2020

Physical-layer secrecy and privacy of wireless communication

Hien Quang Ta
Iowa State University

Physical-layer secrecy and privacy of wireless communication

by

Hien Ta

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering

Program of Study Committee:
Sang Wu Kim, Major Professor
Ahmed El-Sayed Kamal
Daji Qiao
Yong Guan
Zhengdao Wang

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2020

DEDICATION

I would like to dedicate this thesis to my family, especially to my father who inspired me with his thoughts and wisdom, to my mother who brought me virtue, and to my wife for her constant love, support, and patience.

All of this was, is and always will be for them.

TABLE OF CONTENTS

Page

LIST OF ABBREVIATIONS vi

LIST OF NOTATIONS vii

ACKNOWLEDGEMENTS ix

ABSTRACT x

CHAPTER 1. INTRODUCTION 1
1.1 Secret communication 1
1.1.1 Principle of secrecy 1
1.1.2 Recent Works 3
1.1.3 Secrecy Energy Efficiency 4
1.2 Covert communication 5
1.2.1 Covert communication under channel uncertainty and noise uncertainty . . . 5
1.2.2 Covert non-orthogonal multiple access 6
1.3 Thesis Contributions and Organizations 8

CHAPTER 2. ADAPTING RATE AND POWER FOR MAXIMIZING SECRECY ENERGY EFFICIENCY 10
2.1 System Model 10
2.2 Outage Secrecy Capacity 11
2.3 Secrecy Energy Efficiency 13
2.4 Variable Rate Variable Power Transmission 13
2.5 Suboptimal Adaptive Transmissions 15
2.5.1 Variable Rate On-Off Transmission 15
2.5.2 Variable Rate Fixed Power Transmission 15
2.5.3 Fixed Rate Variable Power Transmission 15
2.6 Numerical Results 16
2.7 Chapter Summary 18

CHAPTER 3. COVERT COMMUNICATION UNDER CHANNEL UNCERTAINTY AND NOISE UNCERTAINTY 19
3.1 System Model 19
3.2 Willie's Detection Strategy and Covert Requirement 21
3.3 Optimum Detection Threshold and Minimum total detection error probability 22
3.3.1 Perfect CSI at Willie 24
3.3.2 No CSI at Willie 26

3.3.3 AWGN Channel . . . 28
3.3.4 Numerical Results . . . 28
3.4 Covert Throughput . . . 29
3.4.1 Perfect CSI at Willie . . . 31
3.4.2 No CSI at Willie . . . 32
3.4.3 AWGN Channel . . . 33
3.4.4 Numerical Results . . . 33
3.5 Chapter Summary . . . 35

CHAPTER 4. COVERT NON-ORTHOGONAL MULTIPLE ACCESS . . . 37
4.1 System Model . . . 38
4.1.1 Achievable Rate . . . 39
4.1.2 Transmission Rate . . . 40
4.2 Willie's Detection Strategy . . . 40
4.3 Optimum Detection Threshold for Willie . . . 42
4.4 Optimum Cover Set $\mathcal{E}$ for Alice . . . 43
4.5 Maximum Total Detection Error Probability . . . 44
4.6 Decoding Outage Probability . . . 46
4.7 Covert Rate . . . 47
4.8 Non-covert Rate . . . 48
4.9 Channel Adaptation . . . 48
4.9.1 Maximum Total Detection Error Probability . . . 49
4.9.2 Decoding Error Probability . . . 50
4.9.3 Covert Rate . . . 51
4.9.4 Non-covert Rate . . . 52
4.10 Multiple Antenna at Transmitter . . . 52
4.10.1 Optimum Antenna Selection for Alice . . . 52
4.10.2 Maximum total detection error probability . . . 53
4.10.3 Decoding Outage Probability . . . 54
4.11 Numerical Results . . . 54
4.11.1 Maximum total detection error probability . . . 54
4.11.2 Covert rate . . . 56
4.12 Chapter Summary . . . 59

CHAPTER 5. CONCLUSIONS AND FUTURE WORK . . . 60
5.1 Conclusions and Contributions . . . 60
5.2 Future Work . . . 61
5.2.1 AN-aided covert NOMA . . . 62
5.2.2 Cooperative covert NOMA . . . 63

BIBLIOGRAPHY . . . 64

APPENDIX A. ADAPTING RATE AND POWER FOR MAXIMIZING SECRECY ENERGY EFFICIENCY . . . 70

APPENDIX B. COVERT COMMUNICATION UNDER CHANNEL UNCERTAINTY AND NOISE UNCERTAINTY 72
B.1 Equation (3.19) derivation 72
B.2 Proof of pseudo-convexity 72
B.3 Equation (3.38) derivation 74
B.4 Range of interest derivation 74
B.5 Equation (3.55) derivation 75

APPENDIX C. COVERT NON-ORTHOGONAL MULTIPLE ACCESS 77
C.1 Equation (4.37) derivation 77
C.2 Proof of optimum hiding strategy 78
C.3 Equations (4.58)-(4.63) derivation 80
C.4 Equation (4.64) derivation 81
C.5 Equation (4.67) derivation 82
C.6 Equation (4.78) derivation 83
C.7 Equation (4.79) derivation 85
C.8 Equation (4.80) derivation 85

LIST OF ABBREVIATIONS

a.k.a	Also known as
i.e.	That is
e.g.	For example
i.i.d	Independent and Identically Distributed
s.t.	Subject to
w.r.t	With respect to
Hz	Hertz
sec	second
b/s	bits per second
dB	Decibel
SNR	Signal-to-noise ratio
SINR	Signal-to-interference plus noise ratio
AN	Artificial noise
SEE	Secrecy Energy Efficiency
AWGN	the additive white Gaussian noise
DMC	discrete memoryless channel
GSM	Global system for mobile
PDF	Probability density function
CDF	Cumulative density function
CSI	Channel state information
SISOSE	Single-input single-output single-antenna eavesdropper
NOMA	Non-orthogonal multiple access
OMA	Orthogonal multiple access

LIST OF NOTATIONS

$\mathbf{x}$	Boldface lowercase letter denote vector
$\mathbf{X}$	Boldface uppercase letter denotes matrix
$\mathbf{X}^*$	The conjugate of matrix $\mathbf{X}$
$\mathbf{X}^T$	The transpose of matrix $\mathbf{X}$
$\mathbf{X}^H$	The Hermitian transpose of matrix $\mathbf{X}$
$\|x\|$	The absolute value of the scalar x
$\|\| \cdot \|\|$	The Frobenius norm of a vector or a matrix
$[x]^+$	$\max(0, x)$
$\int$	The integral operator
$\sum$	The summation operator
$\mathcal{A}$	The notation of set
$\mathcal{A}^c$	The complement of the set $\mathcal{A}$
$\emptyset$	The empty set
$\cap$	Set intersection
$\cup$	Set union
$\in$	Member of
$\notin$	Not member of
$d^n y/dx^n$	The n-th derivatives of y with respect to x
$[\cdot, \cdot]$	The range of value
$E[\cdot]$	The expectation operator
$CN(m, \sigma^2)$	The complex circularly symmetric Gaussian random variable with mean m and variance σ^2
$\Pr(\cdot)$	The probability

$C_X(\cdot)$	The capacity at node X
$\hat{x}$	The estimated value of scalar x
$\tilde{x}$	The estimation error of scalar x
$f_X(\cdot)$	The probability density function of X
W_0 and W_{-1}	The Lambert-W function with branch 0 and -1, respectively
$\ln(\cdot)$	The natural logarithm
$\log_2(\cdot)$	The logarithm in base two
$\exp(\cdot)$	Exponential function
$\Gamma(\cdot,\cdot)$	Incomplete Gamma function
$B(n,m)$	Beta function
$B(x;n,m)$	Incomplete Beta function

ACKNOWLEDGEMENTS

I am grateful to all people that I have met during my journey of Ph.D. I would like to thank my supervisor Prof. Sang Wu Kim for his continuing guidance, support and encouragement over the past years. I am always inspired by his motivation, wide knowledge and patience. Without his careful supervision, my academic progress would not been possible. I would also like to express my thanks to my friends who helped me with various aspects of conducting research and the writing of this thesis.

ABSTRACT

The motivation of this thesis is to contribute to the improvement of the physical-layer secrecy and privacy of wireless communication. Firstly, the rate and power adaptation technique is investigated to improve the energy efficiency of the physical-layer secrecy. We present the optimum rate and power adaptation rule that maximizes the average secrecy energy efficiency (SEE) subject to an average transmission power constraint. The SEE is defined as the outage secrecy capacity, the largest secrecy rate, such that the outage probability is less than a certain value, divided by the total power consumption (bits per joule). We also characterize the SEE gain provided by varying the rate and/or the power, and discuss the impact of the number of antennas on the optimum adaptation rule. Secondly, the joint impact of imperfect knowledge of the channel gain (channel uncertainty) and noise power (noise uncertainty) at the adversary is investigated to improve the physical-layer privacy. We characterize the covert throughput gain provided by the channel uncertainty as well as the covert throughput loss caused by the channel fading as a function of the noise uncertainty. We also show the impact the channel uncertainty on the total detection error probability and the covert throughput. Our result shows that the channel fading is crucial to hiding the signal transmission, particularly when the noise uncertainty is low and/or the receive SNR is high. The impact of the channel uncertainty on the total detection error probability and the covert throughput is more significant when the noise uncertainty is larger. Finally, hiding a covert (private) message in non-orthogonal multiple access (NOMA) systems by superimposing (embedding) it under other messages is proposed. We determine the total detection error probability (sum of false alarm and missed detection probability), the adversary's optimum detection strategy that minimizes the total detection error probability, and the communicator's optimum message hiding strategy that maximizes the total detection error probability. Additionally, we explore exploiting the channel variations to further increase the total detection error probability. We show that the

total detection error probability increases and converges to 1 as the number of users increases and that the total detection error probability, hence the covert rate, can be increased by increasing the transmission power when the channel variation is exploited.

CHAPTER 1. INTRODUCTION

A tremendously increasing demand of confidential and private data transmission has made security a pivotal issue in the current wireless system network. In tradition, security is realized by cryptography technique using secret key [47]. However, the cryptography technique has showed vulnerable to secure information as it relies on assumption of infinite computational capability at the adversary. It also requires high computational complexity and bandwidth and significant challenges to secret key distribution and management in large-scale decentralized wireless network has been recently addressed in [42]. Therefore, physical-layer security (also known as information-theoretic security) has become attracted as an alternative security solution or additional layer of security [53].

While many studies address security in physical layer by limiting information leaked to the adversary [10], a.k.a secret communication, the threat to users' privacy from the discovery of the existence of the message has not been mitigated. Covert or low probability of detection communication is crucial to protect user privacy and provide a strong security. It has great implications for many practical applications ranging from covert military and national security operations to privacy protection for users of commercial wireless networks. In this chapter, we, therefore, present the overview and recent works of both secret communication and covert communication. We also mention our contributions and organization of the thesis.

1.1 Secret communication

1.1.1 Principle of secrecy

The principle of secret communication was established by Wyner as a single-input single-output single-antenna eavesdropper (SISOSE) model [65] in Fig. 1.1. The source transmits a confidential message W to legitimate receiver in the presence of eavesdropper. The message W is encoded into

n symbols presented as an n-vector X^n. Y^n and Z^n denote the the received signals at the legitimate receiver and eavesdropper, respectively. As Shannon's notation, a perfect secrecy requires

$$I(W, Z^n) = 0, \tag{1.1}$$

where $I(W, Z^n)$ denotes the mutual information between W and Z^n. Different from Shannon, Wyner introduced the wiretap code (R_b, R_s), where $R_b = \frac{1}{n}H(X^n)$ is the transmission rate and $R_s = \frac{1}{n}H(W)$ is the confidential information rate, and the notion of secrecy as

$$\frac{1}{n}I(W, Z^n) = 0. \tag{1.2}$$

Wyner showed that the secrecy exists only if the wiretap channel between source and eavesdropper is a degraded version of the main channel between source and the legitimate receiver.

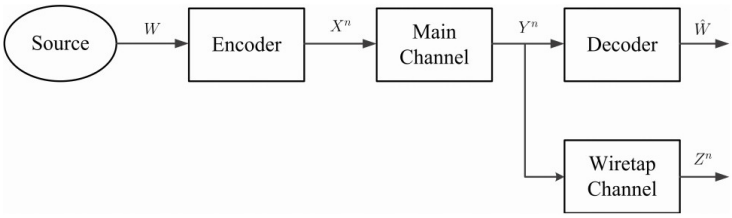

Figure 1.1 The wiretap channel of Wyner [65].

Later, Csiszar and Korner in [19] characterized the secrecy capacity for the case where the main and wiretap channels are independent as

$$C_s = \max_{V \to X \to YZ} I(V;Y) - I(V;Z), \tag{1.3}$$

where V is an input variable. [19] also proved that there exist channel codes guaranteeing both robustness to transmission errors and a prescribed degree of data confidentiality. Then, in [38], the secrecy capacity under AWGN channels is characterized as

$$C_s = [C_b - C_e]^+, \tag{1.4}$$

where C_b and C_e denote the capacity of the main channel and the wiretap channel, respectively. Thereby, to achieve a positive secrecy capacity, it is required in (1.4) that the quality of main channel is better than that of wiretap channel.

1.1.2 Recent Works

The performance of secret communication can be measured in terms of the secrecy throughput, which is the capacity of conveying information to the intended users while keeping it confidential from eavesdroppers. There are two well-known secrecy measurements, namely, ergodic secrecy capacity in fast fading [41, 40] and secrecy outage capacity in slow fading [11].

Recent techniques in secret communication can be listed as

1. Rate and power adaptation for maximizing the average secrecy capacity was presented in [27]. More specifically, [27] concerned the joint rate and power adaptation transmission scheme for the fast fading in order to maximize the ergodic secrecy rate subject to average power constraint.

2. Artificial Noise (AN) degrading the ability of the eavesdropper to intercept the signal destined for the intended recipient was proposed in [25]. AN systems use a fraction of power to transmit data (information) and allocate the remaining power to transmit AN. The high-SNR performance of this type of technique was shown to be nearly optimal [35], and the optimal power allocation between the information signal and AN was examined in [75]. The optimal power allocation between the information signal and AN that maximizes the secrecy throughput have been analyzed in [73, 67, 66]. While the traditional AN is designed based on mutiple antennas at transmitter, an injection AN scheme for single-antenna transmitter has been proposed in [30].

3. Cooperative and jamming relay was studied in [62]. Also, interference channels, multiple access channels or multi-user broadcast channels for secrecy is also studied in [62, 49].

4. Since the secrecy performance heavily relies on the level of CSI knowledge at transmitter, receiver and adversary, some strategies proposed for pilot training transmission to prevent CSI leaked to the adversary has been considered in [43] via the reverse training strategy in a time division duplex system, i.e. pilot transmitted at the legitimate receiver.

For the slow fading, to better assist the system design, an revised secrecy outage formula were developed in [77] to distinguish the security and reliability. [77] also analyzed the fixed rate fixed power transmission scheme that maximizes the average secrecy throughput subject to delay and fixed average secrecy outage probability constraints. [24] proved the optimal beamforming that maximizes the instantaneous secrecy rate and then, derived the maximum instantaneous secrecy rate subject to a fixed secrecy outage probability constraint. Next, the joint rate and power adaptation strategy that maximizes the average secrecy throughput subject to fixed secrecy outage probability and average power constraints is studied in [48] for the single antenna case. Finally, the AN scheme with variable rate and variable power allocation between the information signal and AN transmission that maximizes the average secrecy throughput subject to fixed secrecy outage probability has also been analyzed in [73, 67, 66].

1.1.3 Secrecy Energy Efficiency

As energy use and costs for communications continue to rise, the energy efficiency of secret communications, called secrecy energy efficiency (SEE), is emerging as another important figure-of-merit. SEE is determined by the number of reliably and securely transmitted bits per unit energy (bits/Joule). In recent years, several studies have been conducted to maximize the SEE which can be defined in different ways. In [17], the SEE is defined as the outage secrecy capacity (b/s) divided by the power consumption, while in [33, 71, 70], it is defined as the secrecy capacity (b/s) divided by the power consumption and in [50] as the average outage secrecy capacity divided by the average power consumption. Authors in [17, 33] developed iterative power control algorithms for maximizing the instantaneous SEE. [71] determined the optimal transmit power and beamforming vector that maximizes the instantaneous SEE depending on the availability of channel state information of the eavesdropper. [70] developed a power control algorithm that maximizes the ergodic SEE by means of fractional programming and sequential convex optimization tools. [50] developed an iterative resource allocation algorithm for maximizing the SEE in OFDMA system. In addition, the impact of artificial noise on the SEE is analyzed in [70, 50].

Although the rate and power adaptation strategy has been considered in secrecy, it has not been done in secrecy energy efficiency. In the chapter 2, we, therefore, consider *joint* rate and power adaptation for maximizing secrecy energy efficiency under the constraint of secrecy outage and average transmission power.

1.2 Covert communication

1.2.1 Covert communication under channel uncertainty and noise uncertainty

The broadcast nature of the wireless medium allows wireless networks to be easily monitored, which creates a serious concern about the privacy of wireless communications. The vast majority of research in the past has focused on protecting the message content through cryptography [52] or physical-layer security [10]. While these approaches address security in many domains by protecting the content of the message, they do not mitigate the threat to users' privacy from the discovery of the existence of the message. Covert or low probability of detection communication is crucial to protect user privacy and provide a strong security. It has great implications for many practical applications ranging from covert military and national security operations to privacy protection for users of commercial wireless networks.

Covert communication is governed by the square root law (SRL): $\mathcal{O}(\sqrt{N})$ bits can be reliably transmitted in N channel uses without being detected by the adversary; transmission of more bits results in either detection or uncorrectable decoding errors. The SRL was first proven for the classical wireless channels subject to the additive white Gaussian noise (AWGN) [9], with follow-on works extending this result to discrete memoryless channels (DMCs) and fully characterizing the constant hidden by the Big-O notation [15, 12, 64].

Other extensions have attempted to identify scenarios in which the SRL may be overcome. For instance, authors in [61] showed that robust detection of signal transmission is impossible, even if the detector takes an infinite number of samples, if the detector's signal-to-noise ratio (SNR) is below a threshold, known as the SNR wall [61]. This SNR wall, caused by the inherent mismatch between the true noise power and its estimate, called noise uncertainty, can be leveraged to hide the

signal transmission. In the realistic situation of uncertain knowledge of the noise power, a positive covert rate, i.e. reliable transmission of $\mathcal{O}(N)$ bits in N channel uses, is possible while guaranteeing that the adversary cannot detect the signal transmission [36, 37]. The idea of exploiting the noise uncertainty was extended to jamming the adversary by varying the power [59]. Most recently, [31] has examined the impact of noise uncertainty on covert communication by considering two practical uncertainty models. Other works have analyzed covert communication under block fading channels, where users experience uncertainty about their channel knowledge [57].

In the chapter 3, we, therefore, analyze the *joint* impact of imperfect knowledge of the channel gain (channel uncertainty) and noise power (noise uncertainty) at the adversary on the total detection error probability (sum of the probability of false alarm and missed detection) and the covert throughput in Rayleigh fading channel. We determine the optimum detection threshold for the energy detector that minimizes the total detection error probability as a function of the channel gain estimate. Then, we determine the maximum allowed transmission power for the total detection error probability to be no less than a threshold. Based on this, we determine the maximum average transmission rate (bits/s/Hz) subject to a covert communication constraint, hereafter referred to as the covert throughput. We characterize the covert throughput gain provided by imperfect knowledge of the channel gain and noise power at the adversary and the covert throughput loss caused by the channel fading as a function of the noise uncertainty.

1.2.2 Covert non-orthogonal multiple access

Non-orthogonal multiple access (NOMA) allows users to share the same spectrum and thus is envisaged to address high spectral efficiency challenge in the fifth generation (5G) networks. In [21], the authors investigated the spectrum efficiency of NOMA and its potential gain over the conventional orthogonal multiple access (OMA). An extension to multiple input multiple output (MIMO) was considered in [60]. More recently, the secrecy aspects of NOMA systems have been studied. The secrecy sum rate has been investigated for single input single output NOMA [29, 74] and multiple antenna NOMA [39, 45]. However, to the best of our knowledge, the privacy aspect

of NOMA has not been studied. Covert or low probability of detection (LPD) communication is crucial to protect user privacy and provide a strong security for many practical applications ranging from military and national security operations to privacy protection for users of commercial wireless networks.

Recent study of covert communication has considered embedding the covert signal into an existing non-covert transmission [5, 32]. More specifically, an information theoretic analysis of embedding the covert signal in an innocent signal transmission has been developed in [5]. This work is motivated by [23] where a dirty constellation (hardware imperfection) is exploited to hide the transmission of information. Other work considered covertly sending a covert message in amplify-and-forward relay network while forwarding the source message to the destination [32]. Although the secrecy has been heavily studied in NOMA system, there is no prior work considering the privacy (covertness) aspect of NOMA system.

In the chapter 4, we, therefore, study the privacy (covertness) aspect of NOMA system. The covert message is superimposed onto K non-covert (public) messages in NOMA system such that the total transmission power remains the same whether or not the covert message is transmitted. We show that the covert message can be detected only when the non-covert message, where the covert message is superimposed onto, can be decoded. This suggests hiding the covert message under the non-covert message that is most difficult to decode. Hence, the effectiveness of hiding the covert message can be improved by exploiting the multiplicity of users in NOMA system. We determine the total detection error probability (sum of false alarm and missed detection probability) as a function of the number of users in Rayleigh fading channel. We show that it increases and converges to 1 as the number of non-covert users increases. This means that the covert transmission is undetectable if the number of non-covert users is sufficiently large. We also show that the total detection error probability can be increased as the transmit power is increased, thereby increasing the covert rate, by adapting the superposition rule to the channel variations.

1.3 Thesis Contributions and Organizations

The contributions are organized as follows

- In chapter 2, we present the optimum rate and power adaptation rule that maximizes the average SEE subject to an average power constraint. We compare the average SEE provided by the optimum rate and power adaptation with that provided by three suboptimal transmission rules: variable rate on-off transmission, variable rate fixed power transmission, and fixed rate variable power transmission. We characterize the SEE gain provided by varying the rate and/or the power, and discuss the impact of the number of antennas and the circuit power consumption on the optimum adaptation rule.

- In chapter 3, we show that the channel fading is crucial to hiding the signal transmission, particularly when the noise uncertainty is low and/or the receive SNR is high. We determine the optimum detection threshold for the energy detector that minimizes the total detection error probability as a function of the channel gain estimate. Then, we determine the maximum allowed transmission power for the total detection error probability to be no less than a threshold. Based on this, we determine the maximum average transmission rate (bits/s/Hz) subject to a covert communication constraint, hereafter referred to as the covert throughput. We characterize the covert throughput gain provided by imperfect knowledge of the channel gain and noise power at the adversary and the covert throughput loss caused by the channel fading as a function of the noise uncertainty. Our analysis shows that the channel fading is crucial to hiding the signal transmission, particularly when the noise uncertainty is low and/or the receive SNR is high. The impact of the channel uncertainty on the total detection error probability and the covert throughput is particularly noticeable when the noise uncertainty is large. The channel uncertainty provides a covert throughput gain of 12% $\sim$ 19% over the case that perfect channel knowledge is available at the adversary when the noise uncertainty is in the range of 1 $\sim$ 2 dB. However, if the noise uncertainty is small, the channel uncertainty does not help much increase the total detection error probability and the covert throughput.

- In chapter 4, we study the physical-layer privacy (covertness) in NOMA system. The covert message is superimposed onto K non-covert (public) messages in NOMA system such that the total transmission power remains the same whether or not the covert message is transmitted. We show that the covert message can be detected only when the non-covert message, where the covert message is superimposed, can be decoded. This suggests hiding the covert message under the non-covert message that is most difficult to decode. Hence, the effectiveness of hiding the covert message can be improved by exploiting the multiplicity of users in NOMA system. We determine the total detection error probability (sum of false alarm and missed detection probability) as a function of the number of users in Rayleigh fading channel. We show that it increases and converges to 1 as the number of non-covert users increases. This means that the covert transmission is undetectable if the number of non-covert users is sufficiently large. We also show that the total detection error probability can be increased as the transmit power is increased, thereby increasing the covert rate, by adapting the superposition rule to the channel variations.

The remainder of this report is organized as follows

- Chapter 2 considers the problem of adapting power and rate to maximize the secrecy energy efficiency.
- Chapter 3 studies *joint* impact of imperfect knowledge of the channel gain and noise power on the detection error probability at the adversary and the covert throughput in Rayleigh fading channel.
- Chapter 4 studies the physical-layer privacy (covertness) in NOMA system.
- Chaper 5 concludes this work and outline the main contributions. Future work is also presented.

CHAPTER 2. ADAPTING RATE AND POWER FOR MAXIMIZING SECRECY ENERGY EFFICIENCY

In this chapter, we present the optimum rate and power adaptation rule that maximizes the average SEE subject to an average power constraint, where the SEE is defined as the outage secrecy capacity divided by the power consumption (bits per Joule). We compare the average SEE provided by the optimum rate and power adaptation with that provided by three suboptimal transmission rules: variable rate on-off transmission, variable rate fixed power transmission, and fixed rate variable power transmission. We characterize the SEE gain provided by varying the rate and/or the power, and discuss the impact of the number of antennas on the optimum adaptation rule.

The rest of this chapter is organized as follows. Section 2.1 describes the system model. Section 2.2 derives secrecy outage capacity. Section 2.3 develops the adaptive power and rate transmission scheme and derives the average secrecy-energy efficiency. Section 2.6 shows the numerical results and section 2.7 concludes the chapter.

2.1 System Model

We consider sending secret information from a transmitter (Alice) equipped with N antennas to a receiver (Bob) equipped with one antenna in the presence of an eavesdropper (Eve) equipped with one antenna. The transmitted signal vector is given by

$$\mathbf{x} = \mathbf{w}u, \tag{2.1}$$

where $\mathbf{w}$ is the unit norm ($\|\mathbf{w}\|^2 = 1$) precoding vector and $u \sim CN\left(0, \sigma_u^2\right)$ is the information signal, where $CN\left(m, \sigma^2\right)$ denotes the complex circularly symmetric Gaussian random variable with mean m and variance σ^2.

The received signals at Bob and Eve are given by

$$y_B = \mathbf{h}^T\mathbf{x} + n_B = \mathbf{h}^T\mathbf{w}u + n_B \tag{2.2}$$

$$y_E = \mathbf{g}^T\mathbf{x} + n_E = \mathbf{g}^T\mathbf{w}u + n_E, \tag{2.3}$$

respectively, where $\mathbf{h} \sim CN\left(0, \sigma_h^2\mathbf{I}_N\right)$ denotes the channel gain vector between Alice and Bob, $\mathbf{g} \sim CN\left(0, \sigma_g^2\mathbf{I}_N\right)$ denotes the channel gain vector between Alice and Eve, and n_B and n_E denote the complex Gaussian noise at Bob and Eve, respectively. We assume n_B and n_E are independent with mean zero and variance σ_n^2.

We assume that both the main channel (Alice to Bob) and the eavesdropper's channel (Alice to Eve) are quasi-static fading channels. That is, the fading coefficients, albeit random, are constant during the transmission of an entire codeword and independent from codeword to codeword. This corresponds to a situation where the coherence time of the channel is large. We assume that Bob has perfect knowledge of $\mathbf{h}$ from the pilot signal sent by Alice and that $\mathbf{h}$ is fedback to Alice for adaptation. We also assume that Eve has perfect knowledge of $\mathbf{h}$ and $\mathbf{g}$. These assumptions are realistic for the slow-fading wireless environment under consideration.

2.2 Outage Secrecy Capacity

The channel capacity (b/s/Hz) between Alice and Bob and that between Alice and Eve are given by

$$C_B(\mathbf{h}, \mathbf{w}) = \log_2\left(1 + |\mathbf{h}^T\mathbf{w}|^2\sigma_u^2/\sigma_n^2\right) \tag{2.4}$$

$$C_E(\mathbf{g}, \mathbf{w}) = \log_2\left(1 + |\mathbf{g}^T\mathbf{w}|^2\sigma_u^2/\sigma_n^2\right) \tag{2.5}$$

Since $|\mathbf{g}^T\mathbf{w}|^2$ has exponential distribution with mean σ_g^2, the outage probability for a given secrecy rate R_S is

$$P_O(\mathbf{h},\mathbf{w}) = \Pr\left(C_B(\mathbf{h},\mathbf{w}) - C_E(\mathbf{g},\mathbf{w}) < R_S\right) \tag{2.6}$$

$$= \Pr\left(|\mathbf{g}^T\mathbf{w}|^2 > \frac{2^{-R_S}\left(1+|\mathbf{h}^T\mathbf{w}|^2\frac{\sigma_u^2}{\sigma_n^2}\right)-1}{\sigma_u^2/\sigma_n^2}\right) \tag{2.7}$$

$$= \exp\left(-\frac{2^{-R_S}(1+|\mathbf{h}^T\mathbf{w}|^2\sigma_u^2/\sigma_n^2)-1}{\sigma_g^2\sigma_u^2/\sigma_n^2}\right). \tag{2.8}$$

This is the probability that secrecy condition, which depends on $\mathbf{g}$, is not satisfied because Alice does not know $\mathbf{g}$ and thus, her transmission is independent of $\mathbf{g}$.

