July 2017

Covert Wireless Communications in a Dynamic Environment

Tamara V. Sobers

COVERT WIRELESS COMMUNICATIONS IN A DYNAMIC ENVIRONMENT

A Dissertation Presented

by

TAMARA V. SOBERS

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2017

Electrical & Computer Engineering

COVERT WIRELESS COMMUNICATIONS IN A DYNAMIC ENVIRONMENT

A Dissertation Presented

by

TAMARA V. SOBERS

Approved as to style and content by:

Patrick Kelly, Co-chair

Dennis Goeckel, Co-chair

Robert W. Jackson, Member

Donald F. Towsley, Member

C. V. Hollot, Department Chair
Electrical & Computer Engineering

DEDICATION

To my family and friends...near, far, extended and in memory.

ACKNOWLEDGMENTS

I would like to thank my committee members Professor Dennis Goeckel, Professor Patrick Kelly, Professor Robert Jackson and Professor Donald Towsley. I have had the pleasure of working with all my committee members on different research projects during my tenure here at UMass and have benefited from their expertise and insightful questions.

I appreciate the guidance from Professor Kelly during the completion of my Masters Thesis as well as the opportunity to take various classes he has taught. Some of the same methodologies I learned in his courses applied to the work presented in this dissertation. I am also grateful to Prof. Jackson for the time I have worked with him on research projects. His critical questions and insight have impacted the way in which I consider potential covert communication applications. A majority of my interactions with Prof. Towsley centered around the covert communications research presented in this dissertation. He helped with simplifying the presentations of some of the proofs in this work. I thank Prof. Goeckel for serving as my co-advisor. These past few years I have worked on various projects that have allowed me to further expand my background related to statistical signal processing. Professor Goeckel's passion to study new and challenging problems has also served as a great example.

Special thanks to Dr. Boulat Bash and Dr. Saikat Guha from Raytheon BBN for discussions related to the research presented in this dissertation and general topics regarding professional development. This collaboration also allowed me the opportunity to present my research at Raytheon BBN which helped me gain experience in communicating my research.

I am also appreciative of my labmates and peers, past and present: Dr. Çağatay Çapar, Dr. Adam Polak, Dr. Ali Rakhshan, Miyong Ko, Mostafa Dehghan, Hamid Dadkhahi, Dian Mo, Dr. Martin Muthee, Prasana Ravindran, Randy Kwende, Ramin Soltani and Ke Li. Although I have spent most of my time finishing up my dissertation outside the lab this past year, I do appreciate the times spent discussing research, potential post-graduation plans, and other general topics. I am also thankful to the staff who support the work of the Electrical and Computer Engineering Department: Joan Kermensky and Mary McCulloch.

There are numerous persons across the College of Engineering that I have interacted with and that have been supportive during my graduate career. Thank you to: Paul Jones, Dr. Paula Rees, Dr. Cheryl Brooks, Sally Darby and Dr. Cheryl Nicholas. I have shared my love of engineering by engaging in outreach programs that helped sustain my drive and also communicated my research with various age groups. I also acknowledge the many undergraduates I have had the pleasure of meeting in student groups or through teaching opportunities.

The ECE PhD Womens group is another association that provided the opportunity for great lunch discussions with peers: Shermin Hamzehei, Shirin Montazeri, Nirupama Ravi, Meenakshi Upadhyaya and other members. Thank you to Prof. Ganz, Prof. Gao and Prof. Christopher Hollot in helping support this group. I am also grateful to the various offices on campus which have provided graduate support: Phaedra Davis from the Graduate School and the staff from the Office of Professional Development.

The financial support I have received during my tenure here at UMass has also been vital towards the completion of my dissertation. Special thanks to the Northeast Alliance for Graduate Education and the Professoriate (NEAGEP) and the coordinators Dr. Sandy Petersen, Vanessa Hill, Pat Lehouillier and past coordinators Dr. Marlina Duncan and Dr. Heyda Martinez. I also thank the GEM Fellowship and

the MITRE Corporation for the many opportunities to gain experience through internships: Jeff P. Long, Cho Li, Rama, Dave Demoulpied and my other co-workers. The National Science Foundation (NSF) has also provided support through grants: CNS-1564067, ECCS-1309573, and ECCS-1341979. A portion of the work presented in this dissertation was also supported by the DARPA Quiet program.

I believe mentorship is very important and I would be remiss if I did not mention my mentors from my undergraduate institution, Rensselaer Polytechnic Institute: Dr. Durgans, Mrs. Durgans, Dafney Amilcar, Dr. Michael Silas and Mrs. Farmer. Your words of encouragement to think bigger were instrumental during my time at RPI as well as here at UMass Amherst. I also thank my classmates and peers from RPI: Dr. Allyce Caines, Kimara Freeman, Dr. James Morris-King, Bryan Finck, George Ponick, Richard Rodney, Esuasi Segbefia, Travis Carless any many more. Special thanks to Dr. Kyle Morrison for his friendship and a fellow labmate during my tenure at UMass Amherst. Connecting with Dr. Morrison also created the opportunity to share our group's research at MIT Lincoln Laboratory which provided fruitful discussions.

Thank you to my many many friends that I have met here at UMass as well as the various accountability groups I have participated in: Adaeze Egwuatu, Charisse Pickron, Christina Chisholm and Steele Valenzuela. Special thanks to Radhameris Gómez for her continued friendship since Rensselaer and energetic support.

"It takes a village to raise a child," and I think I may have had a city. I would like to thank my home church, the Episcopal Church of the Holy Spirit Mattapan, MA for their steadfast encouragement and prayers. Thank you also to family friends and neighbors for their support throughout my life. I am also grateful to the teachers and friends that were supportive of me while I was growing up. Also, my basketball coach, Coach Cecily Charsky, for showing me the importance of setting goals.

My world-wide family has been beyond awesome: Locally in the Boston, MA area; Toronto, Canada; Maryland, MA; Washington D.C.; England, U.K.; and Barbados. All of you have had a hand in shaping my life and we have never allowed physical distance to overshadow our connections and love of one another. Your support throughout my life has felt awesome, warming and beyond words I can describe. Whether babysitting me, playing video games, beating me in scrabble, taking me to the beach, waking me up for school, just sitting around talking and so much more. I also thank Cheryl Barrett, my sister from another mother, for showing me what it means to live with a big heart. Your selflessness and unbounded love is the ideal anyone could strive to attain.

Last but certainly not least, I would like to thank my mother, my father and my brother. To my parents, I thank you for encouraging me to further my curiosity. You provided various opportunities for me to grow with music lessons, wooden logic puzzles, pick ups from basketball practices and so much more. My brother, thank you for your ears to listen as well as reminding this little sister that I must listen as well. Your own path to college has had an impact on my own educational journey. To all my loved ones, thank you for your patience over the past few years as I worked towards completing my dissertation.

I thank all of my loved ones who are in memory. Pursuing a doctorate has been a blessing and I humbly appreciate the opportunities afforded to me due to the sacrifices of those who have walked before me. One of my grandmother's once said, "I have a masters and a Ph.D. A masters in common sense and a Ph.D. in survival." At the time, I did not have any plans for after high school and I was also unaware of what a Ph.D. entailed. However, I hope this document is a reflection of the drive both of my grandmothers have demonstrated to me in my lifetime. There is a Barbadian saying, "The hardest thing is to know." Now, I think I know.

ABSTRACT

COVERT WIRELESS COMMUNICATIONS IN A DYNAMIC ENVIRONMENT

MAY 2017

TAMARA V. SOBERS

B.Sc., RENSSELAER POLYTECHNIC INSTITUTE

M.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Patrick Kelly and Professor Dennis Goeckel

This dissertation investigates covert communication in dynamic wireless communication environments. A key goal is to provide insight about the capabilities of a transmitter desiring to remain covert and analogously, the capabilities of the party attempting to detect covert communications. The first chapter provides background on covert communications prior to this work. The second chapter studies the theoretical limits of covert communication and proves that positive rate is achievable when a jammer is added to the classical Alice/Bob/Warden Willie model. The third chapter expands on the second chapter by considering more generally the impact of the dynamics of the environment on the Alice/Bob/Warden Willie model. The dynamics of the environment generate uncertainty at Willie even if the jammer does not vary his/her power or even if Willie employs an antenna array to mitigate the jamming. The fourth and fifth chapters investigate the impact of considering the exact

continuous-time model rather than a discrete-time model approximation. In particular, detectors at Willie which leverage information in the continuous-time domain outperform detectors based on the discrete-time model approximation. The fourth and fifth chapters consider the continuous-time model of the Alice/Bob/Willie scenario and the Alice/Bob/Willie/Jammer scenarios respectively. The fourth and fifth chapters may appear to question the results of Chapter 2, Chapter 3 and prior wireless covert communication related research. However, these final chapters provide insight about different detectors available to Willie and the importance of Alice implementing communication schemes which do not contain features that significantly differ from Willie's observation under the null hypothesis. Our work has demonstrated how the covert throughput critically depends on Willie's knowledge of the environment and how the covert transmitter, allies in the area, or the dynamics of the environment itself might impact that knowledge. Future work will continue to move covert communications closer to practice by integrating further aspects of practical communication system design.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

Prior research in wireless covert communications analyzed the fundamental theoretical limits of covert communications when a wireless transmitter wants to reliably communicate to a legitimate recipient without risk of detection by a watchful adversary. A key finding is that the transmitter can only transmit $\mathcal{O}(\sqrt{n})$ covert bits in n channel uses in order to maintain covert and reliable communication to the intended recipient [1]. This dissertation builds upon this prior work in covert communications. Chapter 2 considers the addition of a jammer and demonstrates that the addition of the jammer creates uncertainty at the adversary's receiver. This additional uncertainty allows the legitimate transmitter to send $\mathcal{O}(n)$ covert bits in n channel uses reliably to the legitimate receiver. Both additive white Gaussian noise and finite block fading channels are considered in Chapter 2.

The third chapter expands on results in Chapter 2 by considering more generally a dynamic environment which generates fading variations on the adversary's observations. Research presented in Chapter 2 assumes that all devices in the environment are stationary and that fading variations are finite over the duration of the codeword length which the transmitter sends. However, dynamic environments such as urban environments or electronic warfare scenarios, may cause the adversary to observe large degrees of fading variations due to the channel conditions or the movement of entities in the model. The ability for a legitimate transmitter to communicate covertly in dynamic channel conditions is considered in the third chapter. The fading

variations observed at the adversary varies as a function of the codeword length of the transmitters codeword.

The fourth and fifth chapters consider the continuous-time models. The fourth chapter considers the Alice/Bob/Willie scenario and the fifth chapter considers the Alice/Bob/Willie/Jammer scenario. Prior research as well as work presented in Chapter 2 and Chapter 3 analyze covert communications based on the discrete-time model. The equivalent discrete-time model is an approximation of the continuous-time model and is generally assumed to contain sufficient information to represent the continuous-time model. However, the discrete-time model assumes Willie can determine the exact time instances to sample his continuous-time observation. This assumption is not always valid and the continuous-time model is capable of modeling when Willie does not know when to sample. For example, in Chapter 2, a power detector is proven to be an optimal detector based on the equivalent discrete-time model; however, the power detector is not always the optimal detector based on the continuous-time model. Therefore, the fourth and fifth chapters investigate whether the "equivalent discrete-time model" is in fact an equivalent discrete-time model when analyzing covert communications.

The sixth and final chapter concludes with a summary of the work presented in this dissertation as well as suggestions for future work based on the findings herein.

1.1 Motivation

The desire for two parties to communicate without a third party (an adversary) understanding the content of their communication has existed for many ages. To prevent adversaries from understanding messages between legitimate transmitters and receivers, legitimate parties can use encryption to hide the content of their messages. Encryption is the act of encoding original messages (plaintext) into a ciphertext using a secret key with the aim that an adversary is not capable of extracting the

original message. Some of the earliest instances of encryption can be traced back to 1900 BC [2]. Since 1900 BC, the sophistication of encryption schemes have improved drastically. However, classic cryptography has an underlying consistent design characteristic that may hinder maintaining the privacy of messages that use classic cryptography. Classic cryptography is designed so the decoding process is "easy" for the intended recipient and preferably "hard" for any eavesdropper. However, advancements in technology have allowed adversaries to develop strategies to decode (i.e. break) encrypted messages. As a result, legitimate transmitters and receivers make further technological improvements by designing more sophisticated encryption protocols. Standards for secure crypto systems are constantly evolving due to the cyclic nature of legitimate users attempting to maintain successful encrypted communications and adversaries attempting to break encryptions.

A popular historical example of an adversary breaking an encryption scheme is the decoding of the German Code Enigma by Allied powers in the 1940s [2]. Researchers and mathematicians were able to break the Engima code using ciphertext obtained from German radio transmissions. Each day, ciphertext messages were observed in the morning and analyzed using machines to determine the ciphertext key. As a result, Allied powers were able to decode encrypted German transmissions that were later sent throughout the day. There are many instances throughout history when adversaries decode messages using a known ciphertext or advances in computational power to develop deciphering algorithms. In recent decades, some cryptographic schemes such as Wired Equivalent Privacy (WEP) [3] and Data Encryption Standard (DES) [4] have become obsolete due to advances in computational power. Such technological advances align with Moore's Law which states that advances in hardware are expected and so the ability of adversaries to break encryption schemes is not unexpected. Quantum computing is also a maturing research area and once successful may make all modern day cryptography obsolete [5–8]. However, it is assumed that

new encryption schemes would be developed to operate under quantum conditions which is also currently an open area of research. Some other notable cases of adversaries breaking encrypted messages can be found in *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography* by Signh [2] and *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet* by Kahn [9] which may be of further historical interest.

Even if adversaries cannot acquire the ciphertext or exploit computational resources to decipher encrypted messages, there are still strategies adversaries can employ to learn information about message content hidden in a ciphertext. These types of attacks are called side-channel attacks and an example attack strategy is the use of *meta-data* to infer message content. Exploiting meta-data to learn about message content has been well established by Edward Snowden [10]. Edward Snowden disclosed that even if the government of United States of America does not have access to the content within a communication, government agencies are still capable of extracting information about the message content by leveraging meta-data. For example, knowing the two parties involved in a phone call, the length of the call, and the frequency at which two parties converse may be used to learn about what is discussed during a phone call without actually hearing any of the conversation.

To summarize, there are two potential drawbacks of classical cryptography that are highlighted thus far: 1) technological advances by adversaries may place current and past encrypted communications at risk; and 2) meta-data can be exploited as a side-channel attack to infer message content without direct access to content within a communication. Given these vulnerabilities, developing strategies for legitimate parties to send messages covertly without detection by an adversary is of great interest.

Covert communication occurs when two legitimate parties are able to communicate without an adversary detecting their communications. If users communicate covertly, adversaries cannot exploit meta-data or collect ciphertext with the goal of developing

strategies to decode the encrypted messages at a later date. Cryptography should still be used in conjunction with covert strategies communications, however, research is needed to determine under what conditions covert communication can be achieved or thwarted.

There are many relevant applications for the adoption of covert wireless communication in modern day wireless communications. Covert wireless communication has obvious military applications such as providing additional cover when soldiers are deployed in theater and need to keep their presence hidden. There are also instances when governments may want to prevent or monitor social unrest by monitoring wireless communications. Determining the limits of covert communication helps adversaries evaluate their detection capabilities and analogously, communicating parties can evaluate their ability to achieve covert and reliable communication.

1.2 Background

Modern day wireless covert communications research has been revitalized due to work by Bash, Goeckel and Towsley in [11]. In [11], Bash *et al.* consider the communication scenario (shown in Figure 1.1) where a transmitter (Alice) would like to send a message to legitimate receiver (Bob) in the presence of an eavesdropper (Willie the Warden). The nomenclature Warden is used instead of Eve because the Warden's role differs from Eve's in classic cryptographic scenarios. Eve's role is to extract the hidden content in encrypted messages that are shared between Alice and Bob. However, the Warden's sole task is to detect if any communication is occurring between Alice and Bob. If all channels experience additive white Gaussian noise (AWGN), the Square Root Law (SRL) presented in [11] states that Alice can only transmit $\mathcal{O}(\sqrt{n})$ covert bits reliably to Bob in n channel uses. If Alice tries to transmit bits at a higher rate, then she risks being detected by Willie with high probability. If she tries to transmit any fewer bits, then Bob cannot successfully

decode her messages with small probability of error. Throughout the remainder of this work, covert communication refers to covert and reliable communication such that Bob can successfully reconstruct Alice's message with minimum error without Warden Willie detecting she transmitted a message.

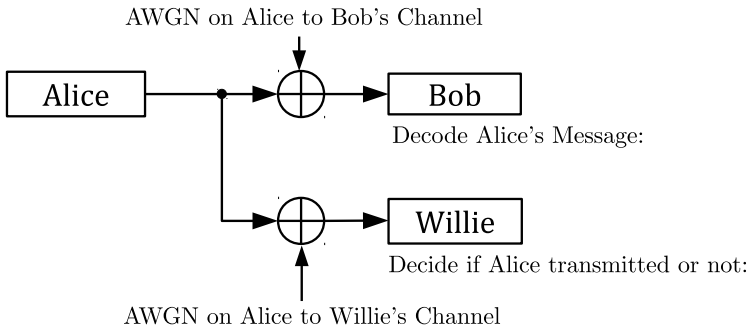

Figure 1.1. Alice, Bob and Warden Willie model under AWGN channel conditions.

Essentially, there is a balance that must be maintained for Alice to achieve covert communication. As an example, consider a very simple situation such as Alice talking with her voice. Alice must talk "quiet" enough such that Willie cannot detect her communications, yet "loud" enough so Bob can understand what she is saying.

Since [11], there have been many research contributions that study modern day covert wireless communication. Bash *et al.* in [1] and [12] performed optical experiments which support the theoretical results found in [11]. The SRL is also valid when an adversary has access to quantum technology which was demonstrated by Bash *et al.* in [13]. Additionally, covert communication over channel models such as Binary Symmetric Channels (BSCs) are considered by Che *et al.* in [14, 15]. More generically, Discrete Memoryless Channels (DMCs) are considered by Wang *et al.* in [16,17]. These works demonstrate that the SRL also holds in BSCs and DMCs.

The models presented in [1,11–13,16,17] assume Alice and Bob pre-share a secret key that is unknown to Willie. The necessary key length to covertly communicate in a generalized channel model is studied further by Bloch in [18]. Results show that

if the Alice-to-Bob channel is better than the Alice-to-Willie channel, then the SRL holds without requiring Alice and Bob to pre-share a secret key. However, if such channel conditions are not satisfied, then Alice can satisfy the SRL with a key of length $\mathcal{O}(\sqrt{n})$.

1.2.1 Increasing the Rate of Covert Communications

All of the references noted thus far have shown that Alice can only transmit $\mathcal{O}(\sqrt{n})$ covert bits in n channel uses for various channel conditions. Therefore, there is great interest to determine how covert communication at a positive rate ($\mathcal{O}(n)$ bits in n channel uses) can be achieved. To achieve positive rate, researchers started assuming that Willie has uncertainty in the system model.

In [19], Bash *et al.* consider if Willie has uncertainty about *when* Alice transmits in $T(n)$ available slots. Per [19], Alice can communicate covertly and reliably $\mathcal{O}(\min\{\sqrt{n\log(T(n))}, n\})$ bits in n channel uses if Alice and Bob pre-arrange a time to communicate. These results show that Alice and Bob can increase their rate of covert and reliable communication; however, the rate is still not positive. The desire to achieve positive rate ($\mathcal{O}(n)$ bits in n channel uses) while also maintaining covert communication motivated future research efforts. Lee and Baxley found that Alice can achieve positive rate if: 1) Willie employs a power detector (i.e. a radiometer) and 2) Willie has uncertainty about the noise power at his receiver [20–22].

Lee and Baxley's work makes a positive and important contribution; however, their results rely on assumptions about Willie's receiver and do not hold in practical scenarios. For example, assume Alice is assigned a single time slot to transmit out of many available time slots and that no other communication is occurring by any parties during the time slots that Alice does not transmit to Bob. Goeckel *et al.* in [23] show that even if Willie has uncertainty about his noise variance, Willie can use the time slots when Alice does not transmit to estimate his noise variance and reduce

his uncertainty. As a result, Alice's covert throughput is limited to the same as when Willie knows his noise variance. This result holds even when Willie does not know the time slot that Alice and Bob agree to communicate in. Therefore, adding uncertainty at Willie is important for achieving positive rate, but, practical scenarios must be considered. Such scenarios are considered in this work by including a jammer in the communication model. Also in this work, covert communication over fading channels is investigated in addition to AWGN channels when a jammer is included in the model.

1.2.2 Considering the Continuous-Time Model

As more complex channels are considered, this work also removes some important assumptions at Willie's receiver. For example, current research on covert wireless communications assumes that Willie employs a discrete-time model to approximate his continuous-time observations. This assumption only holds if Willie knows the correct time instances to sample his continuous-time observations. However, if Willie does not know the correct time instances to sample, then existing research proves that a power detector is not optimal. For example, wireless waveforms often have periodic features and a detector designed to identify specific frequencies allows for waveform specific detectors.

Cyclostationary Detectors (CSDs) are a set of detectors that exploit the periodic features of signals of interest to perform detection. William Gardner performed extensive research on CSDs and the optimal CSDs for various modulation schemes [24–27]. If Willie has limited resources, a power detector is a feasible solution [28]. However, if Willie has knowledge of Alice's waveform structure or the capability to search over different cyclic frequencies, a power detector is not optimal. Before analyzing the covert rate of communications in a continuous-time model, it is first important to understand the optimal detector to detect continuous-time signals.

Carrara and Adams in [29] assume that a power detector is optimal if Alice transmits signals which are continuous and band-limited. In [29], the authors do not state that Alice employs a Gaussian codebook or what specific type of continuous-time signal she transmits. There is also no consideration for how Alice maps her discrete-time symbols to the continuous-time domain. Modeling the mapping is important because wireless signals in continuous-time often exhibit features, which Willie can leverage to design detectors. For example, continuous-time wireless waveforms are often pulse shaped to reduce intersymbol interference (ISI) or modulated using a carrier frequency [30, Chapter 3.3, Chapter 4]. Therefore, when considering continuous-time signals, the detector which best exploits any unique features in Alice's signal should be employed instead of assuming a power detector is the optimal detector. Also, Bash *et al.* in [12] acknowledged the importance of considering continuous-time model in early work. This work investigates the impact of detectors based on the continuous-time model. Both the Alice/Bob/Willie and the Alice/Bob/Willie/Jammer scenarios are re-evaluated using the continuous-time model.

1.3 Contributions

1.3.1 Adding a Jammer to the Communication Model

In this work, the addition of a jammer to the Alice, Bob, Willie model is considered to help facilitate covert communication under various channel conditions. The jammer is uninformed and does not know when or even if Alice transmits in her agreed upon time slot to transmit to Bob. In addition, all parties are synchronized. Results in this work show that the inclusion of the jammer allows Alice to achieve positive rate covert communication for any detector at Willie. In particular, prior work in covert wireless communication assumed that a radiometer (i.e. power detector) is an optimal detector. In contrast, we prove that a radiometer is indeed optimal in AWGN and single block fading models. However, if the Alice-to-Willie channel or the

jammer-to-Willie channel is subject to a finite number of multiple fading coefficients in a single time slot, the radiometer is no longer an optimal detector. But we model the structure of an optimal detector and show that Alice can still achieve positive rate.

1.3.2 Considering Dynamic Channels in the Communication Model

Covert communication proofs presented in Chapter 2 assume the jammer is stationary. However, a moving jammer that transmits with constant power in a single block fading environment also generates variations in the power observed by Willie according to the jammer's movement. Additionally, there are instances when a dynamic environment also causes fading variations. The mathematical formulation of dynamic channels may first appear similar to the finite multiple block fading scenario presented in Chapter 2. However, this portion of the dissertation generalizes the dynamics to include when the fading variations are a function of the total number of symbol slots observed by Willie. Generalizing the number of variations as a function of the codeword length, n, helps provide insight about the rate Alice should employ in order to achieve covert communication. For example, Alice must abide by the SRL in order to remain covert when the jammer's power observed by Willie does not vary at all. However, the work presented in Chapter 2 demonstrates that a fixed number of variations which occur over n allow Alice to communicate covertly. Therefore, modeling the number of fading variations as a function of the codeword length allows for the analysis of covert communications in various wireless environments.

1.3.3 Covert Communication on the Continuous-Time Model

Work presented in Chapter 2, Chapter 3 and research presented in the Background Section are based on a discrete-time communications model. The underlying assumption in prior work is that the discrete-time model is an equivalent approxi-

mation of the continuous-time model. Chapter 4 re-evaluates the Alice/Bob/Willie covert communication scenario employing the continuous-time model instead of the discrete-time model to determine if in fact, the discrete-time model is an "equivalent" discrete-time model. Covert communications in the continuous-time model is re-evaluated by assuming that Willie employs a cyclostationary detector (CSD) instead of a power detector to detect Alice's signal. Results demonstrate that the CSD designed based on the continuous-time model outperforms the power detector. Based on these results, the discrete-time model employed in prior work is not an equivalent representation of the continuous-time model.

The Alice/Bob/Willie/Jammer presented in Chapter 2 is also re-evaluated based on the continuous-time model in Chapter 5. The new model assumes that both Alice and the jammer transmit pulse shaped signals with different timing offsets when their signals arrive at Willie. A detector is then proposed that exploits the timing offset differences to detect if Alice is transmitting. Results again demonstrate that if Willie is unaware of Alice's timing, the standard power detector is not optimal. Instead, there exists a method for Willie estimate the jammer's timing offset and develop a detector that significantly outperforms a power detector. The goal of this work is to help provide insight into how Alice can achieve covert communication by adding uncertainty into Willie's detector.

CHAPTER 2

COVERT COMMUNICATION WITH THE ASSISTANCE OF AN UNINFORMED JAMMER

2.1 Introduction

Much of secure communications centers on preventing an adversary from determining the content of the message. However, there are circumstances when communicating parties Alice and Bob may want *covert* communication: hiding the very existence of their communication from a watchful adversary Willie. Examples include communicating in the presence of an authoritarian government who may want to curtail any organization by certain entities, or military communications where detection might inform an adversary that there is activity in a given geographical area.

As defined precisely below, recent work has studied reliable covert communication, which requires: (i) Willie's error in detecting that Alice transmitted a message to Bob be arbitrarily close to random guessing; and (ii) Bob's error of recovering Alice's message be arbitrarily small. When the Alice-to-Bob and Alice-to-Willie channels are additive white Gaussian noise (AWGN) channels, [11] and [12] showed a *square root law* (SRL): provided Alice and Bob share a secret of sufficient length prior to transmission, Alice can communicate covertly to Bob if and only if she employs a per-symbol power of no more than $\mathcal{O}(1/\sqrt{n})$, which decreases to 0 in the limit of large n. Thus, $\mathcal{O}(\sqrt{n})$ bits (and no more) can be transmitted in n channel uses [12]. Follow-on work has considered the length of the pre-shared secret in [14] and [31], characterization of the constant hidden by Big-$\mathcal{O}$ notation in [31] and [32], and both the theory and experimental verification of covert communication over quantum

channels in [13] and [33]. Additional research by Soltani *et al.* in [34–36] has also considered covert communication over networks.

Subsequent work considered whether positive rate covert communications, which requires the transmission of $O(n)$ bits in n channel uses, is possible. Lee *et al.* in [22] demonstrated that positive rate is indeed achievable over AWGN channels if Willie has uncertainty about the statistics of the background noise and is restricted to a receiver that employs a threshold on the received power when attempting to detect Alice. Che *et al.* in [15] proved that positive rate is achievable if Willie has uncertainty in the parameters of the binary symmetric channel between Alice and himself. In [23], the authors re-visit the results of [15] and [22]. Rather than starting with parametric uncertainty in Willie's knowledge of the noise statistics, [23] allows Willie to have access to a large collection of inputs spanning many possible codeword slots and to employ them in any way that he deems suitable. Then, the lack of knowledge of channel statistics at Willie does not increase the order of the covert throughput from Alice to Bob [23]. This is because Willie is able to use any "quiet" periods to estimate the noise statistics of his receiver accurately and then detect if Alice is transmitting, even if he does not know a priori the time at which Alice might transmit.

In this work, we allow Willie to have a general receiver, as in [23], but we seek conditions under which Alice can transmit with power not decreasing in the block-length n; in the case of an AWGN channel between Alice and Bob, this then achieves the transmission of $O(n)$ bits covertly in n channel uses. To do such, we add another node to the environment, the "jammer", who Willie knows is transmitting. For example, this might be a jammer in an electronic warfare (EW) environment placed by Alice and Bob, or, as discussed in Section 2.5, a jammer placed in the environment by Willie for other security objectives. If this jammer randomly varies his/her transmit power appropriately or if time-varying multipath fading causes sufficient variation, channel estimation during periods outside the time period when Willie is attempting

to detect Alice's transmission cannot be used to estimate the statistics of the noise impacting Willie's receiver during the period of interest. Hence, the results of [23] do not apply; rather, we arrive at a similar mathematical problem to that considered in [22]. A limitation of the achievability results of [22] is that the power detector is not established to be the optimal receiver for Willie; in fact, in the case of block fading channels with multiple fading blocks per codeword, it is known to be sub-optimal. Here, in contrast to [22], we establish covert communication against any detector that Willie might employ.

We consider both additive white Gaussian noise (AWGN) and standard block fading channels. Note that the problem is readily solved if the jammer and Alice are closely coordinated (i.e. , an "informed" jammer) by the following construction. Alice generates a codebook by drawing codeword symbols independently from a Gaussian distribution, and provides this codebook only to Bob as the shared secret. At the time Alice starts to transmit a codeword, the jammer turns down the power of his transmission of Gaussian noise, and then he turns it back up at the moment Alice finishes transmitting. Willie is then unable to determine that any change has taken place when Alice is transmitting. We are interested in the case where the jammer and Alice do not coordinate. In the AWGN case, our construction has the jammer randomly change his/her power of the Gaussian noise in each "slot" of n symbols, where n is the codeword length used by Alice. By doing such, Willie is unaware of the background noise to expect and it is plausible, particularly based on the work of [22], that Alice should be able to achieve positive rate covert communication to Bob. To establish this result rigorously against an arbitrary receiver at Willie, we first establish that Willie's optimal receiver is indeed a comparison of the received power to a threshold, from which the achievability of positive rate covert communication follows.

We then consider a block fading channel with M fading blocks per codeword of length n. If $M = 1$, we demonstrate that a threshold test on the total received power in the codeword slot is the optimal detector at Willie, from which covert transmission by Alice with power not decreasing in the blocklength n follows. When $M > 1$, a threshold test on the total received power at Willie is sub-optimal. Thus, we first establish a technical property on the structure of Willie's optimal detector and then show that this property suffices to establish the ultimate goal when the jammer-to-Willie channel is an $M > 1$ block fading channel: Alice can covertly transmit with a power that does not decrease with her blocklength n.

The main contributions o this chapter are:

1. The consideration of covert communication in the presence of an uninformed jammer.

2. The demonstration of the optimality of a power detector at Willie for the AWGN and $M = 1$ block fading cases, from which the ability of Alice to transmit covertly with a power that does not decrease with her blocklength follows.

3. The demonstration of the ability for Alice to transmit covertly with a power that does not decrease with her blocklength in the $M > 1$ block fading scenario, even when Willie uses an optimal detector (which is not a power detector in this case).

2.2 System Model and Metrics

2.2.1 System Model

Consider a scenario where Alice ("a") would like to communicate covertly to Bob ("b") without detection by a warden Willie ("w"), and suppose a jammer ("j") is active in the environment who is willing to assist with this communication. The geographic model is shown in Figure 2.1. The distances from Alice to Willie and Alice

to Bob are denoted by $d_{\mathrm{a,w}}$ and $d_{\mathrm{a,b}}$ respectively. The distances from the jammer to Willie and the jammer to Bob are $d_{\mathrm{j,w}}$ and $d_{\mathrm{j,b}}$ respectively.

