{g}'_2$ according to (4.108).

As a result, the public channel from Bob to Alice can be considered as a Gaussian wiretap channel[7] according to Figure 1.5, where Bob is the sender and Alice and Eve are the receivers. Also, the transmission signal of the Gaussian wiretap channel is $\mathbf{u}$, and Alice and Eve receive $\mathbf{u} + \mathbf{g}_1$ and $\mathbf{u} + \mathbf{g}'_1 - \mathbf{g}'_2$, respectively. Hence, Bob can achieve secure rate

$$R = \frac{1}{2} \log \left(\frac{1 + \frac{\Omega}{\sigma_1^2}}{1 + \frac{\Omega}{\sigma_1^2 + \sigma_2^2 - 2\varrho\sigma_1\sigma_2}} \right) \tag{4.109}$$

according to (1.14), where Ω is the variance of U ($\mathbf{u}$ is generated i.i.d. according to $U \sim \mathcal{N}(0, \Omega)$). This secure rate can be treated as an achievable key rate between Alice and Bob, i.e., $R_K(\infty, \infty) = R$. When the capacity of public channel is unlimited in the backward direction, $\Omega \to \infty$ gives the maximum achievable key rate by the proposed scheme, and the result of this theorem is concluded.

According to the proposed key agreement code, $\mathbf{p}_2 - \mathbf{z}$ is a sufficient statistic of $\mathbf{z}, \mathbf{p}_2$ for key K conditioned on key agreement codebook $\mathbf{C}$, i.e., $I(K; \mathbf{p}_2 - \mathbf{z}|\mathbf{C}) = I(K; \mathbf{p}_2, \mathbf{z}|\mathbf{C})$. This fact, which is proved below, approves that the justification, as explained above, for the security of the code is correct.

$$\begin{aligned}
\mathcal{H}(K|\mathbf{z}, \mathbf{p}_2, \mathbf{C}) &= \mathcal{H}(K|\mathbf{p}_2 - \mathbf{z}, \mathbf{p}_2, \mathbf{C}) \\
&= \mathcal{H}(K, \mathbf{p}_2|\mathbf{p}_2 - \mathbf{z}, \mathbf{C}) - \mathcal{H}(\mathbf{p}_2|\mathbf{p}_2 - \mathbf{z}, \mathbf{C}) \\
&= \mathcal{H}(K|\mathbf{p}_2 - \mathbf{z}, \mathbf{C}) + \mathcal{H}(\mathbf{p}_2|K, \mathbf{p}_2 - \mathbf{z}, \mathbf{C}) - \mathcal{H}(\mathbf{p}_2|\mathbf{p}_2 - \mathbf{z}, \mathbf{C}) \\
&= \mathcal{H}(K|\mathbf{p}_2 - \mathbf{z}, \mathbf{C})
\end{aligned}$$

[7] The Gaussian wiretap channel is introduced as a physically degraded channel in paper [13]; however, the secrecy capacity given in (1.14) still holds if the wiretap channel is stochastically degraded.

Chapter 5

Conclusions

In this work, we have investigated two information-theoretic key agreement problems: the key agreement over a discrete memoryless (DM) wiretap model, and the key agreement over a Gaussian wiretap model. The key is to be agreed between the sender (Alice) and the legitimate receiver (Bob) in presence of eavesdropper (Eve). Each model consists of a state-dependent wiretap channel with non-causal side information in parallel with a public channel. The CSI is fully known at Alice as well. The ultimate aim of this research is to characterize the key capacity as a function of (C_{P_1}, C_{P_2}), where C_{P_1} and C_{P_2} are the capacity of public channel in the forward direction and backward direction, respectively (see Definition 3.1 for more details).

For the DM model, we have derived two UBs on the forward key capacity. The UB given in Theorem 3.2 is valid for any DM model; however, the UB given in Theorem 3.3 is established for a DM model in which the wiretap channel is less noisy in Bob's favor. Each of these upper bounds is in a form of a maximization problem subject to a constraint which is imposed by the limited capacity of the public channel.

In Theorems 3.1, 3.2, and 3.3, X is drawn according to conditional PMF $\mathcal{P}_{X|US}$ when (U, S) is given, and so X is a stochastic function of (U, S) in general. Hence, $\mathcal{H}(X|U, S) \neq 0$ in general in Theorem 3.2. This fact is in contrast with the fact that X suffices to be a deterministic function of (U, S) in (1.32) according to [49]. In other words, randomized generation of transmitted signal X from (U, S) can generally enhance the key capacity, but it has no effect on the main channel capacity.

For each of the DM model and Gaussian model, we have achieved an LB on the forward key capacity, i.e., when $C_{P_2} = 0$, as a function of $C_{P_1} \in [0, \infty)$. The achievable key agreement code exploits two resources in key generation: the first resource is a random generator at Alice, and the second one is the random channel state sequences. As demonstrated in Section 2.2, the CSI is capable of enhancing the secrecy capacity of a wiretap channel with random states. However, we have shown that the positive effect of the CSI on the forward key capacity is generally more than that of the secrecy capacity because the random CSI, which contains no information about the message in the secrecy problem, is a valuable resource in key generation. Hence, the forward key capacity of a wiretap channel with random states is generally larger than its secrecy capacity. With the use of the public channel, contribution of the CSI in key generation may be intensified according to our achievable key agreement code.