For $P_O(\mathbf{h},\mathbf{w}) < \epsilon$, we require

$$R_S < \left[\log_2\left(\frac{1+|\mathbf{h}^T\mathbf{w}|^2\sigma_u^2/\sigma_n^2}{1+\sigma_g^2\ln(\epsilon^{-1})\,\sigma_u^2/\sigma_n^2}\right)\right]^+ \tag{2.9}$$

$$:= R_S(\mathbf{h},\mathbf{w}) \tag{2.10}$$

where $[\cdot]^+ := \max(\cdot, 0)$. The largest secrecy rate $R_S(\mathbf{h},\mathbf{w})$ such that $P_O(\mathbf{h},\mathbf{w}) < \epsilon$ is called outage secrecy capacity.

$R_S(\mathbf{h},\mathbf{w})$ is maximized when $\mathbf{w} = \mathbf{h}^*/\|\mathbf{h}\|$, which yields

$$R_S(\gamma) = \left[\log_2\left(\frac{1+\|\mathbf{h}\|^2\sigma_u^2/\sigma_n^2}{1+\sigma_g^2\ln(\epsilon^{-1})\,\sigma_u^2/\sigma_n^2}\right)\right]^+ \tag{2.11}$$

$$= \left[\log_2\left(\frac{1+\gamma P(\gamma)}{1+\bar{\gamma}\sigma_g^2\ln(\epsilon^{-1})\,P(\gamma)}\right)\right]^+ \tag{2.12}$$

where $\gamma := \|\mathbf{h}\|^2\bar{\gamma}$ is the instantaneous received signal-to-noise ratio (SNR) at Bob, $P(\gamma) = \sigma_u^2/P$ is the power adaptation rule, and $\bar{\gamma} := P/\sigma_n^2$ is the average transmit SNR. The probability density function (PDF) of γ is given by

$$f(\gamma) = \frac{\gamma^{N-1}e^{-\gamma/\bar{\gamma}_B}}{(N-1)!\bar{\gamma}_B^N},\ \gamma \geq 0, \tag{2.13}$$

where $\bar{\gamma}_B = \sigma_h^2\bar{\gamma}$.

2.3 Secrecy Energy Efficiency

The instantaneous SEE (bits/J) subject to the secrecy outage constraint of $P_O(\mathbf{h},\mathbf{w}) < \epsilon$ is given by

$$\zeta(\sigma_u^2,\mathbf{h},\mathbf{w}) = \frac{B\cdot R_S(\mathbf{h},\mathbf{w})}{P_C+\sigma_u^2/\mu} \tag{2.14}$$

$$= \frac{B\left[\log_2\left(\frac{1+|\mathbf{h}^T\mathbf{w}|^2\sigma_u^2/\sigma_n^2}{1+\sigma_g^2\ln(\epsilon^{-1})\sigma_u^2/\sigma_n^2}\right)\right]^+}{P_C+\sigma_u^2/\mu}, \tag{2.15}$$

where B is the bandwidth, P_C is the circuit power and μ is the power amplifier efficiency. The circuit power, P_C, is given by NP_A+P_B, where P_A is the circuit power consumption per antenna and P_B is the basic circuit power used by the transmitter [51]. Therefore, the instantaneous SEE is maximized when $\mathbf{w}=\mathbf{h}^*/\|\mathbf{h}\|$, which yields

$$\zeta(P(\gamma),\gamma) = \frac{B\left[\log_2\left(\frac{1+\gamma P(\gamma)}{1+\alpha P(\gamma)}\right)\right]^+}{P_C+P\cdot P(\gamma)/\mu}, \tag{2.16}$$

where $\alpha=\bar{\gamma}\sigma_g^2\ln\left(\epsilon^{-1}\right)$.

2.4 Variable Rate Variable Power Transmission

In this section, we determine the optimum power adaptation rule that maximizes the average SEE subject to an average transmission power constraint. Our optimization problem is

$$\max_{P(\gamma)} \quad \zeta=\int_\alpha^\infty \frac{B\log_2\left(\frac{1+\gamma P(\gamma)}{1+\alpha P(\gamma)}\right)}{P_C+P\cdot P(\gamma)/\mu}f(\gamma)d\gamma \tag{2.17}$$

$$\text{subject to} \quad \int_0^\infty P(\gamma)\,f(\gamma)\,d\gamma\le 1. \tag{2.18}$$

Since $\zeta(P(\gamma),\gamma)$ is a ratio of strictly concave function and positive affine function for $\gamma>\alpha$, $\zeta(P(\gamma),\gamma)$ is a strictly pseudo-concave function of $P(\gamma)$ for $\gamma>\alpha$ [14]. Hence, there should exist a unique maxima. Let $P^*(\gamma)$ denote the power adaptation rule that maximizes the instantaneous SEE, $\zeta(P(\gamma),\gamma)$, i.e. the solution of

$$\frac{\partial\zeta(P(\gamma),\gamma)}{\partial P(\gamma)} = \frac{\frac{B\gamma}{1+\gamma P(\gamma)}-\frac{B\alpha}{1+\alpha P(\gamma)}}{\left(P_C+\frac{P\cdot P(\gamma)}{\mu}\right)\ln 2}-\frac{\frac{P}{\mu}B\ln\left(\frac{1+\gamma P(\gamma)}{1+\alpha P(\gamma)}\right)}{\left(P_C+\frac{P\cdot P(\gamma)}{\mu}\right)^2\ln 2}$$

$$= 0, \tag{2.19}$$

for $\gamma > \alpha$[1].

1) $E[P^*(\gamma)] \leq 1$: If $E[P^*(\gamma)] \leq 1$, then the optimal power adaptation rule, $P_{PR}(\gamma)$, that maximizes the average SEE, ζ, is equal to $P^*(\gamma)$.

2) $E[P^*(\gamma)] > 1$: If $E[P^*(\gamma)] > 1$, then $P_{PR}(\gamma)$ should be in the range of $[0, P^*(\gamma)]$. This is because $E[P_{PR}(\gamma)] > 1$, hence (2.18) is not satisfied, if $P_{PR}(\gamma) > P^*(\gamma)$. It can be shown that $\partial^2\zeta(P(\gamma),\gamma)/\partial P(\gamma)^2 < 0$ for $P(\gamma) \in [0, P^*(\gamma)]$ and $\gamma > \alpha$. Hence, the optimization problem in (2.17) can be solved by convex optimization.

Theorem: The solution of the optimization problem in (2.17) under the constraints of (2.18) and $0 \leq P(\gamma) \leq P^*(\gamma)$ is given by

$$P_{PR}(\gamma) = \begin{cases} P^{\dagger}(\gamma), & 0 \leq \lambda \leq B(\gamma-\alpha)/(P_C \ln 2) \\ 0, & \lambda > B(\gamma-\alpha)/(P_C \ln 2), \end{cases} \tag{2.20}$$

for some constant λ, where $P^{\dagger}(\gamma)$ is the solution of $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma) = \lambda$:

$$\frac{\frac{B\gamma}{1+\gamma P(\gamma)} - \frac{B\alpha}{1+\alpha P(\gamma)}}{\left(P_C + \frac{P\cdot P(\gamma)}{\mu}\right)\ln 2} - \frac{\frac{P}{\mu} B \ln\left(\frac{1+\gamma P(\gamma)}{1+\alpha P(\gamma)}\right)}{\left(P_C + \frac{P\cdot P(\gamma)}{\mu}\right)^2 \ln 2} = \lambda. \tag{2.21}$$

The proof of (2.20) is provided in Appendix A.

Since $\zeta(P(\gamma),\gamma)$ is strictly concave and has a unique maxima, it is an increasing function of $P(\gamma)$ for $P(\gamma) \in [0, P^*(\gamma)]$. Therefore, λ is determined from the average power constraint:

$$\int_0^{\infty} P_{PR}(\gamma) f(\gamma) d\gamma = 1. \tag{2.22}$$

3) Summary: From 1) and 2), the optimal power adaptation rule, $P_{PR}(\gamma)$, that maximizes the average SEE, ζ, subject to an average transmission power constraint of (2.18) is given by

$$P_{PR}(\gamma) = \begin{cases} P^*(\gamma), & E[P^*(\gamma)] \leq 1 \\ P^{\dagger}(\gamma), & E[P^*(\gamma)] > 1, 0 \leq \lambda \leq \frac{B(\gamma-\alpha)}{P_C \ln 2} \\ 0, & E[P^*(\gamma)] > 1, \lambda > \frac{B(\gamma-\alpha)}{P_C \ln 2}. \end{cases} \tag{2.23}$$

[1]Since $\zeta(P(\gamma),\gamma)$ is 0 for $\gamma \leq \alpha$, $P^*(\gamma)$ should be 0 for $\gamma \leq \alpha$ to save the power.

Then, the theoretical limit of the average SEE provided by the variable rate variable power transmission is given by

$$\zeta_{PR} = \int_{\alpha}^{\infty} \frac{B\log_2\left(\frac{1+\gamma P_{PR}(\gamma)}{1+\alpha P_{PR}(\gamma)}\right)}{P_C + P\cdot P(\gamma)/\mu} f(\gamma)\, d\gamma. \tag{2.24}$$

2.5 Suboptimal Adaptive Transmissions

In this section we present three suboptimal, but simpler, adaptive transmission schemes which are special cases of the variable rate variable power transmission.

2.5.1 Variable Rate On-Off Transmission

The power adaptation rule of the variable rate on-off transmission is given by

$$P_{OR}(\gamma) = \begin{cases} \min\left(P_{OR}^*, \frac{1}{Q(N,\alpha/\overline{\gamma}_B)}\right), & \gamma \geq \alpha \\ 0, & \gamma < \alpha, \end{cases} \tag{2.25}$$

where $P_{OR}^* = \arg\max_{P_{OR}} \int_{\alpha}^{\infty} \zeta(P_{OR},\gamma) f(\gamma) d\gamma$ and $Q(n,\, x) := \int_x^{\infty} \frac{u^{n-1}\exp(-u)}{(n-1)!} du$. The average SEE of the variable rate on-off transmission rule is given by

$$\zeta_{OR} = \int_{\alpha}^{\infty} \frac{B\log_2\left(\frac{1+\gamma P_{OR}(\gamma)}{1+\alpha P_{OR}(\gamma)}\right)}{P_C + P\cdot P_{OR}(\gamma)/\mu} f(\gamma)\, d\gamma. \tag{2.26}$$

2.5.2 Variable Rate Fixed Power Transmission

The power adaptation rule of the variable rate fixed power transmission is given by

$$P_R(\gamma) = \min(P_{OR}^*, 1) \tag{2.27}$$

for all $\gamma \geq 0$. The average SEE of the variable rate fixed power transmission is given by

$$\zeta_R = \int_{\alpha}^{\infty} \frac{B\log_2\left(\frac{1+\gamma P_R(\gamma)}{1+\alpha P_R(\gamma)}\right)}{P_C + P\cdot P_R(\gamma)/\mu} f(\gamma) d\gamma. \tag{2.28}$$

2.5.3 Fixed Rate Variable Power Transmission

For a fixed transmission rate, i.e. $R_S(\gamma) = R_S$ for all γ, the maximum allowed transmission power for $P_O(\gamma) < \epsilon$ can be obtained from (2.12) as

$$P_P(\gamma) = \begin{cases} \frac{2^{R_S}-1}{\gamma-\alpha 2^{R_S}}, & \gamma > \alpha 2^{R_S} \\ 0, & \gamma \leq \alpha 2^{R_S}, \end{cases} \tag{2.29}$$

where R_S is determined from the average transmission power constraint

$$\int_{\alpha 2^{R_S}+\delta}^{\infty} \left(\frac{2^{R_S}-1}{\gamma-\alpha 2^{R_S}}\right) f(\gamma)\, d\gamma \tag{2.30}$$

$$= \frac{\left(2^{R_S}-1\right) e^{-\alpha 2^{R_S}/\overline{\gamma}_B}}{(N-1)!} \sum_{k=0}^{N-1} \binom{N-1}{k} \left(\frac{\alpha 2^{R_S}}{\overline{\gamma}_B}\right)^{N-1-k} \frac{\Gamma(k,\, \delta/\overline{\gamma}_B)}{\overline{\gamma}_B}$$

$$\leq 1 \tag{2.31}$$

for arbitrarily small $\delta > 0$ and $\Gamma(n,\, x) := \int_x^{\infty} t^{n-1} e^{-t} dt = (n-1)! Q(n, x)$. The average SEE of the fixed rate variable power transmission is given by

$$\zeta_P = \max_{R_S} \int_{\alpha 2^{R_S}+\delta}^{\infty} \frac{B \cdot R_S f(\gamma)\, d\gamma}{P_C + P \cdot P_P(\gamma)/\mu}. \tag{2.32}$$

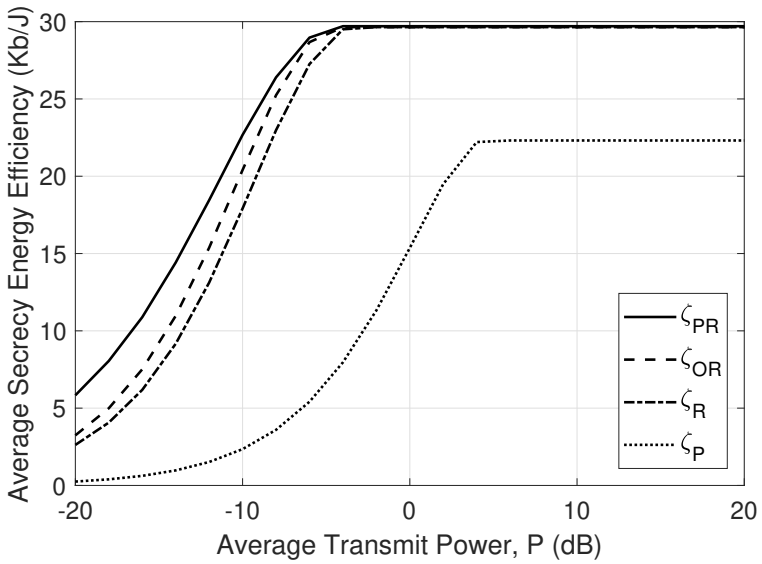

Figure 2.1 Average secrecy energy efficiency versus average transmit power P; $N = 4$ and $\epsilon = 0.1$.

2.6 Numerical Results

This section provides numerical results to evaluate the SEE of various adaptation schemes. The parameters are chosen from GSM-1900 in micro-cell environment [63]: $d = 1$Km, $P_A = 0.36$W,

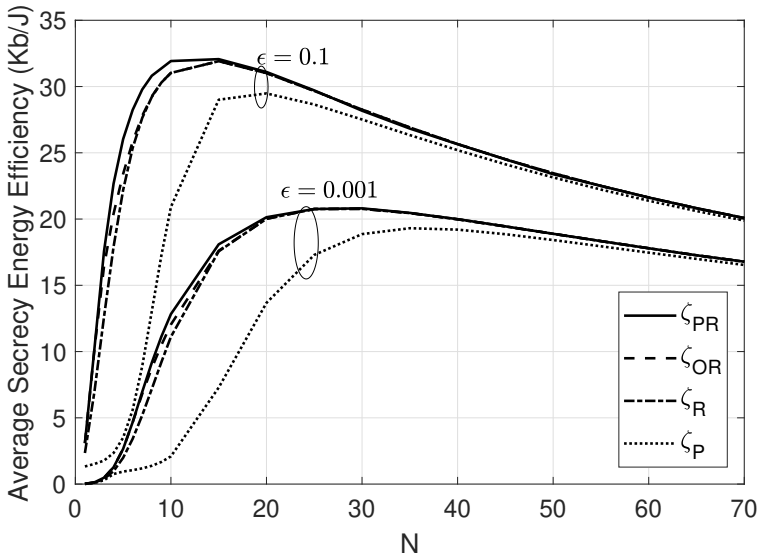

Figure 2.2 Average secrecy energy efficiency versus ϵ; $P = -10dB$.

$P_B = 0.24$W, $\sigma_h^2 = \sigma_g^2 = 10^{-(34.53+38\log_{10}(d))/10}$, $\mu = 0.4$ and $\sigma_n^2 = N_f N_0 B$, where $N_f = 3dB$ (noise figure), $N_0 = -174$dBm/Hz and $B = 200$KHz.

Fig. 2.1 depicts the average SEE of various adaptive transmission schemes versus the average transmit power, P. One observes that the average SEE is limited if only the transmission power is adapted (see ζ_P): the rate adaptation is essential in order to maximize the average SEE. In fact, at high transmit power, the additional gain provided by adapting the transmission power is negligible if the transmission rate is adapted (compare ζ_{PR} vs. ζ_R). However, at low transmit power, the power adaptation provides an additional gain even if the rate is adapted (compare ζ_{PR} vs. ζ_R).

Fig. 2.2 depicts the average SEE of various adaptive transmission schemes versus the number of antennas at the transmitter, N, for different secrecy outage probability constraints. One observes that the highest average SEE that can be provided by adapting both the transmission rate and power can be achieved by adapting either of them if the number of antennas at the transmitter is sufficiently large. One also observes that there exists an optimal number of transmit antennas that maximizes the average SEE. This follows from the diminishing gain in the secrecy rate and linearly increasing circuit power consumption as N increases.

2.7 Chapter Summary

In this chapter, we determined the optimum rate and power adaptation rule that maximizes the average SEE subject to an average transmission power constraint. We found that the average SEE is limited if only the transmission power is adapted: the rate adaptation is essential in order to maximize the average SEE. In fact, at high transmit power, the additional gain provided by adapting the transmission power is negligible if the transmission rate is adapted. However, at low transmit power, the power adaptation provides an additional gain even if the rate is adapted. We also found that the highest average SEE provided by adapting both the transmission rate and power can be achieved by adapting either of them if the number of transmitter's antennas is sufficiently large.

CHAPTER 3. COVERT COMMUNICATION UNDER CHANNEL UNCERTAINTY AND NOISE UNCERTAINTY

In this chapter, we analyze the joint impact of imperfect knowledge of the channel gain (channel uncertainty) and noise power (noise uncertainty) at the adversary on the total detection error probability and the covert throughput in Rayleigh fading channel. We characterize the covert throughput gain provided by the channel uncertainty as well as the covert throughput loss caused by the channel fading as a function of the noise uncertainty. Our result shows that the channel fading is crucial to hiding the signal transmission, particularly when the noise uncertainty is low and/or the receive SNR is high. The impact of the channel uncertainty on the total detection error probability and the covert throughput is more significant when the noise uncertainty is larger.

The remaining part of this chapter is organized as follows. Section 3.1 describes the system model. Section 3.2 describes the detection strategy and the covert requirement. Section 3.3 derives the optimum detection threshold that minimizes the total detection error probability. Section 3.4 derives the covert throughput and characterizes the covert throughput loss caused by the channel fading and the covert throughput gain provided by the channel uncertainty as a function of the noise uncertainty. Section 3.5 concludes the chapter.

3.1 System Model

Consider a scenario where Alice tries to send her message $x[n]$, $n = 1, 2, \cdots, N$, covertly to Bob without being detected by a warden, Willie. The system model is illustrated in Fig.3.1. We assume that Alice, Bob and Willie have single antenna. The received signal at Willie is given by

$$y_W[n] = \begin{cases} v_W[n], & H_0 \\ g\sqrt{P}x[n] + v_W[n], & H_1, \end{cases} \tag{3.1}$$

where H_0 denotes the hypothesis of no transmission, H_1 denotes the hypothesis of transmission, $g \sim CN(0, \sigma_g^2)$ is the channel gain between Alice and Willie, P is the transmit power, and $v_W[n] \sim CN(0, \sigma_w^2)$ is the complex Gaussian noise. We assume $E[|x[n]|^2] = 1$ for all n.

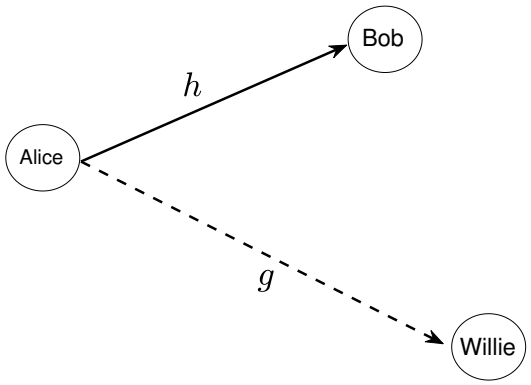

Figure 3.1 Alice attempts to transmit covertly to Bob in the presence of an adversary, Willie, who wants to detect Alice's transmission.

We assume that Willie has an imperfect estimation of the channel gain (channel uncertainty). Willie's estimation of the channel gain and the estimation error are denoted by $\hat{g}$ and $\tilde{g}$, respectively. Thus,

$$g = \hat{g} + \tilde{g}, \tag{3.2}$$

where $\hat{g}$ and $\tilde{g}$ are independent complex Gaussian random variables with mean zero and variance $(1 - \beta)\sigma_g^2$ and $\beta\sigma_g^2$, respectively, and $\beta \in [0, 1]$ represents the channel gain uncertainty. The assumption of Gaussian distributed estimation error arises from using the MMSE estimator [34].

We further assume that Willie has imperfect knowledge of the noise power (noise uncertainty). Noise uncertainty arises due to temperature change, environment noise change, or calibration error [58]. We consider the bounded uncertainty model, where the actual noise power σ_w^2 lies in a finite range around the nominal (estimated) noise power, $\hat{\sigma}_w^2$. We assume that $\sigma_{w,dB}^2 = 10\log_{10}\sigma_w^2$ is uniformly distributed in its uncertainty range $[\hat{\sigma}_{w,dB}^2 - \rho_{dB}, \hat{\sigma}_{w,dB}^2 + \rho_{dB}]$, where $\hat{\sigma}_{w,dB}^2 = 10\log_{10}\hat{\sigma}_w^2$ and ρ_{dB} denotes the noise uncertainty deviation [58]. Then, the probability density function (PDF)

of σ_w^2 is given by

$$f_{\sigma_w^2}(x) = \begin{cases} \frac{1}{2\ln(\rho)x}, & \text{if } \frac{1}{\rho}\hat{\sigma}_w^2 \leq \sigma_w^2 \leq \rho\hat{\sigma}_w^2, \\ 0, & \text{otherwise}, \end{cases} \tag{3.3}$$

where $\rho = 10^{\rho_{dB}/10}$.

3.2 Willie's Detection Strategy and Covert Requirement

Willie is interested in knowing whether Alice transmits or not based on the observation vector $\mathbf{y}_W = (y_W[1], ..., y_W[N])$. The test statistic for detection[1] is given by [59]

$$T(\mathbf{y}_W) = \frac{1}{N}\sum_{n=1}^{N} |y_W[n]|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \lambda, \tag{3.4}$$

where λ is the detection threshold. If Willie is allowed to observe an infinite number of samples, which minimizes the total detection error probability, the test statistic converges to

$$T(\mathbf{y}_W) \rightarrow E[|y_W[n]|^2] \tag{3.5}$$

$$= \begin{cases} \sigma_w^2, & H_0, \\ |g|^2 P + \sigma_w^2, & H_1. \end{cases} \tag{3.6}$$

Given $\hat{g}$, the probability of false alarm and that of missed detection are given by

$$P_F = \Pr(T(\mathbf{y}_W) > \lambda | H_0) \tag{3.7}$$

$$= \Pr(\sigma_w^2 > \lambda) \tag{3.8}$$

and

$$P_M = \Pr(T(\mathbf{y}_W) < \lambda | H_1) \tag{3.9}$$

$$= \Pr(|\hat{g} + \tilde{g}|^2 P + \sigma_w^2 < \lambda), \tag{3.10}$$

respectively. Hence, the total detection error probability, $P_M + P_F$, is given by

$$\xi(\hat{g}) = P_F + P_M \tag{3.11}$$

$$= 1 - \Pr(\lambda - |\hat{g} + \tilde{g}|^2 P < \sigma_w^2 < \lambda) \tag{3.12}$$

[1] If prior knowledge of the transmitter signal is unknown, the optimal detector is the energy detector [55].

as $N \to \infty$. Willie attempts to choose the optimum detection threshold, λ^*, that minimizes the total detection error probability, i.e.

$$\lambda^* = \arg\min_{\lambda} \xi(\hat{g}). \tag{3.13}$$

We consider Alice achieving covert communication if, for any $\epsilon > 0$, a communication scheme exists so that $E_{\hat{g}}[\min_\lambda(P_F + P_M)] \geq 1 - \epsilon$ as $N \to \infty$. Here, ϵ signifies the covert requirement, since a sufficiently small ϵ renders any detector employed at Willie to be ineffective.

3.3 Optimum Detection Threshold and Minimum total detection error probability

In this section, we study the optimum detection threshold that minimizes the total detection error probability and the resulting minimum total detection error probability. If we let $X = |\hat{g}+\tilde{g}|^2$, then X is a non-central Chi-square random variable with non-centrality parameter $|\hat{g}|^2$ and variance $\beta\sigma_g^2$. The conditional PDF of X given $\hat{g}$ is given by [54]

$$f_X(x|\hat{g}) = \frac{e^{-(x+|\hat{g}|^2)/(\beta\sigma_g^2)}}{\beta\sigma_g^2} I_0\left(\frac{2|\hat{g}|}{\beta\sigma_g^2}\sqrt{x}\right), \tag{3.14}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind with zeroth order[2]. Therefore, it follows from (3.3) and (3.12) that

$$\xi(\hat{g}) = 1 - \int_0^\infty f_X(x|\hat{g}) \int_{\max\{\lambda - xP, \frac{1}{\rho}\hat{\sigma}_w^2\}}^{\min\{\lambda, \rho\hat{\sigma}_w^2\}} f_{\sigma_w^2}(y) dy dx. \tag{3.15}$$

As a special case, if there is no uncertainty about the noise power, i.e. $f_{\sigma_w^2}(y) = \delta(y - \hat{\sigma}_w^2)$ where $\delta(x)$ is the Dirac delta function, then (3.15) reduces to

$$\xi(\hat{g}) = \begin{cases} 1 - Q\left(\sqrt{\frac{|\hat{g}|^2}{\beta\sigma_g^2}}, \sqrt{\frac{\lambda - \hat{\sigma}_w^2}{\beta\sigma_g^2 P}}\right), & \lambda \geq \hat{\sigma}_w^2, \\ 1, & \lambda < \hat{\sigma}_w^2, \end{cases} \tag{3.16}$$

which matches with the result in [3] for the case of infinite number of samples.

[2] In [57], X is replaced by $|\hat{g}|^2 + |\tilde{g}|^2$.

If $\lambda < \hat{\sigma}_w^2/\rho$, the second integral of (3.15) is 0 and thus $\xi(\hat{g}) = 1$. If $\lambda \geq \rho\hat{\sigma}_w^2$, $\xi(\hat{g})$ is an increasing function of λ. Therefore, the optimum detection threshold that minimizes $\xi(\hat{g})$ should be inside the range $[\frac{1}{\rho}\hat{\sigma}_w^2, \rho\hat{\sigma}_w^2]$. For $\lambda \in [\frac{1}{\rho}\hat{\sigma}_w^2, \rho\hat{\sigma}_w^2]$, it can be shown from (3.3) and (3.15) that

$$\xi(\hat{g}) = 1 - \int_0^{\infty} f_X(x|\hat{g}) \int_{\max\{\lambda - xP, \frac{1}{\rho}\hat{\sigma}_w^2\}}^{\lambda} \frac{1}{2\ln(\rho)} \frac{dy}{y} dx \tag{3.17}$$

$$= 1 - \frac{\ln(\lambda)}{2\ln(\rho)} + \frac{1}{2\ln(\rho)} \int_0^{\infty} f_X(x|\hat{g}) \ln\left(\max\{\lambda - xP, \frac{1}{\rho}\hat{\sigma}_w^2\}\right) dx \tag{3.18}$$

$$= 1 - \frac{\ln(\rho\lambda/\hat{\sigma}_w^2)}{2\ln(\rho)} + \frac{1}{2\ln(\rho)} \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)\frac{1}{\gamma}} \ln\left(\rho\left(\frac{\lambda}{\hat{\sigma}_w^2} - x\gamma\right)\right) f_X(x|\hat{g}) dx, \tag{3.19}$$

where $\gamma = P/\hat{\sigma}_w^2$. The derivation of (3.19) is provided in Appendix B.1. Also, it is shown in Appendix B.2 that $\xi(\hat{g})$ is a strictly pseudo-convex function of λ for $\lambda \geq \hat{\sigma}_w^2/\rho$. Therefore, there should exist unique minima. The optimum λ that minimizes $\xi(\hat{g})$ can be found by taking the derivative of $\xi(\hat{g})$ with respect to λ and setting it to zero:

$$2\ln(\rho)\frac{d\xi(\hat{g})}{d\lambda} = -\frac{1}{\lambda} + \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)\frac{1}{\gamma}} \frac{f_X(x|\hat{g})}{\lambda - xP} dx = 0. \tag{3.20}$$

Let $\lambda^{\perp}$ denote the solution of (3.20). Since the optimum λ that minimizes $\xi(\hat{g})$ lies inside $[\frac{1}{\rho}\hat{\sigma}_w^2, \rho\hat{\sigma}_w^2]$, it is given by

$$\lambda^* = \min\left\{\lambda^{\perp}, \rho\hat{\sigma}_w^2\right\}, \tag{3.21}$$

and the minimum total detection error probability can be obtained by applying (3.21) to (3.19):

$$\xi_{min}(\hat{g}) = 1 - \frac{\ln(\rho\lambda^*/\hat{\sigma}_w^2)}{2\ln(\rho)} + \frac{1}{2\ln(\rho)} \int_0^{\left(\frac{\lambda^*}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)\frac{1}{\gamma}} \ln\left(\rho\left(\frac{\lambda^*}{\hat{\sigma}_w^2} - x\gamma\right)\right) f_X(x|\hat{g}) dx, \tag{3.22}$$

where the integration can be computed by using Trapezoidal method. Then, averaging (3.22) over $\hat{g}$ (averaging over multiple coherence time intervals) yields the minimum total detection error probability, ξ_{min}. In the remaining part of this section, we derive the minimum total detection error probability for several special cases.