This work considers Alice's ability to transmit covertly in a slot equal to the codeword length n and Willie's ability to detect such a transmission in that slot. For integer constant $T > 0$, consider a discrete-time channel with T slots, each of length n symbols, as shown in Figure 2.2, with the nT symbols indexed by $k = -\frac{T}{2}n+1, -\frac{T}{2}n+2 \ldots, -2, -1, 0, 1, 2, \ldots, \frac{T}{2}n-1, \frac{T}{2}n$. Assume that the slot of interest is slot $t = 0$; hence, Alice may (or may not) transmit for a duration of n symbols starting at time $k = 1$, and Willie's goal is to detect whether or not such a transmission took place using observations for all $k = -\frac{T}{2}n+1, -\frac{T}{2}n+2 \ldots, -2, -1, 0, 1, 2, \ldots, \frac{T}{2}n-1, \frac{T}{2}n$, since observations outside of $k = 1, 2, \ldots, n$ might be useful to Willie in estimating aspects of the environment [23]. The jammer is "uninformed" in the sense that it does not know if Alice transmits, and if Alice transmits, the jammer does not know that Alice is going to use a slot starting at time $k = 1$.

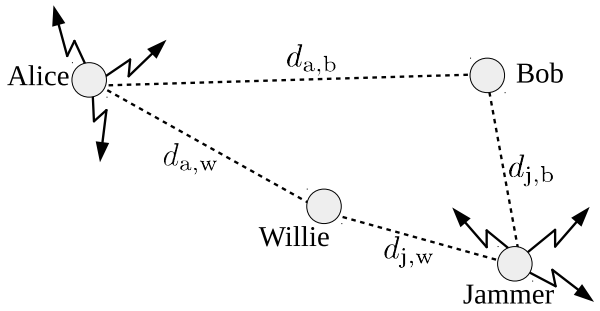

Figure 2.1. Alice, Bob, Willie and Jammer Wireless communication scenario. With the help of a jammer, Alice attempts to transmit covertly to Bob in the presence of a watchful adversary Willie.

Alice transmits a message with probability p and if she decides to transmit, she maps her message to the complex symbol sequence $\mathbf{f} = [f_1,\ f_2,\ \ldots\ , f_n]$ and sends it in the $t = 0$ slot corresponding to symbols $k = 1, 2, \ldots, n$. The jammer is allowed to transmit continuously (in all symbols of all slots) subject only to an average power

limitation of $P_{\max}$ per symbol. Let the (complex) signal transmitted by the jammer for all time slots be given by $\{\mathbf{g}_t\}_{t=-\frac{T}{2}}^{\frac{T}{2}-1}$, where $\mathbf{g}_t = [g_{tn+1}, g_{tn+2}, \ldots, g_{tn+n}]$ is the vector of transmitted jamming signals sent during the t^{th} slot, with the per symbol power constraint $E[|g_k|^2] \leq P_{\max}$.

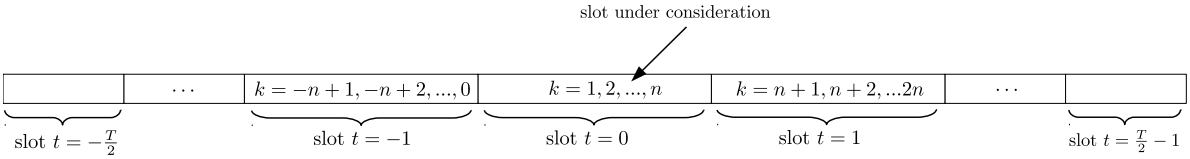

Figure 2.2. Slot model diagram of Willie's observations in AWGN channel conditions. Representation of the indexing of nT symbol periods in T slots, each of length n. Alice decides to transmit in slot $t = 0$ with probability p, and Willie attempts to detect a transmission in that slot.

2.2.1.1 AWGN channel

Consider first the AWGN channel. Denote the collection of channel outputs at Willie over all time slots as: $\{\mathbf{Z}_t\}_{t=-\frac{T}{2}}^{\frac{T}{2}-1}$, where $\mathbf{Z}_t = [Z_{tn+1}, Z_{tn+2}, \ldots, Z_{tn+n}]$ is the vector of observations collected during the t^{th} slot. Hence, for slot t, $i = 1, 2, \ldots, n$:

$$Z_{tn+i} = \begin{cases} \frac{f_i}{d_{\text{a,w}}^{\alpha/2}} + \frac{g_{tn+i}}{d_{\text{j,w}}^{\alpha/2}} + N_{tn+i}^{(\text{w})}, & \text{Alice transmits and } \ t = 0 \\ \frac{g_{tn+i}}{d_{\text{j,w}}^{\alpha/2}} + N_{tn+i}^{(\text{w})}, & \text{else,} \end{cases} \tag{2.1}$$

where α is the path-loss exponent, and

$$\left\{ N_k^{(\text{w})}, k = -\frac{T}{2}n + 1, -\frac{T}{2}n + 2 \ldots, -2, -1, 0, 1, 2, \ldots, \frac{T}{2}n - 1, \frac{T}{2}n \right\} \tag{2.2}$$

is a set of independent and identically distributed (i.i.d.) zero-mean complex Gaussian random variables, each with variance $E[|N_k^{(\text{w})}|^2] = \sigma_{\text{w}}^2$.

Similarly, denote the collection of channel outputs at Bob over all time slots as: $\{\mathbf{Y}_t\}_{t=-\frac{T}{2}}^{\frac{T}{2}-1}$, where $\mathbf{Y}_t = [Y_{tn+1}, Y_{tn+2}, \ldots, Y_{tn+n}]$ is the vector of observations collected during the t^{th} slot. Hence, for slot t, $i = 1, 2, \ldots, n$:

$$Y_{tn+i} = \begin{cases} \frac{f_i}{d_{\text{a,b}}^{\alpha/2}} + \frac{g_{tn+i}}{d_{\text{j,b}}^{\alpha/2}} + N_{tn+i}^{(\text{b})}, & \text{Alice transmits and } \ t = 0 \\ \frac{g_{tn+i}}{d_{\text{j,b}}^{\alpha/2}} + N_{tn+i}^{(\text{b})}, & \text{else,} \end{cases} \tag{2.3}$$

where

$$\left\{ N_k^{(\text{b})}, k = -\frac{T}{2}n + 1, -\frac{T}{2}n + 2 \ldots, -2, -1, 0, 1, 2, \ldots, \frac{T}{2}n - 1, \frac{T}{2}n \right\} \tag{2.4}$$

is a set of i.i.d. zero-mean complex Gaussian random variables, each with variance $E[|N_k^{(\text{b})}|^2] = \sigma_{\text{b}}^2$.

2.2.1.2 Block fading channel

Consider next the standard Rayleigh block fading channel, as shown in Figure 2.3. The fading is constant for a block of n/M symbols but changes independently to a different value for the next block, where M is the number of fading blocks per codeword slot [37]. Denote $h_{t,m}^{(x,y)}$, $m = 1, \ldots, M$ as the (complex) fading coefficient for the m^{th} block during slot t between transmitter x and receiver y, where x is either "a" (Alice) or "j" (jammer), and y is either "w" (Willie) or "b" (Bob). By the Rayleigh fading assumption, $h_{t,m}^{(x,y)}$, $m = 1, \ldots, M$ is assumed to be a zero mean complex Gaussian random variable with $E[|h_{t,m}^{(x,y)}|^2] = 1$ for all channels. The fading processes affecting different transmitter-receiver pairs are assumed to be independent of each other. For slot t, $i = 1, 2, \ldots, n$, Willie observes:

$$Z_{tn+i} = \begin{cases} \frac{h_{t,\lfloor (i-1)\frac{M}{n} \rfloor + 1}^{(\text{a,w})} f_i}{d_{\text{a,w}}^{\alpha/2}} + \frac{h_{t,\lfloor (i-1)\frac{M}{n} \rfloor + 1}^{(\text{j,w})} g_{tn+i}}{d_{\text{j,w}}^{\alpha/2}} + N_{tn+i}^{(\text{w})}, & \text{Alice transmits and } \ t = 0 \\ \frac{h_{t,\lfloor (i-1)\frac{M}{n} \rfloor + 1}^{(\text{j,w})} g_{tn+i}}{d_{\text{j,w}}^{\alpha/2}} + N_{tn+i}^{(\text{w})}, & \text{else.} \end{cases} \tag{2.5}$$

For slot t, $i = 1, 2, \ldots, n$, Bob observes:

$$Y_{tn+i} = \begin{cases} \frac{h^{(\mathrm{a,b})}_{t,\lfloor (i-1)\frac{M}{n} \rfloor + 1} f_i}{d^{\alpha/2}_{\mathrm{a,b}}} + \frac{h^{(\mathrm{j,b})}_{t,\lfloor (i-1)\frac{M}{n} \rfloor + 1} g_{tn+i}}{d^{\alpha/2}_{\mathrm{j,b}}} + N^{(\mathrm{b})}_{tn+i}, & \text{Alice transmits and } \ t = 0 \\ \frac{h^{(\mathrm{j,b})}_{t,\lfloor (i-1)\frac{M}{n} \rfloor + 1} g_{tn+i}}{d^{\alpha/2}_{\mathrm{j,b}}} + N^{(\mathrm{b})}_{tn+i}, & \text{else.} \end{cases} \tag{2.6}$$

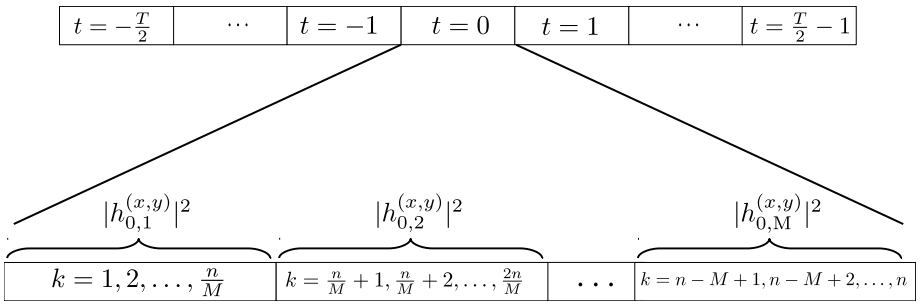

Figure 2.3. Slot model diagram of Willie's observations in multiple-block fading channel conditions where x is either Alice or the jammer and y is either Willie or Bob.

2.2.2 Metrics

Based on his observations over all time slots, Willie must determine whether Alice transmitted in time slot $t = 0$. The null hypothesis (H_0) is that Alice did not transmit and the alternative hypothesis (H_1) is that that Alice transmitted a message. Define $P(H_0) = 1 - p$ as the probability that Alice does not transmit and $P(H_1) = p$ as the probability that Alice transmits in time slot $t = 0$, where the assumption is that p is known to Willie (pessimistically) . Willie seeks to minimize his probability of error $\mathbb{P}_e = (1-p)\cdot\mathbb{P}_{\mathrm{FA}} + p\cdot\mathbb{P}_{\mathrm{MD}}$, where $\mathbb{P}_{\mathrm{MD}}$ and $\mathbb{P}_{\mathrm{FA}}$ are the probabilities of missed detection and false alarm at Willie, respectively. Per [38], $\mathbb{P}_e \geq \min(p, 1-p) \cdot (\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}})$. Hence, Alice achieves *covert* communication if, for any $\epsilon > 0$, $\mathbb{P}_{\mathrm{MD}} + \mathbb{P}_{\mathrm{FA}} > 1 - \epsilon$ for

n sufficiently large.[1] Alice can transmit covertly with power not decreasing in n if, for any $\epsilon > 0$, there exists $P_{\mathrm{f}} > 0$ not dependent on n (but possibly dependent on ϵ) such that, as $n \to \infty$, a system employing power P_{f} is covert. Bob should also be capable of *reliably* decoding Alice's message [12]. Bob can reliably decode messages from Alice if, for any $\delta > 0$, his probability of error is less than δ for n sufficiently large.

Assume that Willie has full knowledge of the statistical model: the parameters for Alice's random codebook generation and the jammer's random interference generation, the noise variance σ_{w}^2, and in the case of fading on the Alice-to-Willie channel or jammer-to-Willie link, the statistics of that fading. Thus, Willie's test is between two simple hypotheses for Alice's transmission state, and he has complete statistical knowledge of his observations when either hypothesis is true. Therefore, by applying the Neyman-Pearson (NP) criterion, the optimal test for Willie to minimize his probability of error is the *likelihood ratio test (LRT)* [39, Chapter 3.3],

$$\Lambda(\tilde{\mathbf{Z}}) = \frac{f_{\tilde{\mathbf{Z}}|H_1}(\tilde{\mathbf{Z}}|H_1)}{f_{\tilde{\mathbf{Z}}|H_0}(\tilde{\mathbf{Z}}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \tag{2.7}$$

where $\gamma = P(H_0)/P(H_1)$, and $f_{\tilde{\mathbf{Z}}|H_1}(\cdot|H_1)$ and $f_{\tilde{\mathbf{Z}}|H_0}(\cdot|H_0)$ are the probability density functions (pdfs) for Willie's observations over all slots given Alice transmitted in the $t = 0$ slot or given Alice did not transmit in the $t = 0$ slot, respectively. As can be inferred by the assumption of a power detector for Willie's receiver in [22] and made precise in the proof of Theorem 1 below, a desirable property for the likelihood ratio $\Lambda(\cdot)$ to exhibit is *monotonicity*. In the remainder of this section, the approach for establishing such a property that applies in our context is described.

---

[1] This guarantees that Willie's probability of error is within ϵ of the probability of error $\min(p, 1-p)$ obtained if he ignores his observations and chooses the hypothesis H_0 and H_1 that was most likely a priori.

The concept of *stochastic ordering* [40] is employed to derive the desired monotonicity results in a more streamlined fashion relative to our preliminary work in [41]. A random variable X is smaller than W in the likelihood ratio order (written as $X \leq_{\text{lr}} W$) when $f_W(x)/f_X(x)$ is non-decreasing over the union of their supports, where $f_W(x)$ and $f_X(x)$ are their respective probability density functions. Consider a family of pdfs $\{g_\theta(\cdot), \theta \in \mathcal{X}\}$ where $\mathcal{X}$ is a subset of the real line. Let $X(\theta)$ denote a random variable with density $g_\theta(\cdot)$ for fixed parameter θ. Let Θ denote a random variable with support $\mathcal{X}$ and probability distribution function $F_\Theta(\cdot)$; denote $X(\Theta)$ as the random variable that is the mixture of the random variables $X(\theta)$ under distribution $F_\Theta(\theta)$; that is, the probability density function of $X(\Theta)$ is given by:

$$f_{X(\Theta)}(x) = \int_{\theta \in \mathcal{X}} g_\theta(x) dF(\theta), \quad x \in \mathbb{R}. \tag{2.8}$$

The following result regarding mixtures of random variables is employed to prove the power detector is optimal for AWGN and single-block fading channel models.

Lemma 1. *[Theorem 1.C.11 in [40]] Consider a family of probability density functions $\{g_\theta(\cdot), \theta \in \mathcal{X}\}$ with $\mathcal{X}$ a subset of the real line. Let Θ_0 and Θ_1 denote random variables with support in $\mathcal{X}$ and probability distribution functions $F_0(\theta)$ and $F_1(\theta)$, respectively. Let W_0 and W_1 be random variables such that $W_i =_{\text{d}} X(\Theta_i)$, $i = 0, 1$, (where $=_{\text{d}}$ is defined as equality in distribution or law):*

$$f_{W_i}(x) = \int_{\theta \in \mathcal{X}} g_\theta(x) dF_i(\theta), \quad i = 0, 1; x \in \mathbb{R}. \tag{2.9}$$

If

$$X(\theta) \leq_{\text{lr}} X(\theta'), \quad \theta \leq \theta' \tag{2.10}$$

and

$$\Theta_0 \leq_{\text{lr}} \Theta_1, \tag{2.11}$$

then

$$W_0 \leq_{\text{lr}} W_1. \tag{2.12}$$

2.3 Adding Uncertainty at Willie with a Jammer

This subsection demonstrates that Alice and Bob can achieve covert communication with positive rate when a jammer is added to the communication model by generating uncertainty at Willie's receiver. Even if Willie employs a radiometer, which is shown to be an optimal detector in AWGN and single block fading scenarios, Alice and Bob can still achieve covert and reliable communication. In the multiple fading block scenarios, the radiometer is no longer an optimal test statistic, however covert communication at a positive rate can still be achieved if Willie employs an optimal detector.

2.3.1 Achievability for the AWGN Model

Consider the case of additive white Gaussian noise (AWGN) channels between all nodes, with the slot boundaries between Alice, Willie, and the jammer synchronized, and, as in [12], assume that Alice and Bob share a secret of unlimited length. A construction for Alice and the jammer is provided, and then a power detector is shown to be Willie's optimal detector based on the construction. The transmission of $\mathcal{O}(n)$ bits in n channel uses is then demonstrated. It is assumed that $d_{\text{a,w}}$ and $d_{\text{j,w}}$ are known to Alice, although it is readily apparent that a lower-bound to $d_{\text{a,w}}$ and an upper-bound to $d_{\text{j,w}}$ are sufficient to establish the results.

Construction: Random coding arguments are employed to generate K codewords, each of length n, by independently drawing symbols from a zero-mean complex Gaussian distribution with variance P_{f}, where P_{f} is determined later. This codebook is revealed to Alice and Bob, is used only once, and comprises the shared secret unknown to Willie (and the jammer). If Alice decides to transmit in slot $t = 0$, she selects the codeword corresponding to her message, sets f_i to the i^{th} symbol of that codeword, and transmits the sequence $f_1, f_2, \ldots, f_n$. The jammer, with knowledge of the slot boundaries but without knowledge of whether Alice transmits in a given slot (or at all), transmits a symbol drawn independently from a zero-mean complex Gaussian distribution during each symbol period. However, the variance of this Gaussian distribution is not constant; in particular, during the t^{th} slot, the jammer draws each of its symbols independently from a zero-mean Gaussian distribution with variance $E[|g_{tn+i}|^2] = P_t^{(\mathrm{j})}$, $i = 1, 2, \ldots, n$, with $P_t^{(\mathrm{j})}$ changing between slots. The sequence of variances employed across the slots, $P_t^{(\mathrm{j})}$, $t = -\frac{T}{2}, -\frac{T}{2}+1, \ldots, -1, 0, 1, \ldots, \frac{T}{2}-2, \frac{T}{2}-1$ is an i.i.d. sequence of uniform random variables on $[0, P_{\max}]$, where $P_{\max}$, as defined in Section 2.2.1, is the maximum average power per symbol that the jammer can employ.

Per above, Alice's codebook is only shared with Bob and thus is unknown to Willie. However, Willie knows everything else about how the system is constructed, including the length of the codeword n, the distribution from which the codeword symbols are drawn (including P_{f}), the distribution of the jamming power (including $P_{\max}$), the time of Alice's potential transmission, and his distances from Alice and the jammer. Next, the power detector is established as Willie's optimal strategy for detecting Alice's transmission.

Lemma 2. *Under assumptions of the AWGN model in Section 2.2.1.1, Willie's optimal detector compares the total received power in slot $t = 0$ to a threshold.*

Proof: Consider Willie's attempt to detect Alice during the slot $t = 0$ of interest. Since the jammer's power outside of this slot is independent of the jammer's power within the slot and since Willie knows σ_{w}^2, it is sufficient for Willie to consider the vector of observations $\mathbf{Z}_0$ only within slot $t = 0$, as defined in Section 2.2.1. Hence, to simplify notation, the slot index is dropped and we denote the input to Willie's receiver as $\mathbf{Z} = [Z_1, Z_2, \ldots, Z_n]$.

Given the assumptions of the lemma, the distribution of $\mathbf{Z}$ is complex Gaussian. Under H_0, Willie observes only the jamming signal in addition to background noise. Under H_1, Willie observes both the jamming signal and Alice's transmission in addition to background noise. Let θ denote the variance of the power observed due to Alice's transmissions and the jammer's signal and thus define $\mathbf{Z}(\theta) = [Z_1(\theta), Z_2(\theta), \ldots, Z_n(\theta)]$, where $Z_i(\theta) \sim \mathcal{CN}(0, \sigma_{\mathrm{w}}^2 + \theta)$. Thus, H_0 and H_1 are distinguished by introducing two non-negative valued random variables Θ_0 and Θ_1 with probability density functions:

$$f_{\Theta_\rho}(\theta) = \begin{cases} 1/\zeta, & 0 < \theta \leq P_{\max}/d_{\mathrm{j,w}}^{\alpha}, \rho = 0 \\ 1/\zeta, & \sigma_{\mathrm{a}}^2 < \theta \leq \sigma_{\mathrm{a}}^2 + P_{\max}/d_{\mathrm{j,w}}^{\alpha}, \rho = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{2.13}$$

where $\zeta = P_{\max}/d_{\mathrm{j,w}}^{\alpha}$ and $\sigma_{\mathrm{a}}^2 = P_{\mathrm{f}}/d_{\mathrm{a,w}}^{\alpha}$. The pdf of Willie's observations conditioned on θ is:

$$\begin{aligned} f_{\mathbf{Z}(\theta)}(\mathbf{z}) &= \prod_{i=1}^{n} \frac{1}{\pi(\sigma_{\mathrm{w}}^2 + \theta)} \exp\left(-\frac{|z_i|^2}{(\sigma_{\mathrm{w}}^2 + \theta)}\right) \\ &= \left(\frac{1}{\pi(\sigma_{\mathrm{w}}^2 + \theta)}\right)^n \exp\left(-\frac{z}{(\sigma_{\mathrm{w}}^2 + \theta)}\right), \end{aligned} \tag{2.14}$$

where $z = \sum_{i=1}^{n} |z_i|^2$. Thus, by the Neyman-Fisher Factorization Theorem, the total power $Z(\theta) = \sum_{i=1}^{n} |Z_i(\theta)|^2$ is a sufficient statistic for Willie's test [42, Chapter 5.4].

Let χ_l^2 denote a chi-squared random variable with l degrees of freedom. Then $Z(\theta) = (\sigma_{\text{w}}^2 + \theta)\chi_{2n}^2$. Since Willie does not know either Θ_0 or Θ_1, his LRT becomes:

$$\Lambda(Z) = \frac{E_{\Theta_1}[f_{Z(\theta)}(Z)]}{E_{\Theta_0}[f_{Z(\theta)}(Z)]} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma.$$

The next steps show that $\Lambda(\cdot)$ is monotone. From the definition of a chi-squared random variable, $Z(\theta) \leq_{\text{lr}} Z(\theta')$ whenever $\theta \leq \theta'$. In addition, applying the definition of $\leq_{\text{lr}}$ to the densities of Θ_0, Θ_1 yields that $\Theta_0 \leq_{\text{lr}} \Theta_1$. The application of Lemma 1 then yields that $\Lambda(\cdot)$ is non-decreasing in z. Thus, the LRT is equivalent to the test:

$$Z \underset{H_0}{\overset{H_1}{\gtrless}} \Gamma_n$$

corresponding to a threshold test on the total received power. ∎

Theorem 1. *Under the assumptions of the AWGN model in Section 2.2.1.1, there exists a communication strategy for Alice, Bob, and the jammer whereby Alice transmits $\mathcal{O}(n)$ bits in n channel uses reliably and covertly to Bob in the presence of Willie.*

Proof: *Construction:* Alice and the jammer employ the construction given at the beginning of Section 2.2.1. Per Lemma 2, the optimal detector for Willie is to employ a threshold test $Z \gtrless_{H_0}^{H_1} \Gamma_n$ on the total received power. Dividing both sides by n yields the equivalent test:

$$\frac{Z}{n} \underset{H_0}{\overset{H_1}{\gtrless}} \tau_n, \tag{2.15}$$

where $\tau_n \equiv \Gamma_n/n$. Whereas there is an optimal τ_n for any finite n, this work establishes for any sequence of τ_n that Willie chooses, the detector is asymptotically useless as $n \to \infty$; that is, for any $\epsilon > 0$, there exists a construction such that $\mathbb{P}_{\text{FA}} + \mathbb{P}_{\text{MD}} > 1 - \epsilon$ for sufficiently large n.

Analysis: Note that $\sigma_{\mathrm{j}}^2 = U\zeta$, where U is a uniform random variable on $[0, 1]$. Recall that Willie does not know the value of U. Let $\mathbb{P}_{\mathrm{FA}}(u)$ and $\mathbb{P}_{\mathrm{MD}}(u)$ be Willie's probability of false alarm and probability of missed detection conditioned on $U = u$, respectively. Then,

$$\mathbb{P}_{\mathrm{FA}}(u) = P\left(\frac{Z}{n} \geq \tau_n | U = u, H_0\right). \tag{2.16}$$

Recall that χ_l^2 denotes a chi-squared random variable with l degrees of freedom. Under H_0 and given $U = u$, $Z = (\sigma_{\mathrm{w}}^2 + u\zeta)\chi_{2n}^2$ and $Z/n = (\sigma_{\mathrm{w}}^2 + u\zeta)\chi_{2n}^2/n$. By the weak law of large numbers, χ_{2n}^2/n converges in probability to 1; hence, for any $\delta > 0$, $\exists N_0$ (not dependent on u) such that, for $n \geq N_0$,

$$P\left(\frac{\chi_{2n}^2}{n} \in \left(1 - \frac{\delta}{\sigma_{\mathrm{w}}^2 + \zeta}, 1 + \frac{\delta}{\sigma_{\mathrm{w}}^2 + \zeta}\right)\right) > 1 - \frac{\epsilon}{2}. \tag{2.17}$$

Hence, for any $n > N_0$,

$$P\left(\frac{Z}{n} \in \left((\sigma_{\mathrm{w}}^2 + u\zeta)\left(1 - \frac{\delta}{\sigma_{\mathrm{w}}^2 + \zeta}\right), (\sigma_{\mathrm{w}}^2 + u\zeta)\left(1 + \frac{\delta}{\sigma_{\mathrm{w}}^2 + \zeta}\right)\right)\right) > 1 - \frac{\epsilon}{2}. \tag{2.18}$$

Since $u \leq 1$, $\sigma_{\mathrm{w}}^2 + u\zeta < \sigma_{\mathrm{w}}^2 + \zeta$ and thus,

$$P\left(\frac{Z}{n} \in \left(\sigma_{\mathrm{w}}^2 + u\zeta - \delta, \sigma_{\mathrm{w}}^2 + u\zeta + \delta\right)\right) > 1 - \frac{\epsilon}{2}. \tag{2.19}$$

Therefore, $\mathbb{P}_{\mathrm{FA}}(u) \geq 1 - \epsilon/2$ for any $\tau_n < \sigma_{\mathrm{w}}^2 + u\zeta - \delta$. Likewise, following analogous arguments, there exists N_1 such that, for any $n > N_1$ (not dependent on u):

$$\mathbb{P}_{\mathrm{MD}}(u) = P\left(\frac{Z}{n} \leq \tau_n | U = u, H_1\right) > 1 - \frac{\epsilon}{2} \tag{2.20}$$

for any $\tau_n > \sigma_{\mathrm{w}}^2 + u\zeta + \sigma_{\mathrm{a}}^2 + \delta$. Define the set $\mathcal{A} = \{u : \sigma_{\mathrm{w}}^2 + u\zeta - \delta < \tau_n < \sigma_{\mathrm{w}}^2 + u\zeta + \sigma_{\mathrm{a}}^2 + \delta\}$. This work establishes that, for any $u \in \mathcal{A}^c$ and any $n > \max(N_0, N_1)$, $\mathbb{P}_{\mathrm{FA}}(u) + \mathbb{P}_{\mathrm{MD}}(u) \geq 1 - \frac{\epsilon}{2}$. The probability of event $\mathcal{A}$ is bounded as:

$$\begin{aligned} P(\mathcal{A}) &= P\left(\frac{\tau_n - \sigma_{\mathrm{w}}^2 - \sigma_{\mathrm{a}}^2 - \delta}{\zeta} \leq U \leq \frac{\tau_n - \sigma_{\mathrm{w}}^2 + \delta}{\zeta}\right) \\ &\leq \frac{\sigma_{\mathrm{a}}^2 + 2\delta}{\zeta}. \end{aligned} \tag{2.21}$$

Hence, choosing $\delta = \zeta\epsilon/8$ and $\sigma_{\mathrm{a}}^2 = \zeta\epsilon/4$ yields:

$$P(\mathcal{A}^c) \geq 1 - \frac{\epsilon}{2}. \tag{2.22}$$

Therefore, the summation of Willie's false alarm and missed detection is lower-bounded as:

$$\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} = E_U\left[\mathbb{P}_{\mathrm{FA}}(U) + \mathbb{P}_{\mathrm{MD}}(U)\right] \tag{2.23}$$

$$\geq E_U\left[\mathbb{P}_{\mathrm{FA}}(U) + \mathbb{P}_{\mathrm{MD}}(U) | \mathcal{A}^c\right] P(\mathcal{A}^c) \tag{2.24}$$

$$> 1 - \epsilon. \tag{2.25}$$

Hence, Alice can employ codebooks with power $P_{\mathrm{f}} = \sigma_{\mathrm{a}}^2 d_{\mathrm{a,w}}^{\alpha}$ and remain covert from Willie. Recognizing that the maximum interference caused by the jammer at Bob can be upper-bounded and hence the received signal-to-noise ratio at Bob can be lower-bounded by a constant, Alice can transmit $\mathcal{O}(n)$ bits in n channel uses covertly and reliably to Bob. ∎

2.3.2 Achievability for the Single Block Fading Model ($M = 1$)

Recall that there are four channels in the problem formulation: Alice-to-Bob, Alice-to-Willie, jammer-to-Bob, and jammer-to-Willie. In this section, the channel

model is expanded to consider the situation where one or more of the four channels is a fading channel. As in Section 2.3.1, the problem is investigated by first characterizing how the Alice-to-Willie and jammer-to-Willie channels constrain (or not) the allowable scheme at Alice, in particular the power that she is able to employ while remaining covert. The achievable performance under various metrics when Alice employs that power then follows classical information and communication theory based on the nature of the Alice-to-Bob and jammer-to-Bob channels.

Consider first the case where the Alice-to-Willie channel is an AWGN channel and the jammer-to-Willie channel is a $M = 1$ block fading channel. From an application perspective, this appears at first to be a pessimistic case: the jammer who Alice is counting on to confuse Willie is subject to fading, whereas Willie has a strong direct path from Alice that makes the Alice-to-Willie channel comparatively benign (AWGN). As in the case when all of the channels are AWGN, this work first demonstrate that the optimal receiver at Willie is a power detector. Unlike in Section 2.3.1, here the jammer can transmit Gaussian noise drawn from a distribution with constant variance $P_{\mathrm{j}} = P_{\max}$, since the channel randomizes the power received at Willie from the jammer.