For the DM model, however, the forward key capacity is *not* generally a strictly increasing function of the public channel capacity over interval $C_{P_1} \in [0, \infty)$. For every discrete memoryless state-dependent wiretap channel, there exists a finite public channel capacity $C^*_{P_1}$ such that the forward key capacity does not increase over interval $C_{P_1} \in [C^*_{P_1}, \infty)$. This capacity is determined by the wiretap channel. As a matter of fact, this saturation in key generation is due to the state alphabet set which is finite, and so the entropy rate of the CSI (as a source) can be described with a finite number of bits. Hence, the contribution of the CSI in key generation does not grow if the capacity of the public channel exceeds some finite value, i.e., $C^*_{P_1}$.

On the other hand, the achieved LB on the forward key capacity of a Gaussian model is a strictly increasing function over interval $C_{P_1} \in [0, \infty)$ at any SIR $\frac{\Gamma}{\Lambda} \in \mathbb{R}^+$ because of our simulations. This is due to the fact that the size of state alphabet set is infinite, and the entropy rate of the CSI (as a source) can not be described (error free) with a limited number of bits. Hence, the contribution of the interference (CSI) in key generation can be increased if the public message, which is correlated with the CSI, is sent over the public channel. According to our achievable scheme, the public message improves Bob's knowledge about the interference more than Eve's, and thus it regenerates the achievable key rate. However, neither Bob nor Eve can retrieve the channel state vector (without error) unless the capacity of the public channel goes to infinity. In this special case, Bob can asymptotically retrieve the channel state vector with arbitrarily small error.

When $C_{P_1} \to \infty$, the optimum solution for the key generation is to amplify the interference at the transmitter according to its maximum power Γ and to forward it to the wiretap channel. In this case, the number of codewords exponentially goes to infinity. The codewords are partitioned into subcodebooks such that each subcodebook represents a wiretap codebook of a Gaussian wiretap channel [13] with equivalent power constraint $\Gamma' = (\sqrt{\Gamma} + \sqrt{\Lambda})^2$. A codeword is selected at random among codewords which are weakly typical with the released channel state sequence at Alice. Then, the label of its subcodebook is sent over the public channel. Hence, the achievable key rate equals the securely achievable rate by the subcodebook. We have proven that this strategy achieves the forward key capacity of the Gaussian model. According to (1.14), we have calculated the forward key capacity of the Gaussian model as

$$\forall\, (\frac{\Gamma}{\Lambda}) \in \mathbb{R}^+ : \quad C_K(\infty, 0) = \frac{1}{2} \log \left(\frac{1 + \frac{\Gamma'}{\sigma_1^2}}{1 + \frac{\Gamma'}{\sigma_2^2}} \right) , \tag{5.1}$$

where σ_1^2 and σ_2^2 are noise variance of Bob's channel and that of Eve's channel, respectively.

The LB on the forward key capacity for any $C_{P_1} \in [0, \infty)$ asymptotically converges to $\frac{1}{2} \log \left(\frac{\sigma_2^2}{\sigma_1^2} \right)$ as $\Gamma \to \infty$. This forward key capacity asymptotically equals (5.1) when $\Gamma' \to \infty$. This comparison justifies that the public channel has negligible contribution in the key generation in high SIR regime. On the other hand, that forward key capacity asymptotically equals the secrecy capacity of a Gaussian wiretap channel[1] (with the same parameters and no interference) given in (1.14) as $\Gamma \to \infty$. Hence, the random generator as the first resource of key generation takes the dominant role over the interference as the second resource of key generation in high SIR regime. In low SIR regime, the second resource of key generation becomes dominant over the first one because the transmitted power is considerably less than the interference average power. In this regime, the public channel significantly assists Alice and Bob for the key generation.

In the Gaussian model, we have proven that the forward key capacity is positive if and only if Bob's channel is less noisy than Eve's channel, i.e., $\sigma_2^2 > \sigma_1^2$. When $C_{P_2} = 0$, the correlation coefficient ϱ between noise of Bob's channel and that of Eve's channel has no effect on the forward key capacity. Hence, the forward key capacity of a given Gaussian model is the same as that of an equivalent Gaussian model with a physically degraded wiretap channel (see Chapter 4 for more details). Based on this fact, a UB on the forward key capacity of the Gaussian model is

[1] Random generator is the only resource to provide security in a Gaussian wiretap channel (without interference) [13].

calculated, which equals $C_K(\infty, 0)$ as given in (5.1).