Fig. 3.2 illustrates the total detection error probability versus the detection threshold for different values of $\hat{g}$. If $|\hat{g}|^2$ is small (say 0.1), $\lambda^{\perp}$ is smaller than $\rho\hat{\sigma}_w^2$. Hence, the optimum detection threshold that minimizes $\xi(\hat{g})$ is $\lambda^{\perp}$. But if $|\hat{g}|^2$ is large (say 0.5), $\lambda^{\perp}$ becomes larger

than $\rho\hat{\sigma}_w^2$. Hence, the optimum detection threshold is $\rho\hat{\sigma}_w^2$. Below we consider three special cases of $\hat{g} = g$ ($\beta = 0$), perfect CSI; $\hat{g} = 0$ ($\beta = 1$), no CSI; and no fading.

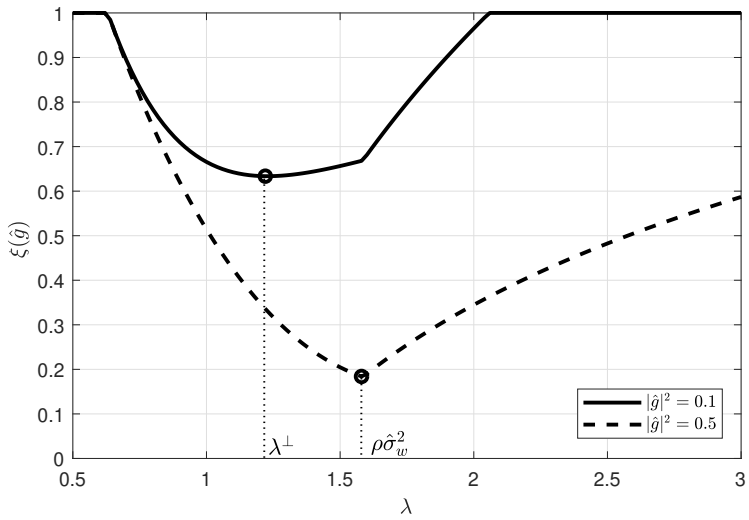

Figure 3.2 Total detection error probability $\xi(\hat{g})$ versus detection threshold λ for different values of $|\hat{g}|^2$; $\sigma_g^2 P/\hat{\sigma}_w^2 = 3$dB, $\beta = 0.1$, and $\rho = 2$dB.

3.3.1 Perfect CSI at Willie

If the channel gain g is known perfectly to Willie, i.e. $\hat{g} = g$ or $\beta = 0$, then the conditional PDF of X given $\hat{g}$ is given by

$$f_X(x|\hat{g}) = \delta(x - |\hat{g}|^2). \tag{3.23}$$

Then, it can be shown from (3.19) and (3.23) that

$$\xi(\hat{g}) = \begin{cases} 1 - \frac{\ln(\rho\lambda/\hat{\sigma}_w^2)}{2\ln(\rho)}, & |\hat{g}|^2 \geq \left(\lambda - \hat{\sigma}_w^2\rho^{-1}\right)/P, \\ 1 + \frac{1}{2\ln(\rho)} \ln\left(1 - \frac{|\hat{g}|^2 P}{\lambda}\right), & |\hat{g}|^2 < \left(\lambda - \hat{\sigma}_w^2\rho^{-1}\right)/P, \end{cases} \tag{3.24}$$

and

$$2\ln(\rho)\frac{d\xi(\hat{g})}{d\lambda} = \begin{cases} -\frac{1}{\lambda}, & \lambda \leq |\hat{g}|^2 P + \hat{\sigma}_w^2/\rho, \\ -\frac{1}{\lambda} + \frac{1}{\lambda - |\hat{g}|^2 P}, & \lambda > |\hat{g}|^2 P + \hat{\sigma}_w^2/\rho. \end{cases} \tag{3.25}$$

Since $d\xi(\hat{g})/d\lambda$ is negative for $\lambda \leq |\hat{g}|^2P+\hat{\sigma}_w^2/\rho$ and positive for $\lambda > |\hat{g}|^2P+\hat{\sigma}_w^2/\rho$, $\xi(\hat{g})$ is minimized when $\lambda = |\hat{g}|^2P + \hat{\sigma}_w^2/\rho$. Since the optimum detection threshold that minimizes $\xi(\hat{g})$ lies inside $[\frac{1}{\rho}\hat{\sigma}_w^2, \rho\hat{\sigma}_w^2]$, we obtain

$$\lambda^* \quad = \quad \min\left\{|\hat{g}|^2P + \frac{1}{\rho}\hat{\sigma}_w^2, \rho\hat{\sigma}_w^2\right\}. \tag{3.26}$$

As a special case, if $\rho = 1$ then we obtain $\lambda^* = \hat{\sigma}_w^2$, which matches with the result in [6].

Applying (3.26) to (3.24) yields the minimum instantaneous total detection error probability

$$\xi_{min}(\hat{g}) = \left(1 - \frac{\ln(1+\rho|\hat{g}|^2\gamma)}{2\ln(\rho)}\right)^+, \tag{3.27}$$

where $(x)^+ = \max(x, 0)$. Note that $\xi_{min}(\hat{g})$ reduces to 0 if the received SNR, $|\hat{g}|^2\gamma$, is above a threshold, $\rho - \rho^{-1}$. Averaging (3.27) over $\hat{g}$ (averaging over multiple coherence time intervals) yields

$$\xi_{min} = E_{\hat{g}}[\xi_{min}(\hat{g})] \tag{3.28}$$

$$= \int_0^{(\rho-\frac{1}{\rho})/\gamma} \left(1 - \frac{\ln(1+x\rho\gamma)}{2\ln(\rho)}\right) \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \tag{3.29}$$

$$= 1 - \frac{e^{\frac{1}{\rho\sigma_g^2\gamma}}}{2\ln(\rho)} \left[\mathrm{Ei}\left(-\frac{\rho}{\sigma_g^2\gamma}\right) - \mathrm{Ei}\left(-\frac{1}{\rho\sigma_g^2\gamma}\right)\right], \tag{3.30}$$

where $\mathrm{Ei}(x) = \int_{-\infty}^{x} t^{-1}e^t dt$ is the exponential integral function.

Graphical Interpretation: Fig. 3.3 illustrates P_M and P_F under channel uncertainty and noise uncertainty in general. Since the PDF $f_{\sigma_w^2}(x)$ is a decreasing function, the sum P_M+P_F is minimized when $\lambda - |\hat{g}+\tilde{g}|^2P$ is equal to $\hat{\sigma}_w^2/\rho$, which makes $P_M = 0$. Therefore, if $\tilde{g} = 0$, i.e. perfect CSI, $P_M + P_F$ is minimized by choosing $\lambda = |\hat{g}|^2P + \hat{\sigma}_w^2/\rho$, where $\xi_{min} = 1 - \frac{1}{2\ln(\rho)}\ln(1+\rho|\hat{g}|^2\gamma)$. If $|\hat{g}|^2P+\hat{\sigma}_w^2/\rho > \rho\hat{\sigma}_w^2$, i.e. $|\hat{g}|^2\gamma > \rho-\rho^{-1}$, then $P_F = 0$ because $\lambda > \rho\hat{\sigma}_w^2$ and consequently $\xi_{min} = 0$.

Low SNR Approximation: Since $\mathrm{Ei}(-x) \simeq -\frac{1}{2}e^{-x}\ln(1+\frac{2}{x})$ for $x \gg 1$ [2], (3.30) can be approximated by

$$\xi_{min} \simeq 1 - \frac{1}{2\ln(\rho)}\left[-\frac{1}{2}e^{-\frac{\rho-\rho^{-1}}{\sigma_g^2\gamma}}\ln\left(1+\frac{2\sigma_g^2\gamma}{\rho}\right) + \frac{1}{2}\ln\left(1+2\rho\sigma_g^2\gamma\right)\right] \tag{3.31}$$

$$\simeq 1 - \frac{1}{4\ln(\rho)}\ln\left(1+2\rho\sigma_g^2\gamma\right), \tag{3.32}$$

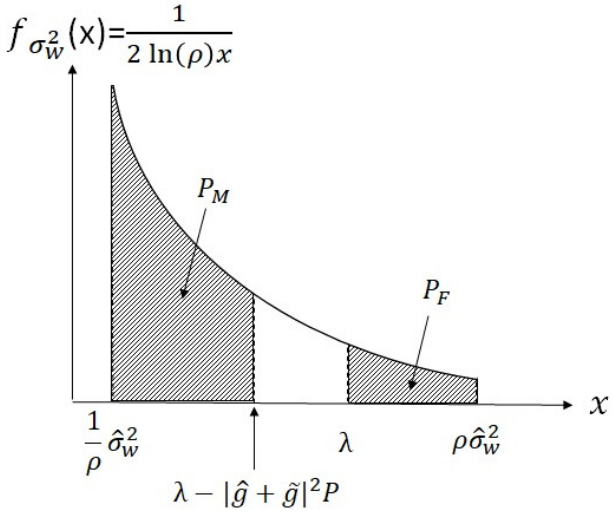

Figure 3.3 Probability of missed detection and false alarm.

for $\sigma_g^2\gamma \ll 1$.

High SNR Approximation: For $x \ll 1$, $e^x \simeq 1+x$ and $\mathrm{Ei}(-x) \simeq c+\ln(x)-x$, where c is the Euler-Mascheroni constant [2]. Therefore, for $\sigma_g^2\gamma \gg 1$, (3.30) can be approximated by

$$\xi_{min} \simeq 1-\frac{1}{2\ln(\rho)}\left(1+\frac{1}{\rho\sigma_g^2\gamma}\right)\left(2\ln(\rho)-\frac{\rho-\rho^{-1}}{\sigma_g^2\gamma}\right) \tag{3.33}$$

$$=\frac{1}{2\ln(\rho)}\frac{\rho-\rho^{-1}}{\sigma_g^2\gamma}\left(1+\frac{1}{\rho\sigma_g^2\gamma}\right)-\frac{1}{\rho\sigma_g^2\gamma} \tag{3.34}$$

$$\simeq \frac{1}{\rho\sigma_g^2\gamma}\left(\frac{\rho^2-1}{2\ln(\rho)}-1\right), \tag{3.35}$$

which decays inverse linearly with the received SNR, $\sigma_g^2\gamma$.

3.3.2 No CSI at Willie

If CSI is not available at Willie, i.e. $\beta = 1$, the conditional PDF of X given $\hat{g}=0$ in (3.14) reduces to

$$f_X(x|\hat{g}) = e^{-x/\sigma_g^2}/\sigma_g^2, \ x \geq 0. \tag{3.36}$$

Then, it can be shown from (3.19) and (3.36) that the total detection error probability is given by

$$\xi(\hat{g}=0) = 1 - \frac{\ln(\rho\lambda/\hat{\sigma}_w^2)}{2\ln(\rho)} + \frac{1}{2\ln(\rho)} \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)\frac{1}{\gamma}} \ln\left(\rho\left(\frac{\lambda}{\hat{\sigma}_w^2} - x\gamma\right)\right) \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \tag{3.37}$$

$$= 1 - \frac{e^{-\frac{\lambda}{\sigma_g^2 P}}}{2\ln(\rho)} \int_{\frac{1}{\rho\sigma_g^2\gamma}}^{\frac{\lambda}{\sigma_g^2 P}} \frac{e^x}{x} dx \tag{3.38}$$

$$= 1 - \frac{e^{-\frac{\lambda}{\sigma_g^2 P}}}{2\ln(\rho)} \left[\mathrm{Ei}\left(\frac{\lambda}{\sigma_g^2 P}\right) - \mathrm{Ei}\left(\frac{1}{\rho\sigma_g^2\gamma}\right)\right], \tag{3.39}$$

for $\lambda \geq \hat{\sigma}_w^2/\rho$. The derivation of (3.38) is provided in Appendix B.3.

By taking the first derivative of $\xi(\hat{g})$ with respect to λ and setting it to zero, we obtain

$$\int_{\frac{1}{\rho\sigma_g^2\gamma}}^{\frac{\lambda}{\sigma_g^2 P}} \frac{e^x}{x} dx - \frac{e^{\frac{\lambda}{\sigma_g^2 P}}}{\frac{\lambda}{\sigma_g^2 P}} = 0, \;\; \lambda \geq \hat{\sigma}_w^2/\rho. \tag{3.40}$$

Let $\lambda^\dagger$ denote the solution of (3.40). Since the optimum λ that minimizes $\xi(\hat{g}=0)$ lies inside $[\frac{1}{\rho}\hat{\sigma}_w^2, \rho\hat{\sigma}_w^2]$, it is given by

$$\lambda^* = \min\{\lambda^\dagger, \rho\hat{\sigma}_w^2\}. \tag{3.41}$$

Therefore, the minimum total detection error probability is given by

$$\xi_{min} = \begin{cases} 1 - \frac{\sigma_g^2 P}{2\lambda^\dagger \ln(\rho)}, & \lambda^\dagger < \rho\hat{\sigma}_w^2, \\ 1 - \frac{e^{-\frac{\rho}{\sigma_g^2\gamma}}}{2\ln(\rho)} \left[\mathrm{Ei}\left(\frac{\rho}{\sigma_g^2\gamma}\right) - \mathrm{Ei}\left(\frac{1}{\rho\sigma_g^2\gamma}\right)\right], & \lambda^\dagger \geq \rho\hat{\sigma}_w^2. \end{cases} \tag{3.42}$$

It should be noted from (3.40) that $\lambda^\dagger$ is proportional to $\sigma_g^2 P$ as the LHS of (3.40) is a function of $\lambda/(\sigma_g^2 P)$. Hence, the condition $\lambda^\dagger < \rho\hat{\sigma}_w^2$ is equivalent to $\sigma_g^2 P/\hat{\sigma}_w^2 < \gamma_g^*$ for some threshold γ_g^*, where $\sigma_g^2 P/\hat{\sigma}_w^2$ represents the received SNR at Willie. Also, the optimum detection threshold in (3.41) assumes the knowledge of the average received power, $\sigma_g^2 P$, by Willie as $\lambda^\dagger$ is proportional to $\sigma_g^2 P$.

Fig.3.4 shows the detection error probability, ξ, versus the received SNR, $\sigma_g^2 P/\hat{\sigma}_w^2$, with two detection thresholds, $\lambda^\dagger$ and $\rho\hat{\sigma}_w^2$. It can be seen that $\lambda^\dagger$ provides a lower ξ than $\rho\hat{\sigma}_w^2$ if the received SNR is less than a threshold, and, otherwise, $\rho\hat{\sigma}_w^2$ provides a lower ξ than $\lambda^\dagger$.

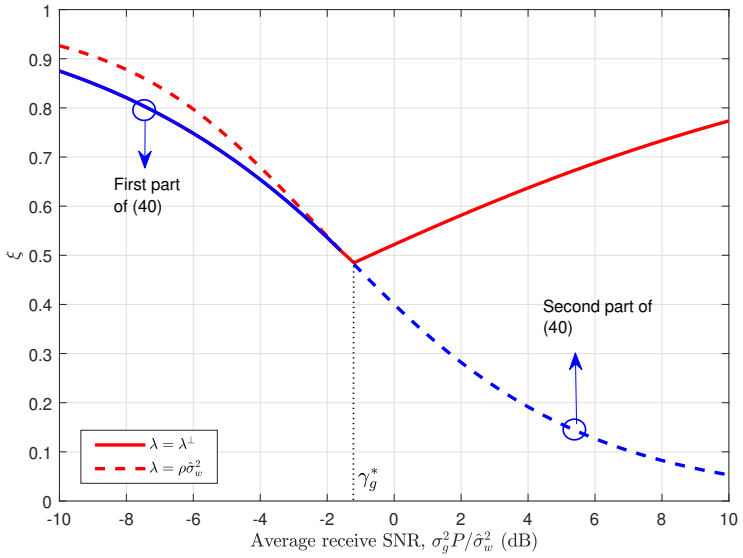

Figure 3.4 Average total detection error probability, ξ, versus the average receive SNR, $\sigma_g^2 P/\hat{\sigma}_w^2$, under no CSI; $\rho = 2$dB.

3.3.3 AWGN Channel

For AWGN channel where $|\hat{g}|^2 = \sigma_g^2$, the minimum total detection error probability in (3.27) reduces to

$$\xi_{min} = \left(1 - \frac{\ln(1+\rho\sigma_g^2\gamma)}{2\ln(\rho)}\right)^+, \tag{3.43}$$

which matches with the result in [31].

3.3.4 Numerical Results

Fig. 3.5 illustrates the minimum total detection error probability, ξ_{min}, versus the average received SNR at Willie, $\sigma_g^2 P/\hat{\sigma}_w^2$, for different channel uncertainty, β. Also shown in the figure is ξ_{min} for AWGN channel. It can be seen that when the received SNR is low, ξ_{min} is close to 1, regardless of the channel fading. However, when the received SNR is high, ξ_{min} drops quickly to zero in AWGN channel if the received SNR is above a threshold (as shown in (3.43)), while in Rayleigh fading channel it decreases gradually (inverse linearly) with the received SNR (as shown

in (3.35)). Therefore, channel fading plays a critical role in hiding the signal transmission when the received SNR is high.

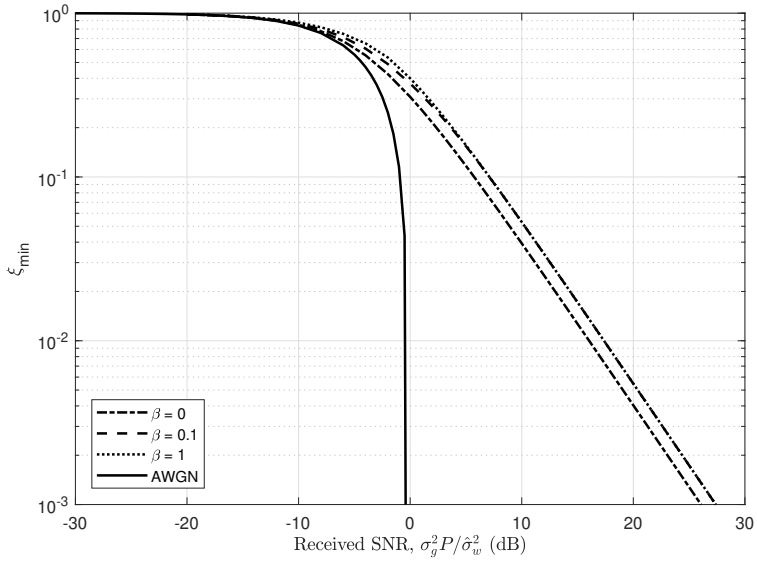

Figure 3.5 The minimum total detection error probability, ξ_{min}, versus average receive SNR, $\sigma_g^2 P/\hat{\sigma}_w^2$, for different values of β; $\rho = 2$dB.

Fig. 3.6 illustrates the minimum total detection error probability, ξ_{min}, versus the noise uncertainty for different channel uncertainty, β. It can be seen that the channel fading is critical to hiding the signal transmission if the noise uncertainty, ρ, is below a threshold where ξ_{min} is 0 in AWGN channel. Imperfect knowledge of the channel gain increases the minimum total detection error probability, particularly when the noise uncertainty is large. However, if the noise uncertainty is small, imperfect knowledge of the channel gain has little impact on the minimum total detection error probability.

3.4 Covert Throughput

In this section we study the covert throughput, defined as the maximum average rate (bits/s/Hz) between Alice and Bob subject to the covert constraint of $\xi_{min} \geq 1 - \epsilon$ as $N \to \infty$. Assuming that Alice is not aware of the channel gain, h, to Bob (due to unavailability of pilot transmission from

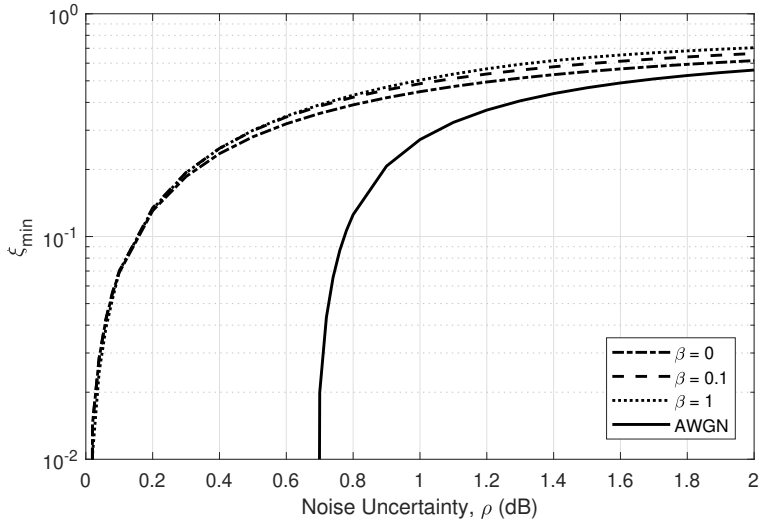

Figure 3.6 The minimum total detection error probability, ξ_{min}, versus noise uncertainty, ρ, for different values of β; $\sigma_g^2 P/\hat{\sigma}_w^2 = -5$dB.

Bob to avoid being detected by Willie), we consider sending the message at a fixed rate, R, in slowly-varying channels, where the instantaneous SNR is constant over a large number of transmissions (a transmission burst) and then changes to a new value based on the fading distribution.

With this model, the correct metric for Alice's performance is capacity with outage [26]. The message is correctly received if the instantaneous received SNR is greater than or equal to $2^R - 1$. If the instantaneous received SNR is below $2^R - 1$ then the bits received over that transmission burst cannot be decoded correctly, and the receiver declares an outage. The probability of decoding outage at Bob is thus

$$P_O = \Pr(\log_2(1 + |h|^2 P/\sigma_b^2) < R) \tag{3.44}$$

$$= 1 - \exp\left(-(2^R - 1)/(\sigma_h^2 P/\sigma_b^2)\right) \tag{3.45}$$

in Rayleigh fading channel, where $\sigma_h^2 = \mathrm{E}[|h|^2]$ and σ_b^2 is the noise power of Bob. The average rate correctly received over many transmission bursts is $R(1 - P_O)$ since the message is only correctly

received on $1-P_O$ transmissions. Then, the covert throughput is defined by[3]

$$\max_{R} \quad R(1-P_O) \tag{3.46}$$

$$\text{s.t.} \quad \xi_{min} \geq 1-\epsilon. \tag{3.47}$$

It can be shown from (3.45) and (3.46) that the optimum transmission rate that maximizes $R(1-P_O)$ is $W_0(\sigma_h^2 P/\sigma_b^2)/\ln 2$, where $W_0(x)$, solution of $x = W_0(x)e^{W_0(x)}$, is the Lambert-W function with branch 0 [18].

Since $\xi_{min}(\hat{g})$ in (3.22) decreases monotonically with increasing $\gamma = P/\hat{\sigma}_w^2$, the covert constraint $\xi_{min} \geq 1-\epsilon$ requires $P \leq P^*$ for some threshold P^* which is computed from $\xi_{min} = 1-\epsilon$. Therefore, the covert throughput, η, is given by

$$\eta = \frac{1}{\ln 2} W_0(\sigma_h^2 P^*/\sigma_b^2) \times \exp\left(-\frac{e^{W_0(\sigma_h^2 P^*/\sigma_b^2)}-1}{\sigma_h^2 P^*/\sigma_b^2}\right). \tag{3.48}$$

Since $W_0(x) \simeq x$ [18] and $(e^{W_0(x)}-1)/x \simeq 1$ for $x \ll 1$, the covert throughput can be approximated by

$$\eta \simeq \frac{\sigma_h^2 P^*}{\sigma_b^2 \ln 2} e^{-1}, \tag{3.49}$$

for $\sigma_h^2 P^*/\sigma_b^2 \ll 1$. In the remaining part of this section, we derive the covert throughput for several special cases.

3.4.1 Perfect CSI at Willie

If Willie knows his channel gain g perfectly, then it follows from (3.32) that P^* can be approximated by

$$P^* \simeq \frac{\rho^{4\epsilon-1}-\rho^{-1}}{2} \frac{\hat{\sigma}_w^2}{\sigma_g^2}, \tag{3.50}$$

and thus the covert throughput is given by

$$\eta_p \simeq \frac{\rho^{4\epsilon-1}-\rho^{-1}}{2\ln 2} \frac{\sigma_h^2 \hat{\sigma}_w^2}{\sigma_g^2 \sigma_b^2} e^{-1} \tag{3.51}$$

for $\epsilon \ll 1$.

[3]The maximization in (3.46) is achieved without constraint on P_O. If P_O needs to be below a threshold, δ, then the covert throughput is $\log_2(1+\sigma_h^2 \gamma \ln(1-\delta)^{-1})$ subject to the covert constraint of $\xi_{min} \geq 1-\epsilon$.

3.4.2 No CSI at Willie

If Willie has no knowledge on his channel gain g, then $\xi_{min} < 0.9$ if $\lambda^\dagger \geq \rho\hat{\sigma}_w^2$ for $1.0002 \leq \rho \leq 3.16$ and hence for $\xi_{min} \geq 0.9$ and $1.0002 \leq \rho \leq 3.16$, which is the range of practical interest, we need to consider the case of $\lambda^\dagger < \rho\hat{\sigma}_w^2$ only in (3.42). The derivation of range of interest is provided in Appendix B.4. Then, for $\xi_{min} \geq 1-\epsilon$, it is required that

$$1 - \frac{1}{2\ln(\rho)} \frac{\sigma_g^2 P}{\lambda^\dagger} \geq 1 - \epsilon, \tag{3.52}$$

for $\lambda^\dagger < \rho\hat{\sigma}_w^2$, which yields

$$\sigma_g^2 P/(2\epsilon \ln(\rho)) \leq \lambda^\dagger < \rho\hat{\sigma}_w^2. \tag{3.53}$$

Therefore, the maximum transmission power for $\xi_{min} \geq 1-\epsilon$ while satisfying (3.53) is given by

$$P^* = \min\{P_n, 2\epsilon\rho\ln(\rho)\}\hat{\sigma}_w^2/\sigma_g^2, \tag{3.54}$$

where P_n is the solution of $\lambda^\dagger = \sigma_g^2 P/(2\epsilon\ln(\rho))$. The covert throughput under no CSI, denoted η_n, can then be obtained from (3.48) and (3.54).

Low SNR approximation: At low SNR, $\sigma_g^2\gamma \ll 1$, it can be shown that the maximum transmission power for $\xi_{min} \geq 1-\epsilon$ is approximately given by

$$P^* \simeq \frac{\rho^{-1}}{-W_{-1}\left(-\frac{1}{(2\epsilon\ln(\rho))^2} e^{-\frac{1}{2\epsilon\ln(\rho)}}\right)} \frac{\hat{\sigma}_w^2}{\sigma_g^2}, \tag{3.55}$$

where $W_{-1}(x)$ is the Lambert-W function with branch -1 [18]. The derivation of (3.55) is provided in Appendix B.5.