Lemma 3. *Under the assumptions of the $M = 1$ block fading model in Section 2.2.1.2 and Alice's construction presented in Section 2.2.1 but with the jammer transmitting Gaussian noise drawn from a distribution with constant variance, Willie's optimal detector for detecting Alice's transmission is to compare the total received power in the slot of interest to a threshold.*

Proof: Let $\zeta = P_{\mathrm{j}}/d_{\mathrm{j,w}}^{\alpha}$. The received jammer power σ_{j}^2 is exponentially distributed with mean ζ. As in Section 2.3.1, note that observations outside of $k = 1, 2, \ldots, n$ do not help Willie to detect a transmission by Alice in slot $t = 0$; hence, it is sufficient to consider $\mathbf{Z}_0$ as the input to Willie's receiver. Therefore, the slot index is suppressed and denote Willie's observation conditioned on θ by $\mathbf{Z}(\theta) = [Z_1(\theta), Z_2(\theta), \ldots, Z_n(\theta)]$

where $Z_i(\theta) \sim \mathcal{CN}(0, \sigma_{\mathrm{w}}^2 + \theta)$. H_0 and H_1 are distinguished by introducing two non-negative valued random variables Θ_0 and Θ_1 with probability density functions:

$$f_{\Theta_\rho}(\theta) = \begin{cases} \frac{1}{\zeta} e^{-\theta/\zeta}, & 0 < \theta, \rho = 0, \\ \frac{1}{\zeta} e^{-(\theta - \sigma_{\mathrm{a}}^2)/\zeta}, & \sigma_{\mathrm{a}}^2 < \theta, \rho = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.26}$$

Thus, $\Theta_0 \leq_{\mathrm{lr}} \Theta_1$ based on the assumptions presented in Section 2.2.2. The distribution of Willie's observations conditioned on θ is:

$$f_{\mathbf{Z}(\theta)}(\mathbf{z}) = \left(\frac{1}{\pi(\sigma_{\mathrm{w}}^2 + \theta)} \right)^n \exp\Big(- \frac{z}{\sigma_{\mathrm{w}}^2 + \theta} \Big), \tag{2.27}$$

where z is as defined in Section 2.2.1.1. Hence, the LRT test is optimal based on the NP rule and the optimal decision rule for Willie again becomes:

$$\Lambda(Z) = \frac{E_{\Theta_1}[f_{Z(\theta)}(Z)]}{E_{\Theta_0}[f_{Z(\theta)}(Z)]} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{2.28}$$

The monotonicity of $\Lambda(\cdot)$ then follows from Lemma 1 by observing that, as in the proof of Lemma 2, $Z(\theta) \leq_{\mathrm{lr}} Z(\theta')$ whenever $\theta \leq \theta'$, and, as noted above, $\Theta_0 \leq_{\mathrm{lr}} \Theta_1$. Thus, the LRT is equivalent to the power detector: $Z \gtrless_{H_0}^{H_1} \Gamma_n$. ∎

Next, consider the case when the Alice-to-Willie channel is also a $M = 1$ block fading channel. In practice, Willie does not know the value of the fading coefficient $h_{0,1}^{(\mathrm{a,w})}$ on this channel and, indeed, that is our assumption in our achievability result below. However, since the achievability result for covert communication from Alice to Bob is the main point of interest, giving Willie any extra knowledge (say, by a genie) only strengthens the result. Therefore, this work assumes Willie knows $h_{0,1}^{(\mathrm{a,w})}$ and thus Corollary 3.1 is employed to establish Theorem 2.

Corollary 3.1. *Consider the assumptions of the single block fading model in Section 2.2.1.2 and assume that Willie knows the value of* $h_{0,1}^{(\mathrm{a,w})}$. *Then, given Alice's construction in Section 2.2.1.1 but with the jammer transmitting Gaussian noise drawn from a distribution with constant variance, Willie's optimal detector for detecting a transmission by Alice is to compare the total received power in the slot of interest to a threshold.*

Proof: Knowing $h_{0,1}^{(\mathrm{a,w})}$ and $d_{\mathrm{a,w}}$, Willie knows σ_{a}^2, and the proof follows from Lemma 3. ∎

Theorem 2. *Under the assumptions of the single block fading model in Section 2.2.1.2, there exists a communication strategy for Alice, Bob, and the jammer whereby Alice transmits with a power that does not decrease with the blocklength while remaining covert from warden Willie.*

Proof: This proof follows along the lines of Theorem 1 and is provided in Appendix A. ∎

2.3.3 The Number of Covert Bits Transmitted Reliably

Theorem 2 establishes that Alice can transmit with power not decreasing in the blocklength n while maintaining covertness. In the case of AWGN channels on both the Alice-to-Bob and jammer-to-Bob channels, the covert and reliable communication of $\mathcal{O}(n)$ bits in n channel uses can be achieved. However, when the Alice-to-Bob or jammer-to-Bob channels are M-block fading channels, $M \geq 1$, the problem is analogous to the standard problem of communication over slowly fading channels [37, Section 5.4]. Strictly speaking, reliable communication as defined in Section 2.2.2 of $\mathcal{O}(n)$ bits is not possible. In particular, if Alice transmits nR_0 bits for any given constant $R_0 > 0$, there always exists some nonzero probability, not diminishing in n, that the instantiations of $|h_{0,m}^{(\mathrm{a,b})}|$ and $|h_{0,m}^{(\mathrm{j,b})}|$, $m = 1, 2, \ldots M$, leads to a received signal-to-interference-plus-noise ratio (SINR) such that the communication is not reliable.

However, the presence of the jammer, which allows Alice to transmit at per-symbol power $P_{\mathrm{f}} > 0$ not dependent on n (versus $\mathcal{O}(\frac{1}{\sqrt{n}})$ power per symbol when there is no jammer [12]), greatly improves system performance even in the case when the Alice-to-Bob or jammer-to-Bob channels are M-block fading channels. This can be seen via multiple metrics. First, if the metric of Section 2.2.2 is still of pertinent interest, covert and reliable communication of $o(n)$ bits is possible, as demonstrated for $M = 1$ in Appendix B. Second, and probably of more interest, is that the analog of the ϵ-outage capacity (see [37]) is non-zero, whereas it would be zero for any transmission power at Alice that decreases to 0 as $n \to \infty$.

2.3.4 Achievability Proofs for $M > 1$ Block Fading Channel Models

Here, consider the case of an $M > 1$ block fading channel on the jammer-to-Willie link. In contrast to the results of Lemma 2 and Lemma 3 for the AWGN and $M = 1$ block fading channels on the Alice-to-Willie link, respectively, a power detector is not the optimal detector for Willie. Instead, an important property of the optimal detector in Lemma 4 is established: that, if a given vector of observed powers for the M blocks encompassing a slot results in a point on the boundary between Willie's decision regions, an increase in any component of that vector results in a decision of H_1. Whereas this does not explicitly identify the optimal receiver, it does guarantee an important property of the dividing "curve" between the two decision regions: for any given $M - 1$ components of the vector of observed powers, there is at most one solution for the remaining component that falls on this curve between H_0 and H_1, as defined precisely below. In particular, this is then sufficient to establish the result of interest: that Alice can transmit covertly at power that does not decrease with the blocklength n.

2.3.5 Properties of the Optimal Detector at Willie

With $t = 0$ the slot of interest, observations outside of $k = 1, 2, \ldots, n$ do not help Willie detect transmissions by Alice in slot $t = 0$. Therefore, the slot index is suppressed, and denote Willie's observations by $\hat{\mathbf{Z}} = [\hat{Z}_1, \hat{Z}_2, \ldots, \hat{Z}_n]$. Conditioned on the fading coefficients on the jammer-to-Willie channel, measurements within each fading block of length n/M are i.i.d., but the measurements from different blocks come from different distributions determined by the sequence of block fading variables. Therefore, when Alice does not transmit, Willie's observations have the distribution:

$$f_{\hat{\mathbf{Z}}|H_0}(\hat{\mathbf{z}}|H_0) = E_{\mathbf{h}^{(\mathrm{j,w})}}\left[\prod_{m=1}^{M}\prod_{i=1}^{n/M}\frac{1}{\pi(\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{j},m}^2)}\cdot e^{-\frac{|\hat{z}_{(m-1)\frac{n}{M}+i}|^2}{(\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{j},m}^2)}}\right] \tag{2.29a}$$

$$= \prod_{m=1}^{M} E_{h_m^{(\mathrm{j,w})}}\left[\left(\frac{1}{\pi(\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{j},m}^2)}\right)^{\frac{n}{M}} e^{-\frac{z_m}{(\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{j},m}^2)}}\right], \tag{2.29b}$$

where $\mathbf{h}^{(\mathrm{j,w})} = [h_1^{(\mathrm{j,w})}, h_2^{(\mathrm{j,w})}, \ldots, h_M^{(\mathrm{j,w})}]$ is the vector of (complex) fading coefficients on the jammer-to-Willie channel, $z_m = \sum_{i=1}^{n/M}|\hat{z}_{(m-1)\frac{n}{M}+i}|^2$, and $\sigma_{\mathrm{j},m}^2 = \frac{P_{\mathrm{j}}^{(t)}|h_m^{(\mathrm{j,w})}|^2}{d_{\mathrm{j,w}}^{\alpha}}$. Let $\zeta = P_{\mathrm{j}}^{(t)}/d_{\mathrm{j,w}}^{\alpha}$ and $\mathbf{Z} = [Z_1, Z_2, \ldots, Z_M]$, where $Z_m = \sum_{i=1}^{n/M}|\hat{Z}_{(m-1)\frac{n}{M}+i}|^2$ is the power measured in the m^{th} block. The distribution of the vector $\mathbf{Z}$ of received powers across the M blocks under H_0 is:

$$f_{\mathbf{Z}|H_0}(\mathbf{z}|H_0) = \frac{1}{\pi^n}\prod_{m=1}^{M}\int_0^{\infty}\left(\frac{1}{\sigma_{\mathrm{w}}^2+u}\right)^{\frac{n}{M}} e^{-\frac{z_m}{(\sigma_{\mathrm{w}}^2+u)}}e^{-\frac{u}{\zeta}}du \tag{2.30}$$

$$= \frac{e^{\frac{M\sigma_{\mathrm{w}}^2}{\zeta}}}{\pi^n}\prod_{m=1}^{M}\int_{\sigma_{\mathrm{w}}^2}^{\infty}\left(\frac{1}{v}\right)^{\frac{n}{M}} e^{-\frac{z_m}{v}}e^{-\frac{v}{\zeta}}dv. \tag{2.31}$$

Similarly, the distribution under H_1 is:

$$f_{\mathbf{Z}|H_1}(\mathbf{z}|H_1) = \frac{e^{\frac{M(\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2)}{\zeta}}}{\pi^n}\prod_{m=1}^{M}\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty}\left(\frac{1}{v}\right)^{\frac{n}{M}} e^{-\frac{z_m}{v}}e^{-\frac{v}{\zeta}}dv. \tag{2.32}$$

The LRT test is then:

$$\Lambda(\mathbf{Z}) = \frac{e^{\frac{M\sigma_{\mathrm{a}}^2}{\zeta}} \prod_{m=1}^{M} \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^{\frac{n}{M}} e^{-\frac{Z_m}{v}} e^{-\frac{v}{\zeta}} dv}{\prod_{m=1}^{M} \int_{\sigma_{\mathrm{w}}^2}^{\infty} \left(\frac{1}{v}\right)^{\frac{n}{M}} e^{-\frac{Z_m}{v}} e^{-\frac{v}{\zeta}} dv} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{2.33}$$

The LRT in (2.33) shows that $\mathbf{Z}$ forms a sufficient statistic for the optimal test for Willie to determine whether Alice transmits in that slot or not. Lemma 4 then establishes that $\Lambda(\cdot)$ is monotone increasing in each of its components.

Lemma 4. *Consider the assumptions of the multiple block fading channel model in Section 2.2.1.2 and Alice's construction presented in Section 2.2.1 but with the jammer transmitting Gaussian noise drawn from a distribution with constant variance. When the Alice-to-Willie channel is AWGN and the jammer-to-Willie channel is faded,* $\Lambda(\mathbf{Z})$ *is monotonically increasing in each of the components of* $\mathbf{Z}$*.*

Proof: $\Lambda(Z)$ (defined in (2.28)) monotonically increases in Z in the $M = 1$ case as shown in Appendix C. The proof then follows from the observation that $\Lambda(\mathbf{Z})$ in the $M > 1$ case can be expressed as:

$$\Lambda(\mathbf{Z}) = \prod_{i=1}^{M} \Lambda(Z_i). \qquad \blacksquare \tag{2.34}$$

Corollary 4.1. *Consider the assumptions of the multiple block fading model in Section 2.2.1.2 and Alice's construction presented in Section 2.2.1.1 but with the jammer transmitting Gaussian noise drawn from a distribution with constant variance. Additionally, assume that Willie knows* $h_{0,m}^{(\mathrm{a,w})}, m = 1, 2, \ldots, M$*. When fading exists on both the jammer-to-Willie channel and the Alice-to-Willie channel, then the likelihood ratio* $\Lambda(\mathbf{Z})$ *is monotonically increasing in each of the components of* $\mathbf{Z}$*.*

Proof: Conditioned on Willie's knowledge of $h_{0,m}^{(\mathrm{a,w})}, m = 1, 2, \ldots, M$, the channel from Alice-to-Willie is an AWGN channel with a different signal power for Alice per block; hence, the result follows similarly to that of Lemma 4. $\blacksquare$

2.3.6 Covertness with Transmit Power not Decreasing in the Blocklength

Next, Lemma 4 is leveraged on the structure of the optimal receiver at Willie to demonstrate the ability for Alice to employ power not decreasing in the blocklength for the case where there exists $M > 1$ block fading on the jammer-to-Willie channel. The general concept of the proof is similar to Theorem 1: demonstrate that the optimal detector at Willie works poorly on a set of fading instantiations of the jammer's signal that has high probability.

Before outlining the proof, a number of regions are defined that characterize Willie's detector. Recall that a sufficient statistic for Willie's optimal detector is given by $\mathbf{Z} = [Z_1, Z_2, \ldots, Z_M]$, where Z_i is the power measured in the i^{th} block. A normalized version corresponding to the average observed power per symbol within a block is also a sufficient statistic for the optimal detector: $\mathbf{X} = [X_1, X_2, \ldots, X_M]$, where $X_i = \frac{Z_i}{n/M}, i = 1, 2, \ldots M$. A detector for Willie is defined by the regions $R_{H_0}(n)$ and $R_{H_1}(n)$, each in $\mathcal{R}^M$, where H_0 is chosen if $\mathbf{X} \in R_{H_0}(n)$, and H_1 is chosen if $\mathbf{X} \in R_{H_1}(n)$. For the optimal detector at Willie, as given in (2.33), a vector $\mathbf{x}$ is in $R_{H_1}(n)$ if and only if $\Lambda(\frac{n}{M}\mathbf{x}) > \gamma$; otherwise $\mathbf{x}$ is in $R_{H_0}(n)$. Hence, define the boundary curve dividing $R_{H_0}(n)$ and $R_{H_1}(n)$ as $\mathrm{C}(n) = \{\mathbf{x} : \Lambda(\frac{n}{M}\mathbf{x}) = \gamma\}$. Finally, define a boundary region, $R_{\mathrm{B}}^{\delta}(n)$ as the set of points that are within distance δ in at least one dimension from the dividing curve between the regions.

Define the M-dimensional vectors $\boldsymbol{\sigma}_{\mathrm{j}}^2 = [\sigma_{\mathrm{j},1}^2, \sigma_{\mathrm{j},2}^2, \ldots, \sigma_{\mathrm{j},M}^2]$ and $\boldsymbol{\sigma}_{\mathrm{w}}^2 = \sigma_{\mathrm{w}}^2[1, 1, \ldots, 1]$. Note that $\boldsymbol{\sigma}_{\mathrm{j}}^2$ is random, since it depends on the fading from the jammer to Willie, whereas $\boldsymbol{\sigma}_{\mathrm{w}}^2$ is deterministic and known to Willie. The proof then proceeds, as follows. Given the instantiation of the block fading values between the jammer and Willie, which determines the expected jammer power per symbol $\sigma_{\mathrm{j},i}^2$ for the i^{th} fading block, the i^{th} element of the vector $\mathbf{X}$ has the expected value $\sigma_{\mathrm{j},i}^2 + \sigma_{\mathrm{w}}^2$ (under H_0) or $\sigma_{\mathrm{j},i}^2 + \sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{a}}^2$ (under H_1). The proof then begins with Lemma 5, which leverages Lemma 4 to show that the probability of fading instantiations that result in

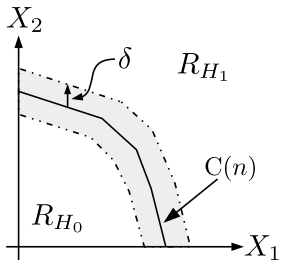

Figure 2.4. An example diagram of decision regions for an arbitrary detector under $M = 2$ block fading conditions. X_1 and X_2 are the normalized power measurements in the first and second block respectively. The solid line $(-)$ represents the boundary curve $\mathrm{C}(n)$ and the dashed lines $(- \cdot\cdot -)$ represent the edge of the boundary region $R_{\mathrm{B}}^{\delta}(n)$.

$\boldsymbol{\sigma}_{\mathrm{j}}^2 + \boldsymbol{\sigma}_{\mathrm{w}}^2 \in R_{\mathrm{B}}^{\delta}(n)$ can be made arbitrarily small by choosing δ small enough; hence, the probability that the jamming is such that the average power received per symbol when Alice is not transmitting is in the boundary region can be made arbitrarily small. The theorem then follows by considering what happens for the (highly probable) event that the instantiation of the block fading values yields $\boldsymbol{\sigma}_{\mathrm{j}}^2 + \boldsymbol{\sigma}_{\mathrm{w}}^2 \notin R_{\mathrm{B}}^{\delta}(n)$; in this case, for σ_{a}^2 sufficiently small, the probability of missed detection or the probability of false alarm is near one. Hence, Alice can employ power that does not decrease with n and still achieve covertness. Essentially, Willie is not able to set a boundary curve that works for a large set of $\boldsymbol{\sigma}_{\mathrm{j}}^2$, and thus his detector is only effective in the unlikely event that $\boldsymbol{\sigma}_{\mathrm{j}}^2 + \boldsymbol{\sigma}_{\mathrm{w}}^2$ is near the boundary curve between his decision regions.

Lemma 5. *Under the assumptions of the multiple block fading model in Section 2.2.1.2, for Willie's optimal detector, with $R_{\mathrm{B}}^{\delta}(n)$ as defined above, there exists $\delta > 0$ s.t. $P(\mathbf{h} : \boldsymbol{\sigma}_{\mathrm{j}}^2 + \boldsymbol{\sigma}_{\mathrm{w}}^2 \in R_{\mathrm{B}}^{\delta}(n)) < \epsilon$ for any $\epsilon > 0$.*

Proof: Consider solving for the values (if there are any) of x_m, the m^{th} component of the vector $\mathbf{x} = [x_1, x_2, \ldots, x_M]$, for which $\mathbf{x} \in \mathrm{C}(n)$, with the other components fixed. By Lemma 4, for a given $[x_1, x_2, x_{m-1}, x_{m+1}, \ldots x_M]$, it is known that the set of x_m such that $\mathbf{x} \in \mathrm{C}(n)$ consists of no points or a single point. Define the

$(M-1)$-dimensional vector $\mathbf{x}_{\sim m} = [x_1, x_2, \ldots, x_{m-1}, x_{m+1}, \ldots, x_M]$ as the vector $\mathbf{x}$ with the m^{th} component removed and define the following function on this vector, which specifies the x_m on the boundary (if there is one) when the rest of the elements of $\mathbf{x}$ are set to the values in $\mathbf{x}_{\sim m}$:

$$g_m(\mathbf{x}_{\sim m}) = \begin{cases} 0, & \text{no } x_m \text{ s.t. } \mathbf{x} \in \mathrm{C}(n) \\ x_m, & \text{a single } x_m \text{ s.t. } \mathbf{x} \in \mathrm{C}(n). \end{cases} \tag{2.35}$$

Define the boundary region $R_{\mathrm{B}}^{\delta}(n)$ as:

$$R_{\mathrm{B}}^{\delta}(n) = \bigcup_{m=1}^{M} \{\mathbf{x} : x_m \in (g_m(\mathbf{x}_{\sim m}) - \delta, g_m(\mathbf{x}_{\sim m}) + \delta)\}. \tag{2.36}$$

Now, applying a union bound yields:

$$P(\boldsymbol{\sigma}_{\mathrm{j}}^2 + \boldsymbol{\sigma}_{\mathrm{w}}^2 \in R_{\mathrm{B}}^{\delta}(n)) = \int_{R_{\mathrm{B}}^{\delta}(n)} \prod_{i=1}^{M} f_{\sigma_{\mathrm{j},i}^2+\sigma_{\mathrm{w}}^2}(x_i) dx_i \tag{2.37}$$

$$\leq \sum_{m=1}^{M} \int_{\mathbf{x}_{\sim m}} \int_{g_m(\mathbf{x}_{\sim m})-\delta}^{g_m(\mathbf{x}_{\sim m})+\delta} \prod_{i=1}^{M} f_{\sigma_{\mathrm{j},i}^2+\sigma_{\mathrm{w}}^2}(x_i) dx_i \; d(\mathbf{x}_{\sim m}) \tag{2.38}$$

$$= \sum_{m=1}^{M} \int_{\mathbf{x}_{\sim m}} \prod_{\substack{i=1 \\ i\neq m}}^{M} f_{\sigma_{\mathrm{j},i}^2+\sigma_{\mathrm{w}}^2}(x_i) \cdot \left[\int_{g_m(\mathbf{x}_{\sim m})-\delta}^{g_m(\mathbf{x}_{\sim m})+\delta} f_{\sigma_{\mathrm{j},m}^2+\sigma_{\mathrm{w}}^2}(x_m) dx_m \right] d(\mathbf{x}_{\sim m}) \tag{2.39}$$

$$\leq \sum_{m=1}^{M} \int_{\mathbf{x}_{\sim m}} \prod_{\substack{i=1 \\ i\neq m}}^{M} f_{\sigma_{\mathrm{j},i}^2+\sigma_{\mathrm{w}}^2}(x_i) \; [2\delta \sup_x f_{\sigma_{\mathrm{j},m}^2+\sigma_{\mathrm{w}}^2}(x)] \; d(\mathbf{x}_{\sim m}) \tag{2.40}$$

$$= 2M\delta \sup_x f_{\sigma_{\mathrm{j},1}^2+\sigma_{\mathrm{w}}^2}(x), \tag{2.41}$$

Noting that $\sup_x f_{\sigma_{\mathrm{j},1}^2+\sigma_{\mathrm{w}}^2}(x)$ is finite, a choice of $\delta = \epsilon/(2M \sup_x f_{\sigma_{\mathrm{j},1}^2+\sigma_{\mathrm{w}}^2}(x))$ yields $P(\mathbf{h} : \boldsymbol{\sigma}_{\mathrm{j}}^2 + \boldsymbol{\sigma}_{\mathrm{w}}^2 \in R_{\mathrm{B}}^{\delta}(n)) < \epsilon$. ∎

Theorem 3. *Consider the assumptions of the multiple block fading model and Alice's construction in Section 2.2.1 but with the jammer transmitting Gaussian noise drawn from a distribution with constant variance. Then, there exists a communication strategy for Alice, Bob, and the jammer whereby Alice transmits with a power that does not decrease with the blocklength while being covert from Willie.*

Proof: Consider a covertness criterion $\mathbb{P}_{\text{MD}} + \mathbb{P}_{\text{FA}} > 1 - \epsilon$. By Lemma 5, choose $\delta > 0$ s.t.:

$$P(\mathbf{h} : \boldsymbol{\sigma}_{\text{j}}^2 + \boldsymbol{\sigma}_{\text{w}}^2 \in R_{\text{B}}^{2\delta}(n)) < \frac{\epsilon}{4}. \tag{2.42}$$

If the Alice-to-Willie channel is AWGN, choose constant $P_{\text{f}} > 0$ such that $\sigma_{\text{a}}^2 < \delta$. If the Alice-to-Willie channel is a $M \geq 1$ block fading channel, choose $P_{\text{f}} > 0$ such that the average received power from Alice is less than δ for all fading blocks with high probability. We proceed with the proof for the case when the Alice-to-Willie channel is AWGN, but the modifications for when the Alice-to-Willie channel is a $M \geq 1$ block fading channel follow similar steps to those shown in the second part of the proof of Theorem 2 in Appendix A.

Consider an optimal detector at Willie for blocklength n, with associated decision regions $R_{H_0}(n)$ and $R_{H_1}(n)$. First, a sketch of the proof idea is presented. Consider the case where $\boldsymbol{\sigma}_{\text{w}}^2 + \boldsymbol{\sigma}_{\text{j}}^2 \in R_{H_0}(n) \setminus R_{\text{B}}^{2\delta}(n)$. If Alice is employing $\sigma_{\text{a}}^2 < \delta$, the probability of Willie's test result being in $R_{H_1}(n)$ occurs with small probability for large n, regardless of whether H_0 or H_1 is true. Thus, Willie's $\mathbb{P}_{\text{MD}}$ is large and $\mathbb{P}_{\text{FA}}$ is small. Likewise, if $\boldsymbol{\sigma}_{\text{w}}^2 + \boldsymbol{\sigma}_{\text{j}}^2 \in R_{H_1}(n) \setminus R_{\text{B}}^{2\delta}(n)$, then Willie's $\mathbb{P}_{\text{FA}}$ is large and $\mathbb{P}_{\text{MD}}$ is small for large n.

The rigorous proof is the vector extension of that of Theorem 2. Recall that $[\sigma_{\text{j},1}^2, \sigma_{\text{j},2}^2, \ldots, \sigma_{\text{j},M}^2]$ is an i.i.d. vector, where each component is exponentially distributed with mean ζ. Hence, there exists a constant c s.t.

$$P\left(\max_{i=1,2,\ldots,M}\ \sigma_{\mathrm{j},i}^2 > c\right) < \frac{\epsilon}{4}. \tag{2.43}$$

Let

$$\mathbb{P}_{\mathrm{FA}}(\mathbf{u}) = P(\mathbf{X} \in R_{H_1}(n) | \boldsymbol{\sigma}_{\mathrm{j}}^2 + \boldsymbol{\sigma}_{\mathrm{w}}^2 = \mathbf{u}, H_0). \tag{2.44}$$

Under H_0, $X_i = (\sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{j},i}^2)\chi^2_{\frac{2n}{M},i}, i = 1, 2, \ldots, M$, where $\{\chi^2_{\frac{2n}{M},i}, i = 1, 2, \ldots, M\}$ is an i.i.d. collection of (central) chi-squared random variables, each with $2n/M$ degrees of freedom. By the weak law of large numbers, each converges in probability to 1; since M is finite, this implies $\exists N_0$ s.t. $\forall n \geq N_0$,

$$P\left(\bigcap_{i=1}^{M}\left\{\chi^2_{\frac{2n}{M},i} \in \left(1 - \frac{\delta}{\sigma_{\mathrm{w}}^2 + c}, 1 + \frac{\delta}{\sigma_{\mathrm{w}}^2 + c}\right)\right\}\right) > 1 - \frac{\epsilon}{2}, \tag{2.45}$$

and

$$P\left(\bigcap_{i=1}^{M}\left\{X_i \in \left((\sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{j},i}^2)\left(1 - \frac{\delta}{\sigma_{\mathrm{w}}^2 + c}\right), (\sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{j},i}^2)\left(1 + \frac{\delta}{\sigma_{\mathrm{w}}^2 + c}\right)\right)\right\}\right) > 1 - \frac{\epsilon}{2}. \tag{2.46}$$

Now, if $\max_{i=1,2,\ldots,M}\ \sigma_{\mathrm{j},i}^2 \leq c$, then $\sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{j},i}^2 < \sigma_{\mathrm{w}}^2 + c$, and thus, for $n \geq N_0$:

$$P\left(\bigcap_{i=1}^{M}\left\{X_i \in (\sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{j},i}^2 - \delta, \sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{j},i}^2 + \delta)\right\}\right) > 1 - \frac{\epsilon}{2}. \tag{2.47}$$

Thus, if $\mathbf{u} \in R_{H_1} \setminus R_{\mathrm{B}}^{2\delta}(n)$, then $P(\mathbf{X} \in R_{H_1}) > 1 - \frac{\epsilon}{2}$ and

$$\mathbb{P}_{\mathrm{FA}}(\mathbf{u}) > 1 - \frac{\epsilon}{2}. \tag{2.48}$$

Next consider any $\mathbf{u} \in R_{H_0} \backslash R_{\mathrm{B}}^{2\delta}(n)$. Then, recalling $\sigma_{\mathrm{a}}^2 < \delta$, the vector $\mathbf{u}+\sigma_{\mathrm{a}}^2[1\ 1\ \ldots\ 1]$ cannot have any element within δ of $\mathrm{C}(n)$. Then, following analogous arguments to those above, $\exists N_1$ s.t. for $n \geq N_1$,

$$\mathbb{P}_{\mathrm{MD}}(\mathbf{u}) = P(\mathbf{X} \in R_{H_0}(n) | \boldsymbol{\sigma}_{\mathrm{j}}^2 + \boldsymbol{\sigma}_{\mathrm{w}}^2 = \mathbf{u}, H_1) \tag{2.49}$$

$$> 1 - \frac{\epsilon}{2} \tag{2.50}$$

for $\mathbf{u} \in R_{H_0} \setminus R_{\mathrm{B}}^{2\delta}(n)$ whenever $\max_{i=1,2,\ldots,M}\ \sigma_{\mathrm{j},i}^2 \leq c$. Thus, unless

$$\mathcal{A} = \{\mathbf{u} \in R_{\mathrm{B}}^{2\delta}(n)\} \cup \{\max_{i=1,2,\ldots,M}\ \sigma_{\mathrm{j},i}^2 > c\} \tag{2.51}$$

occurs,

$$\mathbb{P}_{\mathrm{FA}}(\mathbf{u}) + \mathbb{P}_{\mathrm{MD}}(\mathbf{u}) > 1 - \frac{\epsilon}{2}. \tag{2.52}$$

By construction, $P(\mathcal{A}) < \epsilon/2$, and thus

$$\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} = E_U[\mathbb{P}_{\mathrm{FA}}(\mathbf{U}) + \mathbb{P}_{\mathrm{MD}}(\mathbf{U})] \tag{2.53}$$

$$\geq E_U[\mathbb{P}_{\mathrm{FA}}(\mathbf{U}) + \mathbb{P}_{\mathrm{MD}}(\mathbf{U}) | \mathcal{A}^c] P(\mathcal{A}^c) \tag{2.54}$$

$$> 1 - \epsilon. \tag{2.55}$$

■

The implications on reliable throughput are then analogous to those discussed in Section 2.3.3.

The implications of a constant transmit power at Alice on reliable throughput are then identical to those discussed in Section 2.3.3 for $M = 1$ and the proof is presented in Appendix D.

2.4 Relationship with Steganography

Steganography is the discipline of hiding messages in innocuous objects. Typical steganographic systems modify fixed-size finite-alphabet covertext objects into stegotext containing hidden information, and are subject to a similar square root law (SRL) as non-jammer assisted covert communication: $\mathcal{O}(\sqrt{n})$ symbols in size n covertext may safely be altered to hide an $\mathcal{O}(\sqrt{n}\log n)$-bit message [43]. As explained in [12], the mathematics of statistical hypothesis testing are responsible for both SRLs while the extra $\log n$ factor is from the lack of noise in the steganographic context. However, arguably the earliest work on SRL shows that it is achievable without the $\log n$ factor when an active adversary corrupts stegotext with AWGN [44][2]. That being said, [45] shows that, because Alice in the steganographic setting has write-access to covertext, the SRL can be broken and $\mathcal{O}(n)$ bits can be embedded in size n covertext using careful selection of the subset of the covertext to be overwritten [45]. Thus, unlike the scenario considered here, breaking the steganographic SRL does not require Willie to be uncertain about the distribution of his observations.

2.5 Summary

This chapter investigated the addition of a Jammer to the Alice, Bob, Willie model in such a way that the jammer creates uncertainty at Willie's detector. By adding the uncertainty at Willie, Alice and Bob can achieve covert and reliable communication at a positive rate. In addition, a radiometer is proven to be an optimal detector for Warden Willie under AWGN and single-block fading channel conditions. For the multiple-block fading channel model a radiometer is not an optimal detector.

---

[2]We note that the results of [11] and [12] were developed independently of [44]. While [44] provides the proof of the SRL when Alice is average-power constrained, [11] and [12] also develop the achievability of SRL for the peak-power constained covert communication and the converse to the SRL.