On the other hand, we have shown that the key capacity is a function of ϱ in the Gaussian model. Specially, the key capacity is zero for any $(C_{P_1}, C_{P_2}) \in \mathbb{R}^+ \times \mathbb{R}^+$ if the wiretap channel is physically degraded in Eve's favor, or equivalently

$$\forall\, (\frac{\Gamma}{\Lambda}) \in \mathbb{R}^+ : \quad C_K(\infty, \infty) = 0\,, \tag{5.2}$$

when $\varrho = \frac{\sigma_2}{\sigma_1}$. However, the key capacity is *positive* provided $\varrho < \frac{\sigma_2}{2\sigma_1}$ even if Eve's channel is less noisy than Bob's. For this special case, we have extended Maurer's strategy [17], which is given for a binary symmetric wiretap channel, to the Gaussian model.

For the achievable forward key rate of the Gaussian model, we have applied a Gaussian input distribution on the Gaussian wiretap channel in our achievable key agreement code. The input sequence $\mathbf{x}$ is correlated with the state sequence $\mathbf{s}$, where $\mathbf{x}$ is generated according to the generalized DPC strategy. The correlation coefficient ρ between X and S is a function of C_{P_1}. In other words, sequence $\mathbf{x}$ can be expressed as sum of two sequences, i.e., $\mathbf{x} = \mathbf{t} + \beta\mathbf{s}$ from (4.41). The first sequence, $\mathbf{t}$, which is (asymptotically) orthogonal to $\mathbf{s}$, conveys the information about the output of the random generator (the first resource of key generation at Alice) to the receivers. The second sequence, $\beta\mathbf{s}$, conveys the information about the interference (the second resource of key generation at Alice) to the receivers. These pieces of information assist Alice and Bob, who is suffering from less noise than Eve, in key generation as mentioned above. Although the second sequence ($\beta\mathbf{s}$) is beneficial in the key agreement problem, it is useless in Costa's problem [55]. The reason is the agreed key is dependent on the CSI, but a message is independent of the CSI, and so $\beta = 0$ in [55].

5.1 Future Work Directions

In this section, the future research directions based on this research are offered, which can be followed by ambitious researchers in the field of information-theoretic security.

5.1.1 Bounds on the Forward Key Capacity of the Gaussian Model

By simulations, we have shown that the LB on forward key capacity of the Gaussian model is a strictly increasing function of the public channel capacity. However, the mathematical proof of this conjecture is still required. Also, the LB is achieved by applying Gaussian input distribution to the wiretap channel. We have proven that this input distribution is optimal when $C_{P_1} \to \infty$ as well as when $\frac{\Gamma}{\Lambda} \to \infty$. However, the Gaussian distribution may not be optimal for all model parameters.

On the other hand, the UB on the forward key capacity is loose in low SIR regime. To derive a tighter UB, we suggest treating supremum of equation (4.91) over set of eligible values of ρ. The set of eligible ρ's is a subset of set $[0, 1]$ which is to be determined by a constraint imposed by public channel capacity C_{P_1} (when it is finite), and thus obtaining this set is our next future work.

5.1.2 Gaussian Models with Side Information at All Parties

In the Gaussian model, we assumed no SI at Bob and Eve for simplicity of calculations. An extended key agreement problem of this work is to suppose Gaussian SI correlated with the CSI at both Bob and Eve, e.g., the SI is a noisy version of the CSI. We suggest using a similar achievable key agreement scheme that is given in Chapter 4 for this new model. To do this, we can assume a Gaussian wiretap channel with augmented outputs in the same way that we applied this method in Chapter 3.

In this thesis, the CSI is also assumed to be fully known at Alice. A generalized version of this key agreement problem is to suppose that the SI at Alice is dependent on the CSI but it is not equal to the CSI (the CSI is partially known at Alice).

The generalization of the results of this thesis for the Gaussian model in which noisy versions of the interference is available at parties is offered as future work.

5.1.3 Key Agreement over Wiretap Models with a Two-Way Public Channel

The main focus of this thesis is on the forward key capacity. However, we will continue this research to characterize (bound) the key capacity of the wiretap models as a function of (C_{P_1}, C_{P_2}).

For the DM model, the achieved LB on the forward key capacity given in Theorem 3.1 is also an LB on the key capacity. When the capacity of the public channel is unlimited in both directions, the key agreement problem was studied by Khisti [73]. As explained in Section 1.8, an LB on the key capacity of this model is obtained in [73, Thm. 3]. Using the strategy of this theorem, an improved LB on the key capacity of the DM model is proposed as

$$C_K(\infty, \infty) = \max\{\max_{\mathcal{P}_{X|S}}[\mathcal{I}(X, S; \check{Y}) - \mathcal{I}(\check{Y}; \check{Z})], C_K(\infty, 0)\} , \tag{5.3}$$

where $C_K(\infty, 0)$ is given in Theorem 3.4.

For the Gaussian model, we have studied the key capacity when Eve's channel is less noisy than Bob's channel in this work. As given in Theorem 4.3, we have proven that the key capacity is a function of noise correlation coefficient ϱ. We have also calculated an LB on $C_K(\infty, \infty)$ when $\varrho \leq \frac{\sigma_2}{2\sigma_1}$. Further, for case $\varrho = \frac{\sigma_2}{\sigma_1}$, we have established that the key capacity vanishes as given in (5.2). However, bounds on the key capacity is still open for other values of ϱ as well as for finite values of C_{P_2}.