Since $W_{-1}(x) \simeq \ln(-x) - \ln(-\ln(-x))$ for $x \to 0^-$ [18], we obtain

$$-W_{-1}\left(-\frac{e^{-\frac{1}{2\epsilon\ln(\rho)}}}{(2\epsilon\ln(\rho))^2}\right) \simeq \frac{1}{2\epsilon\ln(\rho)} + 2\ln\left(2\epsilon\ln(\rho)\right) + \ln\left(\frac{1}{2\epsilon\ln(\rho)} + 2\ln\left(2\epsilon\ln(\rho)\right)\right) \tag{3.56}$$

$$\simeq \frac{1}{2\epsilon\ln(\rho)} + \ln\left(2\epsilon\ln(\rho)\right), \tag{3.57}$$

for $\epsilon \ll 1$. Therefore, it follows from (3.49), (3.55) and (3.57) that the covert throughput under no CSI can be approximated by

$$\eta_n \simeq \frac{1}{\ln 2} \frac{\rho^{-1}}{\frac{1}{2\epsilon\ln(\rho)} + \ln\left(2\epsilon\ln(\rho)\right)} \frac{\sigma_h^2\hat{\sigma}_w^2}{\sigma_g^2\sigma_b^2} e^{-1}. \tag{3.58}$$

Remark: It can be seen from (3.50) and (3.58) that the channel uncertainty increases the covert throughput by a factor of

$$\frac{\eta_n}{\eta_p} \simeq \frac{2}{(\rho^{4\epsilon}-1)\left(\frac{1}{2\epsilon\ln(\rho)}+\ln\left(2\epsilon\ln(\rho)\right)\right)}. \tag{3.59}$$

3.4.3 AWGN Channel

It follows from (3.43) that the maximum transmission power for $\xi_{min} \geq 1-\epsilon$ is given by

$$P^* = (\rho^{2\epsilon-1}-\rho^{-1})\hat{\sigma}_w^2/\sigma_g^2. \tag{3.60}$$

Therefore, the covert throughput in AWGN channel is given by

$$\eta_a = \log_2(1+\sigma_h^2 P^*/\sigma_b^2) \tag{3.61}$$

$$= \log_2\left(1+(\rho^{2\epsilon-1}-\rho^{-1})\frac{\sigma_h^2\hat{\sigma}_w^2}{\sigma_g^2\sigma_b^2}\right). \tag{3.62}$$

For $\epsilon \ll 1$, η_a can be approximated by

$$\eta_a \simeq \frac{\rho^{2\epsilon-1}-\rho^{-1}}{\ln 2}\frac{\sigma_h^2\hat{\sigma}_w^2}{\sigma_g^2\sigma_b^2}, \tag{3.63}$$

by applying $\ln(1+x) \simeq x$ for $x \ll 1$.

Remark: It can be seen from (3.50) and (3.60) that channel fading allows the maximum transmission power to be increased by a factor of $(\rho^{2\epsilon}+1)/2$ over the AWGN channel while still satisfying the same covert constraint $\xi_{min} \geq 1-\epsilon$. But the channel fading reduces the covert throughput by a factor of

$$\frac{\eta_p}{\eta_a} = \frac{(\rho^{2\epsilon}+1)}{2}e^{-1}, \tag{3.64}$$

which converges to e^{-1} as $\rho \to 1$ (no noise uncertainty) or $\epsilon \to 0$ (perfect privacy). It can also be seen from (3.64) that the covert throughput loss caused by channel fading is more significant if the noise uncertainty ρ and/or the covert constraint ϵ is smaller.

3.4.4 Numerical Results

Fig. 3.7 illustrates the covert throughput versus the noise uncertainty, ρ, for different values of ϵ that represents the covertness of communication. It can be seen that the covert throughput increases

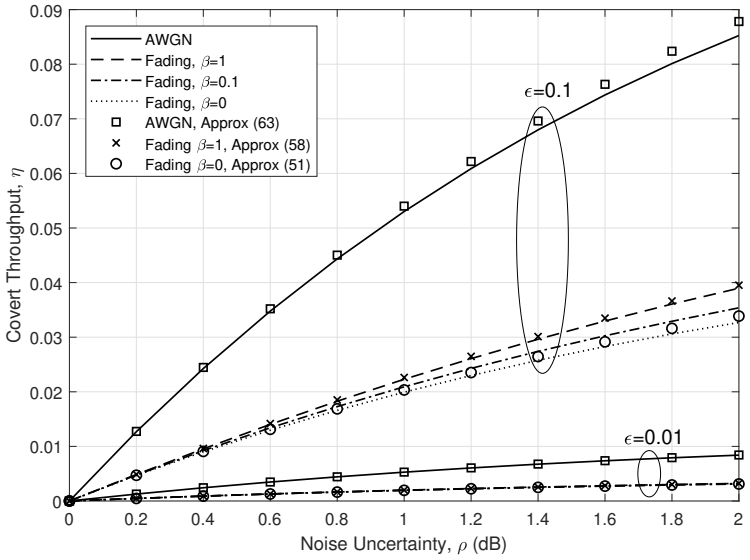

Figure 3.7 Covert throughput, η, versus noise uncertainty, ρ, for different values of ϵ; $\sigma_h^2 = \sigma_g^2$ and $\sigma_b^2 = \hat{\sigma}_w^2$.

monotonically with ρ and that the approximate covert throughput in (3.51), (3.58), and (3.63) are fairly close to the exact one. The channel uncertainty helps increase the covert throughput, particularly when the noise uncertainty is large, but the improvement is limited when the covert constraint is strict, e.g. $\epsilon = 0.01$. Even though the channel fading increases ξ_{min}, as seen in Figs. 3.5 and 3.6, it decreases the covert throughput. This is mainly due to a high decoding outage probability at low signal-to-noise ratio (SNR) in Rayleigh fading channel.

Fig. 3.8 illustrates the covert throughput, η, versus the covert constraint, ϵ. It can be seen that η increases as the covert constraint is relaxed, i.e. ϵ is increased. If ϵ is close to 1, the channel fading can help increase the covert throughput. This is because ξ_{min} decreases gradually (inverse linearly) with the received SNR in Rayleigh fading channel while it drops sharply to zero in AWGN channel (see Fig. 3.5). Therefore, if ξ_{min} is allowed to be small, i.e. ϵ is close to 1, the maximum allowed transmit power can be much higher in fading channel than in AWGN channel, which results in a higher (covert) throughput in fading channel.

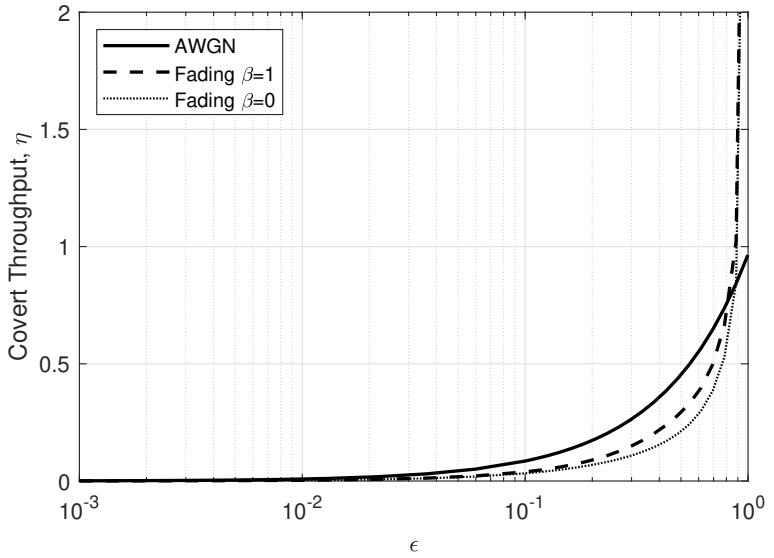

Figure 3.8 The covert throughput, η, versus ϵ; $\rho = 2$ (dB), $\sigma_h^2 = \sigma_g^2$ and $\sigma_b^2 = \hat{\sigma}_w^2$.

Fig. 3.9 illustrates the covert throughput gain, η_n/η_p, provided by the channel uncertainty at the adversary versus the covert constraint, ϵ, for different values of ρ. One can see that the covert throughput gain is more significant when the noise uncertainty ρ is larger and ϵ is larger. For $\epsilon = 0.1$, the covert throughput gain provided by the channel uncertainty is $12\% \sim 19\%$ for $\rho = 1 \sim 2$ dB.

3.5 Chapter Summary

In this chapter, we analyzed the joint impact of imperfect knowledge of the channel gain and noise power at the adversary on the total detection error probability and the covert throughput in Rayleigh fading channel. We found that the channel fading is crucial to hiding the signal transmission, particularly when the noise uncertainty is low or the receive SNR is high. We also found that the impact of the channel uncertainty on the total detection error probability and the covert throughput is particularly noticeable when the noise uncertainty is large. Imperfect knowledge of the channel gain at the adversary provides a covert throughput gain of $12\% \sim 19\%$ over the perfect channel knowledge when the noise uncertainty is in the range of $1 \sim 2$ dB.

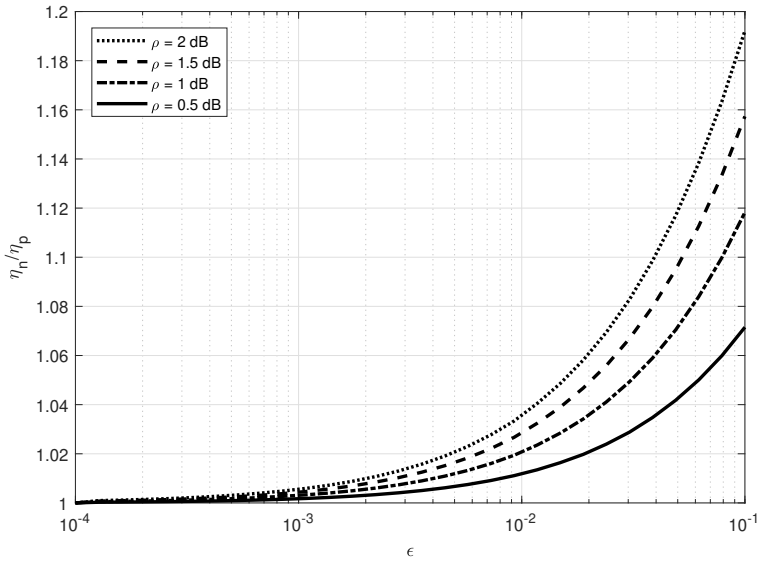

Figure 3.9 η_n/η_p versus ϵ for different values of ρ.

CHAPTER 4. COVERT NON-ORTHOGONAL MULTIPLE ACCESS

In this chapter, we study the privacy (covertness) in NOMA system. The covert message is superimposed onto K non-covert (public) messages in NOMA system such that the total transmission power remains the same whether or not the covert message is transmitted. We show that the covert message can be detected only when the non-covert message, where the covert message is superimposed onto, can be decoded. This suggests hiding the covert message under the non-covert message that is most difficult to decode. Hence, the effectiveness of hiding the covert message can be improved by exploiting the multiplicity of users in NOMA system. We determine the total detection error probability (sum of false alarm and missed detection probability) as a function of the number of users in Rayleigh fading channel. We show that it increases and converges to 1 as the number of non-covert users increases. This means that the covert transmission is undetectable if the number of non-covert users is sufficiently large. We also show that the total detection error probability can be increased as the transmit power is increased, thereby increasing the covert rate, by adapting the superposition rule to the channel variations.

The remaining part of this chapter is organized as follows. Section 4.1 describes the system model. Section 4.2 describes the optimum detection strategy at the adversary. Section 4.3 derives the optimum detection threshold that minimizes the total detection error probability. Section 4.4 derives the optimum cover set for the transmitter to maximize the total detection error probability and Section 4.5 derives the resulting maximum total detection error probability. Section 4.6 derives the decoding outage probability. Sections 4.7 and 4.8 describe the covert rate and the no-covert rate, respectively. Sections 4.9 and 4.10 describes the channel adaptation and multiple antenna at the transmitter, respectively. Section 4.11 shows the numerical results and 4.12 concludes the chapter.

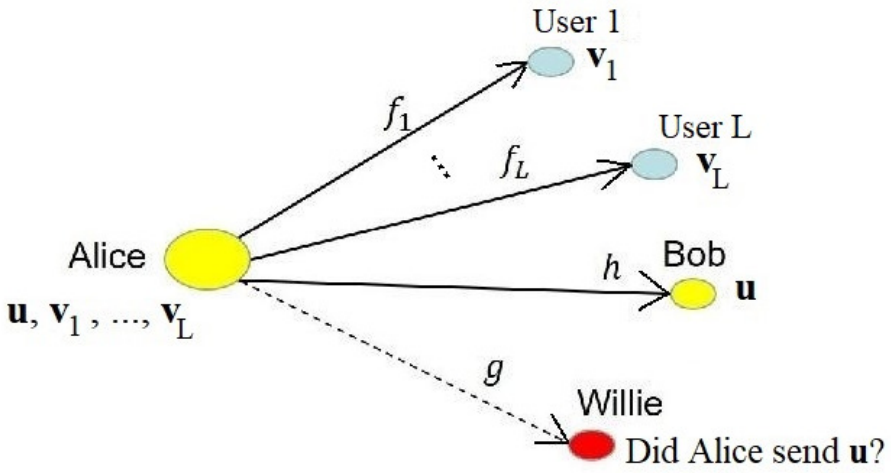

Figure 4.1 Alice attempts to hide the transmission of covert message $\mathbf{u}$ against Willie by using non-covert messages $\mathbf{v}_1, ..., \mathbf{v}_L$ as a cover (camouflage).

4.1 System Model

We consider a downlink NOMA system in which the transmitter (Alice) sends a covert message $\mathbf{u} = (u_1, ..., u_n)$ to a covert user (Bob) and non-covert messages $\mathbf{v}_1, ..., \mathbf{v}_L$, where $\mathbf{v}_j = (v_{1,j}, ..., v_{n,j})$, to L non-covert users. The system model is illustrated in Fig.4.1. A warden (Willie) is interested in finding whether any message other than $\mathbf{v}_1, ..., \mathbf{v}_L$ is sent from Alice. Willie's objectives is not to decode $\mathbf{u}$, but merely to detect the transmission of $\mathbf{u}$.

The transmitter hides the transmission of $\mathbf{u}$ by superimposing it onto K non-covert messages, where $K \leq L$. Let $\mathcal{E}$ denote the set of indices of those K non-covert messages and $\mathcal{E}^c$ denote the complement of $\mathcal{E}$. For example, if $\mathbf{u}$ is superimposed onto $\mathbf{v}_1$, $\mathbf{v}_5$, and $\mathbf{v}_9$, then $\mathcal{E} = \{1, 5, 9\}$. The optimum choice of $\mathcal{E}$ will be discussed in Section 4.4. Then, the transmitted signal from Willie's point of view can be expressed as

$$\mathbf{x} = \begin{cases} \sum_{l=1}^{L} \mathbf{v}_l, & H_0, \\ \sqrt{\alpha} \sum_{l \in \mathcal{E}} \mathbf{v}_l + \sqrt{1-\alpha}\mathbf{u} + \sum_{l \in \mathcal{E}^c} \mathbf{v}_l, & H_1, \end{cases} \tag{4.1}$$

where $\alpha \in (0,1)$ is the power allocation between the covert message and K non-covert messages where $\mathbf{u}$ is superimposed. H_0 denotes the hypothesis that Alice did not send the covert message $\mathbf{u}$ and H_1 denotes the alternative hypothesis that Alice did send $\mathbf{u}$. We assume that $\mathbf{v}_l$, $l = 1, ..., L$, and $\mathbf{u}$ are independent complex Gaussian random vectors with mean zero and variance P_l and $\sum_{l\in\mathcal{E}} P_l$, respectively, where $\sum_{l=1}^{L} P_l = P$. Note that the transmission power of $\mathbf{x}$ is equal to P regardless of the transmission of $\mathbf{u}$.

We assume that all nodes are equipped with single antenna. Let f_j, h and g denote the channel gain between Alice and the j-th non-covert user, that between Alice and Bob, and that between Alice and Willie, respectively. We assume they are independent complex Gaussian random variables with mean zero and variance σ_f^2, σ_h^2, and σ_g^2, respectively. Then, the received signal at the j-th user is given by

$$\mathbf{y}_j = \begin{cases} f_j \sum_{l=1}^{L} \mathbf{v}_l + \mathbf{n}_j, & H_0, \\ f_j(\sqrt{\alpha}\sum_{l\in\mathcal{E}} \mathbf{v}_l + \sqrt{1-\alpha}\mathbf{u} + \sum_{l\in\mathcal{E}^c} \mathbf{v}_l) + \mathbf{n}_j, & H_1, \end{cases} \tag{4.2}$$

where $\mathbf{n}_j = \{n_{1,j}, ..., n_{n,j}\}$ is the complex Gaussian noise vector with mean 0 and variance σ_n^2.

We assume that Alice knows f_l, $l = 1, ..., L$, perfectly via the forward channel training (pilot signal is sent by Alice and then the channel gains are estimated by the non-covert users and fedback to Alice) or the reverse channel training (pilot signals are sent by the non-covert users and Alice estimates the channel gains using the channel reciprocity). However, we assume Alice does not know h, because Bob does not want to reveal his presence and hence does not send his channel gain h to Alice.

4.1.1 Achievable Rate

Without loss of generality, we assume $|f_1|^2 \geq |f_2|^2 \geq ... \geq |f_L|^2$. Each user applies the successive interference cancellation (SIC) to decode its own message: the j-th user will first decode $\mathbf{v}_l$, $l > j$, and then eliminate it from $\mathbf{y}_j$ in a successive manner. $\mathbf{v}_l$, $l < j$, will be treated as noise. The achievable rate of the j-th user under H_0 is given by [21]

$$R_{j,0} = \log_2\left(1 + \frac{|f_j|^2\gamma_j}{1 + |f_j|^2\sum_{l<j}\gamma_l}\right), \tag{4.3}$$

where $\gamma_l = P_l/\sigma_n^2$. The achievable rate of the j-th user under H_1 depends on whether $\mathbf{u}$ is superimposed onto $\mathbf{v}_j$ ($j \in \mathcal{E}$) or not ($j \in \mathcal{E}^c$). If $j \in \mathcal{E}$, the j-th user, after cancelling $\mathbf{v}_l$, $l > j$, from $\mathbf{y}_j$ obtains

$$\mathbf{y}'_j = f_j \left(\sqrt{\alpha}\mathbf{v}_j + (\sqrt{\alpha} - 1) \sum_{l \in \mathcal{E}, l<j} \mathbf{v}_l + \sqrt{1-\alpha}\mathbf{u} + \sum_{l \in \mathcal{E}^c, l<j} \mathbf{v}_l \right) + \mathbf{n}_j. \tag{4.4}$$

Then, the achievable rate of the j-th user under H_1 is given by

$$R_{j,1} = \log_2 \left(1 + \frac{|f_j|^2 \alpha \gamma_j}{1 + |f_j|^2 \big((\sqrt{\alpha} - 1)^2 \sum\limits_{l \in \mathcal{E}, l<j} \gamma_l + (1-\alpha) \sum\limits_{l \in \mathcal{E}} \gamma_l + \sum\limits_{l \in \mathcal{E}^c, l<j} \gamma_l \big)} \right) \tag{4.5}$$

for $j \in \mathcal{E}$.

4.1.2 Transmission Rate

We assume Alice sends $\mathbf{v}_j$, $j \in \mathcal{E}$, at the rate of $R_{j,1}$ in order to guarantee that the j-th user can decode $\mathbf{v}_j$ regardless of the transmission of $\mathbf{u}$.

4.2 Willie's Detection Strategy

The received signal at Willie is given by

$$\mathbf{y}_w = \begin{cases} g \sum_{l=1}^{L} \mathbf{v}_l + \mathbf{n}_w, & H_0, \\ g(\sqrt{\alpha} \sum_{l \in \mathcal{E}} \mathbf{v}_l + \sqrt{1-\alpha}\mathbf{u} + \sum_{l \in \mathcal{E}^c} \mathbf{v}_l) + \mathbf{n}_w, & H_1, \end{cases} \tag{4.6}$$

where $\mathbf{n}_w = \{n_{1,w}, ..., n_{n,w}\}$ is the complex Gaussian vector with mean 0 and variance σ_n^2, Based on his observation vector $\mathbf{y}_w$ of length n, Willie has to decide between the hypotheses, H_0 and H_1, regarding the transmission of $\mathbf{u}$.

We assume Willie knows the complete statistics of his observations under both hypotheses. This means he knows all parameters, such as noise variance, signal power and channel coefficient g. He uses a radio meter (energy detector), which is optimal when the message and the noise are modeled as white Gaussian processes[1] [61, 8], to detect the covert message by comparing the radio meter

[1]For other models, such as zero-mean finite signal constellation and orthogonal frequency-division multiplexing (OFDM) signal, energy detection perform close to the optimal detector [56, 7].

output to a pre-determined threshold λ, namely

$$T = \frac{1}{n}||\mathbf{y}_w||^2 \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda, \tag{4.7}$$

where λ is a detection threshold. If Willie is allowed to observe an infinite number of samples, which minimizes the total detection error probability, the test statistic T converges to

$$T \to \mathrm{E}[|y_{i,n}|^2] = |g|^2 \sum_{l=1}^{L} P_l + \sigma_n^2, \tag{4.8}$$

by the law of large numbers under both H_0 and H_1. The event of false alarm occurs if $T > \lambda$ when Alice did not send $\mathbf{u}$ and missed detection occurs if $T \leq \lambda$ when Alice did send $\mathbf{u}$. Since the test statistic T are identical under H_0 and H_1, the transmission of the covert message cannot be detected from $\mathbf{y}_w$.

However, if Willie succeeds in decoding some $\mathbf{v}_l$'s and subtracts them from $\mathbf{y}_w$ to get

$$\mathbf{y}'_w = \mathbf{y}_w - g\sum_{l\in\mathcal{D}} \mathbf{v}_l \tag{4.9}$$

$$= \begin{cases} g\sum_{l\in\mathcal{D}^c} \mathbf{v}_l + \mathbf{n}_W, & H_0, \\ g(\sqrt{\alpha}\sum_{l\in\mathcal{E}\cap\mathcal{D}^c} \mathbf{v}_l + (\sqrt{\alpha}-1)\sum_{l\in\mathcal{E}\cap\mathcal{D}} \mathbf{v}_l + \sqrt{1-\alpha}\mathbf{u} + \sum_{l\in\mathcal{E}^c\cap\mathcal{D}^c} \mathbf{v}_l) + \mathbf{n}_W, & H_1, \end{cases} \tag{4.10}$$

where $\mathcal{D}$ denotes the index set of the non-covert messages that are decoded by Willie after the SIC is completed and $\mathcal{D}^c$ denotes the complement of $\mathcal{D}$, then the presence of $\mathbf{u}$ can be determined from

$$T' = \frac{1}{n}||\mathbf{y}'_w||^2 \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda', \tag{4.11}$$

where λ' is a detection threshold. It can be shown that the test statistic T' converges to

$$T' \to \begin{cases} |g|^2 \sum_{l\in\mathcal{D}^c} P_l + \sigma_n^2, & H_0, \\ |g|^2(\sum_{l\in\mathcal{D}^c} P_l + 2(1-\sqrt{\alpha})\sum_{l\in\mathcal{E}\cap\mathcal{D}} P_l) + \sigma_n^2, & H_1 \end{cases} \tag{4.12}$$

as $n \to \infty$.

Remark 1: T' under H_1 is different from that under H_0 if $\mathcal{E}\cap\mathcal{D} \neq \emptyset$. Hence, the transmission of the covert message can be detected perfectly from T' (when $n \to \infty$) if $\mathcal{E}\cap\mathcal{D} \neq \emptyset$, i.e. any $\mathbf{v}_l$, $l \in \mathcal{E}$, is decoded by Willie.

4.3 Optimum Detection Threshold for Willie

The total detection error probability, averaged over the events of $\mathcal{E}\cap\mathcal{D}=\emptyset$ and $\mathcal{E}\cap\mathcal{D}\neq\emptyset$, is given by

$$\begin{aligned}P_F+P_M=&\Pr(T>\lambda,\mathcal{E}\cap\mathcal{D}=\emptyset|H_0)+\Pr(T'>\lambda',\mathcal{E}\cap\mathcal{D}\neq\emptyset|H_0)\\&+\Pr(T\leq\lambda,\mathcal{E}\cap\mathcal{D}=\emptyset|H_1)+\Pr(T'\leq\lambda',\mathcal{E}\cap\mathcal{D}\neq\emptyset|H_1).\end{aligned}\tag{4.13}$$

Willie's goal is to minimize P_F+P_M by choosing the detection thresholds λ and λ' properly.

i) $\boldsymbol{\lambda'}$: Willie can choose $\lambda'\in[|g|^2\sum_{l\in\mathcal{D}^c}P_l+\sigma_n^2,|g|^2(\sum_{l\in\mathcal{D}^c}P_l+2(1-\sqrt{\alpha})\sum_{l\in\mathcal{E}\cap\mathcal{D}}P_l)+\sigma_n^2)$ to make the second and fourth term in (4.13) zero.

ii) $\boldsymbol{\lambda}$: Willie can choose $\lambda<T$, namely $\lambda<|g|^2\sum_{l=1}^{L}P_l+\sigma_n^2$, if $\Pr(\mathcal{E}\cap\mathcal{D}=\emptyset|H_0)<\Pr(\mathcal{E}\cap\mathcal{D}=\emptyset|H_1)$ and, otherwise, choose $\lambda\geq T$ to minimize P_F+P_M.

The resulting minimum total detection error probability is given by

$$\xi_{min}:=\min_{\lambda,\lambda'}\ P_F+P_M \tag{4.14}$$

$$=\min\{\Pr(\mathcal{E}\cap\mathcal{D}=\emptyset|H_0),\Pr(\mathcal{E}\cap\mathcal{D}=\emptyset|H_1)\} \tag{4.15}$$

$$=\min\left\{\Pr\left(\bigcap_{j\in\mathcal{E}}I(\mathbf{v}_j;\mathbf{y}_{w,j})<R_{j,1}\middle|H_0\right),\Pr\left(\bigcap_{j\in\mathcal{E}}I(\mathbf{v}_j;\mathbf{y}_{w,j})<R_{j,1}\middle|H_1\right)\right\}, \tag{4.16}$$

where $\mathbf{y}_{w,j}=\mathbf{y}_w-g\sum_{l\in\mathcal{D}_j}\mathbf{v}_l$, $\mathcal{D}_j$ is the decoding set prior to decoding $\mathbf{v}_j$, and

$$I(\mathbf{v}_j;\mathbf{y}_{w,j})=\begin{cases}\log_2\left(1+\frac{|g|^2\gamma_j}{1+|g|^2(\sum_{l\in\mathcal{D}_j^c}\gamma_l-\gamma_j)}\right), & H_0,\\ \log_2\left(1+\frac{|g|^2\alpha\gamma_j}{1+|g|^2(\sum_{l\in\mathcal{D}_j^c}\gamma_l+2(1-\sqrt{\alpha})\sum_{l\in\mathcal{E}\cap\mathcal{D}_j}\gamma_l-\alpha\gamma_j)}\right), & H_1\end{cases}\tag{4.17}$$

is the mutual information between $\mathbf{v}_j$ and $\mathbf{y}_{w,j}$. (4.17) can be obtained from (4.10) with $\mathcal{D}$ replaced by $\mathcal{D}_j$. It can be shown from (4.17) that $I(\mathbf{v}_j;\mathbf{y}_{w,j})$ under H_1 is smaller than that under H_0. Therefore, Willie will choose $\lambda<T$, i.e. $\lambda<|g|^2\sum_{l=1}^{L}P_l+\sigma_n^2$, to minimize P_F+P_M, which yields

$$\xi_{min}=\Pr\left(\bigcap_{j\in\mathcal{E}}I(\mathbf{v}_j;\mathbf{y}_{w,j})<R_{j,1}\middle|H_0\right). \tag{4.18}$$

4.4 Optimum Cover Set $\mathcal{E}$ for Alice

Alice's goal is to maximize ξ_{min} by choosing the cover set $\mathcal{E}$ properly. It can be shown from (4.18) that the optimum $|\mathcal{E}|$ that maximizes ξ_{min} is 1. This means the covert message $\mathbf{u}$ should be superimposed onto one non-covert message. The resulting maximum total detection error probability is given by

$$\xi_{max} := \max_{\mathcal{E}} \xi_{min} \tag{4.19}$$

$$= \max_{j \in \mathcal{E}} \Pr(I(\mathbf{v}_j; \mathbf{y}_{w,j}) < R_{j,1} | H_0), \tag{4.20}$$

where

$$R_{j,1} = \log_2 \left(1 + \frac{|f_j|^2 \alpha \gamma_j}{1 + |f_j|^2((1-\alpha)\gamma_j + \sum_{l=1}^{j-1} \gamma_l)} \right), \tag{4.21}$$

which is obtained from (4.5) with $\mathcal{E} = \{j\}$. Therefore, it follows from (4.17), (4.20) and (4.21) that

$$\xi_{max} = \max_{j \in \mathcal{E}} \Pr \left(|f_j|^2 > \frac{|g|^2}{(\alpha + |g|^2(\alpha \sum_{l \in \mathcal{D}_j^c} \gamma_l - \sum_{l=1}^{j} \gamma_l))^+} \right), \tag{4.22}$$

where $(x)^+ = \max\{0, x\}$.

When g and $\mathcal{D}_j^c$ are unknown to Alice, the maximum of (4.22) is achieved by maximizing $|f_j|^2$ and minimizing $\sum_{l=1}^{j} \gamma_l$. Since $|f_1|^2 = \max_{1 \le l \le L} |f_l|^2$ and $\sum_{l=1}^{j} \gamma_l$ decreases with decreasing j, the maximum of (4.22) is achieved when $j = 1$, i.e. superimposing $\mathbf{u}$ onto $\mathbf{v}_1$, which yields

$$\xi_{max} = \Pr(I(\mathbf{v}_1; \mathbf{y}_{w,1}) < R_{1,1} | H_0) \tag{4.23}$$

$$= \Pr \left(|f_1|^2 > \frac{|g|^2}{(\alpha + |g|^2(\alpha \sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1))^+} \right). \tag{4.24}$$

Remark 2: The best hiding strategy that maximizes the total detection error probability is to superimpose the covert message $\mathbf{u}$ onto the non-covert message $\mathbf{v}_1$ that experiences the highest channel gain among all non-covert users.