The assumptions presented in Section 2.2.1.1 assume that the jammer is attempting to help Alice and Bob to communicate covertly. However, covert communication may still be possible if an adversarial jammer is placed in the environment to actively try to jam any potential communication by Alice, as is commonly done in electronic warfare. For example, suppose that Willie uses a jammer to inhibit communication by any party; then, whereas this jammer does indeed decrease the rate of any reliable (non-covert) communication, it may actually facilitate covert communication by hurting Willie's ability to determine if Alice is transmitting. In particular, if the jammer-to-Willie channel is faded and Willie's jammer transmits Gaussian noise, then exactly the same interference model as derived for the constructions of Sections 2.3.2 and 2.3.4 applies. This enables covert communication from Alice to Bob in precisely the same manner as in the case of a "friendly" jammer. Note that this assumes that such a jammer generates random Gaussian noise; if that jammer instead generates a noise-like signal that is decodable by Willie (say, using a Gaussian codebook shared by the jammer and Willie), then Willie can conceivably decode the jammer's signal and subtract it from his received signal, subject only to the standard challenges of successive interference cancellation in wireless communication environments.

2.6 Acknowledgment

The work in this chapter was sponsored by the National Science Foundation under grants ECCS-1309573 and CNS-1564067.

CHAPTER 3

COVERT COMMUNICATION IN A DYNAMIC ENVIRONMENT

3.1 Introduction

The block fading scenario described in Chapter 2 assumes that Alice, Bob, Willie and the jammer are stationary, but are in a dynamic environment that causes a finite number of variations over the duration of a codeword. For example, in a dynamic urban environment, there are many people, vehicles, buildings and other objects which contribute to a noisy wireless environment.

The block fading model scenario described in Chapter 2 borrowed from standard models in the communications literature without a careful consideration of how these dynamics might fit various covert communication environments of interest. Here, motivated by the wide range of possible environments of interest, we take a more general look at how the rate of covert communications depends on the dynamics of the environment that Willie observes. In the dynamic model, the jammer does not vary his power and there is no variation in the path loss between transmitter-receiver pairs. The variations any receiver observes is due to the movement of additional nodes or objects in the environment.

In particular, the dynamic model considered in this chapter is based on the scenario shown in Figure 3.1. Assume that there are additional nodes or objects in the environment that move; however, Alice, Bob, Willie and the jammer are stationary. These additional nodes can either represent individual objects or users with no desire to assist Alice or Willie. Therefore, the fading variations on the Alice-to-Willie

channel and the fading variations on the jammer-to-Willie channel are dependent on the movement of the additional nodes in the environment. Define n as the codeword length and $f(n)$ as the number of fading variations that occur over the duration of a codeword.

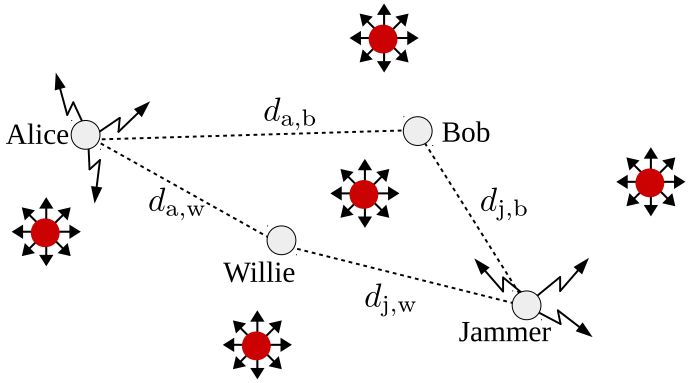

Figure 3.1. Dynamic scenario where Alice, Bob, Willie and the jammer are stationary. The red dots represent additional objects in the environment that move and cause fading variations at Bob and Willie's receivers.

Depending on the nodes' rate of movement, we are able to make some conclusions about Alice's ability to communicate reliably to Bob while remaining covert from Willie based on prior covert communications research. For example, if $f(n) = n$, prior work by Bash *et al.* in [12] suggests a conjecture as to Alice's ability to maintain reliable covert communication to Bob without risking detection by Willie. In particular, if new fading coefficients cause variations in every symbol slot, then potentially, variations in each symbol slot could be averaged out to form an accurate estimate of the null hypothesis statistics, and we expect to be able to achieve $\mathcal{O}(\sqrt{n})$ bits of covert transmission in n channel uses (and no more). Alice's covert communication capabilities are also known when $f(n)$ is finite. Chapter 2 showed that Alice can achieve $\mathcal{O}(n)$ covert bits in n channel uses in this case.

There is also the extreme case when the number of variations are so small that $f(n) < 1$. In this scenario, the movement of the additional nodes causes very little

variation over multiple codeword lengths. For example, assume Willie observes L time slots each with codword length n and that Alice may or may not transmit in a single time slot. The fading variation is so slow that Willie observes the same fading coefficient across multiple time slots. This scenario resembles the results shown by Goeckel *et al.* in [23]. Willie may not know in which time slot Alice transmits; however, he can estimate the channel by using his observations over all $T(n)$ time slots. This result proves that Alice can transmit only $\mathcal{O}(\sqrt{n})$ bits in n channel uses if she desires covert and reliable communication.

A visual representation of Alice's covert rate for $f(n) < 1$, finite $f(n)$ and $f(n) = n$ is shown in Figure 3.2. It is not known what Alice's covert rate is when $f(n)$ is not finite and scales with n. This chapter investigates this to determine the covert communication capabilities and considers situations in which $f(n)$ lies in the unresearched region by considering dynamic environments.

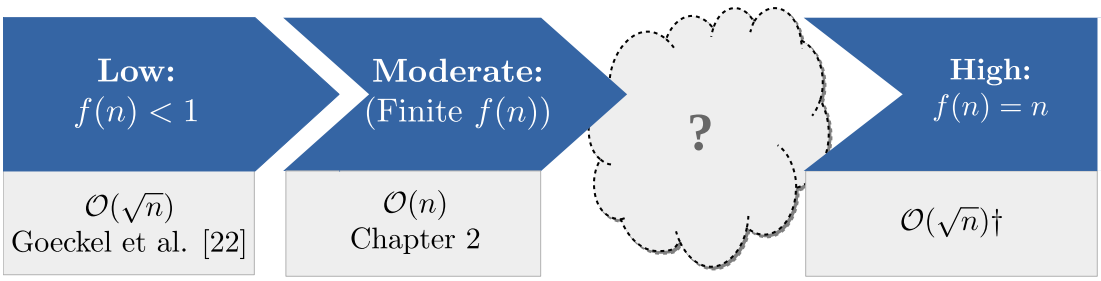

Figure 3.2. Alice's covert communication capabilities based on the number of variations due to fading, $f(n)$, per codeword length n. † was a conjecture based on results in [12] and was not proven prior to work presented in this dissertation.

The main contribution in this chapter is the derivation of the converse. If Alice transmits with power $\omega(1/\sqrt{f(n)})$ then Willie can detect her communications with high probability as $n \to \infty$ where $\omega(\cdot)$ represents little omega notation.

3.2 System Model

Consider Alice ("a") would like to communicate reliably to Bob ("b") without Willie Warden ("w") detecting her communications. A jammer ("j") is present in the environment and assume that Alice, Bob, Willie and the jammer are stationary. Let $d_{\mathrm{a,w}}$ and $d_{\mathrm{a,b}}$ denote the distance from Alice to Willie and Alice to Bob respectively. Similarly, let $d_{\mathrm{j,w}}$ and $d_{\mathrm{j,b}}$ denote the distance from the jammer to Willie and Bob respectively. There are N_U additional users in the environment that do not transmit. However, the additional nodes move in such a way that their movement impacts any fading channels from Alice-to-Willie and Alice-to-Bob as well as any fading channels from the jammer-to-Bob and jammer-to-Willie. Note that although Alice, Bob, Willie and the jammer are stationary, the results in this work extend if any of them are moving. They are assumed stationary to simplify notation. There are limits to the distances from Alice to Willie and from the jammer to Willie. The jammer cannot be too far away, otherwise the range of uncertainty at Willie's receiver is reduced. Alice cannot be too close to Willie, otherwise the Alice's power that Willie observes may be much larger than the range of uncertainty that Willie observes due to the jammer. In such cases, Alice is limited to the SRL. Therefore, when Alice is too close to Willie, the impact of the jammer on the model is negligible.

If Alice chooses to transmit, she first maps her message to a Gaussian codeword $\mathbf{f} = [f_1, f_2, \ldots, f_n]$ such that $E[|f_k|^2] = P_{\max}$. As shown in Figure 3.3, Alice then transmits her codeword in slot $t = 0$ which corresponds to the symbol slots $k = 1, 2, \ldots, n$. The jammer transmits a complex signal $\{\mathbf{g}_t\}_{-T/2}^{T/2-1}$ in the time slots where $\mathbf{g}_t = [g_{tn+1}, g_{tn+2}, \ldots, g_{tn+n}]$ is the jammer's signal in the t^{th} time slot with the power constraint $E[|g_k|^2] = P_{\max}$.

The jammer-to-Willie channel is a fading channel and assume the Alice-to-Willie channel is an AWGN channel as a pessimistic case. The jammer's signal at Willie fades in and out, however, Willie maintains a constant observation of Alice's signal

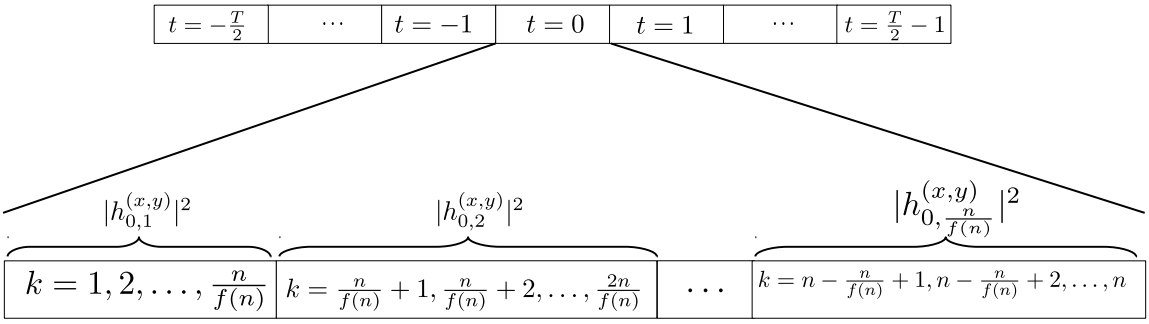

Figure 3.3. Dynamic slot model diagram where each block contains n symbol slots and $f(n)$ fading variations in each time slot. x refers to either Alice or the jammer and y is either Willie or Bob.

if she chooses to transmit. This approach is similar to the modeling employed in Section 2.3.2. The additional nodes in the environment do not transmit, but move in such a way that the fading coefficients on the jammer-to-Willie channel vary. Instead of specifying the effect of each individual node on the power Willie observes from the jammer, assume the collective effect of all the nodes generates a new fading coefficient every $n/f(n)$ symbol slots on the jammer-to-Willie channel as shown in Figure 3.4. Modeling new fading coefficients every $n/f(n)$ symbol slots may at first glance appear similar to the M-Block fading model in Chapter 2.3.2 and Chapter 2.3.4. However, this dynamic scenario models the fading variations as a function of n which invalidates portions of the proof used to prove that positive rate covert communication is achievable in Chapter 2. Portions of the proof in Chapter 2 relied on the fact that the number of fading variations was finite whereas that is not the case when $f(n)$ is a function of n.

When Alice does not transmit, Willie observes:

$$H_0 : \hat{z}_{\frac{mn}{f(n)}+i} = h_m^{(\mathrm{j,w})} g_{\frac{mn}{f(n)}+i} + N_{\frac{mn}{f(n)}+i}^{(\mathrm{w})}, \tag{3.1}$$

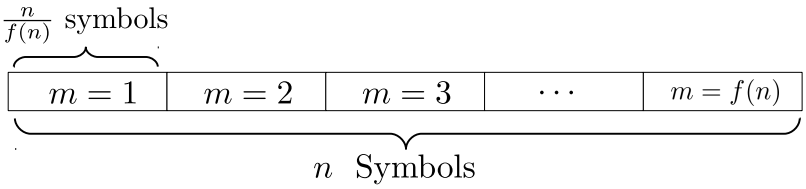

Figure 3.4. Dynamic slot model diagram of a single time slot where each m^{th} block contains $n/f(n)$ symbol slots.

where $h_m^{(\text{j,w})}$ is the fading coefficient between the jammer and Willie in the m^{th} block and $N_k^{(\text{w})}$ are independent and identically distributed zero mean Gaussian complex random variables with variance $E[|N_k^{(\text{w})}|^2] = \sigma_{\text{w}}^2$.

If Alice decides to transmit, she first maps her message to the symbols $\underline{f} = \{f_1, f_2, \ldots, f_n\}$ using a Gaussian codebook and transmits f_k in each k^{th} symbol slot with the power constraint $E[|f_k|^2] = P_{\text{f}}$. Although the channel between the jammer and Willie is a fading channel, this work assumes for now that the channel between Alice and Willie is AWGN. Therefore, Willie's observation when Alice transmits is:

$$H_1 : \hat{z}_{\frac{mn}{f(n)}+i} = f_{\frac{mn}{f(n)}+i} + h^{(\text{j,w})}_{\frac{n}{f(n)}} g_{\frac{mn}{f(n)}+i} + N^{(\text{w})}_{\frac{mn}{f(n)}+i}. \tag{3.2}$$

Define Z as Willie's test result:

$$Z = \sum_{k=1}^{n} |\hat{z}_k|^2. \tag{3.3}$$

Willie then compares his observation Z to some threshold Γ_n to detect if Alice transmitted:

$$Z \underset{H_0}{\overset{H_1}{\gtrless}} \Gamma_n \tag{3.4}$$

where Γ_n is Willie's threshold which depends on n. To generalize Willie's test, Willie normalizes his observation by n:

$$\frac{Z}{n} \underset{H_0}{\overset{H_1}{\gtrless}} \tau_n \tag{3.5}$$

where τ_n is the threshold used to model the asymptotic behavior of Willie's test as $n \to \infty$.

3.3 Converse Proof

Define $S = Z/n$ as Willie's measurement used in his test.

Theorem 4. *Under assumptions of the channel model and the construction given in Section 3.2, then as $n \to \infty$, if Alice transmits with power $\omega(1/\sqrt{f(n)})$, either Willie detects her with high probability or Bob cannot decode her messages with low probability of error.*

The expected value of Willie's observation Z when Alice does not transmit is:

$$E[Z|H_0] = \sum_{k=1}^{n} E[|\hat{z}_k|^2], \tag{3.6}$$

$$= \sum_{k=1}^{n} E\left[\left(h_k g_k + N_k^{(\mathrm{w})}\right)\left(h_k g_k + N_k^{(\mathrm{w})}\right)^*\right], \tag{3.7}$$

$$= \sum_{k=1}^{n} E[|h_k|^2]E[|g_k|^2] + E[|N_k^{(\mathrm{w})}|^2], \tag{3.8}$$

$$= n(P_{\mathrm{j}} + \sigma_{\mathrm{w}}^2) \tag{3.9}$$

where (3.9) follows because $E[|h_k|^2] = 1$ and $E[|g_k|^2] = P_{\mathrm{j}}$. To compute the variance of Z when Alice does not transmit, the term $E[Z^2|H_0]$ is first expanded:

$$E[Z^2|H_0] = \sum_{k=1}^{n}\sum_{l=1}^{n} E\big[(|h_k|^2|g_k|^2 + h_k g_k N_k^* + h_k^* g_k^* N_k + |N_k|^2) \times (|h_l|^2|g_l|^2 + h_l g_l N_l^* + h_l^* g_l^* N_l + |N_l|^2)\big], \tag{3.10}$$

$$= \sum_{k=1}^{n}\sum_{l=1}^{n} E[|h_k|^2|h_l|^2|g_k|^2|g_l|^2] + E[|h_k|^2]E[|g_k|^2]E[|N_l|^2] + E[|N_k|^2]E[|h_l|^2]E[|g_l|^2] + E[h_k^* h_l]E[g_k^* g_l]E[N_k N_l^*] + E[h_k h_l^*]E[g_k g_l^*]E[N_k^* N_l] + E[|N_k|^2]E[|N_l|^2]. \tag{3.11}$$

$E[Z^2|H_0]$ is simplified further by analyzing the individual terms of (3.11) separately.

The first term of (3.11) is simplified by considering the different possible cross-products when: 1) $k = l$; 2) the symbols in the same fading block are cross product terms $\left(l \in [\frac{kn}{f(n)}, \frac{kn}{f(n)} + \frac{n}{f(n)} - 1] \text{ and } l \neq k\right)$; and 3) when symbols that are not in the same fading block are cross products of each other $\left(l \notin [\frac{kn}{f(n)}, \frac{kn}{f(n)} + \frac{n}{f(n)} - 1]\right)$:

$$\sum_{k=1}^{n}\sum_{l=1}^{n} E[|h_k|^2|h_l|^2|g_k|^2|g_l|^2] = \sum_{k=1}^{n}\Bigg(\sum_{l=k} E[|h_k|^4]E[|g_k|^4] + \sum_{l\in[\frac{kn}{f(n)}, \frac{kn}{f(n)}+\frac{n}{f(n)}-1], l\neq k} E[|h_k|^4]E[|g_k|^2]E[|g_l|^2] + \sum_{l\notin[\frac{kn}{f(n)}, \frac{kn}{f(n)}+\frac{n}{f(n)}-1]} E[|h_k|^2]E[|h_l|^2]E[|g_k|^2]E[|g_l|^2]\Bigg), \tag{3.12}$$

$$= 4nP_{\mathrm{j}}^2 + 2P_{\mathrm{j}}^2 f(n)\left(\frac{n}{f(n)} - 1\right)^2 + nP_{\mathrm{j}}^2\left(n - \frac{n}{f(n)}\right), \tag{3.13}$$

$$= 4nP_{\mathrm{j}}^2 + \frac{2n^2P_{\mathrm{j}}^2}{f(n)} + 2f(n)P_{\mathrm{j}}^2 - 2nP_{\mathrm{j}}^2 + n^2P_{\mathrm{j}}^2 - \frac{n^2}{f(n)}P_{\mathrm{j}}^2, \tag{3.14}$$

$$= 2nP_{\mathrm{j}}^2 + \frac{n^2P_{\mathrm{j}}^2}{f(n)} + 2f(n)P_{\mathrm{j}}^2 + n^2P_{\mathrm{j}}^2. \tag{3.15}$$

Equation (3.12) simplifies since $E[|h_k|^4] = 2$, $E[|g_k|^4] = 2P_{\mathrm{j}}^2$, and $E[|g_k|^2|g_l|^2] = P_{\mathrm{j}}^2$ for $k \neq j$. Therefore, the variance of Willie's test result when Alice does not transmit is:

$$\begin{aligned}
\mathrm{Var}(Z|H_0) &= E[Z^2|H_0] - \left(E[Z|H_0]\right)^2, && (3.16)\\
&= 2n^2P_{\mathrm{j}}\sigma_{\mathrm{w}}^2 + 2nP_{\mathrm{j}}\sigma_{\mathrm{w}}^2 + 2n\sigma_{\mathrm{w}}^2 + n^2\sigma_{\mathrm{w}}^4 - n\sigma_{\mathrm{w}}^4 + 2nP_{\mathrm{j}}^2 + \frac{n^2P_{\mathrm{j}}^2}{f(n)} \\
&\qquad + 2f(n)P_{\mathrm{j}}^2 + n^2P_{\mathrm{j}}^2 - n^2P_{\mathrm{j}}^2 - n^2\sigma_{\mathrm{w}}^4 - 2n^2P_{\mathrm{j}}\sigma_{\mathrm{w}}^2, && (3.17)\\
&= 2nP_{\mathrm{j}}\sigma_{\mathrm{w}}^2 + 2n\sigma_{\mathrm{w}}^2 - n\sigma_{\mathrm{w}}^4 + 2nP_{\mathrm{j}}^2 + 2P_{\mathrm{j}}^2f(n) + \frac{n^2P_{\mathrm{j}}^2}{f(n)}. && (3.18)
\end{aligned}$$

The second and third terms of (3.11) do not have cross-product terms, hence, these two terms simplify as:

$$\sum_{k=1}^{n}\sum_{l=1}^{n} E[|h_k|^2]E[|g_k|^2]E[|N_l|^2] + \sum_{k=1}^{n}\sum_{l=1}^{n} E[|N_k|^2]E[|h_l|^2]E[|g_l|^2] = 2n^2P_{\mathrm{j}}\sigma_{\mathrm{w}}^2. \quad (3.19)$$

The fourth and fifth terms of (3.11) are non-zero when $k = l$ and so these terms are written as:

$$\sum_{i=k}^{n}\sum_{l=1}^{n} E[h_k^*h_l]E[g_k^*g_l]E[N_kN_l^*] + \sum_{k=1}^{n}\sum_{l=1}^{n} E[h_kh_l^*]E[g_kg_l^*]E[N_k^*N_l] = 2nP_{\mathrm{j}}\sigma_{\mathrm{w}}^2. \quad (3.20)$$

The sixth term of (3.11) is expanded:

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{l=1}^{n} E[|N_k|^2]E[|N_l|^2] &= \sum_{k=1}^{n}\left(E[|N_k|^4] + \sum_{l\neq k}^{n} E[|N_k|^2]E[|N_l|^2]\right), && (3.21)\\
&= n(2\sigma_{\mathrm{w}}^2 + (n-1)\sigma_{\mathrm{w}}^4), && (3.22)\\
&= 2n\sigma_{\mathrm{w}}^2 + n^2\sigma_{\mathrm{w}}^4 - n\sigma_{\mathrm{w}}^4. && (3.23)
\end{aligned}$$

After normalizing Willie's power measurement, the expected value of his normalized measurement when Alice does not transmit is:

$$E[S|H_0] = P_{\mathrm{j}} + \sigma_{\mathrm{w}}^2 \tag{3.24}$$

and the variance of S is:

$$\mathrm{Var}[S|H_0] = \frac{1}{n}\left[2P_{\mathrm{j}}\sigma_{\mathrm{w}}^2 + 2\sigma_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^4 + 2P_{\mathrm{j}}^2\right] + \frac{2P_{\mathrm{j}}^2 f(n)}{n^2} + \frac{P_{\mathrm{j}}^2}{f(n)}. \tag{3.25}$$

Implementing analogous steps (shown in Appendix E), the expected value of Willie's normalized measurement is:

$$E[S|H_1] = P_{\mathrm{j}} + \sigma_{\mathrm{w}}^2 + P_{\mathrm{f}} \tag{3.26}$$

and the variance of S when Alice transmits is:

$$\begin{aligned}\mathrm{Var}[S|H_1] = \frac{1}{n}\Big[2P_{\mathrm{j}}\sigma_{\mathrm{w}}^2 + 2\sigma_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^4 + 2P_{\mathrm{j}}^2 + 2P_{\mathrm{j}}P_{\mathrm{f}} + 2P_{\mathrm{f}}\sigma_{\mathrm{w}}^2 + 2P_{\mathrm{f}} - P_{\mathrm{f}}^2\Big] \\ + \frac{P_{\mathrm{j}}^2}{f(n)} + \frac{2f(n)P_{\mathrm{j}}^2}{n^2}\ .\end{aligned} \tag{3.27}$$

Now, consider the case when $f(n) < n$ and assume that Willie sets his threshold such that he chooses the null hypothesis for any observation less than $\sigma_{\mathrm{w}}^2 + t + P_{\mathrm{j}}$. Willie then chooses the alternative hypothesis for any observation greater than $\sigma_{\mathrm{w}}^2 + t + P_{\mathrm{j}}$. Therefore, when Alice does not transmit, Willie's probability of false alarm is defined as:

$$\mathbb{P}_{\mathrm{FA}} = P_0(S > \sigma_{\mathrm{w}}^2 + P_{\mathrm{j}} + t) \tag{3.28}$$

where $P_0(\cdot)$ refers to the pdf of Willie's measurement S when Alice does not transmit. Employing Chebyshev's inequality [46, Equation 3.32], the probability of false alarm is bounded:

$$\mathbb{P}_{\text{FA}} \leq P_0(|S - \sigma_{\text{w}}^2 - P_{\text{j}}| > t), \tag{3.29}$$

$$\leq \frac{1}{t^2}\left[\frac{2P_{\text{j}}\sigma_{\text{w}}^2 + 2\sigma_{\text{w}}^2 - \sigma_{\text{w}}^4 + 2P_{\text{j}}^2}{n} + \frac{2P_{\text{j}}^2 f(n)}{n^2} + \frac{P_{\text{j}}^2}{f(n)}\right], \tag{3.30}$$

$$\leq \frac{P_{\text{j}}^2}{f(n)t^2}. \tag{3.31}$$

Equation (3.31) follows because $\frac{P_{\text{j}}^2}{f(n)}$ is the dominant term in (3.30) since all other terms go to zero as $n \to \infty$ for $f(n) < n$. Furthermore, assume Willie defines $t = d/\sqrt{f(n)}$, for some constant d.

Willie misses Alice's communication if his observation, S, is less than $\sigma_{\text{w}}^2 + P_{\text{j}} + t$ when Alice transmits. Thus, Willie's probability of missed detection is:

$$\mathbb{P}_{\text{MD}} = P_1(S < \sigma_{\text{w}}^2 + P_{\text{j}} + t), \tag{3.32}$$

$$\leq P_1(|S - \sigma_{\text{w}}^2 - P_{\text{j}} - P_{\text{f}}| > P_{\text{f}} - t), \tag{3.33}$$

$$\leq \frac{\text{Var}[S|H_1]}{(P_{\text{f}} - t)^2}, \tag{3.34}$$

$$\leq \frac{P_{\text{j}}^2}{(P_{\text{f}}\sqrt{f(n)} - d)^2} \tag{3.35}$$

where $P_1(\cdot)$ in (3.32) represents Willie's distribution when Alice transmits and (3.35) follows because $\frac{P_{\text{j}}^2}{f(n)}$ is the dominant term in $\text{Var}[S|H_1]$. The covert criteria requires that Willie's probability of error satisfies the constraint $\mathbb{P}_{\text{FA}} + \mathbb{P}_{\text{MD}} \geq 1/2 - \epsilon$ where ϵ is small. However, if Alice transmits with power that is $\omega(1/\sqrt{f(n)})$, then $\mathbb{P}_{\text{FA}} + \mathbb{P}_{\text{MD}}$ goes to zero as $n \to \infty$ and the covert criteria is not satisfied. As a result, Alice's throughput cannot transmit $\omega(n/\sqrt{f(n)})$ covert bits reliably in n channel uses, otherwise, she risks Willie detecting her with low probability of error.

3.4 Summary

This chapter builds upon the Alice/Bob/Willie/Jammer scenario presented in Chapter 2 by considering a dynamic fading model. In Chapter 2, Alice's ability to communicate covertly and reliably in a fading channel environment is analyzed for single block fading channels and the finite M-block fading channel model. This chapter generalizes the fading and considers that the fading on the channel between the jammer and Willie is a function of the total codeword length. Results prove that if Alice's transmit power is $\omega(1/\sqrt{f(n)})$, then Willie's probability of error converges to zero as $n \to \infty$ and $f(n) \to n$.

3.5 Acknowledgment

The work in this chapter was sponsored by the DARPA Quiet program.

CHAPTER 4

COVERT COMMUNICATION ON THE CONTINUOUS-TIME MODEL: CYCLOSTATIONARY DETECTORS

4.1 Introduction

Traditional digital communication analyses often model the signal at the receiver with an equivalent discrete-time model. The discrete-time model is equivalent in the sense that analyses and receiver operations based on it produce the same results as those in the true continuous-time model. For example, consider the scenario shown in Figure 4.1 where Alice communicates to Bob and Willie attempts to detect if Alice is transmitting. Both the Alice-to-Willie and Alice-to-Bob channels are additive white Gaussian noise (AWGN) channels.

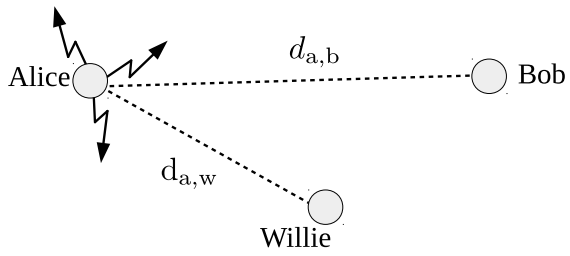

Figure 4.1. Alice, Bob and Warden Willie model under AWGN channel conditions. $d_{\mathrm{a,b}}$ and $d_{\mathrm{a,w}}$ represent the distance from Alice to Bob and Alice to Willie respectively.

Assume that Alice transmits to Bob as shown in Figure 4.2, where f_k represents the k^{th} symbol that Alice transmits. Alice employs pulse shaping with the pulse shape $p(t)$. Denote $N^{(\mathrm{b})}(t)$ as the AWGN on the channel between Alice and Bob, and define $y(t)$ as Bob's observation at his receiver which he then passes through a

matched filter. Bob then samples the filtered signal $y_{\text{mf}}(t)$ at the symbol rate, T_b, and according to the timing offset τ_{a}. Bob then estimates $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_n$ as the original message sent by Alice.

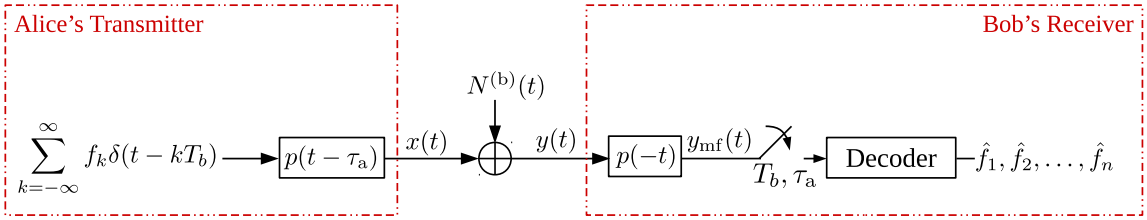

Figure 4.2. Alice to Bob communication diagram when Alice transmits BPSK symbols.

For data decoding, Bob can model the end-to-end process shown in Figure 4.2 as an equivalent discrete-time model as long as his sampling of $y_{\text{mf}}(t)$ occurs at the proper time instances. In particular, let $y[n] = y(nT_b)$ represent Bob's observation in the equivalent discrete-time model corresponding to the n^{th} sample and assuming he knows the timing offset τ_{a}:

$$y[n] = f_n + W[n] \tag{4.1}$$

where $W[n]$ represents the noise from the channel and any system processes that generate noise in the n^{th} sample. We assume that Bob is capable of ascertaining τ_{a} by sharing knowledge with Alice about their timing offsets along with their shared codebook. Since Alice's goal is to transmit symbols to Bob, the equivalent discrete-time model focuses on Bob's ability to estimate f_n and does not model processes such as pulse shaping in the model. Furthermore, the equivalent discrete-time model in (4.1) can be used to exactly model the error rates between communicating parties [47, Chapter. 10.1.2].

Early covert communication research presented by Bash *et al.* in [12] and subsequent work modeled both the Alice-to-Bob and Alice-to-Willie observations using a

discrete-time model. Although there was no reference to the underlying continuous-time model, the clear implication is that the discrete-time model captures the salient characteristics of the underlying continuous-time model. Hence, the equivalent discrete-time model is satisfactory for analyzing the communications between Alice and Bob, but the equivalent discrete-time model for Willie's channel is not designed to consider all detectors that are available to Willie. Therefore, Willie's model should allow for the exploitation of any information available (or not) at his receiver which can assist with detecting Alice's communications.