5.1.4 Reliability-Exponent and Security-Exponent of the Key Agreement

The reliability-exponent and security-exponent are defined in Definition 1.12. The reliability-exponent and security-exponent determine how fast the average probability of error $\mathcal{P}_{error}(n)$ given in (1.15) and security index $\mathcal{S}(n)$ given in (1.30) goes to zero, respectively.

In this work, we have used joint typicality (see Section 2.1.1) for encoding and decoding in Chapter 3 and Chapter 4. However, a jointly typical decoder is not optimal in the sense of minimizing the average probability of error. In fact, this suboptimal decoder can achieve the key capacity in some special cases (see Chapter 3 and Chapter 4). A maximum likelihood (ML) decoder [6, Page 72], on the other hand, is an optimal decoder to obtain minimum average probability of error. This means that the reliability-exponent[2] of a ML decoder is generally better than that of a jointly typical decoder. However, jointly typical decoding is preferred for ease of analysis as long as the block length is sufficiently large. This is why Cover and Thomas [2] developed this method for the multi-user information theory.

The lemmas used for (jointly) typicality to analyze the AR condition and the AS condition in Chapters 3 and 4 justify the asymptotic behavior of those conditions. In other words, although typical sequences are very intuitive and efficient for sufficiently large block length, they can not be used to derive the exponents [81].

Large deviation theory [2, Ch. 11] deals with small probability events with (usually) an exponentially vanishing probability. The method of types [3, 81] is a powerful tool in large deviation theory to derive the security-exponent and reliability-exponent in the DM model. According to this method, sequences with a given length n are partitioned into classes according to type (empirical distribution). Next, an (error) event can be decomposed into its intersections with the type classes. Adding the probability of these intersections, the probability of the (error) event is acquired. The exponent of this probability is determined by the exponential asymptote of the largest intersection probability because the number of the type classes increases polynomially with n (the number of sequences of each type grows exponentially with n). To bound an intersection probability, it is sufficient to bound the cardinality of the corresponding intersection because of the equiprobable property (according to a memoryless probabilistic model) of sequences inside each type class (see Sanov's theorem [2, Thm. 11.4.1] for specific details).

Motivated by Gallager [1], Tzu-Han *et al.* [82] applied an ML decoder to derive the reliability-exponent for his key agreement model.

As an extension of this work, we suggest evaluation of the reliability-exponent and the security-exponent using the demonstrated methods. Also, publications [82, 83] are constructive for this objective.

[2] When the public channel is available in the forward direction, the security-exponent is independent of the decoding rule at Bob.

Appendix A

Relation between Entropy Functions of Agreed Keys

Lemma A.1. *Let $(K, \hat{K}) \in \mathbb{K}^2$ be a pair of agreed keys that satisfy the AR condition. Also, assume that $|\mathbb{K}| < 2^{nc}$ for a positive constant c. For any $\epsilon \in (0, 1)$, there exists $n \geq N(\epsilon)$, where N is a function of ϵ, such that*

$$\frac{1}{n}|\mathcal{H}(K|T) - \mathcal{H}(\hat{K}|T)| < \epsilon \,, \tag{A.1}$$

where T is an arbitrary RV.

Proof.

$$\frac{1}{n}\mathcal{H}(K, \hat{K}|T) = \frac{1}{n}\mathcal{H}(K|T) + \frac{1}{n}\mathcal{H}(\hat{K}|K, T) \tag{A.2}$$

$$= \frac{1}{n}\mathcal{H}(\hat{K}|T) + \frac{1}{n}\mathcal{H}(K|\hat{K}, T) \tag{A.3}$$

On the other hand,

$$\frac{1}{n}\mathcal{H}(\hat{K}|K, T) \leq \frac{1}{n}\mathcal{H}(\hat{K}|K) \,, \tag{A.4}$$

$$\leq \frac{1}{n}(\mathcal{B}(\nu) + \nu \log(|\mathbb{K}| - 1)) \,, \tag{A.5}$$

$$\leq \frac{1}{n}(\mathcal{B}(\epsilon_1) + \epsilon_1 \log(|\mathbb{K}|)) \,, \tag{A.6}$$

$$\leq \frac{1}{n}(\epsilon_2 + cn\epsilon_1) \,, \tag{A.7}$$

$$\leq \epsilon \,, \tag{A.8}$$

where

- (A.5) follows from Fano's inequality [2, Thm. 2.11.1], and $\nu \triangleq Pr\{K \neq \hat{K}\}$;
- (A.6) holds due to the AR condition, i.e., $\nu < \epsilon_1$ for any $\epsilon_1 > 0$ and $n \geq N(\epsilon_1)$;
- (A.7) follows by selection of a small enough ϵ_2 such that $\mathcal{B}(\epsilon_1) < \epsilon_2$; also, it follows from assumption $|\mathbb{K}| < 2^{nc}$;
- (A.8) is valid by appropriate selection of ϵ_2 for a given $\epsilon \in (0,1)$ and $n \geq N(\epsilon) \geq N(\epsilon_1)$.