4.5 Maximum Total Detection Error Probability

In this section, we determine the maximum total detection error probability, ξ_{max}. By the law of total probability, ξ_{max} can be expressed as

$$\xi_{max} = 1 - \sum_{k=0}^{L} \Pr\left(|f_1|^2 \leq \frac{|g|^2}{(\alpha + |g|^2(\alpha \sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1))^+}, |f_{k+1}|^2 \leq |g|^2 < |f_k|^2 \right), \tag{4.25}$$

where $|f_0|^2 = \infty$ and $|f_{L+1}|^2 = 0$.

i) $k = 0$, i.e. $|f_1|^2 \leq |g|^2 < \infty$: Willie can decode $\mathbf{v}_2, ..., \mathbf{v}_L$ if $|f_1|^2 \leq |g|^2$. Hence, $\mathcal{D}_1^c = \{1\}$. Since $|g|^2/(\alpha + |g|^2(\alpha - 1)\gamma_1)^+ > |g|^2$ for $\alpha \in (0, 1)$, we obtain

$$\Pr\left(|f_1|^2 \leq \frac{|g|^2}{(\alpha + |g|^2(\alpha - 1)\gamma_1)^+}, |f_1|^2 \leq |g|^2 \right) = \Pr(|f_1|^2 \leq |g|^2) \tag{4.26}$$

$$= \Pr\left(\bigcap_{l=1}^{L} \{|f_l|^2 \leq |g|^2\} \right) \tag{4.27}$$

$$= \int_0^\infty \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^L \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \tag{4.28}$$

$$= \frac{\sigma_f^2}{\sigma_g^2} B\left(\frac{\sigma_f^2}{\sigma_g^2}, L + 1\right), \tag{4.29}$$

where $B(n, m) = \int_0^1 z^{n-1}(1 - z)^{m-1} dz = \int_0^\infty (1 - e^{-t})^{m-1}(e^{-t})^n dt$ is the Beta function [4].

ii) $1 \leq k \leq L$: Willie can decode $\mathbf{v}_{k+1}, ..., \mathbf{v}_L$ if $|f_{k+1}|^2 \leq |g|^2 < |f_k|^2$. Hence, $\mathcal{D}_1^c = \{1, ..., k\}$. Let

$$\nu_k(|g|^2) = \frac{|g|^2}{\left(\alpha + |g|^2(\alpha \sum_{l=1}^{k} \gamma_l - \gamma_1)\right)^+}. \tag{4.30}$$

Since $\Pr(|f_1|^2 \leq \nu_k(|g|^2), |g|^2 < |f_k|^2) = 0$ if $|g|^2 \geq \nu_k(|g|^2)$ or equivalently $|g|^2 \geq \alpha_k$ where

$$\alpha_k = \frac{1 - \alpha}{(\alpha \sum_{l=1}^{k} \gamma_l - \gamma_1)^+}, \tag{4.31}$$

we obtain

$$\Pr(|f_1|^2 \leq \nu_k(|g|^2), |f_{k+1}|^2 \leq |g|^2 < |f_k|^2)$$

$$= \int_0^{\alpha_k} \Pr(|f_1|^2 \leq \nu_k(x), |f_{k+1}|^2 \leq x < |f_k|^2) \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \tag{4.32}$$

$$= \int_0^{\alpha_k} \binom{L}{k} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{\nu_k(x)}{\sigma_f^2}}\right)^k \frac{e^{-\frac{x}{\sigma_g^2}}}{\sigma_g^2} dx. \tag{4.33}$$

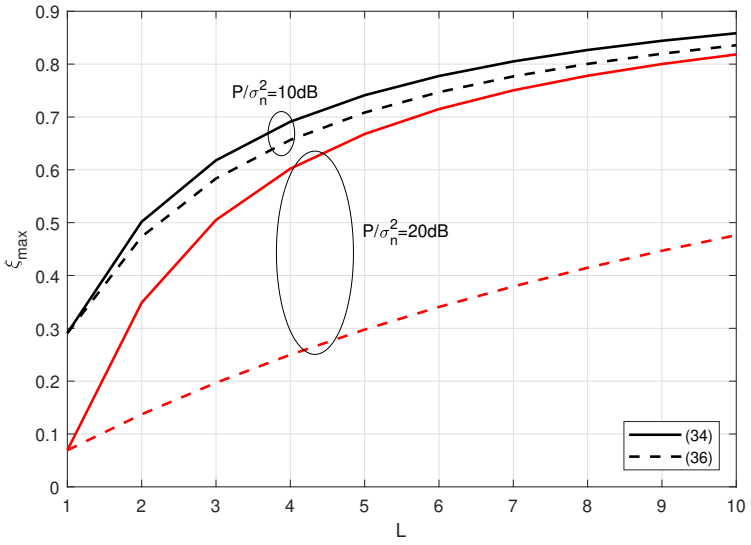

Figure 4.2 The maximum total detection error probability, ξ_{max}, versus L for different values of P/σ_n^2; $\alpha = 0.9$, $P_l = P/L$ for $1 \leq l \leq L$ and $\sigma_f^2 = \sigma_g^2 = 1$.

Therefore, it follows from (4.25), (4.29) and (4.33) that the maximum total detection error probability is given by

$$\xi_{max} = 1 - \frac{\sigma_f^2}{\sigma_g^2} B\left(\frac{\sigma_f^2}{\sigma_g^2}, L+1\right) - \sum_{k=1}^{L} \binom{L}{k} \int_0^{\alpha_k} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{\nu_k(x)}{\sigma_f^2}}\right)^k \frac{e^{-\frac{x}{\sigma_g^2}}}{\sigma_g^2} dx. \tag{4.34}$$

Remark 3: When $\mathcal{D}_1^c = \{1\}$, which is assumed in the analysis of secrecy in NOMA [29, 74], it follows from (4.25) that

$$\xi_{max} = 1 - \Pr\left(|f_1|^2 \leq \frac{|g|^2}{(\alpha + |g|^2(\alpha-1)\gamma_1)^+}\right) \tag{4.35}$$

$$= 1 - \int_0^{\infty} \left(1 - e^{-\frac{x}{\sigma_f^2(\alpha + x(\alpha-1)\gamma_1)^+}}\right)^L \frac{e^{-\frac{x}{\sigma_g^2}}}{\sigma_g^2} dx. \tag{4.36}$$

It can be seen in Fig. 4.2 that ξ_{max} in (4.36) is less than that in (4.34) and their difference is more significant for larger SNR or L. This indicates that the assumption of $\mathcal{D}_1^c = \{1\}$, which simplifies the analysis, overestimates Willie's ability of decoding and then causes the incorrect result.

Remark 4: ξ_{max} in (4.34) converges to

$$\xi_{max} \to 1 \tag{4.37}$$

as $L \to \infty$ regardless of the SNR. This means the transmission of $\mathbf{u}$ is undetectable if L is sufficiently large. The proof is provided in Appendix C.1.

4.6 Decoding Outage Probability

In this section, we determine the decoding outage probability of the covert message at Bob (covert user). The received signal at Bob when $\mathbf{u}$ is superimposed onto $\mathbf{v}_1$ is given by

$$\mathbf{y}_b = h\left(\sqrt{\alpha}\mathbf{v}_1 + \sqrt{1-\alpha}\mathbf{u} + \sum_{l=2}^{L}\mathbf{v}_l\right) + \mathbf{n}_b, \tag{4.38}$$

where $\mathbf{n}_b \sim CN(0, \sigma_n^2)$. Since the transmitter does not know the channel gain of the covert user, the covert message is sent at a fixed rate R_u. Then, the decoding outage probability of the covert message is given by

$$P_{o,u} = \Pr(I(\mathbf{u}; \mathbf{y}_b) < R_u). \tag{4.39}$$

By the law of total probability, $P_{o,u}$ can be expressed as

$$P_{o,u} = 1 - \sum_{k=0}^{L} \Pr(I(\mathbf{u}; \mathbf{y}_b) \geq R_u, |f_{k+1}|^2 \leq |h|^2 < |f_k|^2). \tag{4.40}$$

If $|f_{k+1}|^2 \leq |h|^2 < |f_k|^2$, Bob can decode and remove $\mathbf{v}_{k+1}, ..., \mathbf{v}_L$ from $\mathbf{y}_b$ to get

$$\mathbf{y}_b' = \begin{cases} h\left(\sqrt{\alpha}\mathbf{v}_1 + \sqrt{1-\alpha}\mathbf{u} + \sum_{l=2}^{k}\mathbf{v}_l\right) + \mathbf{n}_b, & 1 \leq k \leq L, \\ h\sqrt{1-\alpha}\mathbf{u} + \mathbf{n}_b, & k = 0 \end{cases} \tag{4.41}$$

since $R_{1,1} = \log_2\left(1 + \frac{|f_1|^2\alpha\gamma_1}{1+|f_1|^2(1-\alpha)\gamma_1}\right)$, $R_{j,0} = \log_2\left(1 + \frac{|f_j|^2\gamma_j}{1+|f_j|^2\sum_{l<j}\gamma_l}\right)$ for $2 \leq j \leq L$, and

$$I(\mathbf{v}_j; \mathbf{y}_b) = \begin{cases} \log_2\left(1 + \frac{|h|^2\alpha\gamma_1}{1+|h|^2(1-\alpha)\gamma_1}\right), & j = 1, \\ \log_2\left(1 + \frac{|h|^2\gamma_j}{1+|h|^2\sum_{l<j}\gamma_l}\right), & j \geq 1. \end{cases} \tag{4.42}$$

Then, the achievable rate of the covert message for Bob given $|f_{k+1}|^2 \leq |h|^2 < |f_k|^2$ is given by

$$I(\mathbf{u}; \mathbf{y}_b')_k = \begin{cases} \log_2\left(1 + \frac{|h|^2(1-\alpha)\gamma_1}{1+|h|^2(\sum_{l=2}^{k}\gamma_l + \alpha\gamma_1)}\right), & 1 \leq k \leq L, \\ \log_2\left(1 + |h|^2(1-\alpha)\gamma_1\right), & k = 0. \end{cases} \tag{4.43}$$

Therefore, we obtain

$$P_{o,u} = 1 - \sum_{k=0}^{L} \Pr(I(\mathbf{u}; \mathbf{y}_b')_k \geq R_u, |f_{k+1}|^2 \leq |h|^2 < |f_k|^2) \tag{4.44}$$

$$= 1 - \sum_{k=0}^{L} \Pr(|h|^2 \geq \mu_k, |f_{k+1}|^2 \leq |h|^2 < |f_k|^2), \tag{4.45}$$

where

$$\mu_k = \begin{cases} \frac{2^{R_u}-1}{\left((1-\alpha)\gamma_1 - (2^{R_u}-1)(\sum_{l=2}^{k}\gamma_l + \alpha\gamma_1)\right)^+}, & 1 \leq k \leq L, \\ (2^{R_u}-1)/((1-\alpha)\gamma_1), & k = 0. \end{cases} \tag{4.46}$$

Since $|h|^2$ is an exponential random variable with mean σ_h^2, we obtain

$$P_{o,u} = 1 - \sum_{k=0}^{L} \int_{\mu_k}^{\infty} \binom{L}{k} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \left(e^{-\frac{x}{\sigma_f^2}}\right)^k \frac{e^{-x/\sigma_h^2}}{\sigma_h^2} dx \tag{4.47}$$

$$= 1 - \frac{\sigma_f^2}{\sigma_h^2} \sum_{k=0}^{L} \binom{L}{k} B\left(e^{-\frac{\mu_k}{\sigma_f^2}}; k + \frac{\sigma_f^2}{\sigma_h^2}, L - k + 1\right), \tag{4.48}$$

where $B(e^{-x}; n, m) = \int_0^{e^{-x}} z^{n-1}(1-z)^{m-1} dz = \int_x^{\infty} (1-e^{-t})^{m-1}(e^{-t})^n dt$ is the incomplete Beta function and $B(0; n, m) = 0$ and $B(1; n, m) = B(n, m)$ [4].

4.7 Covert Rate

In this section, we analyze the covert rate, defined as the maximum reliable transmission rate (bits/s/Hz) between Alice and Bob subject to the covert constraint of $\xi_{max} \geq 1 - \epsilon$ for some ϵ which represents the covertness requirement. The average rate received over many transmission bursts is $R_u(1 - P_{o,u})$ since the covert message is correctly received on $1 - P_{o,u}$ transmissions. Therefore, the covert rate (bits/Hz/s) is given by

$$\eta(R_u) = R_u(1 - P_{o,u}) \tag{4.49}$$

$$\text{s.t.} \quad \xi_{max} \geq 1 - \epsilon \tag{4.50}$$

for a positive constant ϵ. One can then optimize over R_u to get the maximum covert rate

$$\eta_u := \max_{R_u} \eta(R_u). \tag{4.51}$$

4.8 Non-covert Rate

In this section, we determine the rate of the non-covert message $\mathbf{v}_1$ where the covert message $\mathbf{u}$ is superimposed. It follows from (4.21) that the non-covert rate (bits/Hz/s) of $\mathbf{v}_1$ is given by

$$\eta_v = \int_0^\infty \log_2\left(1 + \frac{x\alpha\gamma_1}{1 + x(1-\alpha)\gamma_1}\right) f_{|f_1|^2}(x)dx, \tag{4.52}$$

where

$$f_{|f_1|^2}(x) = L(1 - e^{-x/\sigma_f^2})^{L-1} e^{-x/\sigma_f^2}/\sigma_f^2 \tag{4.53}$$

is the probability density function (PDF) of $|f_1|^2$ [20].

4.9 Channel Adaptation

In this section, we exploit the channel variation to minimize the throughput loss for $\mathbf{v}_j$, $j \in \mathcal{E}$, caused by the superposition (transmission) of $\mathbf{u}$. We consider sending $\mathbf{u}$ only when the channel gain f_j of $\mathbf{v}_j$ is below a threshold, τ, i.e. $|f_j|^2 < \tau$, where the transmission rate of $\mathbf{v}_j$ is low. Therefore, the transmitted signal is given by

$$\mathbf{x} = \begin{cases} \sum_{l=1}^{L} \mathbf{v}_l, & |f_j|^2 \geq \tau, \\ \sqrt{\alpha}\mathbf{v}_j + \sqrt{1-\alpha}\mathbf{u} + \sum_{l\neq j} \mathbf{v}_l, & |f_j|^2 < \tau. \end{cases} \tag{4.54}$$

We assume the reverse channel training is employed to guarantee that Willie cannot estimate f_l's, and hence cannot detect the transmission of $\mathbf{u}$ based on f_l's.

We consider sending $\mathbf{v}_j$, $j \in \mathcal{E}$, at the rate $R_{j,0}$ which guarantees the successful decoding $\mathbf{v}_j$ by the intended receiver when $\mathbf{u}$ is not transmitted, i.e. $|f_j|^2 \geq \tau$. When $\mathbf{u}$ is transmitted, i.e. $|f_j|^2 < \tau$, then the intended receiver of $\mathbf{v}_j$ will fail to decode it due to the interference of $\mathbf{u}$. However, throughput loss of $\mathbf{v}_j$ can be small because its transmission rate is small when $|f_j|^2 < \tau$. Therefore, we obtain from (4.16) with $|\mathcal{E}| = 1$ and $R_{j,1}$ replaced by $R_{j,0}$ that

$$\xi_{min} = \min\{\Pr\left(I(\mathbf{v}_j; \mathbf{y}_{w,j}) < R_{j,0} | \, |f_j|^2 \geq \tau\right), \Pr\left(I(\mathbf{v}_j; \mathbf{y}_{w,j}) < R_{j,0} | \, |f_j|^2 < \tau\right)\}. \tag{4.55}$$

4.9.1 Maximum Total Detection Error Probability

Similar to the case of no channel adaptation in Section 4.4, it can be shown that ξ_{min} is maximized when $j = 1$, i.e. $\mathbf{u}$ is superimposed onto $\mathbf{v}_1$. The proof is provided in Appendix C.2. Then, the maximum detection error probability is given by

$$\xi_{max} = \min\{\xi_0, \xi_1\}, \tag{4.56}$$

where

$$\xi_0 = \Pr(I(\mathbf{v}_1; \mathbf{y}_{w,1}) < R_{1,0} | \, |f_1|^2 \geq \tau) \tag{4.57}$$

$$= \Pr\left(|f_1|^2 > \frac{|g|^2}{1 + |g|^2(\sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1)} \middle| |f_1|^2 \geq \tau \right) \tag{4.58}$$

$$= 1 - \frac{\frac{\sigma_f^2}{\sigma_g^2} B\left(e^{-\frac{\tau}{\sigma_f^2}}; \frac{\sigma_f^2}{\sigma_g^2}, L+1\right)}{1 - (1 - e^{-\tau/\sigma_f^2})^L} + \frac{\left(1 - e^{-\frac{\tau}{\sigma_f^2}}\right)^L e^{-\frac{\tau}{\sigma_g^2}}}{1 - (1 - e^{-\tau/\sigma_f^2})^L}, \tag{4.59}$$

where $B(e^{-x}; n, m) = \int_0^{e^{-x}} z^{n-1}(1-z)^{m-1} dz = \int_x^{\infty} (1 - e^{-t})^{m-1} (e^{-t})^n dt$ is the incomplete Beta function [4], and

$$\xi_1 = \Pr(I(\mathbf{v}_1; \mathbf{y}_{w,1}) < R_{1,0} | \, |f_1|^2 < \tau) \tag{4.60}$$

$$= \Pr\left(|f_1|^2 > \frac{|g|^2 \alpha}{1 + |g|^2(\sum_{l \in \mathcal{D}_1^c} \gamma_l - \alpha\gamma_1)} \middle| |f_1|^2 < \tau \right) \tag{4.61}$$

$$= 1 - \frac{\int_0^{\infty} \left(1 - e^{-\frac{x\alpha/\sigma_f^2}{1 + x(1-\alpha)\gamma_1}}\right)^L \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx}{(1 - e^{-\tau/\sigma_f^2})^L}, \tag{4.62}$$

for $\tau > \frac{\alpha}{(1-\alpha)\gamma_1}$ and

$$\xi_1 = 1 - e^{-\frac{a_1}{\sigma_f^2}} - \frac{\int_0^{a_1} \left(1 - e^{-\frac{x\alpha/\sigma_f^2}{1 + x(1-\alpha)\gamma_1}}\right)^L \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx}{(1 - e^{-\tau/\sigma_f^2})^L}. \tag{4.63}$$

for $\tau \leq \frac{\alpha}{(1-\alpha)\gamma_1}$, where $a_1 = \frac{\tau}{\alpha - (1-\alpha)\gamma_1 \tau}$. The proof of (4.58)-(4.63) is provided in Appendix C.3.

Remark 5: ξ_{max} converges to

$$\xi_{max} \to 1 \tag{4.64}$$

as $L \to \infty$ if $\tau > \frac{\alpha}{(1-\alpha)\gamma_1}$. This means the transmission of the covert message is completely undetectable if the number of non-covert users is sufficiently large. The proof is provided in Appendix C.4.

4.9.2 Decoding Error Probability

In this subsection, we determine the decoding outage probability of the covert message $\mathbf{u}$ and the non-covert message $\mathbf{v}_1$ where $\mathbf{u}$ is superimposed.

4.9.2.1 Covert message

Since we transmit $\mathbf{u}$ when $|f_1|^2 < \tau$, the decoding outage probability of $\mathbf{u}$ for Bob is given by

$$P_{o,u} = \Pr(I(\mathbf{u};\mathbf{y}_b) < R_u \,|\, |f_1|^2 < \tau) \tag{4.65}$$

By the law of total probability, $P_{o,u}$ can be expressed as

$$\begin{aligned} P_{o,u} = & 1 - \Pr\left(I(\mathbf{u};\mathbf{y}_b')_0 \geq R_u, |h|^2 \geq \frac{|f_1|^2}{(\alpha - |f_1|^2(1-\alpha)\gamma_1)^+} \middle| |f_1|^2 < \tau\right) \\ & - \Pr\left(I(\mathbf{u};\mathbf{y}_b')_1 \geq R_u, |f_2|^2 \leq |h|^2 < \frac{|f_1|^2}{(\alpha - |f_1|^2(1-\alpha)\gamma_1)^+} \middle| |f_1|^2 < \tau\right) \\ & - \sum_{k=2}^{L} \Pr(I(\mathbf{u};\mathbf{y}_b')_k \geq R_u, |f_{k+1}|^2 \leq |h|^2 < |f_k|^2 || f_1|^2 < \tau), \end{aligned} \tag{4.66}$$

where $I(\mathbf{u};\mathbf{y}_b')_k$ is given in (4.43). This is because, from (4.38) and $R_{j,0} = \log_2\left(1 + \frac{|f_j|^2\gamma_j}{1+|f_j|^2\sum_{l<j}\gamma_l}\right)$ for $1 \leq j \leq L$, we obtain that Bob can decode and remove $\mathbf{v}_{k+1}, ..., \mathbf{v}_L$ if $|h|^2 \geq \frac{|f_1|^2}{(\alpha - |f_1|^2(1-\alpha)\gamma_1)^+}$ for $k = 0$, if $|f_2|^2 \leq |h|^2 < \frac{|f_1|^2}{(\alpha - |f_1|^2(1-\alpha)\gamma_1)^+}$ for $k = 1$, and if $|f_{k+1}|^2 \leq |h|^2 < |f_k|^2$ for $2 \leq k \leq L$.

Therefore, we obtain

$$P_{o,u}=1-\frac{\int_{\mu_0}^{\infty}\left(1-e^{-\min\left\{\frac{\tau}{\sigma_f^2},\frac{x\alpha}{\sigma_f^2(1+x(1-\alpha)\gamma_1)}\right\}}\right)^L\frac{e^{-x/\sigma_h^2}}{\sigma_h^2}dx}{(1-e^{-\tau/\sigma_h^2})^L}$$
$$-\frac{\int_{\mu_1}^{\infty}\left(1-e^{-\min\left\{\frac{\tau}{\sigma_f^2},\frac{x}{\sigma_f^2}\right\}}\right)^L\frac{e^{-x/\sigma_h^2}}{\sigma_h^2}dx-\int_{\mu_1}^{\infty}\left(1-e^{-\min\left\{\frac{\tau}{\sigma_f^2},\frac{x\alpha}{\sigma_f^2(1+x(1-\alpha)\gamma_1)}\right\}}\right)^L\frac{e^{-x/\sigma_h^2}}{\sigma_h^2}dx}{(1-e^{-\tau/\sigma_h^2})^L}$$
$$-\sum_{k=1}^{L}\binom{L}{k}\frac{\int_{\mu_k}^{\tau}\left(1-e^{-\frac{x}{\sigma_f^2}}\right)^{L-k}\left(e^{-\frac{x}{\sigma_f^2}}-e^{-\frac{\tau}{\sigma_f^2}}\right)^k\frac{e^{-x/\sigma_h^2}}{\sigma_h^2}dx}{(1-e^{-\tau/\sigma_h^2})^L}. \tag{4.67}$$

The proof of (4.67) is provided in Appendix C.5.

4.9.2.2 Non-covert message

Since the transmission rate of $R_{1,0}$ guarantees the successful decoding of $\mathbf{v}_1$ only when $\mathbf{u}$ is not sent, i.e. $|f_1|^2 \geq \tau$, the intended receiver of $\mathbf{v}_1$ will fail to decode it with probability

$$P_{o,v} = \Pr(|f_1|^2 < \tau) \tag{4.68}$$
$$= (1-e^{-\tau/\sigma_f^2})^L. \tag{4.69}$$

For given $P_{o,v}$, the required τ is

$$\tau = \sigma_f^2 \ln\left(\frac{1}{1-P_{o,v}^{1/L}}\right), \tag{4.70}$$

which is an increasing function of L.

4.9.3 Covert Rate

The covert rate is given by

$$\eta(R_u) = R_u(1-P_{o,u})P_{o,v} \tag{4.71}$$
$$\text{s.t.} \quad \xi_{max} \geq 1-\epsilon, \tag{4.72}$$

where the factor $P_{o,v}$ is to account for the fact that the covert message is sent only when $|f_1|^2 > \tau$.

4.9.4 Non-covert Rate

The intended receiver of $\mathbf{v}_1$ will fail to decode it if $|f_1|^2 \geq \tau$ due to the interference from the covert message. Therefore, it follows from (4.3) that the non-covert rate of $\mathbf{v}_1$ is given by

$$\eta_v = \int_{\tau}^{\infty} \log_2(1 + x\gamma_1) f_{|f_1|^2}(x) dx. \tag{4.73}$$

4.10 Multiple Antenna at Transmitter

In this section, we consider that the transmitter has M antennas and select one out of M antennas to transmit message information. In this case, only one RF chain is needed such that the hardware cost, the power consumption of circuit and the computational complexity is reduced while the diversity of transmitter's multiple antennas is preserved [21, 69]. We assume the m-th antenna is selected at the transmitter. For convenience, we change the denotation of channel gains: $g \to g_m$, $h \to h_m$ and $f_j \to f_{m,j}$ for $1 \leq j \leq L$ and $1 \leq m \leq M$.

When the m-th antenna is selected, it follows from Remark 2 that superimposing $\mathbf{u}$ onto the non-covert message which experiences the highest-channel gain among all non-covert users is also optimal to maximize the total detection error probability. Hence, we further change the denotation of the total detection error probability in (4.24) to $\xi_{max}(m)$.

4.10.1 Optimum Antenna Selection for Alice

Alice's goal is to select the antenna properly to maximize the total detection error probability,

$$\xi_{max} = \max_{1 \leq m \leq M} \xi_{max}(m) \tag{4.74}$$

$$= \max_{1 \leq m \leq M} \Pr\left(|f_{m,1}|^2 > \frac{|g_m|^2}{(\alpha + |g_m|^2(\alpha \sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1))^+} \right). \tag{4.75}$$

Assuming $|f_{1,1}|^2 = \max_{1 \leq m \leq M, 1 \leq l \leq L} |f_{m,l}|^2$, since $\xi_{max}(m)$ increases with increasing $|f_{m,1}|^2$, then $\xi_{max}(m)$ is maximized when $m = 1$, i.e.

$$\xi_{max} = \Pr\left(|f_{1,1}|^2 > \frac{|g_m|^2}{(\alpha + |g_m|^2(\alpha \sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1))^+} \right). \tag{4.76}$$

For the case of channel adaptation, we also obtain that $\xi_{max}(m)$ is maximized when $m = 1$.

Remark 6: The optimum antenna selection strategy that maximizes the total detection error probability is to select the antenna that provides the highest channel gain among all users and all transmitter's antennas[2]. Note that the antenna chosen to maximize the total detection error probability is to maximize the rate of $\mathbf{v}_1$ where $\mathbf{u}$ is superimposed. Hence, the order statistic of $|f_{1,1}|^2$ is known while the order statistic of $|f_{1,2}|^2, ..., |f_{1,L}|^2$ is unknown. Therefore, we will recompute the maximum total detection error probability and decoding outage probability in the next subsections.

4.10.2 Maximum total detection error probability

For the case with channel adaptation, it follows from (4.25) with $|g|^2$ replaced by $|g_m|^2$ and $|f_k|^2$ replaced by $|f_{1,k}|^2$ for $k \in [0, L]$ that

$$\xi_{max} = 1 - \sum_{k=0}^{L} \Pr\left(|f_{1,1}|^2 \leq \frac{|g_m|^2}{(\alpha + |g_m|^2(\alpha \sum_{l \in \mathcal{D}_1^c} \gamma_l - \gamma_1))^+}, |f_{1,k+1}|^2 \leq |g_m|^2 < |f_{1,k}|^2 \right) \quad (4.77)$$

$$= 1 - \frac{\sigma_f^2}{\sigma_g^2} B\left(LM + 1, \frac{\sigma_f^2}{\sigma_g^2}\right) - \sum_{k=1}^{L} \binom{L-1}{k-1} \sum_{l=0}^{L(M-1)} \frac{LM}{k+l} \binom{L(M-1)}{l}$$

$$\times \int_0^{\alpha_k} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{LM-k-l} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{\nu_k(x)}{\sigma_f^2}}\right)^{k+l} \frac{e^{-\frac{x}{\sigma_g^2}}}{\sigma_g^2} dx. \quad (4.78)$$

The proof of (4.78) is provided in Appendix C.6.

For the case with channel adaptation, the maximum total detection error probability can be obtained from (4.56), (4.59), (4.62) and (4.63) with L replaced by LM.