The majority of covert communications research in the background Section 1.2 of this work assumes that Willie employs a power detector. Additionally, work presented in Chapter 2 of this dissertation demonstrates that a power detector is often optimal for the equivalent discrete-time model. However, modeling Willie's observations in the equivalent discrete-time model does not allow for the consideration of various assumptions and of all potential detectors available to Willie. To consider a broader range of detectors, a continuous-time model of Willie's observations should be implemented.

Thus, this chapter revisits the original covert communication model presented in [12]. The work in Chapter 2 proves that a power detector is optimal for the discrete-time model under AWGN and single block fading channel conditions. However, if Alice's signal contains periodic features such as the example shown in Figure 4.2, cyclostationary detectors can outperform standard power detectors [25]. Since a power detector is not always optimal depending on the type of periodic signal Alice transmits, this leads to the question whether the classical "equivalent discrete-time model" is in fact an equivalent discrete-time model for covert communications. This chapter investigates this question.

In particular, human-generated signals do not resemble the random noise that occurs in nature, as generated signals often have periodic features. Therefore, if Willie's

noise at this receiver resembles Gaussian noise (as in the system models of Chapter 2 and [12]), then any features in Alice's transmission that do not resemble Gaussian noise provides Willie with additional information to help detect Alice's signal. For example, consider the scenario presented in Figure 4.2 where Alice transmits her message with pulse shaping.

The first scenario considered here is the Alice/Bob/Willie scenario and assumes that Alice transmits a standard periodic signal. Periodic signals generate cyclic frequencies which are defined as the frequency of the periodic feature of interest. Cyclostationary detectors (CSDs) are designed to exploit the periodicity of a signal by attempting to measure the power observed at cyclic frequencies associated with the signal of interest. This chapter compares Alice's ability to communicate covertly if Willie employs a CSD versus a power detector assuming Alice transmits a binary phase shift keying (BPSK) signal. Before comparing the two detectors, a simple motivating example for a CSD is presented in Section 4.2 and further background is provided on Gardner's method used to construct CSDs. A CSD designed to detect baseband BPSK signals and the detector's statistics are discussed in section 4.3. The Kullback−Leibler (KL) distance between the two hypotheses at the output of a cyclic detector is then compared to the KL distance between the two hypotheses at the output of a power detector. Results show that the CSD presented does outperform a power detector of the full continuous-time model.

The main contributions in this chapter are:

1. The background of cyclostationary detectors is presented.

2. A CSD is proposed based on the continuous-time model instead of the equivalent discrete-time model approximation, and a novel performance analysis of the CSD is presented.

3. Simulations and proofs are derived to demonstrate how well the cyclostationary detector outperforms the power detector in the continuous-time model, suggesting the discrete-time model must be employed cautiously in covert communications.

4.2 Constructing Cyclostationary Detectors

Gardner in [24] and [25] conducted extensive research on CSDs and demonstrated that for very low signal-to-noise ratios that CSDs outperform power detectors. Gardner also derived various CSDs based on the cyclostationary properties of potential signals of interest (e.g. BPSK, QAM, length of pulse, etc.) in [26] and [27]. Since Gardner's early work, additional CSD variations have been presented. For example, Zeng *et al.* in [48] derived CSDs that require less computational processing power while still outperforming power detectors. However, the basic design process of a CSD follows similar procedures to that of Gardner.

Let $z(t)$ represent the observed signal at Willie. Willie's goal is to determine if he is observing noise or the sum of noise and a signal with periodic behavior at his receiver. If Willie wants to use a CSD, he first makes a certain computation on $z(t)$, and we let $v(t)$ represent the result of this computation. If $z(t)$ is cyclostationary, then $v(t)$ also exhibits cyclic behavior. However, if $z(t)$ does not contain a periodic signal, then $v(t)$ does not exhibit any cyclic features. Therefore, once $v(t)$ is computed, the power at the cyclic frequencies of interest in $v(t)$ are measured and a threshold test is used to determine if a cyclic signal corresponding to the frequencies of interest (e.g. the baud rate of Alice) are present.

4.2.1 The Intuition Behind Cyclostationary Detectors: A Sinusoid Example

Before presenting the mathematical basis of Gardner's work, this sub-section provides a simple motivating example to demonstrate how CSDs may outperform power detectors when the signal-of-interest contains periodic features.

Consider the Alice, Bob, Willie scenario shown in Figure 4.1. Assume Willie observes the period of time $[0, T]$ and his goal is to determine if Alice transmitted during his observation. Let $N^{(\mathrm{w})}(t)$ represent Willie's noise which is a zero-mean Gaussian random process with power spectral density $S_{N^{(\mathrm{w})}}(f) = \sigma_{\mathrm{w}}^2$. If Alice chooses to transmit, Alice transmits $x(t) = \cos(2\pi F_c t)$ during the time period T, where F_c is the carrier frequency. Let p represent the probability of Alice transmitting and assume $p = 1/2$.

Define $z(t)$ as Willie's observation:

$$z(t) = \begin{cases} N^{(\mathrm{w})}(t), & \text{Alice does not transmit} \\ x(t) + N^{(\mathrm{w})}(t), & \text{Alice transmits.} \end{cases} \tag{4.2}$$

A typical power detector measures the power observed in $z(t)$ and compares the measurement to some threshold γ to determine if Alice transmitted:

$$\frac{1}{T}\int_0^T z^2(t)dt \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{4.3}$$

The CSD examples that are presented are simulated in Matlab. Therefore, an oversampled discrete representation is used in the following steps instead of continuous-time notation. Denote $z[n] = z(nT_s)$ as the sampled version of $z(t)$ where T_s is the sampling period. Let $R_Z(n, \tau)$ represent the autocorrelation function (ACF) of Willie's observation at sample n and with lag τ:

$$R_Z(n,\tau) = E\big[z[n]z[n-\tau]\big]. \tag{4.4}$$

For simplicity in this subsection, the ACF is only computed for $\tau = 0$. Denote $R_{Z|H_0}(n,0)$ as the (time-varying) zero-lag ACF of Willie's observation when Alice does not transmit and $R_{Z|H_1}(n,0)$ as the (time-varying) zero-lag ACF when Alice transmits.

Then, define $\mathbf{F}_{Z|H_0}$ as the Fourier transform of $R_Z(n,0)$ when Alice does not transmit

$$\mathbf{F}_{Z|H_0} = \mathcal{F}\{R_{Z|H_0}(n,0)\} \tag{4.5}$$

where $\mathcal{F}$ represents the discrete Fourier transform (DFT) over the cyclic frequency range $\alpha = \{-N/2, -N/2+1, \ldots, N/2-2, N/2-1\}$, and N is the number of samples observed. Each frequency in $\mathbf{F}_{Z|H_0}$ is called a cyclic frequency because active frequencies in $\mathbf{F}_{Z|H_0}$ are related to the cyclic features which occur in $z[n]$.

Simulations were generated to provide a visual representation of the mathematical expressions described in this section. The simulations represent Willie's observations due to Alice transmitting a cosine signal at frequency $F_c = 2.5$ MHz with an oversampling rate of 100 MHz. The signal-to-noise ratio (SNR) of Alice's signal at Willie's receiver is 0 dB. Therefore, both Willie's noise and the power of Alice's transmitted signal are $\sigma_{\mathrm{w}}^2 = 0.5$. Then $\mathbf{F}_{Z|H_0}$ is generated in Matlab using 1024 discrete samples and the length of the DFT is also set to 1024. $|\mathbf{F}_{Z|H_0}|$ is shown in Figure 4.3 which has a peak when the cyclic frequency is zero and is small for all other frequency values. The behavior of $\mathbf{F}_{Z|H_0}$ is expected because Willie only observes real Gaussian noise at his receiver when Alice does not transmit. Note that in Figure 4.3 results, the lag is fixed and the DFT is computed over the result of the ACF when there is zero lag.

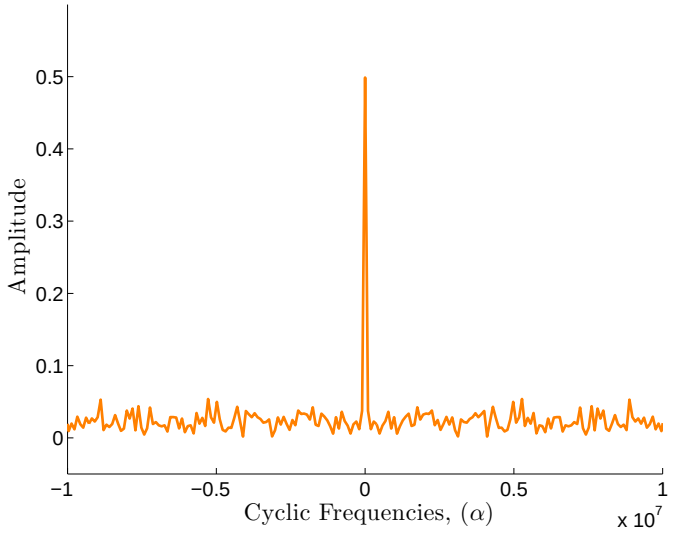

Figure 4.3. Cyclostatinary detector null hypothesis DFT observations. DFT of $R_{Z|H_0}(n, 0)$, $|\mathbf{F}_{Z|H_0}|$, when Alice does not transmit.

Similarly, define $\mathbf{F}_{Z|H_1}$ as the Fourier transform of $R_{Z|H_1}(n, 0)$ when Alice transmits. $|\mathbf{F}_{Z|H_1}|$ is shown in Figure 4.4 and there are peaks at frequencies that correspond to $-2F_c = -5$MHz and $2F_c = 5$MHz in addition to the peak observed at $\alpha = 0$.

A power detector is equivalent to employing Willie's observations at $\alpha = 0$. A threshold detector chooses a threshold to minimize the false alarm rate and maximize the rate of correct detection based on observations in $F_{Z|H_0}(\alpha = 0)$ and $F_{Z|H_1}(\alpha = 0)$. Similarly, a CSD would observe the power at either $-2F_c$ or $2F_c$ and determine a threshold to minimize the rate of error for the CSD detector. The ratio of the powers observed in $F_{Z|H_1}(\alpha = 0)$ and $F_{Z|H_0}(\alpha = 0)$ is 3.11 dB $\left(10\log_{10}\left(\frac{|F_{H_1}(0)|}{|F_{H_0}(0)|}\right)\right)$. However, the ratio of the powers observed at $-2F_c$ is 6.92 dB $\left(10\log_{10}\left(\frac{|F_{H_1}(-2F_c)|}{|F_{H_0}(-2F_c)|}\right)\right)$. Therefore, if Alice transmits a cyclostationary signal with small power, it is easier for her to "hide in the noise" of a power detector than for her to hide her cyclic features when Willie employs a CSD depending on her original transmit power.

In this section, a power detector is compared to a simple CSD when Alice transmits a sinusoid. The CSD presented only considers the spectral content in the auto-

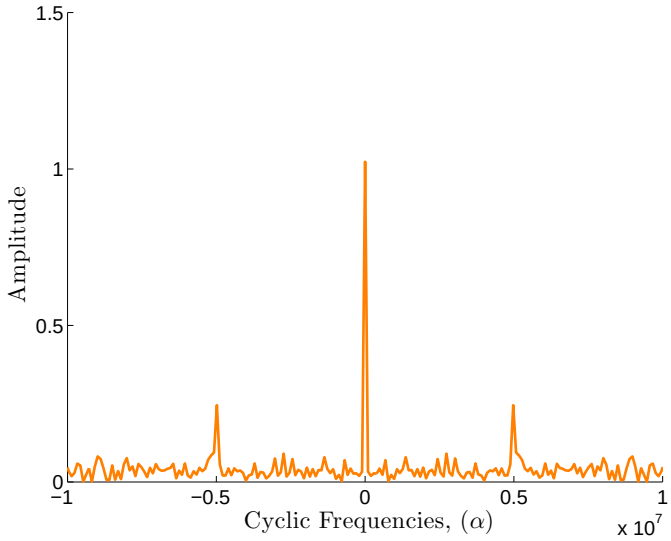

Figure 4.4. Cyclostatinary detector alternative hypothesis DFT observations. DFT of $R_{Z|H_1}(n, 0)$, $|\mathbf{F}_{Z|H_1}|$, when Alice transmits.

correlation function of $z[n]$ when there is no lag. However, considering the average spectral density function of the autocorrelation function with varying delay values τ (i.e. $R_Z(n, \tau)$) can provide a more thorough analysis. The following Section 4.2.2 presents the mathematical framework for CSDs in more detail.

4.2.2 Classical Cyclostationary Detectors

This subsection provides a brief description of the mathematical basis used to construct CSDs. Section 4.2.1 only provided a high-level perspective of CSDs, whereas this subsection presents the mathematical framework based on William Gardner's work in detail [24].

Define $R_Z(t + \tau/2, t - \tau/2)$ as the ACF of an observation $z(t)$ with lag τ

$$R_Z(t + \tau/2, t - \tau/2) = E[z(t + \tau/2)z(t - \tau/2)]. \tag{4.6}$$

As shown in Section 4.2.1, if $z(t)$ is cyclic then the Fourier transform of the ACF is non-zero for cyclic frequencies other than $\alpha = 0$.

Since $R_Z(\cdot,\cdot)$ is periodic in t, $R_Z(\cdot,\cdot)$ can be represented as a Fourier series

$$R_Z(t+\tau/2, t-\tau/2) = \sum_{\{\alpha\}} R_Z^{\alpha}(\tau) e^{i2\pi\alpha t} \tag{4.7}$$

where $\{\alpha\}$ represents the set of cyclic frequencies that are active in $R_Z(t+\tau/2, t-\tau/2)$ and $R_Z^{\alpha}(\tau)$ is the Fourier coefficient of cyclic frequency α defined as:

$$R_z^{\alpha}(\tau) = \lim_{\Delta\to\infty} \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} z(t+\tau/2) z(t-\tau/2) e^{-i2\pi\alpha t} dt. \tag{4.8}$$

$R_z^{\alpha}(\tau)$ is also called the cyclic autocorrelation function at cyclic frequency α. The cyclic frequencies which occur in $R_z(t+\tau/2, t-\tau/2)$ are the result of any periodic behavior in $z(t)$ such as a carrier frequency as shown in Section 4.2.1. Cyclic frequencies are also generated by the symbol rate and other instances in which periodicity may be present.

The next step in the design of a CSD calculates the power observed at the Fourier coefficient corresponding to the cyclic frequency α for various τ values. Define $S_z^{\alpha}(f)$ as the average cyclic autocorrelation function:

$$S_z^{\alpha}(f) = \int_{-\infty}^{\infty} R_z^{\alpha}(\tau) e^{-i2\pi f\tau} d\tau. \tag{4.9}$$

Observing (4.9), $S_z^{\alpha}(f)$ is the power spectral density of R_z^{α} which is calculated by averaging R_z^{α} over all potential values of τ. In the cyclostationary literature, $S_z^{\alpha}(f)$ is called the spectral correlation function or the cyclic spectral density function. Equation (4.9) is a generalization of the high level example presented in Section 4.2.1. The main differences between the high level example and (4.9) is that Section 4.2.1 only considers when $\tau = 0$.

Depending on the cyclic nature of the signal $z(t)$, there are optimal (α, f) pairs to employ when constructing cyclostationary detectors [24–27]. For example, if $z(t)$ is

a BPSK signal transmitted on a carrier frequency F_c, the peaks of $S_z^\alpha(f)$ associated with the cyclostationary behavior of $z(t)$ occur at $(\frac{1}{T_b}, F_c), (2F_c, 0), (2F_c \pm \frac{k}{T_b}, 0), (2F_c \pm \frac{k}{2T_b}, 0)$ where T_b is the symbol period of the pulse and $k \in \mathbb{Z}$.

4.3 Covert Rate of a BPSK Cyclostationary Detector

This section presents a CSD designed to detect Alice's communication when she transmits a baseband BPSK signal. The CSD presented in this section differs from the BPSK detector described in Gardner [24, Chapter 3.6]. Gardner's BPSK detector assumes that a carrier signal is present whereas this work assumes Alice transmits at baseband. Once the CSD statistics are derived, the KL distance between the statistics when Alice does and does not transmit is derived as well. The results of the KL distance of the CSD are then compared to the KL distance of a power detector. Results show that the KL distance of the CSD is larger than the KL distance of a power detector. Monte Carlo simulations are included to support the derivation of the CSD statistics by generating Receiver Operating Characteristic (ROC) curves.

4.3.1 System Model

Consider the Alice, Bob, Willie model where Alice would like to communicate to Bob without Willie detecting her communications as shown in Figure 4.1. The parameters $d_{\text{a,w}}$ and $d_{\text{a,b}}$ represent the distances from Alice to Willie and from Alice to Bob respectively.

Willie's goal is to detect if Alice communicates over the time period $t \in [0, T]$. If Alice does not transmit, Willie observes real Gaussian noise $N^{(\text{w})}(t)$ at his receiver with power spectral density σ_{w}^2. If Alice transmits, she first encodes her message into BPSK symbols $\mathbf{f} = [f_1, f_2, \ldots, f_M]$ where M is the number of symbols and $E[|f_k|^2] = P_{\max}$. Employing $\mathbf{f}$, Alice transmits $x(t)$ which is a pulse shaped BPSK continuous-time signal:

$$x(t) = \sum_{k=-\infty}^{\infty} f_k p(t - kT_b) \tag{4.10}$$

where $p(t)$ is the square root raised cosine pulse and T_b is the symbol rate. Note that f_k is only non-zero for $k \in [1, \ldots, M]$ and assume that Alice transmits such that her entire message is transmitted during Willie's observation over $[0, T]$.

Let $z_{\text{pre}}(t)$ represent Willie's observation:

$$z_{\text{pre}}(t) = \begin{cases} x(t) + N^{(\text{w})}(t), & \text{Alice transmits} \\ N^{(\text{w})}(t), & \text{else,} \end{cases} \tag{4.11}$$

for $t \in [0, T]$. Analogously, let $y(t)$ represent Bob's observation:

$$y(t) = \begin{cases} x(t) + N^{(\text{b})}(t), & \text{Alice transmits} \\ N^{(\text{b})}(t), & \text{else,} \end{cases} \tag{4.12}$$

where $N^{(\text{b})}(t)$ is the noise Bob observes his receiver which is a zero-mean Gaussian process with power spectral density σ_{b}^2.

4.3.2 Deriving the CSD Statistics

Assume that Willie has knowledge of his receiver noise statistics as well as prior knowledge of Alice's transmission scheme including the symbol period T_b; however, he naturally does not know Alice's symbol timing and thus cannot employ a matched filter sampled at the proper time instances. However, Willie can detect if there is power observed at cyclic frequency $1/T_b$ by employing a CSD. Thus, a cyclic detector designed to detect cyclic frequencies at $\alpha = 1/T_b$ is presented in this section. The BPSK detector presented by Gardner is not used in this work because we assume a baseband signal. Willie's CSD presented in this section is also simpler than Gardner's proposed detector; however, the proposed detector follows a strategy similar to

Gardner's BPSK detector. An outline of the detector employed in this work is shown in Figure 4.5.

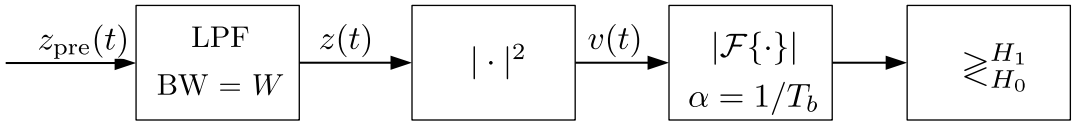

Figure 4.5. Willie's Cyclostationary Detector of baseband BPSK signals.

Willie observes the signal $z_{\text{pre}}(t)$ at his receiver which is his pre-filtered observation. Willie then passes $z_{\text{pre}}(t)$ through a wideband low-pass filter (LPF) with bandwidth W and W is chosen such that $W >> \alpha = 1/T_b$. The bandwidth is limited to reduce the noise entering the quadratic non-linearity, the importance of which is demonstrated in the noise analysis below. Willie then squares his filtered observation $z(t)$:

$$v(t) = z^2(t). \tag{4.13}$$

The frequency representation of $v(t)$ is then computed using the cosine transform:

$$F(\alpha) = \int_0^T v(t)\cos(2\pi\alpha t)dt = \int_0^T z^2(t)\cos(2\pi\alpha t)dt. \tag{4.14}$$

A cosine transform represented by $\mathcal{F}$ in Figure 4.5 is employed since Willie's observations are strictly real. Therefore, depending on whether Alice transmits or not, the frequency component $F(\alpha)$ is:

$$F(\alpha) = \begin{cases} \int_0^T (x(t) + N(t))^2 \cos(2\pi\alpha t)dt, & \text{Alice transmits} \\ \int_0^T N^2(t)\cos(2\pi\alpha t)dt, & \text{Alice does not transmit.} \end{cases} \tag{4.15}$$

where the superscript representing Willie's noise is dropped for convenience from here forward. Willie then measures the magnitude observed at cyclic frequency $F(\alpha = 1/T_b)$ to decide if Alice transmitted or not.

We now conduct a detailed analysis leveraging the wideband nature of the noise $N^{(\mathrm{w})}(t)$ relative to the bandwidth $1/T$ at the output of the frequency detector. First consider the H_0 case when Alice does not transmit. Under H_0, assume $F(\alpha)$ is a real Gaussian random variable. The mean of $F(\alpha)$ is:

$$E[F(\alpha)|H_0] = \int_0^T E[N^2(t)]\cos(2\pi\alpha t)dt, \tag{4.16}$$

$$= c\int_0^T \cos(2\pi\alpha t)dt, \tag{4.17}$$

$$\approx 0 \tag{4.18}$$

from some constant c, and (4.18) is true because $\int_0^T \cos(2\pi\alpha t)dt$ is small relative to the variance for large T. Note that, if $N^{(\mathrm{w})}(t)$ were truly white noise, $c = \infty$, but in reality $N^{(\mathrm{w})}(t)$ is very wideband relative to $1/T$. The variance of $F(\alpha|H_0)$ when Alice does not transmit is:

$$\mathrm{Var}[F(\alpha)|H_0] = E[(F(\alpha)|H_0)^2], \tag{4.19}$$

$$= E\left[\left(\int_0^T N^2(t)\cos(2\pi\alpha t)dt\right)\left(\int_0^T N^2(s)\cos(2\pi\alpha s)ds\right)\right], \tag{4.20}$$

$$= \int_0^T\int_0^T E[N^2(t)N^2(s)]\cos(2\pi\alpha t)\cos(2\pi\alpha s)\ dt\ ds. \tag{4.21}$$

The expected value of the product of four jointly Gaussian random variables is given by:

$$\mathbb{E}[N_1N_2N_3N_4] = E[N_1N_2]E[N_3N_4] + E[N_1N_3]E[N_2N_4] + E[N_1N_4]E[N_2N_3] - E[N_1]E[N_2]E[N_3]E[N_4]. \tag{4.22}$$

Thus,

$$E[N^2(t)N^2(s)] = R_N^2(0) + 2R_N^2(t-s) \tag{4.23}$$

where $R_N(\tau)$ is the ACF of Willie's noise and τ is the delay term. By employing (4.23), (4.21) is simplified:

$$\begin{aligned}
\mathrm{Var}[F(\alpha)|H_0] &= \int_0^T \int_0^T R_N^2(0)\cos(2\pi\alpha t)\cos(2\pi\alpha s)\ dt\ ds \\
&\qquad + 2\int_0^T \int_0^T R_N^2(t-s)\cos(2\pi\alpha t)\cos(2\pi\alpha s)\ dt\ ds, && (4.24)\\
&= R_N^2(0)\int_0^T \cos(2\pi\alpha t)dt \int_0^T \cos(2\pi\alpha s)ds \\
&\qquad + 2\int_0^T \int_0^T R_N^2(t-s)\cos^2(2\pi\alpha s)\ dt\ ds, && (4.25)\\
&= \int_0^T \int_0^T R_N^2(s-t)\ dt\ ds \\
&\qquad + \int_0^T \int_0^T R_N^2(s-t)\cos(4\pi\alpha(s-t+t))\ dt\ ds, && (4.26)\\
&= \int_0^T \int_0^T R_N^2(s-t)\ dt\ ds \\
&\qquad + \int_0^T \int_0^T R_N^2(s-t)\cos(4\pi\alpha(s-t))\cos(4\pi\alpha t)\ dt\ ds \\
&\qquad - \int_0^T \int_0^T R_N^2(s-t)\sin(4\pi\alpha(s-t))\sin(4\pi\alpha t)\ dt\ ds, && (4.27)\\
&= \int_0^T \left(\int_{-u}^{u} R_N^2(v)dv\right)du \\
&\qquad + \int_0^T \left(\int_{-u}^{u} R_N^2(v)\cos(4\pi\alpha v)dv\right)\cos(4\pi\alpha u)du \\
&\qquad - \int_0^T \left(\int_{-u}^{u} R_N^2(v)\sin(4\pi\alpha v)dv\right)\sin(4\pi\alpha u)du, && (4.28)\\
&= \tilde{\sigma}_{\mathrm{w}}^4 T && (4.29)
\end{aligned}$$

where the steps in (4.24)-(4.28) follow the same steps in Appendix A of [49]. The first term in (4.25) goes to zero because $\int_0^T \cos(2\pi\alpha t)dt \approx 0$, (4.26) follows from the half-angle formula trigonometric identity, (4.27) follows from the sum-difference formula and (4.28) is simplified by the assumption that $R_N(\tau) \approx 0$ for $\tau >> 1/W$. The last two terms in (4.28) go away because $\int_0^T \sin(4\pi\alpha u)du = 0$ and $\int_0^T \cos(4\pi\alpha u)du = 0$.

In (4.29), $\tilde{\sigma}_{\mathrm{w}}^2 = 2W\sigma_{\mathrm{w}}^2$ represents the noise power Willie observes after the low pass filter with bandwidth W.

The probability density function (pdf) of $F(\alpha)$ when Alice does not transmit is approximated as an AWGN random variable:

$$\mathbb{P}_{\mathrm{CSD},0} \triangleq \mathcal{N}(0, \tilde{\sigma}_{\mathrm{w}}^4 T). \tag{4.30}$$

Define $S = F(\alpha)/T$ as the normalized observation of Willie's power measurement over the length of time T. The pdf of S when Alice does not transmit is defined as:

$$P(S|H_0) \triangleq \mathcal{N}\left(0, \frac{\tilde{\sigma}_{\mathrm{w}}^4}{T}\right). \tag{4.31}$$

Similarly, when Alice transmits, $F(\alpha)$ is:

$$\begin{aligned} F(\alpha)|H_1 &= \int_0^T (x(t) + N(t))^2 \cos(2\pi\alpha t)dt, && (4.32)\\ &= \int_0^T x^2(t)\cos(2\pi\alpha t)dt + \int_0^T N^2(t)\cos(2\pi\alpha t)dt \\ &\qquad + 2\int_0^T x(t)N(t)\cos(2\pi\alpha t)dt. && (4.33) \end{aligned}$$

Equation (4.33) is then modeled as a Gaussian random variable. Define n_0 as:

$$n_0 = \int_0^T N^2(t)\cos(2\pi\alpha t)dt + 2\int_0^T x(t)N(t)\cos(2\pi\alpha t)dt \tag{4.34}$$

which represents the last two terms in (4.33) and assume n_0 follows an additive white Gaussian noise distribution. The expected value of n_0 is:

$$\begin{aligned} E[n_0] &= \int_0^T E[N^2(t)]\cos(2\pi\alpha t)dt + 2\int_0^T x(t)E[N(t)]\cos(2\pi\alpha t)dt, && (4.35)\\ &= R_N(0)\int_0^T \cos(2\pi\alpha t)dt, && (4.36)\\ &\approx 0. && (4.37) \end{aligned}$$

The variance of n_0 is then:

$$\begin{aligned}
\mathrm{Var}[n_0] &= E[n_0^2], && (4.38)\\
&= \int_0^T \int_0^T E\left[N^2(t)N^2(s)\right] \cos(2\pi\alpha t)\cos(2\pi\alpha s)\ dt\ ds \\
&\quad + 4\int_0^T \int_0^T x(t)x(s)E\left[N(t)N(s)\right]\cos(2\pi\alpha t)\cos(2\pi\alpha s)\ dt\ ds \\
&\quad + 4\int_0^T \int_0^T x(t)x(s)E[N^2(t)N(s)]\cos(2\pi\alpha t)\cos(2\pi\alpha s)\ dt\ ds. && (4.39)
\end{aligned}$$

The first term in (4.39) is equivalent to the variance under H_0 (4.21). The "signal cross noise" term in (4.39) is:

$$\begin{aligned}
&4\int_0^T \int_0^T x(t)x(s)E\left[N(t)N(s)\right]\cos(2\pi\alpha t)\cos(2\pi\alpha s)\ dt\ ds && (4.40)\\
&\qquad = 4\int_0^T \int_0^T R_N(t-s)x(t)x(s)\cos^2(2\pi\alpha s)\ dt\ ds, && (4.41)\\
&\qquad = 2\int_0^T \int_0^T R_N(t-s)x(t)x(s)\ dt\ ds \\
&\qquad\quad + 2\int_0^T \int_0^T R_N(t-s)x(t)x(s)\cos(4\pi\alpha s)\ dt\ ds, && (4.42)\\
&\qquad = 2\int_0^T \int_0^T R_N(t-s)\sum_{k=0}^{n-1} f_k^2 p(t-kT_b)p(s-kT_b)\ dt\ ds, && (4.43)\\
&\qquad = 2\sum_{k=0}^{n-1} f_k^2 \int_0^T \int_0^T R_N(t-s)p(t-kT_b)p(s-kT_b)\ dt\ ds, && (4.44)\\
&\qquad = 2\sum_{k=0}^{n-1} f_k^2 \int_0^T \int_0^T p(s)p(t)R_N(t-s)\ dt\ ds, && (4.45)\\
&\qquad = 2\sum_{k=0}^{n-1} f_k^2 \int_0^T p(s)\int_{-\infty}^{\infty} P(f)S_N(f)e^{j2\pi ft}\ df\ dt, && (4.46)\\
&\qquad = 2P_{\mathrm{f}} n \times \frac{1}{n}\tilde{\sigma}_{\mathrm{w}}^2, && (4.47)\\
&\qquad = 2P_{\mathrm{f}}\tilde{\sigma}_{\mathrm{w}}^2. && (4.48)
\end{aligned}$$

Note that (4.41) is slightly different from prior work in [49] because Alice transmits at baseband in this work. Equation (4.42) follows from employing the half-angle

trigonometric identity, the second term in (4.42) goes to zero based on the assumption that $\int_0^T \cos(4\pi\alpha s)ds = 0$. P_{f} represents the power observed in the symbols that Alice transmits. The last term in (4.39) goes to zero because when $t \neq s$, $E[N^2(t)N(s)] = 0$ and when $t = s$, $E[N^3(t)] = 0$.