Similarly, for every $\epsilon \in (0,1)$ and $n > N(\epsilon)$ we have

$$\frac{1}{n}\mathcal{H}(K|\hat{K}, T) \leq \epsilon. \tag{A.9}$$

Hence, the proof is completed by applying equations (A.8) and (A.9) in (A.3). □

Appendix B

Existence of Maximum Values

In this appendix, we prove that the maximum value given in Theorem 3.1 exists. In fact, the lower bound given in (3.7) was originally in the form of

$$C_K(C_{P_1}, 0) \geq \sup_{\mathcal{P}_{XU|S} \in \mathbb{O}(C_{P_1})} \left[\mathcal{I}(U; \check{Y}) - \mathcal{I}(U; \check{Z})\right]. \tag{B.1}$$

However, the supremum can be replaced by the maximum due to the proof given in this appendix.

For the proof, we need the following lemmas, which are established in book [84].

Lemma B.1. *If $\mathcal{F}$ is a continuous function on a* compact set[1] $\mathbb{O}$*, then $\mathcal{F}$ has an absolute maximum and an absolute minimum on set $\mathbb{O}$.*

Lemma B.2. *If set $\mathbb{O}$ is a closed and bounded set in $\mathbb{R}^N$, where $N \in \mathbb{N}$, then set $\mathbb{O}$ is compact.*

Recalling (3.7), set $\mathbb{O}(C_{P_1})$ is a closed and bounded set in $\mathbb{R}^{|\mathbb{X}|.|\mathbb{U}|.|\mathbb{S}|}$ due to the following facts.

- Each element $\mathcal{P}_{XU|S} \in \mathbb{O}(C_{P_1})$ can be written as a vector in $\mathbb{R}^{|\mathbb{X}|.|\mathbb{U}|.|\mathbb{S}|}$. Hence, set $\mathbb{O}(C_{P_1})$ can be considered as a subset of $\mathbb{R}^{|\mathbb{X}|.|\mathbb{U}|.|\mathbb{S}|}$, where sets $\mathbb{S}$ and $\mathbb{X}$ are finite alphabet sets in the DM model as well as set $\mathbb{U}$ because $|\mathbb{U}| \leq |\mathbb{S}||\mathbb{X}| + 3$. Also, any element $\mathcal{P}_{XU|S} \in \mathbb{O}(C_{P_1})$ is bounded as $0 \leq \mathcal{P}_{XU|S} \leq 1$. Hence, set $\mathbb{O}(C_{P_1})$ is a bounded set in $\mathbb{R}^N$, where $N = |\mathbb{X}|.|\mathbb{U}|.|\mathbb{S}|$.

[1] Definition of a compact set is given in [84, Sec. 4.8]. However, we will prove compactness of set $\mathbb{O}(C_{P_1})$ by using Lemma B.2.

- Function $\mathcal{I}(U;S)-\mathcal{I}(U;\check{Y})$ is a continuous and bounded function because

  $$-\mathcal{H}(\check{Y}) \leq \mathcal{I}(U;S)-\mathcal{I}(U;\check{Y}) \leq \mathcal{H}(S)\,,$$

  where $\mathcal{H}(\check{Y})$ and $\mathcal{H}(S)$ are finite in the DM model. Also, the inequality, which is not in a strict form, in $C_{P_1}+\mathcal{I}(U;\check{Y}) \geq \mathcal{I}(U;S)$ guarantees that boundary points are included in the set $\mathbb{O}(C_{P_1})$.

As a result, set $\mathbb{O}(C_{P_1})$ is a compact set due to Lemma B.2. On the other hand, function $\mathcal{I}(U;\check{Y})-\mathcal{I}(U;\check{Z})$ is a continuous function on compact set $\mathbb{O}(C_{P_1})$, and thus it attains its absolute maximum value in the set $\mathbb{O}(C_{P_1})$ according to Lemma B.1. Hence, supremum in (B.1) can be replaced by maximum as in (3.7).

Appendix C

AEP for Gaussian Random Vectors

This appendix is originally taken from [76, Sec. 2.4] with some modifications.

Lemma C.1. *If* $\mathbf{x}$ *is a random vector drawn i.i.d. according to* $X \sim N(0, \Gamma_0)$*, then the following two statements are equivalent for any* $\epsilon \in (0, 1)$ *and* $\epsilon' = 2 \ln(2) \Gamma_0 \, \epsilon$*:*

(a) $\mathbf{x} \in \mathbb{T}_\epsilon(\mathcal{N}(0, \Gamma_0))$,

(b) $|\frac{1}{n}\mathbf{x}\mathbf{x}^t - \Gamma_0| \leq \epsilon'$.