[2]Note that choosing the antenna providing the highest channel gain among all users and transmitter's antennas is considered in [69] to achieve a nearly-optimal sum rate in NOMA system. When high SNR, i.e. $\gamma_l \to \infty$, $l = 1, ..., L$, it follows from (4.3) with $|f_j|^2$ replaced by $|f_{m,j}|^2$ that the sum rate can be approximated by

$$\sum_{j=1}^{L} R_{j,0} \simeq \log_2(1 + |f_{m,1}|^2 \gamma_1) + \sum_{j=2}^{L} \log_2\left(1 + \frac{\gamma_j}{\sum_{l<j} \gamma_l}\right),$$

which is maximized when $|f_{m,1}|^2$ is maximized. Therefore, choosing the antenna providing the highest channel gain among all users and transmitter's antennas is optimal to maximize the sum rate at high SNR.

4.10.3 Decoding Outage Probability

For the case without channel adaptation, we obtain from Appendix C.7 that

$$P_{o,u}=1-\frac{\sigma_f^2}{\sigma_h^2}B\left(e^{\frac{-\mu_0}{\sigma_f^2}};\frac{\sigma_f^2}{\sigma_h^2},LM+1\right)$$
$$-\frac{\sigma_f^2}{\sigma_h^2}\sum_{k=1}^{L}\binom{L-1}{k-1}\sum_{l=0}^{L(M-1)}\frac{LM}{k+l}\binom{L(M-1)}{l}B\left(e^{-\frac{\mu_k}{\sigma_f^2}};k+l+\frac{\sigma_f^2}{\sigma_h^2},LM-k-l+1\right). \quad (4.79)$$

For the case with channel adaptation, we obtain from Appendix C.8 that

$$P_{o,u}=1-\frac{\int_{\mu_0}^{\infty}\left(1-e^{-\min\left\{\frac{\tau}{\sigma_f^2},\frac{x\alpha}{\sigma_f^2(1+x(1-\alpha)\gamma_1)}\right\}}\right)^{LM}\frac{e^{-\frac{x}{\sigma_h^2}}}{\sigma_h^2}dx}{(1-e^{-\tau/\sigma_h^2})^{LM}}$$
$$-\frac{\int_{\mu_1}^{\infty}\left(1-e^{-\min\left\{\frac{\tau}{\sigma_f^2},\frac{x}{\sigma_f^2}\right\}}\right)^{LM}\frac{e^{-\frac{x}{\sigma_h^2}}}{\sigma_h^2}dx-\int_{\mu_1}^{\infty}\left(1-e^{-\min\left\{\frac{\tau}{\sigma_f^2},\frac{x\alpha}{\sigma_f^2(1+x(1-\alpha)\gamma_1)}\right\}}\right)^{LM}\frac{e^{-\frac{x}{\sigma_h^2}}}{\sigma_h^2}dx}{(1-e^{-\tau/\sigma_h^2})^{LM}}$$
$$-\sum_{k=1}^{L}\binom{L-1}{k-1}\sum_{l=0}^{L(M-1)}\frac{ML}{k+l}\binom{L(M-1)}{l}$$
$$\times\frac{\int_{\mu_k}^{\tau}\left(1-e^{-\frac{x}{\sigma_f^2}}\right)^{LM-k-l}\left(e^{-\frac{x}{\sigma_f^2}}-e^{-\frac{\tau}{\sigma_f^2}}\right)^{k+l}\frac{e^{-\frac{x}{\sigma_h^2}}}{\sigma_h^2}dx}{(1-e^{-\tau/\sigma_h^2})^{LM}}. \quad (4.80)$$

The decoding outage probability of $\mathbf{v}_1$ can be obtained from (4.69) with L replaced by LM.

4.11 Numerical Results

In this section, we present the numerical results. we name the transmission scheme in (4.54) as 'channel adaptation' and that without channel adaptation as 'no channel adaptation'.

4.11.1 Maximum total detection error probability

Fig. 4.3 compares the maximum total detection error probability, ξ_{max}, with channel adaptation and no channel adaptation versus L for different values of M. For the channel adaptation case, the parameter τ is set such that $P_{o,v}=0.01$. One can see that ξ_{max} increases and converges to 1 as L

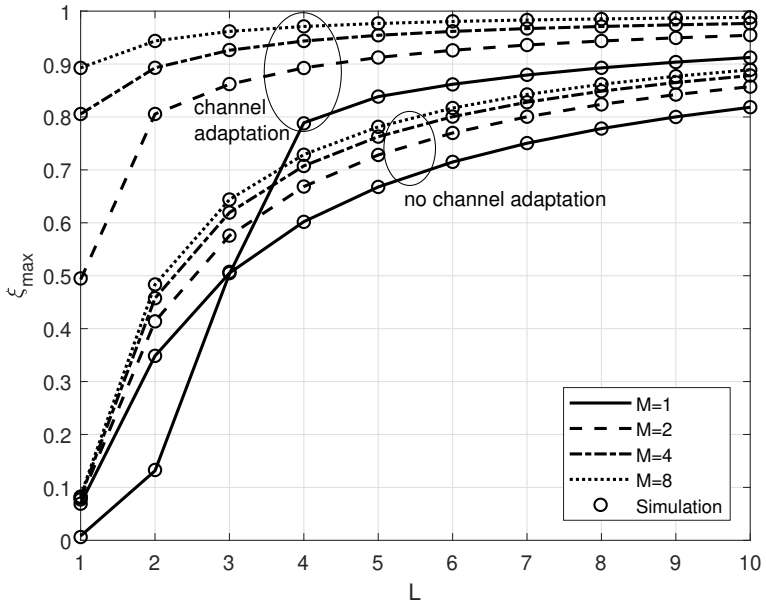

Figure 4.3 The maximum total detection error probability, ξ_{max}, versus L for different values of M; $P/\sigma_n^2 = 20$dB, $\alpha = 0.9$, $P_l = P/L$ for $1 \leq l \leq L$, $\sigma_f^2 = \sigma_g^2 = 1$ and $P_{o,v} = 0.01$.

increases and that the convergence speed is faster for larger M. The simulation results match well with the analytical results.

Fig. 4.4 compares the maximum total detection error probability, ξ_{max}, with channel adaptation and no channel adaptation versus the transmit SNR, P/σ_n^2, for different values of $P_{o,v}$. One can see that ξ_{max} with channel adaptation increases with increasing SNR while that with no channel adaptation decreases with increasing SNR. The reason for the increase of ξ_{max} with increasing SNR is the difference between the transmission rate $R_{1,0}$ of $\mathbf{v}_1$ in (4.3) and the achievable rate $I(\mathbf{v}_1; \mathbf{y}_{w,1})$ of Willie in (4.17) increases with increasing SNR. This suggests switching the transmission mode (channel adaptation and no channel adaptation) depending on the transmit power to maximize ξ_{max}. One can also see that ξ_{max} with channel adaptation increases as $P_{o,v}$ increases. The decrease of ξ_{max} to maintain a low $P_{o,v}$ is small at high SNR.

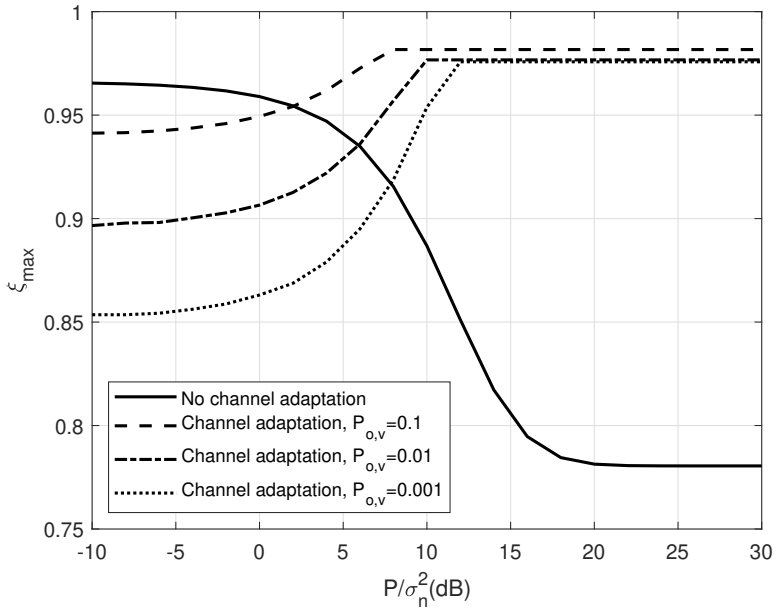

Figure 4.4 The maximum total detection error probability, ξ_{max}, versus the transmit SNR, P/σ_n^2, for different values of $P_{o,v}$; $\alpha = 0.9$, $M = 8$, $L = 5$, $P_l = P/L$ for $1 \leq l \leq L$, and $\sigma_f^2 = \sigma_g^2 = 1$.

4.11.2 Covert rate

Figure 4.5 compares the decoding outage probability, $P_{o,u}$, with channel adaptation ($\tau = \alpha/((1-\alpha)\gamma_1)$) and no channel adaptation ($\tau = \infty$). One can see that the decoding outage probability is higher for larger L, which is because the covert user experiences more interference of the non-covert messages with increasing L. Although an increase of the interference can help increasing the total detection error probability, it will increase the decoding outage probability at the covert user. One can also see that the channel adaptation yields a higher decoding outage probability than no channel adaptation. The simulation result matches well with the analytical result.

Figure 4.6 compares the covert rate (bits/s/Hz), η_u, with channel adaptation and no channel adaptation versus the maximum total detection error probability, ξ_{max}, as the SNR is varied. The decoding outage probability, $P_{o,v}$, of the non-covert message is fixed at 0.3 and 0.4 for the channel adaptation case. One can see that the covert rate can be traded with ξ_{max} for the case of no channel adaptation. The channel adaptation can provide higher covert rate than no channel adaptation

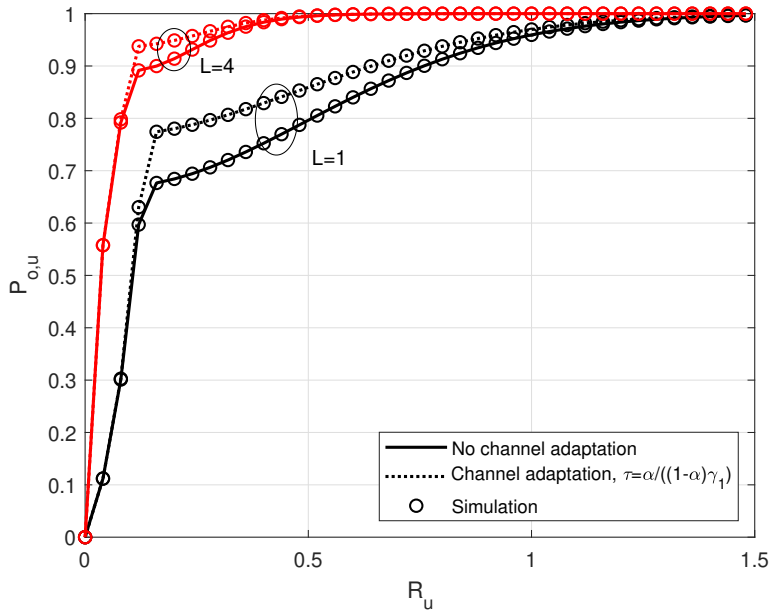

Figure 4.5 Decoding outage probability, $P_{o,u}$, versus R_u for different values of L; $M = 2$, $P/\sigma_n^2 = 5$dB, $\alpha = 0.9$, $P_l = P/L$ for $1 \leq l \leq L$, and $\sigma_f^2 = \sigma_h^2 = 1$.

for high ξ_{max} (close to 1), which is the range of practical interest. However, for low ξ_{max}, the latter provides higher covert rate than the former. This suggests switching the transmission mode (no channel adaptation vs channel adaptation) depending on the covertness requirement, namely ξ_{max}. One can also see that the covert rate for the case of channel adaptation also increases with increasing ξ_{max}, which because the transmission power increases.

Figure 4.7 compares the non-covert rate (bits/s/Hz) with channel adaptation and no channel adaptation versus the maximum total detection error probability, ξ_{max}, as the SNR is varied. One can see that the channel adaptation can provide higher non-covert rate than no channel adaptation for high ξ_{max} (close to 1). One can also see from Figs. 4.6 and 4.7 that there is the tradeoff between the covert and non-covert throughput for high ξ_{max} (close to 1).

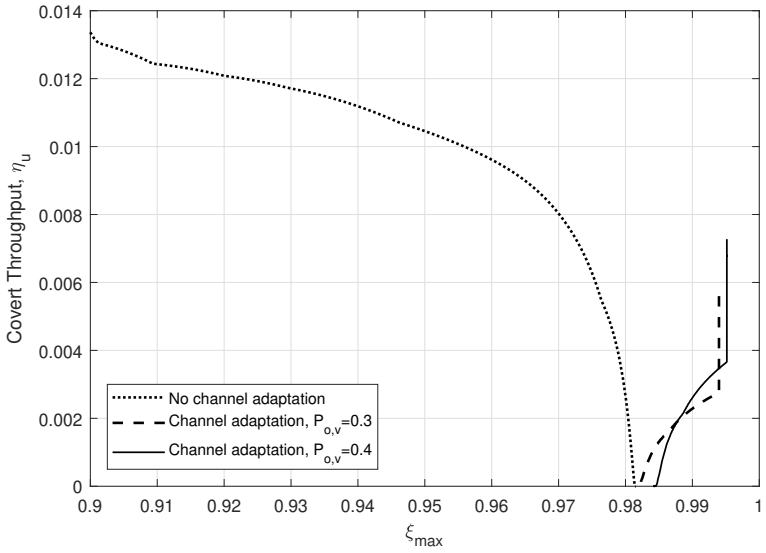

Figure 4.6 covert rate (bits/s/Hz), η_u, versus maximum total detection error probability ξ_{max}; $M = 8$, $\alpha = 0.8$, $L = 10$, $P_l = P/L$ for $1 \leq l \leq L$, and $\sigma_f^2 = \sigma_h^2 = \sigma_g^2 = 1$.

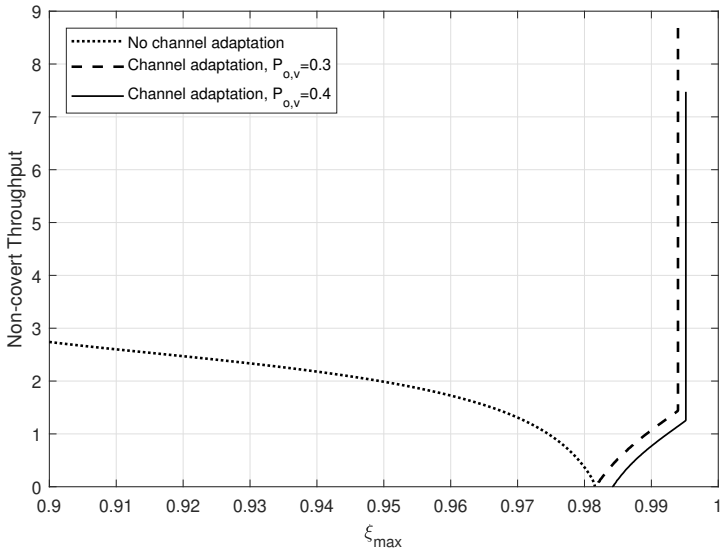

Figure 4.7 Non-covert rate (bits/s/Hz) versus maximum total detection error probability, ξ_{max}; $M = 8$, $\alpha = 0.8$, $L = 10$, $P_l = P/L$ for $1 \leq l \leq L$, and $\sigma_f^2 = \sigma_h^2 = \sigma_g^2 = 1$.

4.12 Chapter Summary

We considered hiding a covert message under other messages in NOMA system. We determined the adversary's optimum detection strategy that minimizes the total detection error probability and the communicator's optimum superposition strategy that maximizes the minimum total detection error probability. We found that the total detection error probability increases and converges to 1 as the number of users (messages) increases. We also found that that the total detection error probability and the covert rate can be increased as the transmission power increases by adapting the superposition rule to the channel variation.

CHAPTER 5. CONCLUSIONS AND FUTURE WORK

5.1 Conclusions and Contributions

This thesis focuses on contributing to the improvement of the physical-layer secrecy and privacy of wireless communication. A brief summary of the main contributions is presented.

In Chapter 2, we investigated *joint* rate and power adaptation to maximize the energy efficiency of physical-layer secrecy. We formulated an optimization problem of maximizing the secrecy energy efficiency (SEE) subject to an average transmission power constraint. We determined the optimum rate and power adaptation rule that maximizes the SEE, and also determined sub-optimal rate and power adaptation rules: on-off variable rate, fixed power variable rate and variable power fixed rate. We characterized the SEE gain by varying rate and/or the power, and the impact of number of antennas on the optimum adaptation rule.

In Chapter 3, we analyzed the *joint* impact of imperfect knowledge of noise power and channel gain at the adversary on the total detection error probability and the covert throughput. We determined the optimum detection threshold for the energy detector that minimizes the total detection error probability as a function of the channel gain estimate. Then, we determined the maximum allowed transmission power for the total detection error probability to be no less than a threshold. Based on this, we determined the maximum average transmission rate (bits/s/Hz) subject to a covert communication constraint, hereafter referred to as the covert throughput. We characterized the covert throughput gain provided by imperfect knowledge of the channel gain and noise power at the adversary and the covert throughput loss caused by the channel fading as a function of the noise uncertainty. Our analysis showed that the channel fading is crucial to hiding the signal transmission, particularly when the noise uncertainty is low and/or the receive SNR is high. The impact of the channel uncertainty on the total detection error probability and the covert throughput is particularly noticeable when the noise uncertainty is large. The channel uncertainty provides a

covert throughput gain of 12% $\sim$ 19% over the case that perfect channel knowledge is available at the adversary when the noise uncertainty is in the range of 1 $\sim$ 2 dB. However, if the noise uncertainty is small, the channel uncertainty does not help much increase the total detection error probability and the covert throughput.

In Chap 4, we studied covertness in NOMA system; that is, hiding the covert message under superposition of other messages in NOMA system. We determined the adversary's optimum detection strategy and the transmitter's optimum hiding message strategy. Then, We analyzed the resulting maximum total detection error probability and decoding outage probability with covert rate. We exploited the channel variation to further increase the total detection error probability. We found that the covert message can be detected only when the non-covert message, where the covert message is superimposed onto, can be decoded. This suggests hiding the covert message under the non-covert message that is most difficult to decode. We also found that the covert message is harder to be decoded if the non-covert rate is transmitted at higher rate. This suggests selecting the best antenna which can provide the highest channel gain among all user and transmitter's antenna. Our results showed that the total detection error probability increases and converges to 1 as the number of users increases, and convergence speed is faster with larger number of transmitter's antenna. This means that the covert transmission is undetectable if the number of non-covert users is sufficiently large. Our results also we also showed that the total detection error probability and the covert rate can be increased by increasing the transmission power when the channel variation is exploited. This indicates that the multiplicity of users which is scattered in wireless network and their mobility (channel variation) can be leveraged to hiding the covert message.

5.2 Future Work

For the current work of covert NOMA, one can see from (4.23) the total detection error probability can be increased if the rate of the non-covert (cover) message is increased or the achievable rate of the non-covert message at the adversary is decreased. Then, we attempt to provide the SNR advantage of the non-covert user over the adversary. In this part of thesis, we, therefore,

present two directions of future work, the Artificial Noise (AN) transmission and the cooperative transmission in NOMA system.

5.2.1 AN-aided covert NOMA

Motivation: In secrecy, the AN scheme has showed its potential to secure information [67, 76]. The AN is transmitted onto the null space of the channel state information of the intended receiver. As such, the AN signal can make more noises at the adversary while does not affect the intended receiver. As a result, the AN can provide the SNR advantage of the intended receiver over the adversary. Although many studies have investigated the AN into NOMA system [46, 72], no prior work considers covertness with aid of Artificial Noise in NOMA system. Therefore, we will exploit the AN scheme in the covert NOMA system.

Observation: Exploiting AN can help degrading the ability of decoding the non-covert message at Willie while it does not affect the intended receiver of the non-covert message, thus increasing the total detection error probability. Besides, the AN can also degrade the ability of decoding the covert message at the covert user. Hence, there should exist an optimal power allocation of AN power to maximize the covert rate. It also follows from Section 4.3 that Willie can detect the covert message with probability 1 if Willie can decode the non-covert message where the covert message is superimposed, i.e. the second and fourth terms in (4.13) are zero. In this extension, we will propose a novel design of null-space such that Willie does not know the AN power and hence cannot detect the covert message with probability 1. i.e. the second and fourth terms in (4.13) are non-zero. This design does not affect the ability of decoding the covert message at the covert user. Therefore, the AN can help increasing the total detection error probability, thereby the covert rate.

Contribution: In this extension, we will study the AN-aided covert NOMA. We will propose a new design of AN and determine the optimum power allocation for AN to maximize the covert rate. We will also characterize the gain of covert rate by having the AN.

5.2.2 Cooperative covert NOMA

Motivation: Recently, the author in [22, 44] has considered the cooperative transmission technique in NOMA system. They showed the significant increase of the sum rate by adopting the cooperative transmission. In secrecy, two types of cooperative transmission, Decode-and-Forward (DF) and Amplify-and-Forward (AF), have been also considered in secure NOMA system [16, 1]. However, no prior work considers cooperative transmission in covert NOMA, which we name it cooperative covert NOMA.

Observation: Similar to the covert NOMA, the covert message is superimposed onto the non-covert message. As the result, the adversary needs to decode the non-covert message in order to detect the presence of covert signal. Since the cooperative NOMA can provide the SNR advantage to the receiver of non-covert message over the adversary, it can degrade the ability of decoding the non-covert message, hence increase the total detection error probability, at the adversary.

Contribution: In this extension, we will study the DF and AF transmission in covert NOMA. We will describe the optimum detection strategy at the adversary and the optimum message hiding strategy and optimum relay selection at the transmitter. We will determine the resulting maximum total detection error probability. We will compare the maximum total detection error probability and covert rate between DF and AF transmission in cooperative covert NOMA.

BIBLIOGRAPHY

[1] M. Abolpour, M. Mirmohseni, and M. R. Aref. Outage performance in secure cooperative NOMA. In *2019 Iran Workshop on Communication and Information Theory (IWCIT)*, pages 1–6. IEEE, 2019.

[2] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1965.

[3] A. A. Alkheir and H. T. Mouftah. An improved energy detector using outdated channel state information. *IEEE Communications Letters*, 19(7):1237–1240, 2015.

[4] G. E. Andrews, R. Askey, and R. Roy. *Special functions*, volume 71. Cambridge university press, 1999.

[5] K. S. K. Arumugam and M. R. Bloch. Embedding covert information in broadcast communications. *arXiv preprint arXiv:1808.09556*, 2018.

[6] S. Atapattu, C. Tellambura, H. Jiang, and N. Rajatheva. Unified analysis of low-SNR energy detection and threshold selection. *IEEE Transactions on Vehicular Technology*, 64(11):5006–5019, 2015.

[7] E. Axell and E. G. Larsson. Optimal and sub-optimal spectrum sensing of OFDM signals in known and unknown noise variance. *IEEE Journal on Selected Areas in Communications*, 29(2):290–304, 2011.

[8] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor. Spectrum sensing for cognitive radio: State-of-the-art and recent advances. *IEEE signal processing magazine (Print)*, 29(3):101–116, 2012.

[9] B. A. Bash, D. Goeckel, and D. Towsley. Limits of reliable communication with low probability of detection on AWGN channels. *IEEE Journal on Selected Areas in Communications*, 31(9):1921–1930, 2013.

[10] M. Bloch and J. Barros. *Physical-layer security: from information theory to security engineering*. Cambridge University Press, 2011.

[11] M. Bloch, J. Barros, M. R. Rodrigues, and S. W. McLaughlin. Wireless information-theoretic security. *IEEE Transactions on Information Theory*, 54(6):2515–2534, 2008.

[12] M. R. Bloch. Covert communication over noisy channels: A resolvability perspective. *IEEE Transactions on Information Theory*, 62(5):2334–2354, 2016.

[13] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[14] A. Cambini and L. Martein. *Generalized Convexity and Optimization: Theory and Applications*. Springer-Verlag Berlin Heidelberg, New York, NY, USA, 2009.

[15] P. H. Che, M. Bakshi, and S. Jaggi. Reliable deniable communication: Hiding messages in noise. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 2945–2949. IEEE, 2013.

[16] J. Chen, L. Yang, and M.-S. Alouini. Physical layer security for cooperative NOMA systems. *IEEE Transactions on Vehicular Technology*, 67(5):4645–4649, 2018.

[17] X. Chen and L. Lei. Energy-efficient optimization for physical layer security in multi-antenna downlink networks with QoS guarantee. *IEEE Communications Letters*, 17(4):637–640, April 2013.

[18] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth. On the LambertW function. *Advances in Computational mathematics*, 5(1):329–359, 1996.

[19] I. Csiszár and J. Korner. Broadcast channels with confidential messages. *IEEE transactions on information theory*, 24(3):339–348, 1978.

[20] H. A. David and H. N. Nagaraja. Order statistics. *Encyclopedia of Statistical Sciences*, 9, 2004.

[21] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava. A Survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends. *IEEE Journal on Selected Areas in Communications*, 35(10):2181–2195, Oct 2017.

[22] Z. Ding, M. Peng, and H. V. Poor. Cooperative non-orthogonal multiple access in 5G systems. *IEEE Communications Letters*, 19(8):1462–1465, Aug 2015.

[23] A. Dutta, D. Saha, D. Grunwald, and D. Sicker. Secret agent radio: Covert communication through dirty constellations. In *International Workshop on Information Hiding*, pages 160–175. Springer, 2012.

[24] S. Gerbracht, C. Scheunert, and E. A. Jorswieck. Secrecy outage in MISO systems with partial channel information. *IEEE Transactions on Information Forensics and Security*, 7(2):704–716, 2012.

[25] S. Goel and R. Negi. Guaranteeing secrecy using artificial noise. *IEEE Transactions on Wireless Communications*, 7(6):2180–2189, 2008.

[26] A. Goldsmith. *Wireless communications*. Cambridge university press, 2005.

[27] P. K. Gopala, L. Lai, and H. El Gamal. On the secrecy capacity of fading channels. *IEEE Transactions on Information Theory*, 54(10):4687–4698, 2008.

[28] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Elsevier/Academic Press, Amsterdam, seventh edition, 2007. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX).

[29] B. He, A. Liu, N. Yang, and V. K. N. Lau. On the design of secure non-orthogonal multiple access systems. *IEEE Journal on Selected Areas in Communications*, 35(10):2196–2206, Oct 2017.

[30] B. He, Y. She, and V. K. N. Lau. Artificial noise injection for securing single-antenna systems. *IEEE Transactions on Vehicular Technology*, 66(10):9577–9581, Oct 2017.

[31] B. He, S. Yan, X. Zhou, and V. K. N. Lau. On covert communication with noise uncertainty. *IEEE Communications Letters*, 21(4):941–944, April 2017.

[32] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J. Wang. Covert communication achieved by a greedy relay in wireless networks. *CoRR*, abs/1708.00905, 2017.

[33] A. Kalantari, S. Maleki, S. Chatzinotas, and B. Ottersten. Secrecy energy efficiency optimization for MISO and SISO communication networks. In *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 21–25, June 2015.

[34] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[35] A. Khisti, G. Wornell, A. Wiesel, and Y. Eldar. On the Gaussian MIMO wiretap channel. In *2007 IEEE International Symposium on Information Theory*, pages 2471–2475. IEEE, 2007.

[36] S. Lee, R. J. Baxley, J. B. McMahon, and R. S. Frazier. Achieving positive rate with undetectable communication over MIMO rayleigh channels. In *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014 IEEE 8th*, pages 257–260. IEEE, 2014.

[37] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst. Achieving undetectable communication. *IEEE Journal of Selected Topics in Signal Processing*, 9(7):1195–1205, Oct 2015.

[38] S. Leung-Yan-Cheong and M. Hellman. The Gaussian wire-tap channel. *IEEE Transactions on Information Theory*, 24(4):451–456, July 1978.

[39] Y. Li, M. Jiang, Q. Zhang, Q. Li, and J. Qin. Secure Beamforming in Downlink MISO Nonorthogonal Multiple Access Systems. *IEEE Transactions on Vehicular Technology*, 66(8):7563–7567, Aug 2017.

[40] Z. Li, R. Yates, and W. Trappe. Secrecy capacity of independent parallel channels. In *Securing Wireless Communications at the Physical Layer*, pages 1–18. Springer, 2009.

[41] Y. Liang, H. V. Poor, and S. Shamai. Secrecy capacity region of fading broadcast channels. In *2007 IEEE International Symposium on Information Theory*, pages 1291–1295. IEEE, 2007.

[42] Y. Liang, H. V. Poor, and S. Shamai (Shitz). Information theoretic security. *Found. Trends Commun. Inf. Theory*, 5(4–5):355–580, Apr. 2009.

[43] T. Liu, P. Lin, S. Lin, Y. . P. Hong, and E. A. Jorswieck. To avoid or not to avoid CSI leakage in physical layer secret communication systems. *IEEE Communications Magazine*, 53(12):19–25, Dec 2015.

[44] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor. Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer. *IEEE Journal on Selected Areas in Communications*, 34(4):938–953, April 2016.

[45] L. Lv, Z. Ding, Q. Ni, and J. Chen. Secure MISO-NOMA Transmission With Artificial Noise. *IEEE Transactions on Vehicular Technology*, 67(7):6700–6705, July 2018.