Since we assume n_0 is AWGN, the distribution of $F(\alpha)$ under H_1 is modeled as:

$$\mathbb{P}_{\mathrm{CSD},1} \triangleq \mathcal{N}(\rho, \tilde{\sigma}_{\mathrm{w}}^4 T + 2\tilde{\sigma}_{\mathrm{w}}^2 P_{\mathrm{f}}) \tag{4.49}$$

where $\rho = \int_0^T x^2(t)\cos(2\pi\alpha t)dt$ and we assume that Willie is capable of computing ρ. If Willie normalizes his measurement by the length of time T, the statistics of his observation are defined by:

$$P(S|H_1) \triangleq \mathcal{N}\left(\frac{\rho}{T}, \frac{\tilde{\sigma}_{\mathrm{w}}^4}{T} + \frac{2\tilde{\sigma}_{\mathrm{w}}^2 P_{\mathrm{f}}}{T^2}\right). \tag{4.50}$$

A threshold detector is employed to determine if Willie's measurement S is classified as either the null or alternative hypothesis. Define γ as Willie's chosen threshold. Willie's probability of false alarm is then defined as the probability that his observation S is greater than γ when Alice does not transmit.

$$\begin{aligned}
\mathbb{P}_{\mathrm{FA}} &= P\left(|S| > \gamma|H_0\right), && (4.51)\\
&= P\left(S > \gamma|H_0\right) + P\left(S < -\gamma|H_0\right), && (4.52)\\
&= 2P\left(S > \gamma|H_0\right), && (4.53)\\
&= 2\int_{\gamma}^{\infty} f_{S|H_0}(t)dt, && (4.54)\\
&= 2Q\left(\frac{\gamma}{\sqrt{\tilde{\sigma}_{\mathrm{w}}^4/T}}\right), && (4.55)\\
&= \mathrm{erfc}\left(\frac{\gamma}{\sqrt{2\tilde{\sigma}_{\mathrm{w}}^4/T}}\right) && (4.56)
\end{aligned}$$

where (4.53) follows because $P(S|H_0)$ is centered at zero. Equation (4.55) follows from the definition of the Q-function:

$$Q\left(\frac{\gamma-\mu}{\sigma}\right) = Q(z) = \int_z^\infty f_S(t)dt \tag{4.57}$$

where μ is the mean and σ is the standard deviation of any normal pdf $f_S(s)$. (4.55) is then written in terms of the complementary error function, erfc$(\cdot)$, because Matlab has a built-in erfc$(\cdot)$ function by employing the definition $Q(z) = \frac{1}{2}\text{erfc}\left(\frac{z}{\sqrt{2}}\right)$ [50, Chapter 3.3].

Similarly, the probability of true detection, $\mathbb{P}_{TD}$, for any threshold γ is:

$$\begin{aligned}
\mathbb{P}_{TD} &= P(|S| > \gamma|H_1), && (4.58)\\
&= P(S < -\gamma|H_1) + P(S > \gamma|H_1), && (4.59)\\
&= 1 - Q\left(\frac{-\gamma-\rho}{\sqrt{\left(\frac{\tilde{\sigma}_{\text{w}}^4}{T} + \frac{2\tilde{\sigma}_{\text{w}}^2 P_{\text{f}}}{T^2}\right)}}\right) + Q\left(\frac{\gamma-\rho}{\sqrt{\left(\frac{\tilde{\sigma}_{\text{w}}^4}{T} + \frac{2\tilde{\sigma}_{\text{w}}^2 P_{\text{f}}}{T^2}\right)}}\right), && (4.60)\\
&= 1 - \frac{1}{2}\text{erfc}\left(\frac{-\gamma-\rho}{\sqrt{2\left(\frac{\tilde{\sigma}_{\text{w}}^4}{T} + \frac{2\tilde{\sigma}_{\text{w}}^2 P_{\text{f}}}{T^2}\right)}}\right) + \frac{1}{2}\text{erfc}\left(\frac{\gamma-\rho}{\sqrt{2\left(\frac{\tilde{\sigma}_{\text{w}}^4}{T} + \frac{2\tilde{\sigma}_{\text{w}}^2 P_{\text{f}}}{T^2}\right)}}\right). && (4.61)
\end{aligned}$$

4.3.3 Simulations

Simulations were conducted in Matlab to verify that the derived statistical approximations match Monte Carlo simulations. Pseudocode which outlines the Matlab script is presented in Appendix F. Five-hundred iterations of both the null hypothesis and the alternative hypothesis were generated using the same noise seed to maintain consistency. As described in the System Model of Section 4.3.1, Alice transmits BPSK symbols with a square root raised cosine pulse with roll-off factor 0.2. Alice's symbol frequency is $1/T_b = 0.699$ MHz. Since the simulations are generated in Matlab, the discretized version of Willie's observations are employed to model Willie's detector statistics and the oversampling rate is set to 100 MHz. Willie's detectors observe

4096 discrete samples. The CSD measures the power observed in $F(\alpha = 1/T_b)$ and his standard power detector measures the power in the discretized version of $z(t)$.

From [51], define $S = \frac{1}{N}\sum_{n=0}^{N-1} z[n]z[n]$ as the power detector test where $z[n]$ is Willie's observation at discrete sample n and N is the total number of discrete samples observed by Willie. S is AWGN when Alice does not transmit; thus, the statistics of S when Alice does not transmit is:

$$P(S|H_0) = \mathcal{N}\left(\tilde{\sigma}_{\mathrm{w}}^2, \frac{2\tilde{\sigma}_{\mathrm{w}}^4}{N}\right) \tag{4.62}$$

where $\tilde{\sigma}_{\mathrm{w}}^2$ represents the noise power Willie observes after filtering his original observation. Analogously, the statistics of Willie's power detector when Alice transmits is:

$$P(S|H_1) = \mathcal{N}\left(\tilde{\sigma}_{\mathrm{w}}^2 + P_{\mathrm{f}}, \frac{2\tilde{\sigma}_{\mathrm{w}}^4 + 4P_{\mathrm{f}}\tilde{\sigma}_{\mathrm{w}}^2}{N}\right). \tag{4.63}$$

The CSD and the power detector results are analyzed by employing Receiver Operating Characteristics (ROCs). ROCs demonstrate the efficiency of detectors by plotting the probability of true detection versus the probability of false alarm [39, Chapter 3.4]. Ineffective detectors generate results where the probability of true detection is equal to the probability of false alarm. Figure 4.6 shows the ROC results when Alice's signal-to-noise ratio (SNR) at Willie's receiver prior to the low-pass filter is -16 dB. The simulated results of the CSD and the power detector are shown along with the theoretical statistical performance of the detectors. As shown in Figure 4.6, the CSD significantly outperforms the power detector. These results demonstrate the importance of Willie designing detectors that exploit any unique features that may occur in Alice's transmitted signal. Note that the effective noise bandwidth must be carefully analyzed in simulations since a low pas filter is employed.

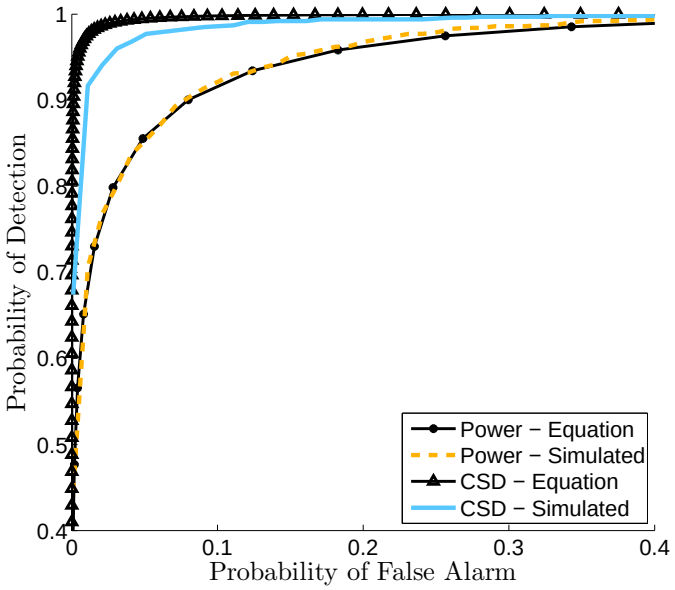

Figure 4.6. ROCs comparing the performance of a CSD and a power detector when the ratio of the power in Alice's signal to the noise power at Willie is -16 dB. The back line with the circle marker (–●–) corresponds to the known statistical performance rate of the power detector and the dashed orange line (– –) corresponds to the simulated detector results of the power detector. The black line with triangle markers (–△–) corresponds to the simulated ROC results of the CSD and the blue line (–) represents the derived performance rate based on the statistics of CSD derived in Section 4.3.2.

4.3.4 Kullback-Leibler Distance

The Kullback-Leibler (KL) distance of two Gaussian random variables p_0 and p_1 is [52, Chpt. 9.1]:

$$\begin{aligned}\mathcal{D}(p_0||p_1) &= \frac{1}{2}\log|\Sigma_1\Sigma_0^{-1}| \\ &\quad + \frac{1}{2}\mathrm{tr}(\Sigma^{-1}((\mu_0-\mu_1)(\mu_0-\mu_1)^T+\Sigma_0-\Sigma_1)), \qquad (4.64)\\ &= \frac{1}{2}\log\left(\frac{\sigma_1^2}{\sigma_0^2}\right)+\frac{(\mu_0-\mu_1)^2+\sigma_0^2-\sigma_1^2}{2\sigma_1^2}, \qquad (4.65)\\ &= \frac{1}{2}\log\left(\frac{\sigma_1^2}{\sigma_0^2}\right)+\frac{(\mu_0-\mu_1)^2+\sigma_0^2}{2\sigma_1^2}-\frac{1}{2} \qquad (4.66)\end{aligned}$$

where $\mathrm{tr}(\cdot)$ represents the trace of a matrix in (4.64). Let $\mathcal{D}_{\mathrm{PD}} = \mathcal{D}(\mathbb{P}_{\mathrm{PD},0}||\mathbb{P}_{\mathrm{PD},1})$ denote the KL distance between the null and alternative statistics of a power detector

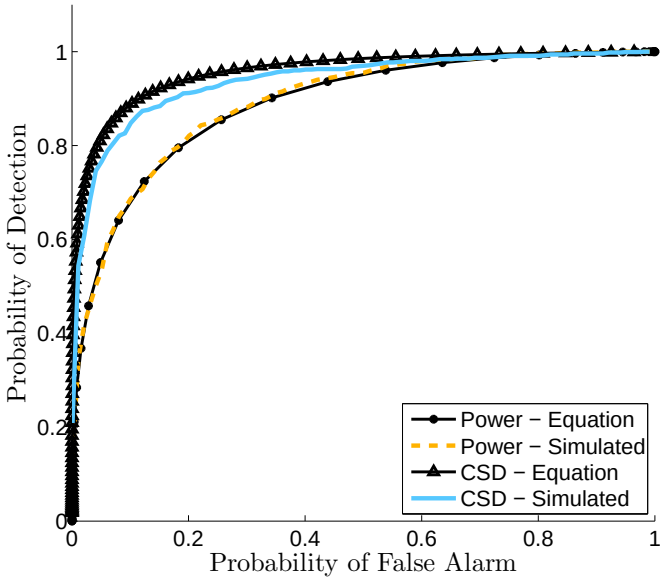

Figure 4.7. ROCs comparing the performance of a CSD and a power detector when the ratio of the power in Alice's signal to the noise power at Willie is -18 dB. The back line with the circle marker (–●–) corresponds to the known statistical performance rate of the power detector and the dashed orange line (−−) corresponds to the simulated detector results of the power detector. The black line with triangle markers (–△–) corresponds to the simulated ROC results of the CSD and the blue line (−) represents the derived performance rate based on the statistics of CSD derived in Section 4.3.2.

for any noise power σ_{w}^2:

$$\mathcal{D}_{\mathrm{PD}} = \frac{1}{2}\log\left(\frac{4P_{\mathrm{f}}\sigma_{\mathrm{w}}^2 + 2\sigma_{\mathrm{w}}^4}{2\sigma_{\mathrm{w}}^4}\right) + \frac{P_{\mathrm{f}}^2 + 2\sigma_{\mathrm{w}}^4/T}{(8P_{\mathrm{f}}\sigma_{\mathrm{w}}^2 + 2\sigma_{\mathrm{w}}^4)/T} - \frac{1}{2}, \tag{4.67}$$

$$= \frac{1}{2}\left[\log(2\sigma_{\mathrm{w}}^2 P_{\mathrm{f}} + \sigma_{\mathrm{w}}^4) - \log(\sigma_{\mathrm{w}}^4)\right] + \frac{P_{\mathrm{f}}^2 T}{8\sigma_{\mathrm{w}}^2 P_{\mathrm{f}} + 4\sigma_{\mathrm{w}}^4} + \frac{\sigma_{\mathrm{w}}^4}{4\sigma_{\mathrm{w}}^2 P_{\mathrm{f}} + 2\sigma_{\mathrm{w}}^4} - \frac{1}{2}. \tag{4.68}$$

Analogously, let $\mathcal{D}_{\mathrm{CSD}} = \mathcal{D}(\mathbb{P}_{\mathrm{CSD},0}||\mathbb{P}_{\mathrm{CSD},1})$ represent the KL distance between the null and alternative statistics when Willie employs a CSD:

$$\mathcal{D}_{\mathrm{CSD}} = \frac{1}{2}\left[\log(2\sigma_{\mathrm{w}}^2 P_{\mathrm{f}} + \sigma_{\mathrm{w}}^4) - \log(\sigma_{\mathrm{w}}^4)\right] + \frac{\rho^2 T}{4\sigma_{\mathrm{w}}^2 P_{\mathrm{f}} + 2\sigma_{\mathrm{w}}^4} + \frac{\sigma_{\mathrm{w}}^4}{4\sigma_{\mathrm{w}}^2 P_{\mathrm{f}} + 2\sigma_{\mathrm{w}}^4} - \frac{1}{2}. \tag{4.69}$$

Removing the common terms in the KL distance of the power detector (4.68) and the KL distance of the CSD (4.69), the term of interest in the power detector KL

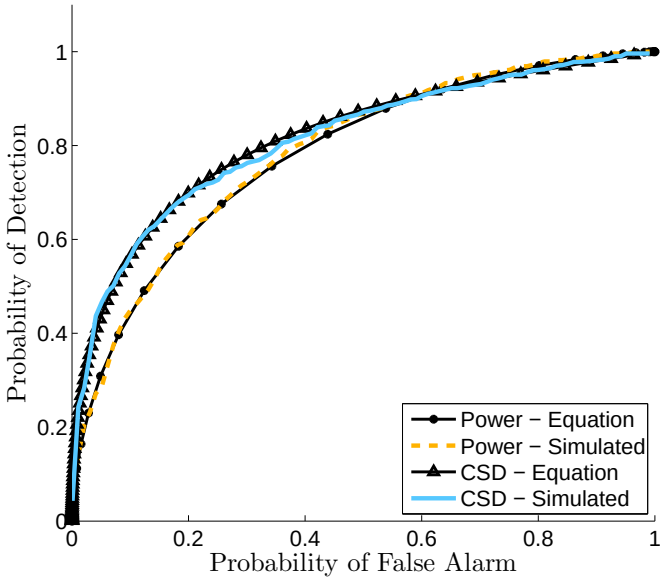

Figure 4.8. ROCs comparing the performance of a CSD and a power detector when the ratio of the Alice's SNR at Willie is -20 dB. The back line with the circle marker (–●–) corresponds to the known statistical performance rate of the power detector and the dashed orange line (– –) corresponds to the simulated detector results of the power detector. The black line with triangle markers (–△–) corresponds to the simulated ROC results of the CSD and the blue line (–) represents the derived performance rate based on the statistics of CSD derived in Section 4.3.2.

distance is:

$$\tilde{\mathcal{D}}_{\mathrm{PD}}(\mathbb{P}_0||\mathbb{P}_1) = \frac{P_{\mathrm{f}}^2 T}{8\sigma_{\mathrm{w}}^2 P_{\mathrm{f}} + 4\sigma_{\mathrm{w}}^4} \tag{4.70}$$

and the term of interest in the CSD KL distance is:

$$\tilde{\mathcal{D}}_{\mathrm{CSD}} = \frac{\rho^2 T}{4\sigma_{\mathrm{w}}^2 P_{\mathrm{f}} + 2\sigma_{\mathrm{w}}^4} \ . \tag{4.71}$$

The ratio of $\tilde{\mathcal{D}}_{\mathrm{CSD}}$ and $\tilde{\mathcal{D}}_{\mathrm{PD}}$ is:

$$\frac{\tilde{\mathcal{D}}_{\mathrm{CSD}}}{\tilde{\mathcal{D}}_{\mathrm{PD}}} = \frac{\rho^2}{P_k^2} \times \frac{2(4P_k\sigma_{\mathrm{w}}^2 + 2\sigma_{\mathrm{w}}^4)}{4P_k\sigma_{\mathrm{w}}^2 + 2\sigma_{\mathrm{w}}^4}, \tag{4.72}$$

$$= \frac{2\rho^2}{P_k^2} \ . \tag{4.73}$$

4.4 Summary

The classic Alice/Bob/Willie scenario considered in [1] is revisited in this chapter under the assumption that Alice transmits a baseband BPSK signal on a continuous-time channel. A cyclostationary detector (CSD) at Willie is proposed to detect Alice's signal when the channels between all entities are AWGN. The CSD detector is designed using the continuous-time model of Willie's observation instead of the discrete-time model. The KL distance of the proposed CSD is compared against the KL distance of a power detector. The KL distance results along with simulated results verify that the CSD outperforms the power detector and demonstrate that a power detector is not an optimal detector for the continuous-time model. Additionally, the equivalent discrete-time model is not an equivalent model because it does not allow for the consideration of all of Willie's observations at his receiver to help him detect Alice's transmissions.

4.5 Acknowledgment

The work in this chapter was sponsored by the National Science Foundation under grants ECCS-1309573 and CNS-1564067.

CHAPTER 5

COVERT COMMUNICATION ON THE CONTINUOUS-TIME MODEL: TIMING OFFSET DETECTORS

5.1 Introduction

Chapter 4 investigates the original Alice/Bob/Willie scenario and demonstrates the importance of considering the continuous-time model instead of the standard discrete-time model when analyzing covert communication. As another way to demonstrate the importance of considering continuous-time models, this chapter revisits the scenario presented in Chapter 2, which considered the discrete-time model. Per Chapter 2, Alice can communicate at a rate that does not scale with her codeword length when there is a jammer in the environment that adds uncertainty to Willie's observations. The uncertainty is added by either the jammer varying his/her power when the jammer-to-Willie channel is AWGN or when there is a fading channel between jammer and Willie which causes variations in the power Willie receives from the jammer. Alice leverages Willie's uncertainty to achieve covert communication at a positive rate.

The discrete-time model presented in Chapter 2 essentially assumes that transmissions arriving from Alice to Willie and arriving from the jammer to Willie are synchronized at Willie's receiver and that Willie samples at the symbol rate at the perfect time. In reality, timing offsets between Alice and the jammer are highly probable since they each use different hardware and are geographically separated without coordination. Although a jammer with varying transmit power in an AWGN channel may cause uncertainty in a power detector, the timing offsets provide Willie with

a unique feature that might mitigate any uncertainty in his received power. These features are not observable in the discrete-time model, but they are present in the continuous-time model.

The main contributions of this chapter are:

1. The considering of timing offsets and the presentation of a detector at Willie that is capable of detecting Alice based on her timing offset from the jammer.

2. A proposed method to overcome the timing offset detector by allowing Alice to vary her timing offset when she transmits.

5.2 System Model

Consider the scenario presented in Figure 5.1 where Alice would like to communicate covertly and reliably to Bob without Warden Willie detecting her communications. An uninformed jammer is also present in the environment and transmits with the same construction as Alice. In Figure 5.1, $d_{x,y}$ represents the distance from a transmitter to a receiver where x is either Alice ("a") or the jammer ("j") and y represents Willie ("w") or Bob ("b").

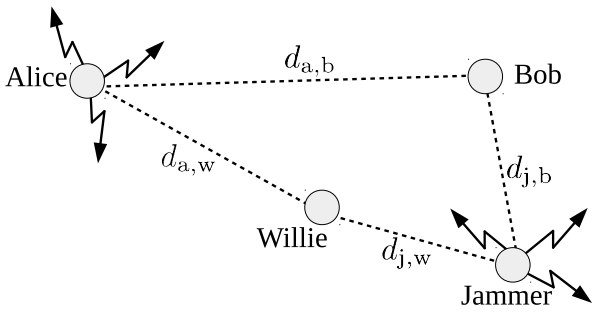

Figure 5.1. Wireless communication scenario with Alice, Bob, Willie and a jammer.

Define H_0 as the null hypothesis which represents Willie's assumption that Alice did not transmit. Let H_1 represent the alternative hypothesis which represents

Willie's assumption that Alice transmitted. Assume Willie observes M time slots each of length T, as shown in Figure 5.2. The jammer transmits n binary phase shift keying (BPSK) symbols $\mathbf{v}^{(s)} = \{v_1^{(s)}, v_2^{(s)}, \ldots, v_n^{(s)}\}$ in slot s. The jammer transmits $\mathbf{v}^{(s)}$ by pulse shaping the symbols with a square root raised cosine (SRRC), which is represented by $p(t)$. The jammer transmits with constant power that is limited to some maximum value, $E[|v_k^{(s)}|^2] = P_{\max}$. Define τ_{j} as the timing offset between the jammer's symbol time and time zero at Willie's receiver and assume that τ_{j} is constant over all M time slots. Let $g(t)$ represent the jammer's transmitted signal, $g(t) = \zeta \sum_{k=1}^{n} v_k^{(s)} p(t - kT_b - \tau_{\mathrm{j}})$ where T_b is the symbol period and ζ is the scaling coefficient used to control the power of the jammer's signal observed at Willie. Assume further that Willie is capable of estimating the timing offset τ_{j} with minimal error by observing all M slots.

If Alice chooses to transmit, she encodes her message into BPSK symbols $\mathbf{f} = \{f_1, f_2, \ldots, f_n\}$ in slot $s = 0$. Define τ_{a} as the timing offset between Alice's pulse $p(t)$ generated at her receiver and time zero. Also assume Alice's symbol period is the same symbol period employed by the jammer, T_b. Alice's transmitted signal is then $x(t) = \lambda \sum_{k=1}^{n} f_k p(t - kT_b - \tau_{\mathrm{a}})$ where λ is the scaling coefficient used to control the power of Alice's signal observed at Willie and $E[|f_k|^2] = P_{\max}$.

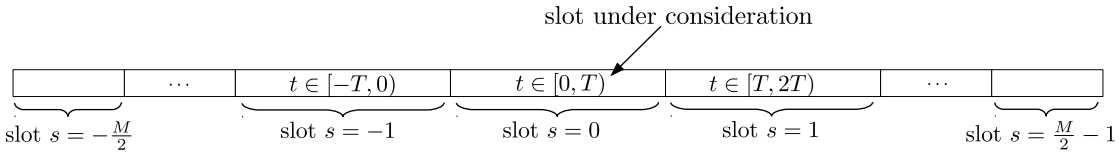

Figure 5.2. Continuous-time slot model diagram where each block contains the time period T and there are M total slots that Willie observes.

Consider Willie's observations in the time slot $s = 0$ and assume that from here forward, all variables apply to the time slot $s = 0$; thus, the index of the slot is dropped for simplicity. When Alice does not transmit, Willie observes:

$$z(t)|H_0 = \zeta \sum_{k=1}^{n} v_k p(t - kT_b - \tau_{\mathrm{j}}) + N^{(\mathrm{w})}(t) \tag{5.1}$$

where $N^{(\mathrm{w})}(t)$ is the noise Willie observes at his receiver which has mean zero and power spectral density $S_N^{(\mathrm{w})}(f) = \sigma_{\mathrm{w}}^2$. If Alice transmits, Willie observes:

$$z(t)|H_1 = \lambda \sum_{k=1}^{n} f_k p(t - kT_b - \tau_{\mathrm{a}}) + \zeta \sum_{k=1}^{n} v_k p(t - kT_b - \tau_{\mathrm{j}}) + N^{(\mathrm{w})}(t). \tag{5.2}$$

If Willie employs a power detector at this point in his receiver, the detector computes the measured power observed in $z(t)$ and compares the measurement to a threshold γ_1 to detect Alice's transmissions:

$$\frac{1}{T} \int_0^T z^2(t) dt \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_1. \tag{5.3}$$

Since Alice and the jammer employ the same symbol spacing, a cyclostationary detector designed to detect frequencies at $1/T_b$ as might be suggested in Chapter 4 has difficulty differentiating between the jammer's transmissions and Alice's transmission. In either the case of a power detector or a cyclostationary detector, Alice can still achieve a covert rate that does not decrease in the block length n. Thus, this chapter considers a detector at Willie that exploits the timing offsets τ_{a} and τ_{j}.

An outline of the system model is shown in Figure 5.3. Willie first passes his observation through a matched filter. Define $z_{\mathrm{mf}}(t) = z(t) * p(-t)$ as the matched filter's result:

$$z_{\mathrm{mf}}(t) = p(-t) * z(t), \tag{5.4}$$

$$= p(-t) * \left(\lambda \sum_{k=1}^{n} f_k p(t - kT_b - \tau_{\mathrm{a}}) + \zeta \sum_{k=1}^{n} v_k p(t - kT_b - \tau_{\mathrm{j}}) + N^{(\mathrm{w})}(t) \right), \tag{5.5}$$

$$= \lambda \sum_{k=1}^{n} f_k p(t - kT_b - \tau_{\mathrm{a}}) * p(-t) + \zeta \sum_{k=1}^{n} v_k p(t - kT_b - \tau_{\mathrm{j}}) * p(-t)$$

$$+ p(-t) * N^{(\mathrm{w})}(t), \tag{5.6}$$

$$= \lambda \sum_{k=1}^{n} f_k q(t - kT_b - \tau_{\mathrm{a}}) + \zeta \sum_{k=1}^{n} v_k q(t - kT_b - \tau_{\mathrm{j}}) + p(-t) * N^{(\mathrm{w})}(t) \tag{5.7}$$

where $q(t) = p(t) * p(-t)$ is the zero inter-symbol interference (ISI) raised cosine pulse.

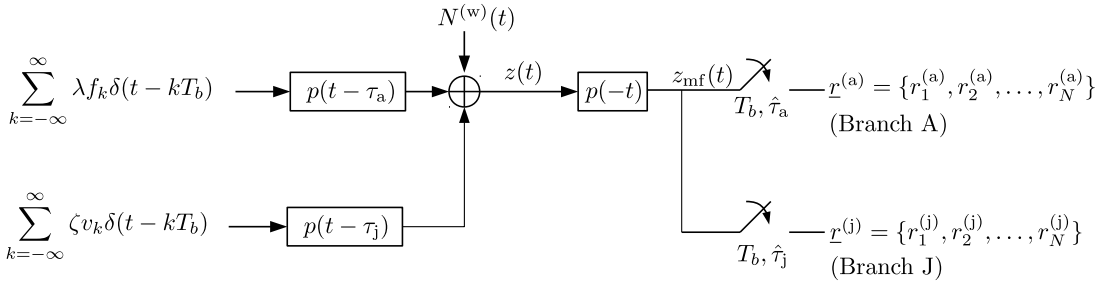

Figure 5.3. Alice/Jammer/Willie scenario where Willie samples his observations at both timing offsets $\hat{\tau}_{\mathrm{j}}$ and $\hat{\tau}_{\mathrm{a}}$.

Willie has two branches after the matched filter. The first branch (Branch A) captures the signal at Alice's estimated offset and the second branch (Branch J) captures the signal at the jammer's estimated offset. We assume that Willie knows the symbol period T_b and needs to form an estimate of the timing offsets for Alice and the Jammer. Let $\hat{\tau}_{\mathrm{a}}$ and $\hat{\tau}_{\mathrm{j}}$ represent the timing offset estimates of Alice and the jammer's transmissions respectively. This work assumes that Willie can estimate the jammer's offset with minimal error by observing all M time slots. This assumption is validated in simulation results shown in Section 5.3.1 and is based on the strategy proposed by Goeckel *et al.* in [23].

Let $\underline{r}^{(\mathrm{j})}$ represent the sampled values on Branch J in Figure 5.3 and assume that Willie observes N samples. Let $\underline{r}^{(\mathrm{a})}$ represent the sampled values Willie observes on Branch A which also contains N samples.

Consider first the matrix representation of $\underline{r}^{(\mathrm{a})}$:

$$\underline{r}^{(\mathrm{a})} = \begin{bmatrix} I_{N\times N} & Q_{N\times N} \end{bmatrix} \begin{bmatrix} \lambda \underline{f} \\ \zeta \underline{v} \end{bmatrix} + \underline{n}^{(\mathrm{a})} \tag{5.8}$$

where $I_{N\times N}$ is an N by N identity matrix because there is zero ISI between Alice's symbols in Branch A at Willie's receiver. $Q_{N\times N}$ is an N by N matrix that models interference from the jammer's signal and $\underline{n}^{(\mathrm{a})}$ represents the noise at Willie's receiver. The diagonal terms in $Q_{N\times N}$ correspond to interference from the jammer's current symbol given by $q(|\hat{\tau}_{\mathrm{a}} - \hat{\tau}_{\mathrm{j}}|)$. The terms that are off the diagonal in $Q_{N\times N}$ model interference from the neighboring symbols of the jammer's message. The number of neighboring terms that are non-zero around the diagonal depends on the pulse shape. As an example, assume that the pulse shape is such that Willie observes appreciable interference from three of the jammer's symbols at the k^{th} sample in $\underline{r}^{(\mathrm{a})}$ and that all other symbols are too far away to impact what Willie observes in $r_k^{(\mathrm{a})}$; then,

$$\begin{aligned} r_k^{(\mathrm{a})} = \lambda f_k + \zeta \Big[& v_{k-1} q(|\hat{\tau}_{\mathrm{a}} - \hat{\tau}_{\mathrm{j}}| - T_b) + v_k q(|\hat{\tau}_{\mathrm{a}} - \hat{\tau}_{\mathrm{j}}|) \\ & + v_{k+1} q(|\hat{\tau}_{\mathrm{a}} - \hat{\tau}_{\mathrm{j}}| + T_b) \Big] + n_k^{(\mathrm{a})} \end{aligned} \tag{5.9}$$

If (5.9) is modeled in $Q_{N\times N}$, the k^{th} row has column entries $[k-1, k, k+1]$ that are non-zero and all other entries in the row are zero. Adjustments need to be made in $Q_{N\times N}$ to the first few rows and last few rows of $Q_{N\times N}$ depending on the pulse shape because only N symbols are observed on each branch.

Next, the noise contribution in Branch A is analyzed. $N^{(\mathrm{w})}(t)$ is a zero-mean Gaussian random process with power spectral density $S_N(f) = \sigma_{\mathrm{w}}^2$. Define $N_{\mathrm{mf}}^{(\mathrm{w})}(t) =$

$p(-t) * N^{(\mathrm{w})}(t)$ and $\underline{n}^{(\mathrm{a})}$ as N T_b-spaced samples of $N_{\mathrm{mf}}^{(\mathrm{w})}(t)$. Therefore, the k^{th} sample in $\underline{n}^{(\mathrm{a})}$ is:

$$n_k^{(\mathrm{a})} = \int_{-\infty}^{\infty} p(t - kT_b) N^{(\mathrm{w})}(t). \tag{5.10}$$

The variables $n_1^{(\mathrm{a})}, n_2^{(\mathrm{a})}, \ldots, n_N^{(\mathrm{a})}$ are jointly Gaussian because linear operations on Gaussian random processes result in jointly Gaussian random variables. The expected value of any sample is:

$$E[n_k^{(\mathrm{a})}] = \int_{-\infty}^{\infty} p(t - kT_b) E[N^{(\mathrm{w})}(t)] dt \tag{5.11}$$

$$= 0 \tag{5.12}$$

where (5.12) follows because $N^{(\mathrm{w})}(t)$ is a zero-mean Gaussian random process. The correlation of the k^{th} and l^{th} entries in the sequence is:

$$E[n_k^{(\mathrm{a})} n_l^{(\mathrm{a})}] = E\left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(t - kT_b) p(s - lT_b) N^{(\mathrm{w})}(t) N^{(\mathrm{w})}(s) \; dt \; ds \right] \tag{5.13}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(t - kT_b) p(s - lT_b) E\left[N^{(\mathrm{w})}(t) N^{(\mathrm{w})}(s) \right] \; dt \; ds \tag{5.14}$$

$$= \int_{-\infty}^{\infty} p(t - kT_b) p(t - lT_b) \sigma_{\mathrm{w}}^2 \; dt \tag{5.15}$$

$$= \delta_{k,l} \sigma_{\mathrm{w}}^2 \tag{5.16}$$

where (5.15) follows because $E[N^{(\mathrm{w})}(t) N^{(\mathrm{w})}(s)] = \sigma_{\mathrm{w}}^2 \delta(t - s)$. The final result (5.16) is due to the zero ISI property of the raised cosine pulse.