Proof.

$$\mathbf{x} \in \mathbb{T}_\epsilon(\mathcal{N}(0, \Gamma_0)) \qquad \Leftrightarrow \tag{C.1}$$

$$\epsilon \geq | - \frac{1}{n} \log \left(\prod_{i=1}^{n} \frac{e^{\frac{-X_i^2}{2\,\Gamma_0}}}{\sqrt{2\pi\Gamma_0}} \right) - \hbar(X)| \qquad \Leftrightarrow \tag{C.2}$$

$$\epsilon \geq | - \frac{1}{n} \log \left(\frac{e^{\frac{-\sum\limits_{i=1}^{n} X_i^2}{2\,\Gamma_0}}}{(2\pi\Gamma_0)^{\frac{n}{2}}} \right) - \frac{1}{2}\log(2\pi e \Gamma_0)| \qquad \Leftrightarrow \tag{C.3}$$

$$\epsilon \geq \left| - \frac{1}{n \ln(2)} \frac{-\sum\limits_{i=1}^{n} X_i^2}{2\Gamma_0} + \frac{1}{2}\log(2\pi\Gamma_0) - \frac{1}{2}\log(2\pi e \Gamma_0) \right| \qquad \Leftrightarrow \tag{C.4}$$

$$\epsilon \geq \left| -\frac{1}{n\ln(2)} \frac{-\sum_{i=1}^{n} X_i^2}{2\,\Gamma_0} - \frac{1}{2\ln(2)} \right| \qquad \Leftrightarrow \tag{C.5}$$

$$2\ln(2)\Gamma_0\,\epsilon \geq |\frac{1}{n}\mathbf{x}\mathbf{x}^t - \Gamma_0| \qquad \Leftrightarrow \tag{C.6}$$

$$\epsilon' \geq |\frac{1}{n}\mathbf{x}\mathbf{x}^t - \Gamma_0| \qquad , \tag{C.7}$$

where equations (C.2) and (C.3) follow from definition of weak typicality [2, Sec. 8.2] and entropy of a normal distribution [2, Thm. 8.4.1], respectively. □

Lemma C.2 (Lemma 2.11 [76])**.** *Let* $\mathbf{t}$ *and* $\mathbf{s}$ *be two sequences of i.i.d. random variables* $T \sim \mathcal{N}(0, \sigma_0^2)$*, and* $S \sim \mathcal{N}(0, \Lambda)$*, respectively, such that* T *is independent of* S*. Let* $\mathbf{u} = \mathbf{t} + \alpha\mathbf{s}$ *for a constant* $\alpha \in \mathbb{R}^+$*. For any* $\epsilon \in (0, 1)$*, if* $(\mathbf{u}, \mathbf{s}) \in \mathbb{T}_\epsilon\left(\mathcal{N}\left((0,0), \begin{bmatrix} \sigma_0^2 + \alpha^2\Lambda & \alpha\Lambda \\ \alpha\Lambda & \Lambda \end{bmatrix}\right)\right)$*, then* $\mathbf{t} \in \mathbb{T}_{2\epsilon}(\mathcal{N}(0, \sigma_0^2))$*, and*

$$|\frac{1}{n}\mathbf{t}\mathbf{s}^t| \leq \frac{\epsilon\ln(2)}{\alpha}(3\sigma_0^2 + 2\alpha^2\Lambda)\,. \tag{C.8}$$

Appendix D

Csiszár-Körner's Sum Identity

This appendix contains an essential identity in the field of network information theory. This identity is credited to Csiszár and Körner according to their publication [12]. This identity is usually beneficial to derive upper bounds in several network information-theoretic problems. The identity is given as the sake of reference in the following lemma.

Lemma D.1 (Csiszár-Körner's sum identity [12])**.** *Let $\mathbb{Y}$, $\mathbb{Z}$, and $\mathbb{T}$ be alphabet sets. Assume that random vector $(Y^n, Z^n, T) \in \mathbb{Y}^n \times \mathbb{Z}^n \times \mathbb{T}$ is generated according to a joint distribution function $\mathcal{P}_{Y^n Z^n T}$. Define $Y_1^0 \triangleq Z_{n+1}^n \triangleq 0$. Then, the following identity holds:*

$$\sum_{i=1}^{n} \mathcal{I}(Z_{i+1}^n; Y_i | Y_1^{i-1}, T) = \sum_{i=1}^{n} \mathcal{I}(Y_1^{i-1}; Z_i | Z_{i+1}^n, T) \,. \tag{D.1}$$

Proof.

$$\begin{aligned}
\sum_{i=1}^{n} \mathcal{I}(Z_{i+1}^n; Y_i | Y_1^{i-1}, T) &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathcal{I}(Z_j; Y_i | Y_1^{i-1}, Z_{j+1}^n, T) && \text{(D.2)} \\
&= \sum_{j=2}^{n} \sum_{i=1}^{j-1} \mathcal{I}(Z_j; Y_i | Y_1^{i-1}, Z_{j+1}^n, T) && \text{(D.3)} \\
&= \sum_{j=2}^{n} \mathcal{I}(Z_j; Y_1^{j-1} | Z_{j+1}^n, T) && \text{(D.4)} \\
&= \sum_{i=1}^{n} \mathcal{I}(Y_1^{i-1}; Z_i | Z_{i+1}^n, T) \,, && \text{(D.5)}
\end{aligned}$$

where

- (D.2) and (D.4) follow from the chain rule for mutual information [2, Thm. 2.5.2];
- (D.3) follows from switching the order of summations;
- (D.5) follows from replacing dummy variable j with i and from the fact that $Y_1^0 = 0$.