[46] L. Lv, Z. Ding, Q. Ni, and J. Chen. Secure MISO-NOMA transmission with artificial noise. *IEEE Transactions on Vehicular Technology*, 67(7):6700–6705, 2018.

[47] J. L. Massey. An introduction to contemporary cryptology. *Proceedings of the IEEE*, 76(5):533–549, May 1988.

[48] P. Mu, P. Yang, B. Wang, H.-M. Wang, and Q. Yin. A new scheme to improve the secrecy throughput under the constraints of secrecy outage probability and average transmit power. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pages 777–782, June 2014.

[49] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst. Principles of physical layer security in multiuser wireless networks: A survey. *IEEE Communications Surveys Tutorials*, 16(3):1550–1573, Third 2014.

[50] D. W. K. Ng, E. S. Lo, and R. Schober. Energy-efficient resource allocation for secure OFDMA systems. *IEEE Transactions on Vehicular Technology*, 61(6):2572–2585, July 2012.

[51] J. Ouyang, M. Lin, W. Zhu, D. Massicotte, and A. L. Swindlehurst. Energy efficient beamforming for secure communication in cognitive radio networks. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3496–3500, March 2016.

[52] C. Paar and J. Pelzl. *Understanding cryptography: a textbook for students and practitioners*. Springer Science & Business Media, 2009.

[53] H. V. Poor. Information and inference in the wireless physical layer. *IEEE Wireless Communications*, 19(1):40–47, 2012.

[54] Proakis. *Digital Communications 5th Edition*. McGraw Hill, 2007.

[55] A. Sahai, N. Hoven, and R. Tandra. Some fundamental limits on cognitive radio. In *Forty-second Allerton Conference on Communication, Control, and Computing*, 2004.

[56] A. Sahai, N. Hoven, and R. Tandra. Some fundamental limits on cognitive radio. In *Allerton Conference on Communication, Control, and Computing*, pages 1662–1671. Monticello, Illinois, 2004.

[57] K. Shahzad, X. Zhou, and S. Yan. Covert communication in fading channels under channel uncertainty. In *Vehicular Technology Conference (VTC Spring), 2017 IEEE 85th*, pages 1–5. IEEE, 2017.

[58] S. Shellhammer. Performance of the power detector with noise uncertainty. *Doc. IEEE 802.22-06/0134r0*, 2006.

[59] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel. Covert communication in the presence of an uninformed jammer. *IEEE Transactions on Wireless Communications*, 16(9):6193–6206, Sept 2017.

[60] Q. Sun, S. Han, C. I, and Z. Pan. On the Ergodic Capacity of MIMO-NOMA Systems. *IEEE Wireless Communications Letters*, 4(4):405–408, Aug 2015.

[61] R. Tandra and A. Sahai. SNR walls for signal detection. *IEEE Journal of selected topics in Signal Processing*, 2(1):4–17, 2008.

[62] E. Tekin and A. Yener. The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming. *IEEE Transactions on Information Theory*, 54(6):2735–2751, June 2008.

[63] A. Y. Wang and C. G. Sodini. On the energy efficiency of wireless transceivers. In *2006 IEEE International Conference on Communications*, volume 8, pages 3783–3788, June 2006.

[64] L. Wang, G. W. Wornell, and L. Zheng. Fundamental limits of communication with low probability of detection. *IEEE Transactions on Information Theory*, 62(6):3493–3503, 2016.

[65] A. D. Wyner. The wire-tap channel. *The bell system technical journal*, 54(8):1355–1387, 1975.

[66] N. Yang, M. Elkashlan, T. Q. Duong, J. Yuan, and R. Malaney. Optimal transmission with artificial noise in MISOME wiretap channels. *IEEE Transactions on Vehicular Technology*, 65(4):2170–2181, 2016.

[67] N. Yang, S. Yan, J. Yuan, R. Malaney, R. Subramanian, and I. Land. Artificial noise: Transmission optimization in multi-input single-output wiretap channels. *IEEE Transactions on Communications*, 63(5):1771–1783, 2015.

[68] Y. Yu. The shape of the noncentral Chi-square density. *ArXiv e-prints*, June 2011.

[69] Y. Yu, H. Chen, Y. Li, Z. Ding, L. Song, and B. Vucetic. Antenna selection for MIMO nonorthogonal multiple access systems. *IEEE Transactions on Vehicular Technology*, 67(4):3158–3171, April 2018.

[70] A. Zappone, P. H. Lin, and E. Jorswieck. Energy efficiency of confidential multi-antenna systems with artificial noise and statistical CSI. *IEEE Journal of Selected Topics in Signal Processing*, 10(8):1462–1477, Dec 2016.

[71] A. Zappone, P. H. Lin, and E. A. Jorswieck. Energy-efficient secure communications in MISO-SE systems. In *2014 48th Asilomar Conference on Signals, Systems and Computers*, pages 1001–1005, Nov 2014.

[72] M. Zeng, N.-P. Nguyen, O. A. Dobre, and H. V. Poor. Securing downlink massive MIMO-NOMA networks with artificial noise. *IEEE Journal of Selected Topics in Signal Processing*, 13(3):685–699, 2019.

[73] X. Zhang, X. Zhou, and M. R. McKay. On the design of artificial-noise-aided secure multi-antenna transmission in slow fading channels. *IEEE Transactions on Vehicular Technology*, 62(5):2170–2181, 2013.

[74] Y. Zhang, H. Wang, Q. Yang, and Z. Ding. Secrecy sum rate maximization in non-orthogonal multiple access. *IEEE Communications Letters*, 20(5):930–933, May 2016.

[75] X. Zhou and M. R. McKay. Secure transmission with artificial noise over fading channels: Achievable rate and optimal power allocation. *IEEE Transactions on Vehicular Technology*, 59(8):3831–3842, 2010.

[76] X. Zhou and M. R. McKay. Secure transmission with artificial noise over fading channels: Achievable rate and optimal power allocation. *IEEE Transactions on Vehicular Technology*, 59(8):3831–3842, Oct 2010.

[77] X. Zhou, M. R. McKay, B. Maham, and A. Hjorungnes. Rethinking the secrecy outage formulation: A secure transmission design perspective. *IEEE Communications Letters*, 15(3):302–304, March 2011.

APPENDIX A. ADAPTING RATE AND POWER FOR MAXIMIZING SECRECY ENERGY EFFICIENCY

Let

$$g_1(P(\gamma)) = -P(\gamma), \tag{A.1}$$

$$g_2(P(\gamma)) = P(\gamma) - P^*(\gamma), \tag{A.2}$$

$$\text{and} \quad g_3(P(\gamma)) = \int_0^\infty P(\gamma) f(\gamma) d\gamma - 1. \tag{A.3}$$

Then, the optimization problem in (2.17) under the constraint of (2.18) and $0 \leq P(\gamma) \leq P^*(\gamma)$ is given by

$$\min_{P(\gamma)} \quad -\int_\alpha^\infty \zeta(P(\gamma), \gamma) f(\gamma) d\gamma \tag{A.4}$$

$$\text{subject to} \quad g_1(P(\gamma)) \leq 0 \tag{A.5}$$

$$g_2(P(\gamma)) \leq 0 \tag{A.6}$$

$$g_3(P(\gamma)) \leq 0. \tag{A.7}$$

The Kuhn-Tucker condition for the solution of the optimization problem is given by [13]

$$\lambda \geq 0 \tag{A.8}$$

$$v_i \geq 0 \tag{A.9}$$

$$\lambda g_3(P(\gamma)) = 0 \tag{A.10}$$

$$v_i g_i(P(\gamma)) = 0 \tag{A.11}$$

$$\frac{\partial \mathcal{L}(P(\gamma))}{\partial P(\gamma)} = 0, \tag{A.12}$$

for i=1,2, where

$$\mathcal{L}(P(\gamma)) = -\int_\alpha^\infty \zeta(P(\gamma), \gamma) f(\gamma) d\gamma + \lambda g_3(P(\gamma)) + v_1 g_1(P(\gamma)) + v_2 g_2(P(\gamma)), \tag{A.13}$$

and λ, v_1, and v_2 are Lagrange multipliers. It follows from (A.1) - (A.3) and (A.13) that (A.12) is equivalent to

$$\left(\frac{\partial\zeta(P(\gamma),\gamma)}{\partial P(\gamma)} - \lambda\right) f(\gamma) \quad = \quad v_2 - v_1. \tag{A.14}$$

Since $\partial^2\zeta(P(\gamma),\gamma)/\partial P(\gamma)^2 < 0$, $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma)$ is a strictly decreasing function of $P(\gamma)$. In addition, $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma)$ is 0 for $P(\gamma) = P^*(\gamma)$ and is $B(\gamma-\alpha)/(P_C \ln 2)$ for $P(\gamma) = 0$ from (2.19). As a result, $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma)$ lies in the range $[0, B(\gamma-\alpha)/(P_C \ln 2)]$ for $P(\gamma) \in [0, P^*(\gamma)]$.

(i) If $\lambda > B(\gamma-\alpha)/(P_C \ln 2)$, then $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma) < \lambda$ for $P(\gamma) \in [0, P^*(\gamma)]$. Hence, it follows from (A.14) that $v_1 > v_2$. Since $v_2 \geq 0$, we obtain $v_1 > 0$. Therefore, it follows from (A.9) that $g_1(P(\gamma)) = 0$. That is, the optimal transmission power, $P_{PR}(\gamma)$, is 0.

(ii) If $0 \leq \lambda \leq B(\gamma-\alpha)/(P_C \ln 2)$, then $v_1 = v_2$. This can be proved by showing that if $v_1 \neq v_2$, then $\lambda \notin [0, B(\gamma-\alpha)/(P_C \ln 2)]$:

- If $v_1 > v_2$, then $v_1 > 0$, hence $P(\gamma) = 0$, and $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma) < \lambda$. The latter follows from (A.14). Since $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma) \leq B(\gamma-\alpha)/(P_C \ln 2)$ for $P(\gamma) \in [0, P^*(\gamma)]$, we obtain $\lambda > B(\gamma-\alpha)/(P_C \ln 2)$.

- If $v_2 > v_1$, then $v_2 > 0$, hence $P(\gamma) = P^*(\gamma)$, and $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma) > \lambda$. The latter follows from (A.14). Since $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma) \geq 0$ for $P(\gamma) \in [0, P^*(\gamma)]$, we obtain $\lambda < 0$. Therefore, if $v_1 \neq v_2$, then $\lambda \notin [0, B(\gamma-\alpha)/(P_C \ln 2)]$.

For $v_1 = v_2$, (A.14) reduces to

$$\frac{\partial\zeta(P(\gamma),\gamma)}{\partial P(\gamma)} - \lambda = 0. \tag{A.15}$$

Since $\partial\zeta(P(\gamma),\gamma)/\partial P(\gamma)$ is a strictly decreasing function of $P(\gamma)$, there should exist a unique solution of (A.15). Let $P^\dagger(\gamma)$ denote the solution of (A.15). Then, the optimal transmission power, $P_{PR}(\gamma)$, is $P^\dagger(\gamma)$ for $0 \leq \lambda \leq B(\gamma-\alpha)/(P_C \ln 2)$.

Therefore, from (i) and (ii), the solution of the problem in (A.4)-(A.7) is given by

$$P_{PR}(\gamma) = \begin{cases} P^\dagger(\gamma), & 0 \leq \lambda \leq B(\gamma-\alpha)/(P_C \ln 2) \\ 0, & \lambda > B(\gamma-\alpha)/(P_C \ln 2). \end{cases} \tag{A.16}$$

This proves (2.20).

APPENDIX B. COVERT COMMUNICATION UNDER CHANNEL UNCERTAINTY AND NOISE UNCERTAINTY

B.1 Equation (3.19) derivation

In this Appendix, we derive (3.19). It follows from (3.18) that we have

$$\xi(\hat{g}) = 1 - \frac{\ln(\lambda)}{2\ln(\rho)} + \frac{1}{2\ln(\rho)} \int_{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}}^{\infty} \ln\left(\frac{1}{\rho}\hat{\sigma}_w^2\right) f_X(x|\hat{g})dx$$

$$+ \frac{1}{2\ln(\rho)} \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \ln(\lambda - xP) f_X(x|\hat{g})dx \tag{B.1}$$

$$= 1 - \frac{\ln(\rho\lambda/\hat{\sigma}_w^2)}{2\ln(\rho)} + \frac{1}{2\ln(\rho)} \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \ln\left(\frac{\lambda - xP}{\hat{\sigma}_w^2/\rho}\right) f_X(x|\hat{g})dx \tag{B.2}$$

$$= 1 - \frac{\ln(\rho\lambda/\hat{\sigma}_w^2)}{2\ln(\rho)} + \frac{1}{2\ln(\rho)} \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \ln\left(\rho\left(\frac{\lambda}{\hat{\sigma}_w^2} - x\gamma\right)\right) f_X(x|\hat{g})dx. \tag{B.3}$$

B.2 Proof of pseudo-convexity

In this Appendix, we prove the strict pseudo-convexity of $\xi(\hat{g})$. We follow the theorem of [14] that $\xi(\hat{g})$ is strictly pseudo-convex if, for any value of $\lambda = \lambda_0$ such that $d\xi(\hat{g})/d\lambda = 0$, we have $d^2\xi(\hat{g})/d\lambda^2 > 0$.

The first and second derivative of $\xi(\hat{g})$ are given by

$$2\ln(\rho)\frac{d\xi(\hat{g})}{d\alpha} = -\frac{1}{\lambda} + \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \frac{f_X(x|\hat{g})dx}{\lambda - xP}, \tag{B.4}$$

and

$$2\ln(\rho)\frac{d^2\xi(\hat{g})}{d\lambda^2} = \frac{1}{\lambda^2} + \frac{\rho}{\hat{\sigma}_w^2} f_X\left(\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)\frac{1}{\gamma}\middle|\hat{g}\right) - \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \frac{P f_X(x|\hat{g})dx}{(\lambda - xP)^2}$$

$$> \frac{\rho}{\hat{\sigma}_w^2} f_X\left(\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)\frac{1}{\gamma}\middle|\hat{g}\right) - \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} f_X(x|\hat{g}) d\left(\frac{1}{\lambda - xP}\right) \tag{B.5}$$

$$= \frac{f_X(0|\hat{g})}{\lambda} - \int_0^{\alpha} \frac{f_X'(x|\hat{g})dx}{\lambda - xP}, \tag{B.6}$$

respectively, where integration by part is applied to derive (B.6).

For any $\lambda = \lambda_0$ such that $d\xi(\hat{g})/d\lambda = 0$, i.e.

$$\frac{1}{\lambda_0} = \int_0^{\left(\frac{\lambda_0}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \frac{f_X(x|\hat{g})dx}{\lambda_0 - xP}, \tag{B.7}$$

it follows from (B.6) that we obtain

$$\left.\frac{d^2\xi(\hat{g})}{d\lambda^2}\right|_{\lambda=\lambda_0} > \int_0^{\left(\frac{\lambda_0}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \frac{f_X(x|\hat{g})dx}{\lambda_0 - xP}\left(f_X(0|\hat{g}) - \frac{f_X'(x|\hat{g})}{f_X(x|\hat{g})}\right). \tag{B.8}$$

From (3.14), we have

$$\frac{f_X'(x|\hat{g})}{f_X(x|\hat{g})} = \frac{d(\ln f_X(x|\hat{g}))}{dx} \tag{B.9}$$

$$= \frac{d\left(-\frac{x+|\hat{g}|^2}{\beta\sigma_g^2} + \ln\left(\frac{1}{\beta\sigma_g^2}\right) + \ln\left(I_0\left(\frac{2|\hat{g}|}{\beta\sigma_g^2}\sqrt{x}\right)\right)\right)}{dx} \tag{B.10}$$

$$= -\frac{1}{\beta\sigma_g^2} + \frac{\left(\sum_{k=0}^{\infty}\frac{1}{(k!)^2}\frac{|\hat{g}|^{2k}x^k}{(\beta\sigma_g^2)^{2k}}\right)'}{I_0\left(\frac{2|\hat{g}|}{\beta\sigma_g^2}\sqrt{x}\right)}, \tag{B.11}$$

where $I_0(z) = \sum_{k=0}^{\infty}(z/2)^{2k}/(k!)^2$ is applied in (B.11). Moreover, since the non-central Chi-square with 2 degree of freedom is a log-concave function [68], i.e. $(\ln(f_X(x|\hat{g})))'' < 0$, $(\ln(f_X(x|\hat{g})))'$ is a decreasing function of x for $x \geq 0$. It follows from (B.11) that we obtain

$$f_X(0|\hat{g}) - \frac{f_X'(x|\hat{g})}{f_X(x|\hat{g})} \geq f_X(0|\hat{g}) - \left.\frac{f_X'(x|\hat{g})}{f_X(x|\hat{g})}\right|_{x=0} \tag{B.12}$$

$$= \frac{1}{\beta\sigma_g^2}e^{-\frac{|\hat{g}|^2}{\beta\sigma_g^2}} - \frac{1}{\beta\sigma_g^2}\left(1 - \frac{|\hat{g}|^2}{\beta\sigma_g^2}\right) \tag{B.13}$$

$$\geq 0, \tag{B.14}$$

where the inequality, $e^{-t} \geq 1 - t$, is applied to derive (B.14). It can be obtained from (B.8) and (B.14) that $d^2\xi/d\lambda^2|_{\lambda=\lambda_0} > 0$. Therefore, $\xi(\hat{g})$ is the strictly pseudo-convex function of λ.

B.3 Equation (3.38) derivation

In this Appendix, we derive (3.38).

$$\xi(\hat{g}=0) = 1 - \frac{\ln(\rho\lambda/\hat{\sigma}_w^2)}{2\ln(\rho)} + \frac{1}{2\ln(\rho)} \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \ln\left(\rho\left(\frac{\lambda}{\hat{\sigma}_w^2} - x\gamma\right)\right) \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \tag{B.15}$$

$$= 1 - \frac{\ln(\rho\lambda/\hat{\sigma}_w^2)}{2\ln(\rho)} - \frac{1}{2\ln(\rho)} \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \ln\left(\rho\left(\frac{\lambda}{\hat{\sigma}_w^2} - x\gamma\right)\right) d\left(e^{-x/\sigma_g^2}\right) \tag{B.16}$$

$$= 1 - \frac{\gamma\hat{\sigma}_w^2}{2\ln(\rho)} \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \frac{e^{-x/\sigma_g^2}}{\lambda - x\gamma\hat{\sigma}_w^2} dx \tag{B.17}$$

$$= 1 - \frac{1}{2\ln(\rho)} \int_0^{\left(\frac{\lambda}{\hat{\sigma}_w^2} - \frac{1}{\rho}\right)^{\frac{1}{\gamma}}} \frac{e^{-x/\sigma_g^2}}{\frac{\lambda}{\gamma\hat{\sigma}_w^2} - x} dx \tag{B.18}$$

$$= 1 - \frac{e^{-\frac{\lambda}{\sigma_g^2 P}}}{2\ln(\rho)} \int_{\frac{1}{\rho\sigma_g^2\gamma}}^{\frac{\lambda}{\sigma_g^2 P}} \frac{e^x}{x} dx. \tag{B.19}$$

B.4 Range of interest derivation

In this Appendix, we prove that if $\lambda^\dagger \geq \rho\hat{\sigma}_w^2$, then $\xi_{min} < 0.9$ for $1.0002 \leq \rho \leq 3.16$.

Since LHS of (A.15) is an increasing function of λ and is 0 when $\lambda = \lambda^\dagger$, we obtain for $\lambda^\dagger \geq \rho\hat{\sigma}_w^2$

$$0 \geq \int_{1/(\rho\sigma_g^2\gamma)}^{\rho/(\sigma_g^2\gamma)} \frac{e^x}{x} dx - \frac{\sigma_g^2\gamma}{\rho} e^{\frac{\rho}{\sigma_g^2\gamma}} \tag{B.20}$$

$$= \frac{e^x}{x}\Big|_{x=1/(\rho\sigma_g^2\gamma)}^{x=\rho/(\sigma_g^2\gamma)} + \int_{1/(\rho\sigma_g^2\gamma)}^{\rho/(\sigma_g^2\gamma)} \frac{e^x}{x^2} dx - \frac{\sigma_g^2\gamma}{\rho} e^{\frac{\rho}{\sigma_g^2\gamma}} \tag{B.21}$$

$$= \int_{1/(\rho\sigma_g^2\gamma)}^{\rho/(\sigma_g^2\gamma)} \frac{e^x}{x^2} dx - \rho\sigma_g^2\gamma e^{\frac{1}{\rho\sigma_g^2\gamma}} \tag{B.22}$$

$$\geq \left(\frac{\sigma_g^2\gamma}{\rho}\right)^2 \left(e^{\frac{\rho}{\sigma_g^2\gamma}} - e^{\frac{1}{\rho\sigma_g^2\gamma}}\right) - \rho\sigma_g^2\gamma e^{\frac{1}{\rho\sigma_g^2\gamma}}. \tag{B.23}$$

In (B.23), we applied $\int_a^b e^x dx/x^2 \geq (e^b - e^a)/b^2$. From (B.23), we obtain

$$e^{\frac{\rho-\rho^{-1}}{\sigma_g^2\gamma}} \leq 1 + \frac{\rho - \rho^{-1}}{\sigma_g^2\gamma} \frac{\rho^4}{\rho^2 - 1}. \tag{B.24}$$

Since $x = -W_{-1}(-ae^{-a}) - a$ is the unique positive solution of $e^x = x/a + 1$ [18], we obtain from (B.24) that

$$\frac{\rho - \rho^{-1}}{\sigma_g^2\gamma} \leq -W_{-1}\left(-\frac{\rho^2 - 1}{\rho^4} e^{-\frac{\rho^2-1}{\rho^4}}\right) - \frac{\rho^2 - 1}{\rho^4}, \tag{B.25}$$

or equivalently,

$$\sigma_g^2\gamma \geq \frac{\rho-\rho^{-1}}{-W_{-1}\left(-\frac{\rho^2-1}{\rho^4}e^{-\frac{\rho^2-1}{\rho^4}}\right)-\frac{\rho^2-1}{\rho^4}}. \tag{B.26}$$

Applying the inequality $\int_a^b e^x dx/x \geq (e^b-e^a)/b$, we have

$$\mathrm{Ei}\left(\frac{\rho}{\sigma_g^2\gamma}\right)-\mathrm{Ei}\left(\frac{1}{\rho\sigma_g^2\gamma}\right)=\int_{1/(\rho\sigma_g^2\gamma)}^{\rho/(\sigma_g^2\gamma)}\frac{e^x}{x}dx \tag{B.27}$$

$$>\frac{\sigma_g^2\gamma}{\rho}\left(e^{\frac{\rho}{\sigma_g^2\gamma}}-e^{\frac{1}{\rho\sigma_g^2\gamma}}\right). \tag{B.28}$$

Then, it follows from (3.42) that for $\lambda^\dagger \geq \rho\hat{\sigma}_w^2$

$$\xi_{min} < 1-\frac{1}{2\ln(\rho)}\frac{\sigma_g^2\gamma}{\rho}\left(1-e^{-\frac{\rho-\rho^{-1}}{\sigma_g^2\gamma}}\right) \tag{B.29}$$

$$<1-\frac{1}{2\rho\ln(\rho)}\frac{\rho-\rho^{-1}}{-W_{-1}\left(-\frac{\rho^2-1}{\rho^4}e^{-\frac{\rho^2-1}{\rho^4}}\right)-\frac{\rho^2-1}{\rho^4}}\left(1-e^{W_{-1}\left(-\frac{\rho^2-1}{\rho^4}e^{-\frac{\rho^2-1}{\rho^4}}\right)+\frac{\rho^2-1}{\rho^4}}\right), \tag{B.30}$$

where (B.30) is derived from (B.26) and the fact that $\sigma_g^2\gamma\left(1-e^{-\frac{\rho-\rho^{-1}}{\sigma_g^2\gamma}}\right)$ is an increasing function of $\sigma_g^2\gamma$. Moreover, it can also be shown that

$$\frac{1}{2\rho\ln(\rho)}\frac{\rho-\rho^{-1}}{-W_{-1}\left(-\frac{\rho^2-1}{\rho^4}e^{-\frac{\rho^2-1}{\rho^4}}\right)-\frac{\rho^2-1}{\rho^4}}\left(1-e^{W_{-1}\left(-\frac{\rho^2-1}{\rho^4}e^{-\frac{\rho^2-1}{\rho^4}}\right)+\frac{\rho^2-1}{\rho^4}}\right)\geq 0.1, \tag{B.31}$$

for $1.0002 \leq \rho \leq 3.16$. Therefore, $\xi_{min} < 0.9$ for $1.0002 \leq \rho \leq 3.16$.

B.5 Equation (3.55) derivation

In this Appendix, we derive (3.55). Since the LHS of (A.15) is an increasing function of λ and is 0 when $\lambda = \lambda^\dagger$, we obtain from (3.53) that

$$\int_{\frac{1}{\rho\sigma_g^2\gamma}}^{\frac{1}{2\epsilon\ln(\rho)}}\frac{e^x}{x}dx-2\epsilon\ln(\rho)e^{\frac{1}{2\epsilon\ln(\rho)}}\leq 0, \tag{B.32}$$

for $\lambda^\dagger \geq \sigma_g^2 P/(2\epsilon \ln(\rho))$. Since $\int e^x x^{-1} dx \simeq e^x(x^{-1} + x^{-2})$ for $x \gg 1$ [28], the LHS of (B.32) can be approximated by

$$\int_{\frac{1}{\rho\sigma_g^2\gamma}}^{\frac{1}{2\epsilon\ln(\rho)}} \frac{e^x}{x} dx - 2\epsilon \ln(\rho) e^{\frac{1}{2\epsilon\ln(\rho)}}$$

$$\simeq (2\epsilon \ln(\rho))^2 e^{\frac{1}{2\epsilon\ln(\rho)}} - (\rho\sigma_g^2\gamma + (\rho\sigma_g^2\gamma)^2) e^{\frac{1}{\rho\sigma_g^2\gamma}} \tag{B.33}$$

$$\simeq (2\epsilon \ln(\rho))^2 e^{\frac{1}{2\epsilon\ln(\rho)}} - \rho\sigma_g^2\gamma e^{\frac{1}{\rho\sigma_g^2\gamma}}, \tag{B.34}$$

for $\sigma_g^2\gamma \ll 1$. Therefore, it follows from (B.32) and (B.34) that

$$\frac{1}{\rho\sigma_g^2\gamma} e^{-\frac{1}{\rho\sigma_g^2\gamma}} \lesssim \frac{1}{(2\epsilon\ln(\rho))^2} e^{-\frac{1}{2\epsilon\ln(\rho)}}. \tag{B.35}$$

Applying the Lambert-W function, W_{-1}, on both sides of (B.35) and $W_{-1}(-xe^{-x}) = -x$ for $x \geq 1$ [18], we obtain (3.55).

APPENDIX C. COVERT NON-ORTHOGONAL MULTIPLE ACCESS

C.1 Equation (4.37) derivation

In this Appendix, we provide the proof of (4.37)

a)

$$B\left(\frac{\sigma_f^2}{\sigma_g^2}, L+1\right) = \frac{\Gamma(L+1)\Gamma(\sigma_f^2/\sigma_g^2)}{\Gamma(L+1+\sigma_f^2/\sigma_g^2)} \tag{C.1}$$

$$\simeq \frac{\Gamma(\sigma_f^2/\sigma_g^2)}{(L+1)^{\sigma_f^2/\sigma_g^2}} \tag{C.2}$$

$$\to \quad 0 \tag{C.3}$$

as $L \to \infty$, where (C.2) follows from $\frac{\Gamma(n)}{\Gamma(n+m)} \simeq n^{-m}$ for large n and $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ [4].