Analogously, the set of sampled observations on the jammer's branch is defined as:

$$\underline{r}^{(\mathrm{j})} = \begin{bmatrix} Q_{N \times N} & I_{N \times N} \end{bmatrix} \begin{bmatrix} \lambda \underline{f} \\ \zeta \underline{v} \end{bmatrix} + \underline{n}^{(\mathrm{j})} \tag{5.17}$$

where Alice's message now acts like an interferer and $\underline{n}^{(\mathrm{j})}$ is an $N \times 1$ vector which represents the noise that Willie observes on Branch J. The expected value and the correlation of the jointly Gaussian sequence $\underline{n}^{(\mathrm{j})}$ are zero and $\delta_{k,l}\sigma_{\mathrm{w}}^2$ respectively.

Willie's observations $\underline{r}^{(\mathrm{a})}$ and $\underline{r}^{(\mathrm{j})}$ can be represented in matrix notation by stacking the matrices (5.8) and (5.17):

$$\begin{bmatrix} \underline{r}^{(\mathrm{a})} \\ \underline{r}^{(\mathrm{j})} \end{bmatrix} = A \begin{bmatrix} \lambda \underline{f} \\ \zeta \underline{v} \end{bmatrix} + \begin{bmatrix} \underline{n}^{(\mathrm{a})} \\ \underline{n}^{(\mathrm{j})} \end{bmatrix} \tag{5.18}$$

where A is a $2N \times 2N$ matrix:

$$A = \begin{bmatrix} I_{N\times N} & Q_{N\times N} \\ Q_{N\times N} & I_{N\times N} \end{bmatrix} \tag{5.19}$$

One plausible detector measures the power observed only in $\underline{r}^{(\mathrm{a})}$ which corresponds to the time samples which contain Alice's transmitted codewords. However, if the jammer's power is large, then the signal energy from the jammer's signal "leaks" into the observations Willie views in $\underline{r}^{(\mathrm{a})}$. However, Willie can estimate the original symbols transmitted by Alice and the jammer by employing the inverse of A:

$$\begin{bmatrix} \underline{b}^{(\mathrm{a})} \\ \underline{b}^{(\mathrm{j})} \end{bmatrix} = A^{-1} \begin{bmatrix} \underline{r}^{(\mathrm{a})} \\ \underline{r}^{(\mathrm{j})} \end{bmatrix} \tag{5.20}$$

where $\underline{b}^{(\mathrm{a})}$ is a $N \times 1$ vector which represents Willie's estimate of Alice's symbols $\lambda \underline{f}$ and $\underline{b}^{(\mathrm{j})}$ is a $N \times 1$ vector which represents Willie's estimate of Alice's symbols $\zeta \underline{v}$.

In particular, as long as A is full rank, which we show occurs if $\tau_{\mathrm{a}} \neq \tau_{\mathrm{j}}$, the inversion is possible, with the power of the residual noise related to the condition number of A. A well-conditioned matrix leads to less noise enhancement.

Next, a simulation is implemented to study the condition number of A when $p(t)$ is the impulse response of the square root raised cosine pulse [53, Chapter 6.7.1]:

$$p(t) = \left[\frac{\frac{4\beta t}{T_b}\cos((1+\beta)\frac{\pi t}{T_b}) + \sin((1-\beta)\frac{\pi t}{T_b})}{\frac{\pi t}{T_b}\left[1 - \left(\frac{4\beta t}{T_b}\right)^2\right]} \right]. \tag{5.21}$$

By employing the SRRC pulse, the combined response of the transmitter and match filter is a raised cosine pulse which produces pulses that have zero intersymbol interference between the symbols (Nyquist Zero-ISI criteria) [30, Chapter 3.3.1].

Figure 5.4 plots the condition number of A versus the timing offset percentage of Alice's pulse and the jammer's pulse, $|\tau_{\mathrm{a}} - \tau_{\mathrm{j}}|/T_b$. Figure 5.4 is generated by assuming Willie obtains 400 samples to construct both $\underline{r}^{(\mathrm{a})}$ and $\underline{r}^{(\mathrm{j})}$. Therefore, A is an 800 by 800 matrix that is constructed for various timing offsets between τ_{a} and τ_{j}. The symbol period is 75 discrete samples, the SRRC pulse's roll-off factor is .2 and a Figure of the SRRC pulse is shown in Figure 5.5. The condition number results show that A is reasonably well-conditioned when the timing offset between Alice and the jammer is greater than 10%. However, the offset at which A is well-conditioned depends on the pulse shape.

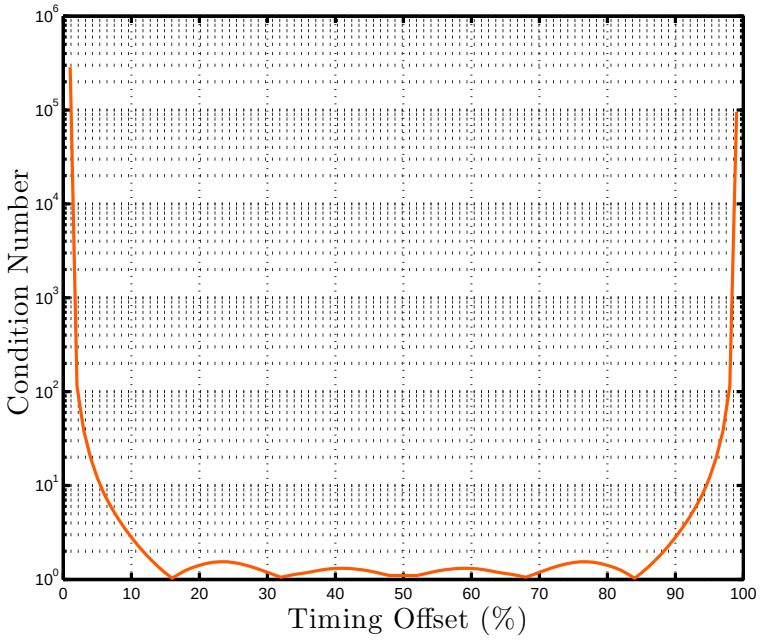

Figure 5.4. Condition number of matrix A (5.19) versus the timing offset between Alice's pulse and the jammer's pulse, $(|\tau_{\mathrm{a}} - \tau_{\mathrm{j}}|/T_b)100$.

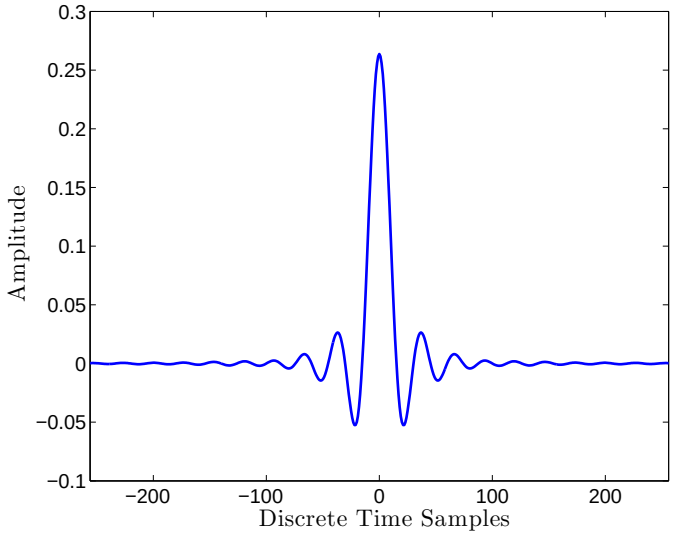

Figure 5.5. Simulated Square Root Raised Cosine (SRRC) pulse.

5.3 Simulation Results of Willie's Timing Offset Detector

5.3.1 Estimating Willie's Timing Offset, τ_{j}

The system model presented in Section 5.2 assumes Willie can estimate the timing offset τ_{j}. The process for such is outlined in this subsection. Since Willie has access to M time slots, assume that Willie considers $M-1$ slots and his initial assumption is that Alice transmits in the M^{th} slot. Willie can measure the power observed in $\underline{r}^{(\mathrm{j})}$ for various estimates of τ_{j} in each of the observed $M-1$ time slots. This method outlined in this work is exhaustive simply to demonstrate the possibility of such an estimation, but there are alternative methods to estimate the timing offset $\hat{\tau}_{\mathrm{j}}$ [47, Chapter 6.3].

Matlab simulations were generated to verify that this estimation process is possible. In simulations, the true jammer timing offset $\tau_{\mathrm{j}} = 0$ and Willie observes 1000 samples to construct $\underline{r}^{(\mathrm{j})}$ for discrete-time offsets $\hat{\tau}_{\mathrm{j}} = \{0, 1, 2, \ldots, 50\}$. Figure 5.6 plots the power observed in $\underline{r}^{(\mathrm{j})}$ for various $\hat{\tau}_{\mathrm{j}}$ values when the jammer-to-noise ratio is -5 dB. The power measured in $\underline{r}^{(\mathrm{j})}$ peaks when $\hat{\tau}_{\mathrm{j}} = 0$, which is the true timing offset. The simulation presented in Figure 5.6 and the symbol period is $T_b = 70$ discrete samples. However, even if the symbol spacing is smaller, Willie can still form an estimate of the timing offset. Figure 5.7 shows the power measured in $\underline{r}^{(\mathrm{j})}$ when the symbol spacing between the jammer's symbols is 16 discrete samples. Willie can still form an estimate of τ_{j} to maximize his observation of the jammer's signal.

5.3.2 Detecting Alice's Signal

Based on results shown in Section 5.3.1, Willie can form an estimate of τ_{j} and then perform detection. Therefore, the simulation results presented are under the assumption that Willie is performing detection in a single time slot. Before presenting Monte Carlo simulations, a single example is provided to justify measuring the power observed in $\underline{b}^{(\mathrm{a})}$ as a detector. Consider the same scenario presented in Section 5.3.1 with Willie's timing offset $\tau_{\mathrm{j}} = 0$. The jammer transmits BPSK symbols that are

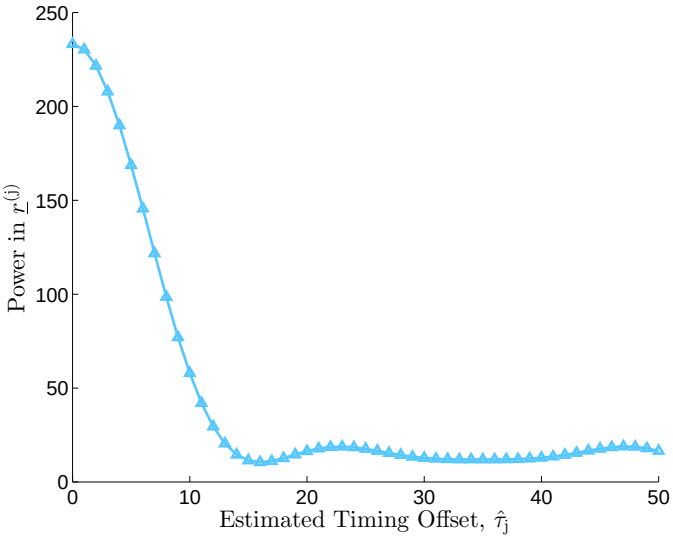

Figure 5.6. Power measured in $\underline{r}^{(j)}$ for different estimated timing offsets $\hat{\tau}_j$ when the jammer's symbol period is 70 discrete samples and the jammer-to-noise ratio is -5 dB.

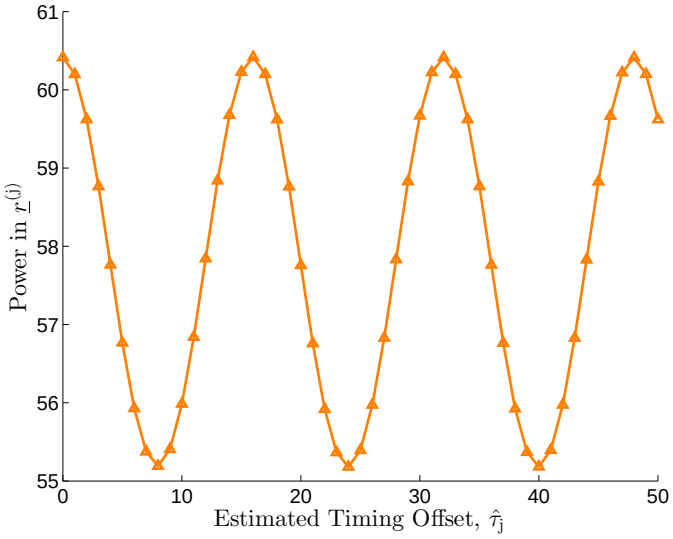

Figure 5.7. Power measured in $\underline{r}^{(j)}$ for different estimated timing offsets $\hat{\tau}_j$ when the jammer's symbol period is 16 discrete samples and the jammer-to-noise ratio is -5 dB.

pulse shaped using a SRRC pulse and the symbol period is 70 discrete samples. Alice also transmits BPSK symbols that have the same symbol period as the jammer and

Alice's true timing offset is $\tau_a = 30$. The SNR is -5 dB and the SJR is 0 dB; therefore, the jammer-to-noise ratio (JNR) is -5 dB as well.

In simulations, Willie uses 1000 samples to construct $\underline{r}^{(j)}$ based on the estimated timing offset $\hat{\tau}_j$ and we assume Willie estimates τ_a as a next step. Willie can employ a similar process used to estimate τ_j by measuring the power observed in $\underline{r}^{(a)}$ for various estimates of τ_a. Figure 5.8 shows the results of Willie's power observed in $\underline{r}^{(a)}$ for various $\hat{\tau}_a$ values based on 1000 samples to construct $\underline{r}^{(a)}$. The power measured in $\underline{r}^{(a)}$ peaks when $\hat{\tau}_a = 30$ discrete samples which is equivalent to a 42.8% timing offset. Therefore, assuming Willie has access to sufficient resources, Willie can measure the power observed in $\underline{r}^{(a)}$ to estimate τ_a.

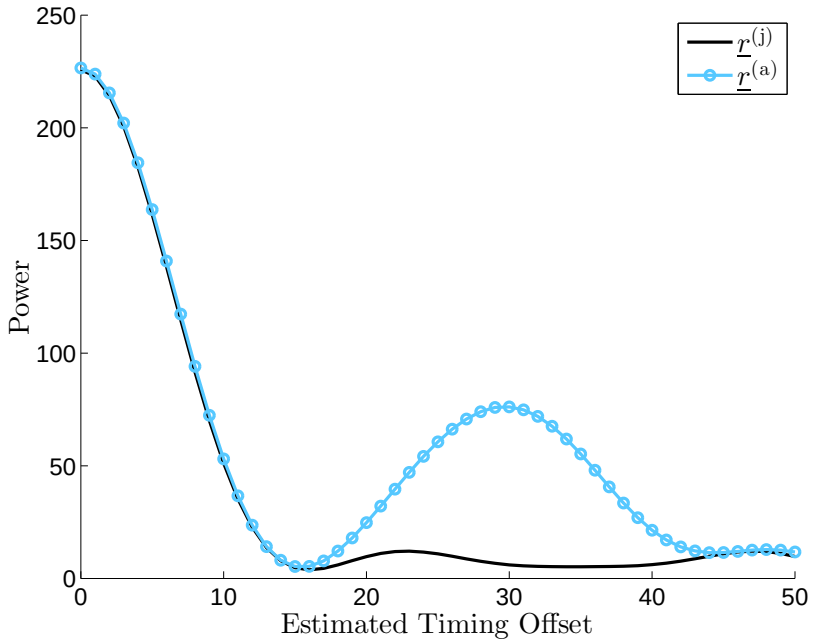

Figure 5.8. Simulated power observed in $\underline{r}^{(a)}$ (—○—) when Alice transmits and the timing offset between the jammer and Alice is 30 discrete samples (42.8% timing offset). The power observed in $\underline{r}^{(j)}$ when Alice does not transmit is also shown for comparison.

Monte Carlo simulations are generated to construct ROC curves based on the assumptions that Willie correctly estimates the timing offsets $\hat{\tau}_j = 0$ and $\hat{\tau}_a = 30$ dis-

crete samples. 800 iterations are generated and in each iteration power measurements of $\underline{b}^{(a)}$ are calculated when Alice does and does not transmit. Figure 5.9 shows the detection results when the SNR is -30dB, the SJR is -30dB and therefore the JNR is 0 dB. Another ROC is shown in Figure 5.10 when the SNR is -30 dB, the SJR is -10 dB and the JNR is -20 dB. The black line represents the standard power detector (5.3) and the blue line represents the timing offset based detector. These results demonstrate that Willie can detect Alice's transmissions with significantly greater success than the power detector by exploiting features which differ between Alice and the jammer.

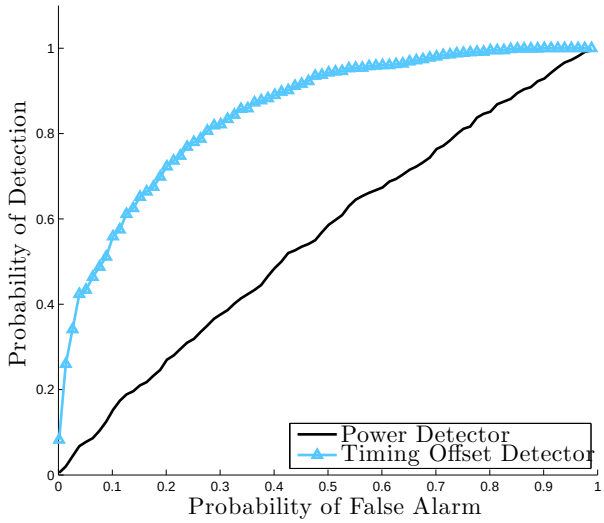

Figure 5.9. ROC detection results when Alice's timing offset is constant, SNR=-30 dB, SJR = -30 dB. The power detector (5.3) is represented by the black line (−) and the timing offset based detector is represented by the blue line (△).

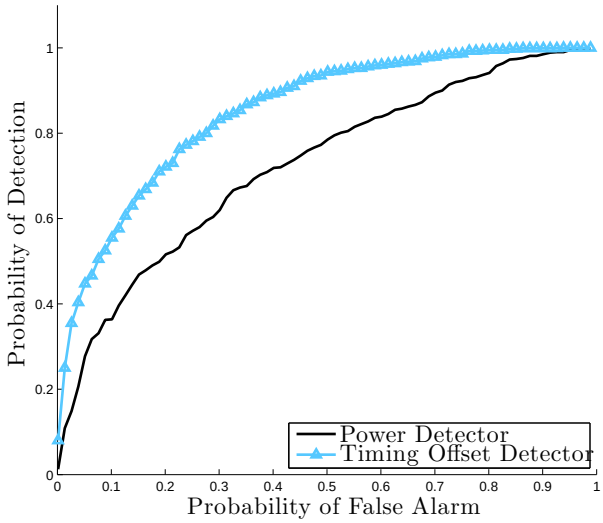

Figure 5.10. ROC detection results of a power detector (5.3) represented by the black line (−) and the detector of the timing offset based detector represented by the blue line (△).

5.4 Adding Uncertainty to the Timing Offset Scenario

This section proposes a different communication strategy at Alice so that it is harder for Willie to detect whether or not Alice is transmitting. The proposed method is similar to frequency hopping techniques that are employed to avoid detection by an adversary [30, Chapter 12].

Consider the system model presented in Section 5.2 and the timing slot model shown in Figure 5.2. The jammer transmits BPSK symbols using pulse shaping in all M time slots. Alice and Bob agree on a pre-assigned time slot to transmit which is unknown to Willie. Without loss of generality, assume that Alice and Bob agree to communicate in slot $s = 0$. In this section, Alice's transmission scheme changes such that she varies her timing offset instead of using a fixed timing offset. Also assume that Bob knows how Alice varies her timing offsets prior to communicating. Therefore, Alice's signal is $x(t) = \lambda \sum_{k=1}^{n} f_k p(t - T_b - \tau_{\mathrm{a},k})$ where $\tau_{\mathrm{a},k}$ represents the timing offset of the k^{th} symbol. Note that Alice does not have to change $\tau_{\mathrm{a},k}$ after

each symbol and the timing offset may be constant for a certain number of symbols that are transmitted successively. Also, note that such a strategy results in Alice incurring intersymbol interference (ISI) at Bob's receiver; however, such ISI does not affect the order of the scaling of the covert throughput.

Assume Willie's proposed detector is designed to measure the power observed in different timing offsets as shown in Section 5.2. This assumption is also true in the simulations implemented in Section 5.3.2. Willie's proposed detector measures the power observed in $\underline{b}^{(\mathrm{a})}$ as he varies his estimate of Alice's timing offset observed at his receiver for different estimates of her timing offset, $\hat{\tau}_{\mathrm{a}}$.

Before generating Monte Carlo simulations, a simple simulation is generated to demonstrate the benefits of Alice varying her timing offset. Assume that Alice varies her timing offset so the timing offset between the jammer and herself at Willie varies by 10, 20, 30 or 40 discrete samples. Alice randomly picks a new timing offset after transmitting every 100 symbols and all timing offsets are chosen with equal probability.

Following the steps outlined in Section 5.2, Willie constructs different sets of $\underline{r}^{(\mathrm{a})}$ by sampling his matched filter result with different timing offset values, $\hat{\tau}_{\mathrm{a}}$. As an example, Figure 5.11 plots the power Willie observes in $\underline{r}^{(\mathrm{a})}$ for different $\hat{\tau}_{\mathrm{a}}$ values when only Alice transmits. The results are generated assuming Willie uses 1000 samples to construct $\underline{r}^{(\mathrm{a})}$. Figure 5.11 also shows the power Willie observes when Alice's true timing offset at Willie is fixed $\tau_{\mathrm{a}} = 30$ for comparison. By varying her timing offset, the power Willie observes in $\underline{b}^{(\mathrm{a})}$ for any estimate $\hat{\tau}_{\mathrm{a}}$ is significantly reduced.

Next, Monte Carlo simulations were generated to observe the impact of Alice varying her timing offset on Willie's detection capabilities. Four detectors at Willie were constructed which measure the power Willie observes in $\underline{b}^{(\mathrm{a})}$ for $\hat{\tau}_{\mathrm{a}} = 10$, 20, 30 and 40 discrete samples. Willie constructs $\underline{r}^{(\mathrm{a})}$ and $\underline{r}^{(\mathrm{j})}$ using 1000 samples. Alice chooses a timing offset of 10, 20, 30 or 40 discrete samples every 100 symbols with

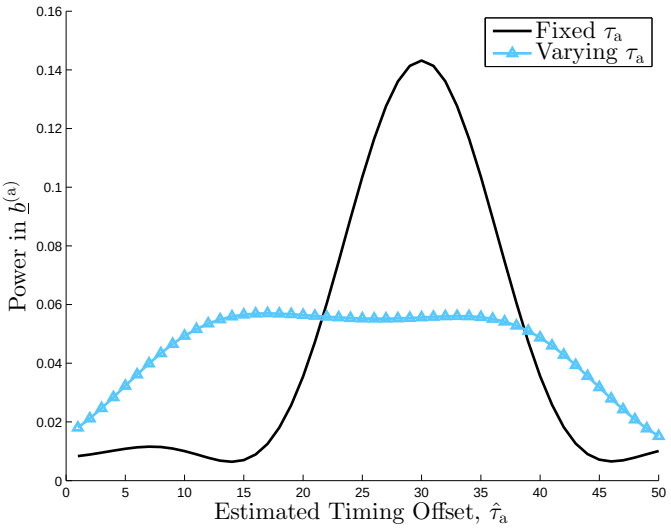

Figure 5.11. Willie's power in $\underline{b}^{(a)}$ when Willie samples his observation at various timing offsets and Alice does not vary her timing offset. The black line (–) represents the power measured in $\underline{b}^{(a)}$ when Alice's timing offset is fixed at $\tau_a = 30$ discrete samples and the blue line (–△–) represents when Alice varies her timing offset by 10, 20, 30 or 40 discrete samples with equal probability. The signal to noise ratio at Willie is -5 dB and the jammer does not transmit in the results shown.

equal probability. The symbol period employed by both Alice and the jammer is $T_b = 70$ discrete samples.

The first simulation assumes Alice transmits with a fixed timing offset, $\tau = 30$. Figure 5.12 shows the ROC curves comparing the different detectors that are constructed based on different estimates of $\hat{\tau}_a$. As shown in Section 5.3.2, Willie's detector performs well when $\hat{\tau}_a = 30$ and his other detectors do not detect any signal at $\hat{\tau}_a =$ 10, 20 and 40 as expected. Figure 5.13 shows the results of Willie's detectors when Alice varies her timing offset by 20 or 30 discrete samples with equal probability. Note that although the detector with $\hat{\tau}_a = 30$ appears to outperform the detector with $\hat{\tau}_a = 20$, these results are dependent on only 1000 samples. Thus, the performance of the detectors are expected to perform similarly as Willie observes larger sample sizes.

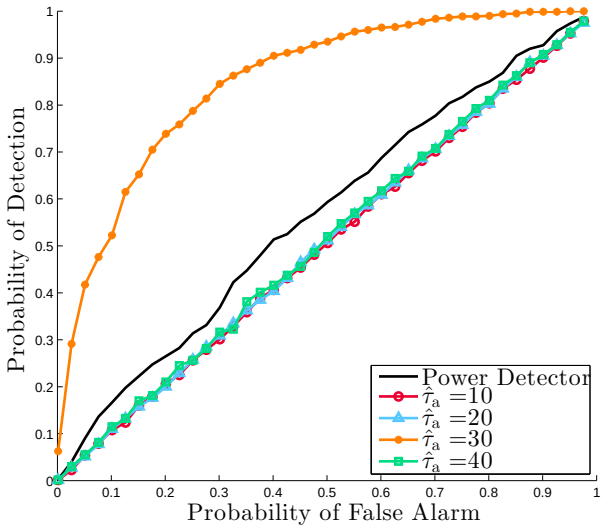

Figure 5.12. ROCs of timing offset detectors when Alice's offset is fixed, SNR = -30 dB, SJR = -30 dB. Willie assumes Alice's timing offset is fixed at $\hat{\tau}_a$ =10 (-○-), 20 (-△-), 30 (-●-) or 40 (-□-) discrete samples. Alice's offset, τ_a, is fixed at 30 discrete samples. Willie constructs $\underline{r}^{(a)}$ and $\underline{r}^{(j)}$ using 1000 samples.

Figure 5.14 shows the results of Willie's detector when Alice varies her timing offset by 10, 20, 30 and 40 samples with equal probability. The performance of each detector is degraded further by the fact that Alice is varying her timing offset. The pseudocode for the simulations presented in this section are detailed in Appendix G.

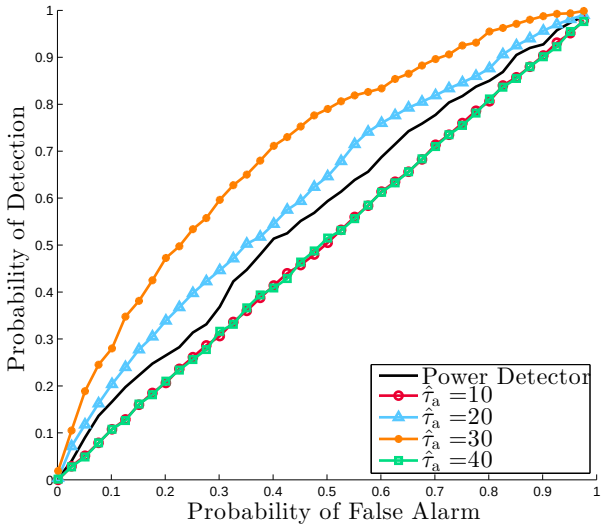

Figure 5.13. ROCs of timing offset detectors when Alice's timing varies slightly, SNR = -30 dB, SJR = -30 dB. Willie assumes that Alice's timing offset is fixed at $\hat{\tau}_a$ =10 (—○—), 20 (—△—), 30 (—●—) or 40 (—□—) discrete samples. The black line (−) represents the performance of a power detector at Willie. Alice varies her offset by either 20 or 30 discrete samples with equal probability. Willie constructs $\underline{r}^{(a)}$ and $\underline{r}^{(j)}$ using 1000 samples.

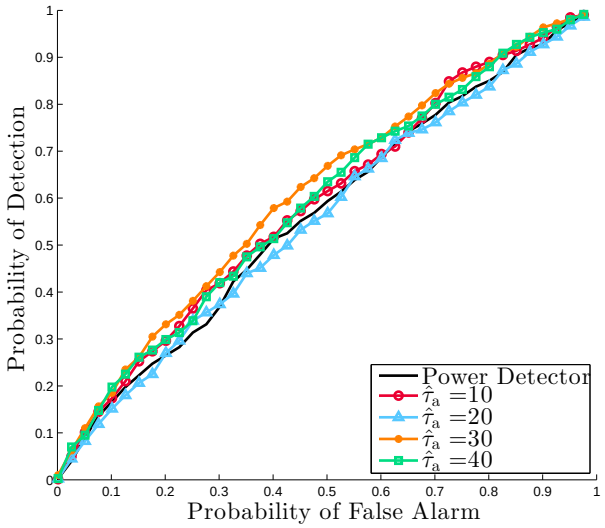

Figure 5.14. ROCs of timing detectors when Alice's offset various significantly, SNR = -30 dB, SJR = -30 dB. Willie assumes that Alice's timing offset is fixed at $\hat{\tau}_a$ =10 (-○-), 20 (-△-), 30 (-●-) or 40 (-□-) discrete samples. The black line (−) represents the performance of a power detector at Willie. Alice varies her offset by either 10, 20, 30 or 40 discrete samples with equal probability. Willie constructs $\underline{r}^{(a)}$ and $\underline{r}^{(j)}$ using 1000 samples.

5.5 Summary

This chapter revisited the Alice/Bob/Willie/Jammer scenario presented in Chapter 2 to investigate the impact of a detector designed based on the continuous-time model instead of the discrete-time model. This is done by assuming that both Alice and the jammer transmit BPSK signals and that the timing of the pulses are such that the symbol boundaries are not aligned at Willie, as would be the case in practice. Simulation results demonstrate that if Alice's timing offset is constant, Willie can collect samples and project away from the jammer, thus mitigating its impact. If Alice varies her timing offsets, she can mitigate such an attack at the expense of Bob's observations incurring intersymbol interference. Thus, the results provide general insight about how Alice can achieve covert communication and how Willie can thwart Alice's ability to remain covert on the true continuous-time channel. Essentially, Alice should transmit messages which resemble Willie's expected background noise statistics for all detectors. Analogously, Willie should leverage any unique features about Alice's transmissions that are different from his expected background noise to design detectors.

5.6 Acknowledgment

The work in this chapter was sponsored by the National Science Foundation under grant CNS-1564067.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

The covert communication research presented in this work builds upon prior research to consider dynamic channel models. The second chapter considered covert communications with the addition of a jammer in the standard AWGN channel model, with single block fading, or with a finite M-block fading model. Results prove that Alice can achieve positive rate covert communication even when Willie employs an optimal detector with the addition of the jammer. The third chapter generalizes the covert communication problem to consider an arbitrary number of fading blocks over the duration of Alice's transmission. The number of fading variations is defined as a potentially increasing function of the codeword length, and thus the number of fading blocks is not finite as in Chapter 2. A converse for covert communication is provided.