□

References

[1] R. G. Gallager, *Information theory and reliable communication.* New York, USA: John Wiley & Sons Inc., 1968.

[2] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New Jersey, USA: John Wiley & Sons Inc., 2006.

[3] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*, 2nd ed. Cambridge, UK: Cambridge University Press, 2011.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.

[5] ——, "The zero-error-capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[6] A. El Gamal and Y. H. Kim, *Network information theory.* Cambridge, UK: Cambridge University Press, 2011.

[7] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.

[8] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography.* Crc-Press series on discrete math. and its app., 1996.

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

[10] G. S. Vernam, "Cipher printing telegraph systems for secret wire and radio telegraphic communications," *Trans. of American Ins. of Elec. Eng.*, vol. 45, pp. 295–301, 1926.

[11] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, 1975.

[12] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[13] S. K. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, 1978.

[14] S. K. Leung-Yan-Cheong, "On a special class of wiretap channels," *IEEE Trans. Inf. Theory*, vol. 23, no. 5, pp. 625–627, 1976.

[15] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2–14, 1972.

[16] J. Körner and K. Marton, "Comparison of two noisy channels," *Topics in Inf. Theory*, pp. 411–423, 1975.

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

[18] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography — Part I: Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[19] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography — Part II: CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, 1998.

[20] U. Maurer and S. Wolf, "Unconditionally secure key agreement and the intrinsic conditional information," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 499–514, 1999.

[21] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 344–366, 2000.

[22] U. Maurer and S. Wolf, "Secret key agreement over a non-authenticated channel — Part I: Definitions and bounds," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 822–831, 2003.

[23] ——, "Secret key agreement over a non-authenticated channel — Part II: The simulatability condition," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 832–838, 2003.

[24] ——, "Secret key agreement over a non-authenticated channel — Part III: Privacy amplification," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 839–851, 2003.

[25] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, 2004.

[26] A. A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals — Part I," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3973–3996, 2010.

[27] ——, "Information-theoretic key agreement of multiple terminals — Part II: Channel model," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3997–4010, 2010.

[28] H. Ahmadi and R. Safavi-Naini, "Common randomness and secret key capacities of two-way channels," *Information-Theoretic Security*, pp. 76–93, 2011.

[29] ——, "Secret keys from channel noise," *Advances in Cryptology–EUROCRYPT 2011*, pp. 266–283, 2011.

[30] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Information-theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4–5, pp. 355–580, 2009.

[31] R. Liu and W. Trappe, *Securing wireless communications at the physical layer.* Springer Verlag, 2009.

[32] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[33] U. Maurer, "Provably secure key distribution based on independent channels," in *IEEE Workshop Inform. Theory*, Eindhoven, The Netherlands, 1990.

[34] A. Khisti, S. N. Diggavi, and G. W. Wornell, "Secret-key generation with correlated sources and noisy channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Toronto, Canada, 2008, pp. 1005–1009.

[35] V. M. Prabhakaran, K. Eswaran, and K. Ramchandran, “Secrecy via sources and channels—A secret key-secret message rate tradeoff region,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Toronto, Canada, 2008, pp. 1010–1014.

[36] K. Eswaran, V. Prabhakaran, and K. Ramchandran, “Secret communication using sources and channels,” in *42nd Asilomar Conference on Signals, Systems and Computers.* Pacific Grove, California, USA: IEEE, 2008, pp. 671–675.

[37] A. El Gamal, “Capacity of the product and sum of two unmatched broadcast channels,” *Probl. Information Transmission*, p. 323, 1980.

[38] S. Nitinawarat, “Secret key generation for correlated Gaussian sources,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Toronto, Canada, 2008, pp. 702–706.

[39] S. Watanabe and Y. Oohama, “Secret key agreement from vector Gaussian sources by rate limited public communication,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Austin, Texas, USA, 2010, pp. 2597–2601.

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

[41] U. Maurer and S. Wolf, “Information-theoretic key agreement: From weak to strong secrecy for free,” in *Proc. of EUROCRYPT, Lecture Notes in Computer Science*, vol. 1807. Springer-Verlag, 2000, pp. 351–368.

[42] C. H. Bennett, G. Brassard, and J. M. Robert, “Privacy amplification by public discussion,” *SIAM journal on Computing*, vol. 17, pp. 210–229, 1988.

[43] J. L. Carter and M. N. Wegman, “Universal classes of hash functions,” *Journal of computer and system sciences*, vol. 18, no. 2, pp. 143–154, 1979.

[44] I. Csiszar and P. Narayan, “The capacity of the arbitrarily varying channel revisited: Positivity, constraints,” *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181–193, 1988.

[45] I. Csiszár and P. Narayan, “Capacity of the Gaussian arbitrarily varying channel,” *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 18–26, 1991.

[46] A. S. Cohen and A. Lapidoth, “The Gaussian watermarking game,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, 2002.

[47] A. J. Goldsmith and P. P. Varaiya, “Capacity of fading channels with channel side information,” *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, 1997.