Let k^* denote the largest integer such that $\alpha \sum_{l=1}^{k} \gamma_l - \gamma_1 \leq 0$ for all $k \leq k^*$. Then,

b) For $1 \leq k \leq k^*$, where $\alpha_k = \infty$, we obtain

$$\begin{pmatrix} L \\ k \end{pmatrix} \int_0^{\alpha_k} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{\nu_k(x)}{\sigma_f^2}}\right)^k \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx$$

$$\leq \begin{pmatrix} L \\ k \end{pmatrix} \int_0^{\infty} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \left(e^{-\frac{x}{\sigma_f^2}}\right)^k \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \tag{C.4}$$

$$= \frac{\sigma_f^2}{\sigma_g^2} \begin{pmatrix} L \\ k \end{pmatrix} B\left(k + \frac{\sigma_f^2}{\sigma_g^2}, L-k+1\right) \tag{C.5}$$

$$= \frac{\sigma_f^2}{\sigma_g^2} \frac{\Gamma\left(k + \frac{\sigma_f^2}{\sigma_g^2}\right)\Gamma(L+1)}{\Gamma(k+1)\Gamma\left(L+1+\frac{\sigma_f^2}{\sigma_g^2}\right)} \tag{C.6}$$

$$\simeq \frac{\sigma_f^2}{\sigma_g^2} \frac{\Gamma(k+\sigma_f^2/\sigma_g^2)}{\Gamma(k+1)} \frac{1}{(L+1)^{\sigma_f^2/\sigma_g^2}} \tag{C.7}$$

$$\to 0 \tag{C.8}$$

as $L \to \infty$.

c) For $k^* < k \leq L$, where $\alpha \sum_{l=1}^{k} \gamma_l - \gamma_1 > 0$: it follows from (4.30) that $\nu(|g|^2) < 1/(\alpha \sum_{l=1}^{k} \gamma_l - \gamma_1)$. Since $1 - e^{-x/\sigma_f^2} < 1 - e^{-\nu_k(x)/\sigma_f^2}$ if $x < \nu_k(x)$ or equivalently $x < \alpha_k$ and $e^{-x/\sigma_f^2} < 1$, we obtain

$$\begin{aligned}
&\binom{L}{k} \int_0^{\alpha_k} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{\nu_k(x)}{\sigma_f^2}}\right)^{k} \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \\
&< \binom{L}{k} \int_0^{\alpha_k} \left(1 - e^{-\frac{\nu_k(x)}{\sigma_f^2}}\right)^{L} \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \qquad \text{(C.9)} \\
&< \binom{L}{k} \int_0^{\alpha_k} \left(1 - e^{-\frac{1}{\sigma_f^2(\alpha \sum_{l=1}^{k} \gamma_l - \gamma_1)}}\right)^{L} \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx \qquad \text{(C.10)} \\
&\to 0 \qquad \text{(C.11)}
\end{aligned}$$

as $L \to \infty$. Therefore,

$$\xi_{max} \to 1 \qquad \text{(C.12)}$$

as $L \to \infty$.

C.2 Proof of optimum hiding strategy

In this Appendix, we prove that ξ_{min} in (4.55) is maximized when $j = 1$. It follows from (4.3) and (4.17) that

$$\Pr(I(\mathbf{v}_j; \mathbf{y}_{w,j}) < R_{j,0} | |f_j|^2 \geq \tau) = \frac{\Pr(|f_j|^2 \geq \tau_0)}{\Pr(|f_j|^2 \geq \tau)}, \qquad \text{(C.13)}$$

$$\Pr(I(\mathbf{v}_j; \mathbf{y}_{w,j}) < R_{j,0} | |f_j|^2 < \tau) = 1 - \frac{\Pr(|f_j|^2 < \tau_1)}{\Pr(|f_j|^2 < \tau)}, \qquad \text{(C.14)}$$

where

$$\tau_0 = \max\left\{\tau, \frac{|g|^2}{1 + |g|^2(\sum_{l \in \mathcal{D}_j^c} \gamma_l - \sum_{l \leq j} \gamma_l)}\right\}, \qquad \text{(C.15)}$$

$$\tau_1 = \min\left\{\tau, \frac{|g|^2 \alpha}{1 + |g|^2(\sum_{l \in \mathcal{D}_j^c} \gamma_l - \alpha \sum_{l \leq j} \gamma_l)}\right\}, \qquad \text{(C.16)}$$

and $\tau_1 \leq \tau \leq \tau_0$. Then, ξ_{min} in (4.55) can be rewritten as

$$\xi_{min} = \min\left\{\frac{\Pr(|f_j|^2 \geq \tau_0)}{\Pr(|f_j|^2 \geq \tau)}, 1 - \frac{\Pr(|f_j|^2 < \tau_1)}{\Pr(|f_j|^2 < \tau)}\right\}. \qquad \text{(C.17)}$$

Let $f(x)$ and $F(x)$ denote the common probability density function (PDF) and cumulative density function (CDF) of the corresponding un-ordered statistics of $|f_1|^2, ..., |f_L|^2$. Then, the PDF, $f_i(x)$, and complement CDF, $\bar{F}_i(x)$, of $|f_j|^2$ are given by [20]

$$f_i(x) = \frac{L!}{(L-i)!(i-1)!}[F(x)]^{L-i}[1-F(x)]^{i-1}f(x) \tag{C.18}$$

$$\bar{F}_i(x) = \sum_{k=i}^{L} \binom{L}{k} [F(x)]^{L-k}[1-F(x)]^{k}, \tag{C.19}$$

respectively. Then,

a) It follows from (C.18) and (C.19) that

$$\frac{d}{dx}\left(\frac{\Pr(|f_j|^2 \geq x)}{\Pr(|f_{j+1}|^2 \geq x)}\right) = \frac{f_{|f_{j+1}|^2}(x)\Pr(|f_j|^2 \geq x) - f_{|f_j|^2}(x)\Pr(|f_{j+1}|^2 \geq x)}{(\Pr(|f_{j+1}|^2 \geq x))^2}, \tag{C.20}$$

and

$$\begin{aligned}
&\Pr(|f_j|^2 \geq x) - \frac{f_{|f_j|^2}(x)}{f_{|f_{j+1}|^2}(x)}\Pr(|f_{j+1}|^2 \geq x) \\
=&\sum_{k=i}^{L}\binom{L}{k}[F(x)]^{L-k}[1-F(x)]^{k} - \frac{i}{L-i}\sum_{k=i+1}^{L}\binom{L}{k}[F(x)]^{L-k+1}[1-F(x)]^{k-1} \qquad \text{(C.21)} \\
>&\sum_{k=i}^{L-1}[F(x)]^{L-k}[1-F(x)]^{k}\left(\binom{L}{k} - \frac{i}{L-i}\binom{L}{k+1}\right) \\
>&0. \qquad \text{(C.22)}
\end{aligned}$$

Then, $\frac{d}{dx}(\frac{\Pr(|f_j|^2 \geq x)}{\Pr(|f_{j+1}|^2 \geq x)}) > 0$, i.e. $\frac{\Pr(|f_j|^2 \geq x)}{\Pr(|f_{j+1}|^2 \geq x)}$ is an increasing function of x. Hence, we have

$$\frac{\Pr(|f_j|^2 \geq \tau_0)}{\Pr(|f_{j+1}|^2 \geq \tau_0)} \geq \frac{\Pr(|f_j|^2 \geq \tau)}{\Pr(|f_{j+1}|^2 \geq \tau)} \tag{C.23}$$

for $\tau_0 \geq \tau$, or equivalently,

$$\frac{\Pr(|f_j|^2 \geq \tau_0)}{\Pr(|f_j|^2 \geq \tau)} \geq \frac{\Pr(|f_{j+1}|^2 \geq \tau_0)}{\Pr(|f_{j+1}|^2 \geq \tau)}, \tag{C.24}$$

That is, $\frac{\Pr(|f_j|^2 \geq \tau_0)}{\Pr(|f_j|^2 \geq \tau)}$ is a decreasing function of j for $\tau \leq \tau_0$. Therefore, $\frac{\Pr(|f_j|^2 \geq \tau_0)}{\Pr(|f_j|^2 \geq \tau)}$ is maximized when $j = 1$.

b) Similarly, it can be proved that $\frac{\Pr(|f_j|^2 < x)}{\Pr(|f_{j+1}|^2 < x)}$ is an increasing function of x. Then, $\frac{\Pr(|f_j|^2 < \tau)}{\Pr(|f_j|^2 < \tau_1)}$ is a decreasing function of j for $\tau \geq \tau_1$. Therefore, $\frac{\Pr(|f_j|^2 < \tau)}{\Pr(|f_j|^2 < \tau_1)}$ is maximized when $j = 1$.

By (C.17), a) and b), ξ_{min} is maximized when $j = 1$.

C.3 Equations (4.58)-(4.63) derivation

In this Appendix, we provide the proof of (4.58)-(4.63).

a) ξ_0: It follows from (4.58) that

$$\xi_0 = 1 - \Pr\left(|f_1|^2 \le \frac{|g|^2}{1+|g|^2(\sum_{l\in\mathcal{D}_1^c}\gamma_l - \gamma_1)}, |g|^2 < |f_1|^2 \middle| |f_1|^2 \ge \tau\right) - \Pr\left(|f_1|^2 \le \frac{|g|^2}{1+|g|^2(\sum_{l\in\mathcal{D}_1^c}\gamma_l - \gamma_1)}, |g|^2 \ge |f_1|^2 \middle| |f_1|^2 \ge \tau\right). \tag{C.25}$$

Since $\frac{|g|^2}{1+|g|^2(\sum_{l\in\mathcal{D}_1^c}\gamma_l-\gamma_1)} \le |g|^2$, we obtain

$$\Pr\left(|f_1|^2 \le \frac{|g|^2}{1+|g|^2(\sum_{l\in\mathcal{D}_1^c}\gamma_l - \gamma_1)}, |g|^2 < |f_1|^2 \middle| |f_1|^2 \ge \tau\right) = 0. \tag{C.26}$$

Also since Bob can decode and remove $\mathbf{v}_2, ..., \mathbf{v}_L$ if $|g|^2 \ge |f_1|^2$, we obtain $\mathcal{D}_1^c = \{1\}$. Hence, it follows from (C.25) and (C.26) that

$$\xi_0 = 1 - \Pr(|f_1|^2 \le |g|^2 | \, |f_1|^2 \ge \tau) \tag{C.27}$$

$$= 1 - \frac{\Pr(\tau \le |f_1|^2 \le |g|^2)}{\Pr(|f_1|^2 \ge \tau)} \tag{C.28}$$

$$= 1 - \frac{\int_\tau^\infty ((1-e^{-x/\sigma_f^2})^L - (1-e^{-\tau/\sigma_f^2})^L)\frac{e^{-x/\sigma_g^2}}{\sigma_g^2}dx}{1-(1-e^{-\tau/\sigma_f^2})^L} \tag{C.29}$$

$$= 1 - \frac{\frac{\sigma_f^2}{\sigma_g^2} B\left(e^{-\frac{\tau}{\sigma_f^2}}; \frac{\sigma_f^2}{\sigma_g^2}, L+1\right)}{1-(1-e^{-\tau/\sigma_f^2})^L} + \frac{\left(1-e^{-\frac{\tau}{\sigma_f^2}}\right)^L e^{-\frac{\tau}{\sigma_g^2}}}{1-(1-e^{-\tau/\sigma_f^2})^L}. \tag{C.30}$$

b) ξ_1: Similarly, we also have

$$\xi_1 = 1 - \Pr\left(|f_1|^2 \le \frac{|g|^2\alpha}{1+|g|^2(1-\alpha)\gamma_1} \middle| |f_1|^2 < \tau\right). \tag{C.31}$$

If $\tau > \frac{\alpha}{(1-\alpha)\gamma_1}$, then $\frac{|g|^2\alpha}{1+|g|^2(1-\alpha)\gamma_1} < \tau$ for all $|g|^2$. Then, we obtain

$$\xi_1 = 1 - \frac{\int_0^\infty \left(1-e^{-\frac{x\alpha/\sigma_f^2}{1+x(1-\alpha)\gamma_1}}\right)^L \frac{e^{-x/\sigma_g^2}}{\sigma_g^2}dx}{(1-e^{-\tau/\sigma_f^2})^L}. \tag{C.32}$$

However, if $\tau \leq \frac{\alpha}{(1-\alpha)\gamma_1}$, then $\frac{|g|^2\alpha}{1+|g|^2(1-\alpha)\gamma_1} \leq \tau$ for $|g|^2 \leq a_1$, where $a_1 = \frac{\tau}{\alpha-(1-\alpha)\gamma_1\tau}$, and $\frac{|g|^2\alpha}{1+|g|^2(1-\alpha)\gamma_1} > \tau$ for $|g|^2 > a_1$. Hence, we obtain

$$\xi_1 = 1 - e^{-\frac{a_1}{\sigma_f^2}} - \frac{\int_0^{a_1} \left(1 - e^{-\frac{x\alpha/\sigma_f^2}{1+x(1-\alpha)\gamma_1}}\right)^L \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx}{(1 - e^{-\tau/\sigma_f^2})^L}. \tag{C.33}$$

C.4 Equation (4.64) derivation

In this Appendix, we provide the proof of (4.64).

a) ξ_0: Since $B(x; n, m) < B(n, m)$, we obtain

$$B\left(e^{-\frac{\tau}{\sigma_f^2}}; \frac{\sigma_f^2}{\sigma_g^2}, L+1\right) \leq B\left(\frac{\sigma_f^2}{\sigma_g^2}, L+1\right) \tag{C.34}$$

$$\simeq \frac{\Gamma(\sigma_f^2/\sigma_g^2)}{(L+1)^{\sigma_f^2/\sigma_g^2}} \tag{C.35}$$

$$\to 0 \tag{C.36}$$

as $L \to \infty$. Also, $(1 - e^{-\tau/\sigma_f^2})^L \to 0$ as $L \to \infty$. Therefore, ξ_0 in (4.59) converges to 1 as $L \to \infty$.

b) ξ_1: Since $x\alpha/(1 + x(1-\alpha)\gamma_1) < \alpha/((1-\alpha)\gamma_1)$ for all x and γ_1, we obtain

$$\frac{\int_0^{\infty} \left(1 - e^{-\frac{x\alpha/\sigma_f^2}{1+x(1-\alpha)\gamma_1}}\right)^L \frac{e^{-x/\sigma_g^2}}{\sigma_g^2} dx}{(1 - e^{-\tau/\sigma_f^2})^L} \leq \left(\frac{1 - e^{-\frac{\alpha}{(1-\alpha)\gamma_1\sigma_f^2}}}{1 - e^{-\tau/\sigma_f^2}}\right)^L \tag{C.37}$$

$$\to 0 \tag{C.38}$$

as $L \to \infty$ for $\tau > \alpha/((1-\alpha)\gamma_1)$. Therefore, ξ_1 in (4.62) converges to 1 as $L \to \infty$.

By a) and b), $\xi_{max} = \min\{\xi_0, \xi_1\}$ converges to 1 as L increases for $\tau > \alpha/((1-\alpha)\gamma_1)$.

C.5 Equation (4.67) derivation

In this Appendix, we provide the proof of (4.67). $P_{o,u}$ in (4.66) can be rewritten as

$$\begin{aligned}P_{o,u}=&1-\Pr\left(|h|^2\geq\mu_0,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}\geq|f_1|^2\middle||f_1|^2<\tau\right)\\&-\Pr\left(|h|^2\geq\mu_1,|f_2|^2\leq|h|^2,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}<|f_1|^2\middle||f_1|^2<\tau\right)\\&-\sum_{k=2}^{L}\Pr(|h|^2\geq\mu_k,|f_{k+1}|^2\leq|h|^2<|f_k|^2||f_1|^2<\tau).\end{aligned}\tag{C.39}$$

Since

$$\begin{aligned}&\Pr\left(|h|^2\geq\mu_1,|f_2|^2\leq|h|^2,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}<|f_1|^2\middle||f_1|^2<\tau\right)\\=&\Pr\left(|h|^2\geq\mu_1,|f_2|^2\leq|h|^2,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}<|f_1|^2,|f_1|^2\leq|h|^2\middle||f_1|^2<\tau\right)\\&+\Pr\left(|h|^2\geq\mu_1,|f_2|^2\leq|h|^2,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}<|f_1|^2,|f_1|^2>|h|^2\middle||f_1|^2<\tau\right)\qquad\text{(C.40)}\\=&\Pr\left(|h|^2\geq\mu_1,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}<|f_1|^2\leq|h|^2\middle||f_1|^2<\tau\right)\\&+\Pr\left(|h|^2\geq\mu_1,|f_2|^2\leq|h|^2<|f_1|^2\middle||f_1|^2<\tau\right),\qquad\text{(C.41)}\end{aligned}$$

then we obtain

$$\begin{aligned}P_{o,u}=&1-\Pr\left(|h|^2\geq\mu_0,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}\geq|f_1|^2\middle||f_1|^2<\tau\right)\\&-\Pr\left(|h|^2\geq\mu_1,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}<|f_1|^2\leq|h|^2\middle||f_1|^2<\tau\right)\\&-\sum_{k=1}^{L}\Pr(|h|^2\geq\mu_k,|f_{k+1}|^2\leq|h|^2<|f_k|^2||f_1|^2<\tau).\end{aligned}\tag{C.42}$$

a)

$$\begin{aligned}&\Pr\left(|h|^2\geq\mu_0,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}\geq|f_1|^2\middle||f_1|^2<\tau\right)\\=&\frac{\Pr\left(|h|^2\geq\mu_0,|f_1|^2\leq\max\left\{\tau,\frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}\right\}\right)}{\Pr(|f_1|^2<\tau)}\qquad\text{(C.43)}\\=&\frac{\int_{\mu_0}^{\infty}\left(1-e^{-\min\left\{\frac{\tau}{\sigma_f^2},\frac{x\alpha}{\sigma_f^2(1+x(1-\alpha)\gamma_1)}\right\}}\right)^L\frac{e^{-x/\sigma_h^2}}{\sigma_h^2}dx}{(1-e^{-\tau/\sigma_h^2})^L}.\qquad\text{(C.44)}\end{aligned}$$

b)

$$\Pr\left(|h|^2 \geq \mu_1, \frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1} < |f_1|^2 \leq |h|^2 \middle| |f_1|^2 < \tau\right)$$

$$= \frac{\Pr\left(|h|^2 \geq \mu_0, |f_1|^2 \leq \max\{\tau, |h|^2\}\right) - \Pr\left(|h|^2 \geq \mu_0, |f_1|^2 \leq \max\left\{\tau, \frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1}\right\}\right)}{\Pr(|f_1|^2 < \tau)} \quad \text{(C.45)}$$

$$= \frac{\int_{\mu_1}^{\infty}\left(1-e^{-\min\left\{\frac{\tau}{\sigma_f^2}, \frac{x}{\sigma_f^2}\right\}}\right)^L \frac{e^{-x/\sigma_h^2}}{\sigma_h^2}dx - \int_{\mu_1}^{\infty}\left(1-e^{-\min\left\{\frac{\tau}{\sigma_f^2}, \frac{x\alpha}{\sigma_f^2(1+x(1-\alpha)\gamma_1)}\right\}}\right)^L \frac{e^{-x/\sigma_h^2}}{\sigma_h^2}dx}{(1-e^{-\tau/\sigma_h^2})^L}. \quad \text{(C.46)}$$

c) For $1 \leq k \leq L$,

$$\Pr(|h|^2 \geq \mu_k, |f_{k+1}|^2 \leq |h|^2 < |f_k|^2 || f_1|^2 < \tau)$$

$$= \frac{\Pr\left(|h|^2 \geq \mu_k, |f_{k+1}|^2 \leq |h|^2 < |f_k|^2, |f_1|^2 < \tau\right)}{\Pr(|f_1|^2 < \tau)} \quad \text{(C.47)}$$

$$= \begin{pmatrix} L \\ k \end{pmatrix} \frac{\int_{\mu_k}^{\tau}\left(1-e^{-\frac{x}{\sigma_f^2}}\right)^{L-k}\left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{\tau}{\sigma_f^2}}\right)^k \frac{e^{-x/\sigma_h^2}}{\sigma_h^2}dx}{(1-e^{-\tau/\sigma_h^2})^L}. \quad \text{(C.48)}$$

By (C.42), a), b) and c), we obtain (4.67).

C.6 Equation (4.78) derivation

In this Appendix, we provide the proof of (4.78). It follows from (4.77) that

a) $k = 0$, i.e. Willie can decode $\mathbf{v}_2, ..., \mathbf{v}_L$ or $\mathcal{D}_1^c = \{1\}$: We obtain from (4.29) with L replaced by LM,

$$\Pr\left(|f_{1,1}|^2 \leq \frac{|g_m|^2}{(\alpha - |g_m|^2(1-\alpha)\gamma_1)^+}, |f_{1,1}|^2 \leq |g_m|^2\right) = \frac{\sigma_f^2}{\sigma_g^2} B\left(\frac{\sigma_f^2}{\sigma_g^2}, LM+1\right). \quad \text{(C.49)}$$

b) $1 \leq k \leq L$, i.e. Willie can decode $\mathbf{v}_{k+1}, ..., \mathbf{v}_L$ or $\mathcal{D}_1^c = \{1, ..., k\}$: For given $|f_{1,1}|^2 = t$, the corresponding un-ordered random variables of the order statistics, $|f_{1,l}|^2$, $l = 1, ..., L$, are i.i.d and have the common CDF of

$$F(x|t) = \frac{1-e^{-x/\sigma_f^2}}{1-e^{-t/\sigma_f^2}} \quad \text{(C.50)}$$

for $0 \leq x \leq t$. Since the probability density function (PDF) of $|f_{1,1}|^2$ is

$$f_{|f_{1,1}|^2}(t) = LM(1-e^{-x/\sigma_f^2})^{LM-1}e^{-x/\sigma_f^2}/\sigma_f^2, \quad \text{(C.51)}$$

it follows from (C.50) and (C.51) that

$$\begin{aligned}
&\Pr(|f_{1,1}|^2 \le \nu_k(x), |f_{1,k+1}|^2 \le x < |f_{1,k}|^2) \\
=&\int_x^{\nu_k(x)} \Pr(|f_{1,k+1}|^2 \le x < |f_{1,k}|^2 |\, |f_{1,1}|^2 = t) f_{|f_{1,1}|^2}(t) dt && \text{(C.52)} \\
=&\int_x^{\nu_k(x)} \binom{L-1}{k-1} (F(x|t))^{L-k} (F(t|t) - F(x|t))^{k-1} f_{|f_{1,1}|^2}(t) dt && \text{(C.53)} \\
=&LM \binom{L-1}{k-1} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \int_x^{\nu_k(x)} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{t}{\sigma_f^2}}\right)^{k-1} \left(1 - e^{-\frac{t}{\sigma_f^2}}\right)^{L(M-1)} \frac{e^{-\frac{t}{\sigma_f^2}}}{\sigma_f^2} dt. && \text{(C.54)}
\end{aligned}$$

Replacing $e^{-x/\sigma_f^2} - e^{-t/\sigma_f^2}$ by y and applying the binomial expansion,

$$\left(y + 1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L(M-1)} = \sum_{l=0}^{L(M-1)} \binom{L(M-1)}{l} y^l \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L(M-1)-l}, \qquad \text{(C.55)}$$

into (C.54) yield

$$\begin{aligned}
&\Pr(|f_{1,1}|^2 \le \nu_k(x), |f_{1,k+1}|^2 \le x < |f_{1,k}|^2) \\
=&LM \binom{L-1}{k-1} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \int_0^{e^{-x/\sigma_f^2} - e^{-\nu_k(x)/\sigma_f^2}} y^{k-1} \left(y + 1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L(M-1)} dy && \text{(C.56)} \\
=&LM \binom{L-1}{k-1} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L-k} \sum_{l=0}^{L(M-1)} \binom{L(M-1)}{l} \\
&\times \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{L(M-1)-l} \int_0^{e^{-x/\sigma_f^2} - e^{-\nu_k(x)/\sigma_f^2}} y^{l+k-1} dy && \text{(C.57)} \\
=&\binom{L-1}{k-1} \sum_{l=0}^{L(M-1)} \frac{LM}{k+l} \binom{L(M-1)}{l} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{LM-k-l} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{\nu_k(x)}{\sigma_f^2}}\right)^{k+l}. && \text{(C.58)}
\end{aligned}$$

Therefore, we obtain from (C.58) that

$$\begin{aligned}
&\Pr(|f_{1,1}|^2 \le \nu_k(|g_m|^2), |f_{1,k+1}|^2 \le |g_m|^2 < |f_{1,k}|^2) \\
=&\binom{L-1}{k-1} \sum_{l=0}^{L(M-1)} \frac{LM}{k+l} \binom{L(M-1)}{l} \int_0^{\alpha_k} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{LM-k-l} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{\nu_k(x)}{\sigma_f^2}}\right)^{k+l} \frac{e^{-\frac{x}{\sigma_g^2}}}{\sigma_g^2} dx. && \text{(C.59)}
\end{aligned}$$

By a) and b), we obtain (4.78).

C.7 Equation (4.79) derivation

In this Appendix, we provide the proof of (4.79). The decoding outage probability can be obtained from (4.45) with $|g|^2$ replaced by $|g_m|^2$ and $|f_k|^2$ replaced by $|f_{1,k}|^2$ for $0 \le k \le L$,

$$P_{o,u} = 1 - \sum_{k=0}^{L} \Pr(|h_m|^2 \ge \mu_k, |f_{1,k+1}|^2 \le |h_m|^2 < |f_{1,k}|^2). \tag{C.60}$$

a) $k = 0$:

$$\Pr(|h_m|^2 \ge \mu_0, |f_{1,1}|^2 \le |h_m|^2) = \int_{\mu_0}^{\infty} (1 - e^{-x/\sigma_f^2})^{LM} \frac{e^{-x/\sigma_h^2}}{\sigma_h^2} dx \tag{C.61}$$

$$= \frac{\sigma_f^2}{\sigma_g^2} B\left(e^{-\frac{\mu_0}{\sigma_h^2}}; \frac{\sigma_f^2}{\sigma_g^2}, LM+1\right). \tag{C.62}$$

b) $k > 0$:

$$\Pr(|h_m|^2 \ge \mu_k, |f_{1,k+1}|^2 \le |h_m|^2 < |f_{1,k}|^2)$$

$$= \int_{\mu_k}^{\infty} \Pr(|f_{1,k+1}|^2 \le |h_m|^2 < |f_{1,k}|^2) \frac{e^{-x/\sigma_h^2}}{\sigma_h^2} dx \tag{C.63}$$

$$= \binom{L-1}{k-1} \sum_{l=0}^{L(M-1)} \frac{LM}{k+l} \binom{L(M-1)}{l} \frac{1}{\sigma_h^2} \int_{\mu_k}^{\infty} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{LM-k-l} \left(e^{-\frac{x}{\sigma_f^2}}\right)^{k+l+\sigma_f^2/\sigma_h^2} dx \tag{C.64}$$

$$= \binom{L-1}{k-1} \sum_{l=0}^{L(M-1)} \frac{LM}{k+l} \binom{L(M-1)}{l} \frac{\sigma_f^2}{\sigma_h^2} B\left(e^{-\frac{\mu_k}{\sigma_f^2}}; k+l+\frac{\sigma_f^2}{\sigma_h^2}, LM-k-l+1\right), \tag{C.65}$$

where (C.64) is derived from (C.58) where $\nu_k(x) = \infty$.

By a) and b), we obtain (4.79).

C.8 Equation (4.80) derivation

In this Appendix, we provide the proof of (4.80). Similar Appendix C.5, replacing $|h|^2$ by $|h_m|^2$ and $|f_k|^2$ by $|f_{1,k}|^2$ for $k \in [0, L]$ into (C.42), we obtain

$$P_{o,u} = 1 - \Pr\left(|h|^2 \ge \mu_0, \frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1} \ge |f_{1,1}|^2 \middle| |f_{1,1}|^2 < \tau\right)$$

$$- \Pr\left(|h|^2 \ge \mu_1, \frac{|h|^2\alpha}{1+|h|^2(1-\alpha)\gamma_1} < |f_{1,1}|^2 \le |h|^2 \middle| |f_{1,1}|^2 < \tau\right)$$

$$- \sum_{k=1}^{L} \Pr(|h|^2 \ge \mu_k, |f_{1,k+1}|^2 \le |h|^2 < |f_{1,k}|^2 || f_{1,1}|^2 < \tau). \tag{C.66}$$

a) The second and third terms of (C.66) can be obtained from (C.44) and (C.46), respectively, with L replaced by LM.

b) The fourth term of (C.66) is given by

$$\begin{aligned}
&\Pr(|h|^2 \geq \mu_k, |f_{1,k+1}|^2 \leq |h|^2 < |f_{1,k}|^2 || f_{1,1}|^2 < \tau) \\
&= \frac{\Pr\left(|h|^2 \geq \mu_k, |f_{1,k+1}|^2 \leq |h|^2 < |f_{1,k}|^2, |f_{1,1}|^2 < \tau\right)}{\Pr(|f_{1,1}|^2 < \tau)} && \text{(C.67)} \\
&= \frac{1}{(1-e^{-\tau/\sigma_h^2})^{LM}} \int_{\mu_k}^{\tau} \Pr(|f_{1,1}|^2 < \tau, |f_{1,k+1}|^2 \leq x < |f_{1,k}|^2) \frac{e^{-x/\sigma_h^2}}{\sigma_h^2} dx && \text{(C.68)} \\
&= \sum_{k=1}^{L} \begin{pmatrix} L-1 \\ k-1 \end{pmatrix} \sum_{l=0}^{L(M-1)} \frac{ML}{k+l} \begin{pmatrix} L(M-1) \\ l \end{pmatrix} \\
&\quad \times \frac{\int_{\mu_k}^{\tau} \left(1 - e^{-\frac{x}{\sigma_f^2}}\right)^{LM-k-l} \left(e^{-\frac{x}{\sigma_f^2}} - e^{-\frac{\tau}{\sigma_f^2}}\right)^{k+l} \frac{e^{-\frac{x}{\sigma_h^2}}}{\sigma_h^2} dx}{(1-e^{-\tau/\sigma_h^2})^{LM}}, && \text{(C.69)}
\end{aligned}$$

which is derived from (C.58) where $\nu_k(x) = \tau$.

By a) and b), we obtain (4.80).