[48] A. V. Kuznetsov and B. S. Tsybakov, “Coding in a memory with defective cells,” *Problemy Peredachi Informatsii*, vol. 10, no. 2, pp. 52–60, 1974.

[49] S. I. Geĺfand and M. S. Pinsker, “Coding for a channel with random parameters,” *Probl. of Control and Inf. Theory*, vol. 9, pp. 19–31, 1980.

[50] C. Heegard and A. El Gamal, “On the capacity of computer memories with defects,” *IEEE Trans. Inf. Theory*, 1983.

[51] C. E. Shannon, “Channels with side information at the transmitter,” *IBM J. of Research and Development*, vol. 2, no. 4, pp. 289–293, 1958.

[52] G. Keshet, Y. Steinberg, and N. Merhav, “Channel coding in the presence of side information,” *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 6, pp. 445–586, 2007.

[53] T. M. Cover and M. Chiang, “Duality between channel capacity and rate distortion with two-sided state information,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1629–1638, 2002.

[54] S. Jafar, “Capacity with causal and noncausal side information: A unified view,” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, 2006.

[55] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, 1983.

[56] U. Erez, S. Shamai, and R. Zamir, “Capacity and lattice strategies for canceling known interference,” *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, 2005.

[57] Y. Liang, K. G., H. V. Poor, and S. Shamai (Shitz), “Compound wire-tap channels,” in *Proc. 45th Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, 2007.

[58] ——, "Recent results on compound wire-tap channels," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Cannes, France, 2008.

[59] T. Liu, V. Prabhakaran, and S. Vishwanath, "The secrecy capacity of a class of parallel Gaussian compound wiretap channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Toronto, Canada, 2008, pp. 116–120.

[60] C. Mitrpant, A. J. Han Vinck, and Y. Luo, "An achievable region for the Gaussian wiretap channel with side information," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2181–2190, 2006.

[61] Y. Chen and A. J. Han Vinck, "Wiretap channel with side information," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 395–402, 2008.

[62] Y. Chen, "Wiretap channel with side information," Ph.D. dissertation, University Duisburg-Essen, Essen, Germany, 2007.

[63] W. Liu and B. Chen, "Wiretap channel with two-sided channel state information," in *Asilomar Conf. on Signals, Systems and Computers (ACSSC 2007)*. IEEE, 2007, pp. 893–897.

[64] A. Khisti, G. W. Wornell, and S. N. Diggavi, "Secret key agreement using asymmetry in channel state knowledge," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Seoul, Korea, 2009, pp. 2286–2290.

[65] Y. K. Chia and A. El Gamal, "Wiretap channel with causal state information," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Austin, Texas, USA, 2010, pp. 2548–2552.

[66] G. Reese, *Cloud application architectures.* O'Reilly Media, Inc., 2009.

[67] M. Schwartz, *Mobile wireless communications.* Cambridge University, 2005.

[68] A. Zibaeenejad, "Key agreement in state-dependent channels with non-causal side information," in *IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, 2011, pp. 658–662.

[69] ——, "Key agreement over Gaussian wiretap models with known interference at the transmitter," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Cambridge, Massachusetts, USA, 2012, pp. 2321–2325.

[70] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, 2005.

[71] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, 1998.

[72] A. Zibaeenejad and A. K. Khandani, "State-dependent wiretap models with side information," Dept. of Elec. and Comp. Eng., University of Waterloo, Waterloo, ON, Canada, Tech. Rep. 2010-08, July 2010.

[73] A. Khisti, "Secret key agreement on wiretap channel with transmitter side information," in *IEEE European Wireless (EW) Conf.*, Lucca, Italy, 2010, pp. 802–809.

[74] A. Khisti, S. Diggavi, and G. Wornell, "Secret-key agreement with channel state information at the transmitter," *IEEE Trans. Inf. Forensics and Security*, vol. 6, no. 3, pp. 672–681, 2011.

[75] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.

[76] C. Mitrpant, "Information hiding–an application of wiretap channels with side information," Ph.D. dissertation, University Duisburg-Essen, Essen, Germany, 2003.

[77] R. M. Gray, *Entropy and information theory.* New York, USA: Springer Verlag, 2010.

[78] T. M. Apostol, *Calculus, volume I*, 2nd ed. New York, USA: John Wiley & Sons Inc., 1967.

[79] Y. Liang, A. Somekh-Baruch, H. V. Poor, S. Shamai (Shitz), and S. Verdú, "Capacity of cognitive interference channels with and without secrecy," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 604–619, 2009.

[80] S. Ross, *A first course in probability*, 8th ed. New York, USA: Prentice Hall, 2009.

[81] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.

[82] C. Tzu-Han, V. Y. F. Tan, and S. C. Draper, "The sender-excited secret key agreement model: Capacity and error exponents," *submitted to IEEE Trans. Inf. Theory*, 2011, availabe online at http://arxiv.org/pdf/1107.4148v2.pdf.

[83] M. Hayashi, “Exponential decreasing rate of leaked information in universal random privacy amplification,” *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3989–4001, 2011.

[84] N. B. Haaser and J. A. Sullivan, *Real analysis.* Dover Publications, 1991.