Chapter 4 revisits the Alice/Bob/Willie covert communication scenario using the continuous-time model instead of the previously considered discrete-time model. Results presented in Chapter 2 prove that a power detector is an optimal detector for Willie for many discrete-time models, including AWGN and single-block fading. However, results in Chapter 4 demonstrate the importance of considering the continuous-time model: the results derived based on the discrete-time model in Chapter 2 are only valid if Willie samples his continuous-time observations at exactly the right time instances. In particular, Chapter 4 demonstrates the importance of the continuous-time model by presenting cyclostationary detectors and demonstrating how their performance outperforms a standard power detector. This work demonstrates that limiting Alice's transmit power is not the only consideration. Alice's transmissions should not

contain any features that differ significantly from Willie's observations when Alice does not transmit.

The Alice/Bob/Willie/Jammer scenario presented in Chapter 2 is revisited in the fifth chapter under a continuous-time model. If Alice and the jammer are unsynchronized, then Willie can exploit their relative time offsets to construct detectors that significantly outperform a power detector and even change the scaling behavior. However, if Alice varies her timing offset, she can reduce the chances of Willie detecting her communications, in fact, causing the timing offset detector to perform similarly as a power detector in the continuous-time model.

As covert communication research continuous to develop, the scenarios under which covert communication are analyzed are also expected to become more complex. Some areas of future work include:

1. Additional strategies Willie can exploit to detect Alice's transmissions with high probability. For example, if Willie employs an antenna array in the AWGN channel model, even if the jammer transmits, Willie can differentiate between the two transmitters by employing an antenna array. If Willie knows in advance where the jammer is located, then Willie can use this knowledge to detect if power observed from other directions is greater than some threshold. The work presented in Chapter 3 provides some insight about what channel conditions can generate uncertainty when Willie employs an antenna array.

2. Future work should also consider how well Bob can reconstruct Alice's messages under more dynamic channel conditions and when Alice changes her transmission strategy.

3. There is already some early work by Yan *et al.* in [54] which determine Alice's exact rate when her codeword length is finite. The proofs presented in this work assume asymptotic conditions. Future work should also consider Alice's exact

rate in dynamic channel conditions based on the models presented in Chapter 5.

APPENDIX A

PROOF OF THEOREM 2

Construction: Alice and the jammer employ the same methods as described in the construction of Lemma 3 in Section 2.3.2. Hence, Willie is aware that the channel gain between the jammer and himself results in σ_{j}^2 being distributed as an exponential random variable with mean ζ. If the Alice-to-Willie channel is AWGN, Lemma 3 establishes that the optimal receiver for Willie to employ is a power detector $Z \gtrless_{H_0}^{H_1} \Gamma_n$ for some threshold Γ_n on the slot of size n, or, equivalently,

$$\frac{Z}{n} \underset{H_0}{\overset{H_1}{\gtrless}} \tau_n, \tag{A.1}$$

where $\tau_n \equiv \Gamma_n/n$. If the Alice-to-Willie channel is an $M = 1$ block fading channel, we assume pessimistically that Willie also knows the value of $h_{0,1}^{(\mathrm{a,w})}$. Then, Corollary 3.1 establishes that the optimal receiver for Willie is again the power detector in (A.1).

Analysis: Consider first the case when the Alice-to-Willie channel is an AWGN channel. Recall that we require $\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} > 1 - \epsilon$ for any $\epsilon > 0$. Thus, consider any $\epsilon > 0$. The unboundedness of the support of σ_{j}^2 requires a slight modification of the proof technique of Theorem 1 in Section 2.3.1. Thus, note that there exists some constant c such that:

$$P(\sigma_{\mathrm{j}}^2 > c) < \frac{\epsilon}{4} \ . \tag{A.2}$$

Consider first the false alarm rate, and, analogously to the proof of Theorem 1, define:

$$\mathbb{P}_{\mathrm{FA}}(u) = P\left(\frac{Z}{n} \geq \tau_n | \sigma_{\mathrm{j}}^2 = u, H_0\right). \tag{A.3}$$

Under H_0, $Z/n = (\sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{j}}^2)\chi_{2n}^2/n$. By the weak law of large numbers, χ_{2n}^2/n converges in probability to 1; hence, for any $\delta > 0$, $\exists N_0$ (not dependent on u) such that, for $n \geq N_0$,

$$P\left(\frac{\chi_{2n}^2}{n} \in \left(1 - \frac{\delta}{\sigma_{\mathrm{w}}^2 + c}, 1 + \frac{\delta}{\sigma_{\mathrm{w}}^2 + c}\right)\right) > 1 - \frac{\epsilon}{2}\,. \tag{A.4}$$

Hence, for any $n > N_0$,

$$P\left(\frac{Z}{n} \in \left((\sigma_{\mathrm{w}}^2 + u)\left(1 - \frac{\delta}{\sigma_{\mathrm{w}}^2 + c}\right), (\sigma_{\mathrm{w}}^2 + u)\left(1 + \frac{\delta}{\sigma_{\mathrm{w}}^2 + c}\right)\right)\right) > 1 - \frac{\epsilon}{2}\,. \tag{A.5}$$

Now, for any $u \leq c$, $\sigma_{\mathrm{w}}^2 + u < \sigma_{\mathrm{w}}^2 + c$ and thus for any $n > N_0$:

$$P\left(\frac{Z}{n} \in \left(\sigma_{\mathrm{w}}^2 + u - \delta, \sigma_{\mathrm{w}}^2 + u + \delta\right)\right) > 1 - \frac{\epsilon}{2} \tag{A.6}$$

and thus $\mathbb{P}_{\mathrm{FA}}(u) \geq 1 - \epsilon/2$ for any $\tau_n < \sigma_{\mathrm{w}}^2 + u - \delta$ as long as $u < c$. Likewise, following analogous arguments, $\exists N_1$ such that, for any $n > N_1$ (not dependent on u):

$$\mathbb{P}_{\mathrm{MD}}(u) = P\left(\frac{Z}{n} \leq \tau_n | \sigma_{\mathrm{j}}^2 = u, H_1\right) > 1 - \frac{\epsilon}{2} \tag{A.7}$$

for any $\tau_n > \sigma_{\mathrm{w}}^2 + u + \sigma_{\mathrm{a}}^2 + \delta$, as long as $u < c$. Combining these results yields that for any $n > \max(N_0, N_1)$:

$$\mathbb{P}_{\mathrm{FA}}(u) + \mathbb{P}_{\mathrm{MD}}(u) \geq 1 - \frac{\epsilon}{2} \tag{A.8}$$

unless $\{u > c\}$ or $u \in \mathcal{A} = \{\sigma_{\mathrm{w}}^2 + u - \delta < \tau_n < \sigma_{\mathrm{w}}^2 + u + \sigma_{\mathrm{a}}^2 + \delta\}$. Now,

$$P(\mathcal{A}) = P(\tau_n - \delta - \sigma_{\mathrm{a}}^2 - \sigma_{\mathrm{w}}^2 < U < \tau_n + \delta - \sigma_{\mathrm{w}}^2), \tag{A.9}$$

$$\leq \frac{\sigma_{\mathrm{a}}^2 + 2\delta}{\zeta} \tag{A.10}$$

where the last line follows by upper bounding the probability density function of σ_{j}^2. A choice of $\delta = \zeta\epsilon/16$ and $\sigma_{\mathrm{a}}^2 = \zeta\epsilon/8$ yields, via the Union Bound:

$$P(\mathcal{A}^c \cap \{\sigma_{\mathrm{j}}^2 \leq c\}) \geq 1 - \frac{\epsilon}{2} \tag{A.11}$$

and then the proof follows analogously to the end of that of Theorem 1. This completes the proof for the case that the Alice-to-Willie channel is an AWGN channel.

Next, consider the case when the Alice-to-Willie channel is a $M = 1$ block fading channel. Let $\epsilon_2 > 0$ be the covertness constraint and set $\epsilon = \epsilon_2/2$. Choose $\tilde{\sigma}_{\mathrm{a}}^2$ according to the AWGN case above such that Alice is covert if the average received power at Willie is $\tilde{\sigma}_{\mathrm{a}}^2$. Finally, choose P_{f} such that:

$$P(\sigma_{\mathrm{a}}^2 < \tilde{\sigma}_{\mathrm{a}}^2) > 1 - \frac{\epsilon_2}{2}. \tag{A.12}$$

Then, Alice can employ (constant) power P_{f} and satisfy the covertness constraint for any $\epsilon > 0$. ∎

APPENDIX B

PROOF OF o(n) COVERT BITS TRANSMITTED FOR M = 1

Consider the assumptions of the $M = 1$ fading model and Alice's construction in Section 2.2.1.1 but with the jammer transmitting Gaussian noise drawn from a distribution with constant variance. If fading channels exist between all parties, there exists a covert communication strategy s.t. Bob can reliably decode Alice's messages if she transmits $o(n)$ bits in n channel uses.

Proof: By Theorem 2, Alice can transmit with $P_{\mathrm{f}} > 0$ not dependent on n while remaining covert. What remains is to demonstrate that Bob can decode the transmission with probability of error less than δ for any $\delta > 0$. Conditioned on the fading variables $h^{(\mathrm{a,b})}$, $h^{(\mathrm{j,b})}$, the channel from Alice to Bob is an AWGN channel with signal-to-noise ratio:

$$\gamma = \frac{|h^{(\mathrm{a,b})}|^2 \frac{P_{\mathrm{f}}}{d_{\mathrm{a,b}}^{\alpha}}}{|h^{(\mathrm{j,b})}|^2 \frac{P_{\mathrm{j}}}{d_{\mathrm{j,b}}^{\alpha}} + \sigma_{\mathrm{b}}^2} \ . \tag{B.1}$$

Hence, given the distributions of $h^{(\mathrm{a,b})}$ and $h^{(\mathrm{j,b})}$, there exists a constant rate R such that the probability that γ is large enough to support communication with reliability greater than $1 - \frac{\delta}{2}$ at rate R is greater than $1 - \frac{\delta}{2}$ (R is the $\frac{\delta}{2}$-outage capacity [37], which is non-zero). Since $o(n) < nR$ for all $n > N_0$ for some N_0, the result follows. ∎

APPENDIX C

PROOF OF INCREASING $\Lambda(\mathbf{Z})$ FOR THE $M = 1$ CASE FOR THE PROOF OF LEMMA 4

Let $\zeta = P_{\mathrm{j}}/d_{\mathrm{j,w}}^{\alpha}$. Hence, in the fading model, the received jammer power σ_{j}^2 is exponentially distributed with mean ζ. As in Section 2.3.1, since the $t = 0$ slot is the slot of interest, observations outside of $k = 1, 2, \ldots, n$ do not help Willie to detect a transmission by Alice in slot $t = 0$. Hence, it is sufficient to consider $\mathbf{Z}_0$ as the input to Willie's receiver. As in Section 2.3.1, we therefore suppress the slot index and denote Willie's observation by $\mathbf{Z} = [Z_1, Z_2, \ldots, Z_n]$. It is then readily established that $Z = \sum_{i=1}^{n} |Z_i|^2$ is a sufficient statistic, with distribution under H_0 given by:

$$\begin{aligned} f_{Z|H_0}(z|H_0) &= E_{\sigma_{\mathrm{j}}^2}\left[\left(\frac{1}{\pi(\sigma_{\mathrm{j}}^2+\sigma_{\mathrm{w}}^2)}\right)^n \exp\left(-\frac{z}{(\sigma_{\mathrm{j}}^2+\sigma_{\mathrm{w}}^2)}\right)\right], \\ &= \frac{1}{\pi^n}\int_0^{\infty}\left(\frac{1}{u+\sigma_{\mathrm{w}}^2}\right)^n e^{-\frac{z}{(u+\sigma_{\mathrm{w}}^2)}} e^{-\frac{u}{\zeta}} du, \\ &= \frac{e^{\frac{\sigma_{\mathrm{w}}^2}{\zeta}}}{\pi^n}\int_{\sigma_{\mathrm{w}}^2}^{\infty}\left(\frac{1}{v}\right)^n e^{-\frac{z}{v}} e^{-\frac{v}{\zeta}} dv. \end{aligned} \tag{C.1}$$

Via analogous arguments, the distribution when Alice transmits is:

$$f_{Z|H_1}(z|H_1) = \frac{e^{\frac{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}{\zeta}}}{\pi^n}\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty}\left(\frac{1}{v}\right)^n e^{-\frac{z}{v}} e^{-\frac{v}{\zeta}} dv. \tag{C.2}$$

Hence, in this case the optimal decision rule for Willie becomes:

$$\Lambda(Z) = \frac{e^{\frac{\sigma_{\mathrm{a}}^2}{\zeta}}\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty}\left(\frac{1}{v}\right)^n e^{-\frac{Z}{v}} e^{-v/\zeta} dv}{\int_{\sigma_{\mathrm{w}}^2}^{\infty}\left(\frac{1}{v}\right)^n e^{-\frac{Z}{v}} e^{-\frac{v}{\zeta}} dv} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{C.3}$$

Now, consider any observation $Z = z^{(0)}$ that falls on the boundary between the decision regions:

$$\Lambda(z^{(0)}) = \frac{e^{\frac{\sigma_{\mathrm{a}}^2}{\zeta}} \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv}{\int_{\sigma_{\mathrm{w}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv} = \gamma, \tag{C.4}$$

and consider the LRT when Willie observes $z^{(0)} + \Delta$:

$$\Lambda(z^{(0)} + \Delta) = \frac{e^{\frac{\sigma_{\mathrm{a}}^2}{\zeta}} \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv}{\int_{\sigma_{\mathrm{w}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv} \ . \tag{C.5}$$

The common integration term in the numerator and denominator of (C.5) is extracted to yield:

$$\Lambda(z^{(0)} + \Delta) = \frac{e^{\frac{\sigma_{\mathrm{a}}^2}{\zeta}} \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv}{\int_{\sigma_{\mathrm{w}}^2}^{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv + \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv}. \tag{C.6}$$

Next, (C.6) is normalized by the common integration range $\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv$ to yield:

$$\Lambda(z^{(0)} + \Delta) = \frac{e^{\frac{\sigma_{\mathrm{a}}^2}{\zeta}}}{\frac{\int_{\sigma_{\mathrm{w}}^2}^{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv}{\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv} + 1}. \tag{C.7}$$

The Second Mean Value Theorem [55, Chapter 4.7] implies that $\exists c_1 \in (\sigma_{\mathrm{w}}^2, \sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{a}}^2)$ such that:

$$e^{-\frac{\Delta}{c_1}} \int_{\sigma_{\mathrm{w}}^2}^{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv = \int_{\sigma_{\mathrm{w}}^2}^{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv. \tag{C.8}$$

Similarly, because $e^{-\frac{\Delta}{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}} \leq e^{-\frac{\Delta}{v}} \leq 1$ for $v \in [\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2, \infty)$,

$$e^{-\frac{\Delta}{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}} \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv \leq \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv, \tag{C.9}$$

$$\leq \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv \tag{C.10}$$

which implies:

$$e^{-\frac{\Delta}{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}} \leq \frac{\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv}{\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv} \leq 1. \tag{C.11}$$

Hence, the ratio of the integrals in (C.11) is either equal to one, or $\exists c_2 \in [\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2, \infty)$ such that:

$$e^{-\frac{\Delta}{c_2}} \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^{\frac{n}{M}} e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv = \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^{\frac{n}{M}} e^{-\frac{(z^{(0)}+\Delta)}{v}} e^{-\frac{v}{\zeta}} dv. \tag{C.12}$$

If there exists such a $c_2 \in [\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2, \infty)$, then:

$$\Lambda(z^{(0)}+\Delta) = \frac{e^{\frac{\sigma_{\mathrm{a}}^2}{\zeta}}}{\frac{e^{-\frac{\Delta}{c_1}} \int_{\sigma_{\mathrm{w}}^2}^{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv}{e^{-\frac{\Delta}{c_2}} \int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv} + 1}, \tag{C.13}$$

$$> \frac{e^{\frac{\sigma_{\mathrm{a}}^2}{\zeta}}}{\frac{\int_{\sigma_{\mathrm{w}}^2}^{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv}{\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty} \left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}} e^{-\frac{v}{\zeta}} dv} + 1} \tag{C.14}$$

where (C.14) follows by noting that $e^{-\frac{\Delta}{x}}$ is monotonically increasing in x and $c_2 > c_1$. And (C.14) also holds if the ratio of the integrals in (C.11) is equal to one, in

which case $e^{-\frac{\Delta}{c_2}}$ is replaced by 1 in (C.13). Multiplying (C.14) through by the term $\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty}\left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}}e^{-\frac{v}{\zeta}}dv$ yields:

$$\Lambda(z^{(0)}+\Delta) > \frac{e^{\frac{\sigma_{\mathrm{a}}^2}{\zeta}}\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty}\left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}}e^{-\frac{v}{\zeta}}dv}{\int_{\sigma_{\mathrm{w}}^2}^{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}\left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}}e^{-\frac{v}{\zeta}}dv+\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty}\left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}}e^{-\frac{v}{\zeta}}dv}, \tag{C.15}$$

$$= \frac{e^{\frac{\sigma_{\mathrm{a}}^2}{\zeta}}\int_{\sigma_{\mathrm{w}}^2+\sigma_{\mathrm{a}}^2}^{\infty}\left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}}e^{-\frac{v}{\zeta}}dv}{\int_{\sigma_{\mathrm{w}}^2}^{\infty}\left(\frac{1}{v}\right)^n e^{-\frac{z^{(0)}}{v}}e^{-\frac{v}{\zeta}}dv}, \tag{C.16}$$

$$= \gamma \tag{C.17}$$

where (C.17) follows from the assumption in (C.4). Hence, if an observation $z^{(0)}$ is such that $\Lambda(z^{(0)})=\gamma$, then an increase in the observed power z results in $\Lambda(z)>\gamma$.

■

APPENDIX D

PROOF OF o(1) COVERT RATE FOR M > 1

Consider the assumptions of the multiple block fading model in Section 2.2.1.2 and Alice's construction presented in Section 2.2.1.1 but with the jammer transmitting Gaussian noise drawn from a distribution with constant variance. Additionally, assume that Willie knows $h_{0,m}^{(\mathrm{a,w})}, m = 1, 2, \ldots, M$. *Then, there exists a covert communication strategy such that Bob can reliably decode Alice's messages if she transmits with rate* $o(1)$.

Proof: *Construction:* The construction is the same as for Lemma 5 and Theorem 3. *Analysis:* By Theorem 3, Alice can transmit with constant power $P_{\mathrm{f}} > 0$ while remaining covert. What remains is to demonstrate that Bob can decode the transmission with probability of error less than ϵ for any $\epsilon > 0$. Conditioned on the fading variables $h_m^{(\mathrm{a,b})}$, $h_m^{(\mathrm{j,b})}$, the channel from Alice to Bob during the m^{th} block is an AWGN channel with signal-to-noise ratio:

$$\gamma_m = \frac{|h_m^{(\mathrm{a,b})}|^2 \frac{P_{\mathrm{f}}}{d_{\mathrm{a,b}}^{\alpha}}}{|h_m^{(\mathrm{j,b})}|^2 \frac{P_{\mathrm{j}}}{d_{\mathrm{j,b}}^{\alpha}} + \sigma_{\mathrm{b}}^2}. \tag{D.1}$$

Now, there exists a constant δ small enough such that:

$$P(\gamma_m > \delta) \geq 1 - \frac{\epsilon}{2M} \tag{D.2}$$

for all $m = 1, 2, \ldots M$. Hence, $P(\min(\gamma_1, \gamma_2, \ldots, \gamma_M) > \delta) > 1 - \frac{\epsilon}{2}$. Now, there exists a constant rate R such that communication is reliable over an AWGN channel with SINR δ; hence, communication at that rate R is reliable here. Finally, since $o(1) < R$ for all $n > N_0$ for some N_0, the result follows. ∎

APPENDIX E

WILLIE'S ALTERNATIVE STATISTICS WHEN ALICE TRANSMITS

The derivations presented in this section are employed to characterize Willie's ability to detect Alice based on the system model presented in Section 3.2. Willie's expected test result when Alice transmits is:

$$E[Z|H_1] = nP_{\mathrm{j}} + n\sigma_{\mathrm{w}}^2 + nP_{\mathrm{f}}. \tag{E.1}$$

The term $E[Z^2|H_1]$ is then evaluated to determine the variance of Willie's observation:

$$\begin{aligned} E[Z^2|H_1] = E[Z^2|H_0] &+ \sum_{k=1}^{n}\sum_{l=1}^{n} E[|h_l|^2]E[|g_l|^2]E[|f_k|^2] + E[|h_k|^2]E[|g_k|^2]E[|f_l|^2] \\ &+ E[|f_l|^2|N_k|^2] + E[|f_k|^2|N_j|^2] \\ &+ E[h_k^* h_l]E[g_k^* g_l]E[f_k f_l^*] + E[h_k^* h_l]E[g_k g_l^*]E[f_k^* f_l] \\ &+ E[f_k^* f_l]E[N_k N_l^*] + E[f_k f_l^*]E[N_k^* N_l] + E[|f_k|^2|f_l|^2]. \end{aligned} \tag{E.2}$$

The second and third terms of (E.2) are:

$$\sum_{k=1}^{n}\sum_{l=1}^{n} E[|h_l|^2]E[|g_l|^2]E[|f_k|^2] + \sum_{k=1}^{n}\sum_{l=1}^{n} E[|h_k|^2]E[|g_k|^2]E[|x_l|^2] = 2n^2 P_{\mathrm{j}} P_{\mathrm{f}} \tag{E.3}$$

and the fourth and fifth terms simplify to:

$$\sum_{k=1}^{n}\sum_{l=1}^{n} E[|f_l|^2]E[|N_k|^2] + E[|f_k|^2]E[|N_l|^2] = 2n^2 P_{\mathrm{f}} \sigma_{\mathrm{w}}^2. \tag{E.4}$$

The sixth and seventh terms of (E.2) are:

$$\sum_{k=1}^{n}\sum_{l=1}^{n} E[h_k^* h_l]E[g_k^* g_l]E[f_k f_l^*] + E[h_k^* h_l]E[g_k g_l^*]E[f_k^* f_l] = 2nP_{\mathrm{j}}P_{\mathrm{f}} \tag{E.5}$$

and the eighth and ninth terms of (E.2) are:

$$\sum_{k=1}^{n}\sum_{l=1}^{n} E[f_k^* f_l]E[N_k N_l^*] + E[f_k f_l^*]E[N_k^* N_l] = 2nP_{\mathrm{f}}\sigma_{\mathrm{w}}^2. \tag{E.6}$$

The final term of (E.2) is:

$$\sum_{k=1}^{n}\sum_{l=1}^{n} E[|f_k|^2|f_l|^2] = \sum_{k=1}^{n}\left(E[|f_k|^4] + \sum_{l\neq k}^{n} E[|f_k|^2]E[|f_l|^2]\right), \tag{E.7}$$

$$= 2nP_{\mathrm{f}} + n^2P_{\mathrm{f}}^2 - nP_{\mathrm{f}}^2. \tag{E.8}$$

Therefore, the full expansion of $E[Z^2|H_1]$ is:

$$\begin{aligned} E[Z^2|H_1] = E[Z^2|H_0] &+ 2n^2P_{\mathrm{j}}P_{\mathrm{f}} + 2n^2P_{\mathrm{f}}\sigma_{\mathrm{w}}^2 + 2nP_{\mathrm{j}}P_{\mathrm{f}} + 2nP_{\mathrm{f}}\sigma_{\mathrm{w}}^2 \\ &+ 2nP_{\mathrm{f}}^2 + n^2P_{\mathrm{f}}^2 - nP_{\mathrm{f}}^2 \end{aligned} \tag{E.9}$$

and the variance of Willie's observation when Alice transmits is:

$$\begin{aligned} E[Z^2|H_1] - (E[Z|H_1])^2 &= 2n^2P_{\mathrm{j}}\sigma_{\mathrm{w}}^2 + 2nP_{\mathrm{j}}\sigma_{\mathrm{w}}^2 + 2n\sigma_{\mathrm{w}}^2 + n^2\sigma_{\mathrm{w}}^4 - n\sigma_{\mathrm{w}}^4 + 2nP_{\mathrm{j}}^2 \\ &\quad + \frac{n^2P_{\mathrm{j}}^2}{f(n)} + 2f(n)P_{\mathrm{j}}^2 + n^2P_{\mathrm{j}}^2 + 2n^2P_{\mathrm{j}}P_{\mathrm{f}} + 2n^2P_{\mathrm{f}}\sigma_{\mathrm{w}}^2 \\ &\quad + 2nP_{\mathrm{j}}P_{\mathrm{f}} + 2nP_{\mathrm{f}}\sigma_{\mathrm{w}}^2 + 2nP_{\mathrm{f}}^2 + n^2P_{\mathrm{f}}^2 - nP_{\mathrm{f}}^2 - n^2P_{\mathrm{f}}^2 \\ &\quad - n^2P_{\mathrm{j}}^2 - n^2\sigma_{\mathrm{w}}^4 - 2n^2P_{\mathrm{f}}\sigma_{\mathrm{w}}^2 - 2n^2P_{\mathrm{j}}\sigma_{\mathrm{w}}^2 - 2n^2P_{\mathrm{f}}P_{\mathrm{j}}, \end{aligned} \tag{E.10}$$

$$\begin{aligned} &= 2nP_{\mathrm{j}}\sigma_{\mathrm{w}}^2 + 2n\sigma_{\mathrm{w}}^2 - n\sigma_{\mathrm{w}}^4 + 2nP_{\mathrm{j}}^2 + \frac{n^2P_{\mathrm{j}}^2}{f(n)} + 2f(n)P_{\mathrm{j}}^2 \\ &\quad + 2nP_{\mathrm{j}}P_{\mathrm{f}} + 2nP_{\mathrm{f}}\sigma_{\mathrm{w}}^2 + 2nP_{\mathrm{f}} - nP_{\mathrm{f}}^2, \end{aligned} \tag{E.11}$$

based on the term:

$$(E[Z|H_1])^2 = n^2(P_{\mathrm{j}}^2 + \sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}^2 + 2P_{\mathrm{j}}P_{\mathrm{f}} + 2P_{\mathrm{f}}\sigma_{\mathrm{w}}^2 + 2P_{\mathrm{j}}\sigma_{\mathrm{w}}^2). \tag{E.12}$$

Therefore, the normalized measurement S has the expected value:

$$E[S|H_1] = P_{\mathrm{j}} + \sigma_{\mathrm{w}}^2 + P_{\mathrm{f}} \tag{E.13}$$

and the variance of S when Alice transmits is:

$$\begin{aligned}\mathrm{Var}[S|H_1] = \frac{1}{n}\Big[2P_{\mathrm{j}}\sigma_{\mathrm{w}}^2 + 2\sigma_{\mathrm{w}}^2 - \sigma_{\mathrm{w}}^4 + 2P_{\mathrm{j}}^2 + 2P_{\mathrm{j}}P_{\mathrm{f}} + 2P_{\mathrm{f}}\sigma_{\mathrm{w}}^2 + 2P_{\mathrm{f}} - P_{\mathrm{f}}^2\Big] \\ + \frac{P_{\mathrm{j}}^2}{f(n)} + \frac{2f(n)P_{\mathrm{j}}^2}{n^2}.\end{aligned} \tag{E.14}$$

APPENDIX F

PSEUDOCODE OF CYCLOSTATIONARY DETECTION SIMULATIONS (SECTION 4.3.3)

```
% Declare Variables
Fs = 100e6                      % Oversample frequency
Ts = 1/Fs                       % Oversample period
pulse                           % Load SRRC pulse
Tb = 143                        % Discrete symbol period
alpha = 1/(Tb*Ts)               % Declare symbol frequency
SNR                             % Declare alice's SNR at willie
SJR                             % Declare alice's SJR at willie
Nsym = 8192                     % Number of symbols alice generates

% Detector Settings
N = 4096                        % Number of samples Willie observes
time_vec = (0:N-1)*Ts
max_iterations                  % Declare max number of iterations
W = 20 MHz                      % Declare bandwidth
lpf = fir1(20,W/Fs)             % Create low-pass filter

rng(0,'twister')                % Initialize random generator seed

for each iteration ii=1:max_iterations
   bpsk = randn(1,Nsym)                    % Generate BPSK symbols
   bpsk(bpsk>=0)=1
   bpsk(bpsk<0)=-1
   bpsk_up = upsample(bpsk,Tb)             % Space by symbol period
   tx_sig = conv(bpsk_up,pulse)            % Convolve with pulse
   % Generate AWGN
   noise = randn(1,length(tx_sig))
   noise = noise/sqrt(power(noise))        % Normalize noise power
   Ps = power(noise)*10^(SNR/10)           % Find signal power
   tx_sig = tx_sig/sqrt(power(tx_sig))     % Normalize
   tx_sig = tx_sig*sqrt(10^(log10(Ps))     % Adjust signal power

   z_pre_h0 = noise(1:N)
   z_pre_h1 = tx_sig(1:N) + noise(1:N)
   z_h0 = filter(lpf,1,z_pre_h0)
```

```
   z_h1 = filter(lpf,1,z_pre_h1)

   % Power Detector
   power_result_h0(ii) = power(z_h0)
   power_result_h1(ii) = power(z_h1)

   % Cyclic Detector
   cyc_result_h0(ii) = abs(sum(z_h0^2*cos(2 π alpha t_vec)))/N
   cyc_result_h1(ii) = abs(sum(z_h1^2*cos(2 π alpha t_vec)))/N
end for loop

% Next, find power detection rates based on derived equations
% First find frequency at which the magnitude response of the
% filter is -3 dB.
[h, w] = freqz(lpf,1)
[v, p] = min(abs(20*log10(abs(h))+3))
W_adj = w(p)/π * Fs
Pna = Pn*W/Fs
Pnb = Pn*W_adj/Fs
mu0 = Pnb
mu1 = Ps + Pnb
std_pow0 = sqrt((Pna^2*2)/N)*2
std_pow1 = sqrt((Pna^2*2+4*Ps*Pna)/N)*2
tau_vec = linspace(0,max(power_result_h1),max_iterations)
pow_eq_fa = 1-.5*erfc((-tau_vec-mu0)/std_pow0/sqrt(2))
             +.5*erfc((tau_vec-mu0)/std_pow0/sqrt(2))
pow_eq_det = 1-.5*erfc((-tau_vec-mu1)/std_pow1/sqrt(2))
             +.5*erfc((tau_vec-mu1)/std_pow1/sqrt(2))

% Cyclic detection rates based on derived equations
x = filter(lpf,1,tx_sig(1:N))
rho = abs(sum(x^2*cos(2 π alpha t_vec)))/N
std_csd0 = sqrt(Pna^2/N)*2
std_csd1 = sqrt((Pna^2/N+2*Pna*Ps/N^2))*2
tau_vec = linspace(0,max(cyc_result_h1),max_iterations)
cyc_eq_fa = erfc(tau_vec./std_csd0/sqrt(2))
cyc_eq_det = 1-.5*erfc((-tau_vec-rho)./std_csd1/sqrt(2))
              +.5*erfc((tau_vec-rho)./std_csd1/sqrt(2))


% Generate simulated ROC results using power_result_h0 and
% power_result_h1
% Generate derived ROC results using pow_eq_fa and pow_eq_fa
% Generate simulated ROC results for cyc_result_h0
% and cyc_result_h1
% Generate derived ROC results using cyc_eq_fa and cyc_eq_fa
```

APPENDIX G

PSEUDOCODE OF TIMING-BASED DETECTOR SIMULATIONS (SECTION 5)

The code below outline's Willie's detector assuming Willie has already estimated the jammer's timing offset.

```
% Declare Variables
Fs = 100e6                          % Oversample frequency
Ts = 1/Fs                           % Oversample period
pulse                               % Load SRRC pulse
Tb = 143                            % Discrete symbol period
alpha = 1/(Tb*Ts)                   % Declare symbol frequency
SNR                                 % Declare Alice's SNR at Willie
SJR                                 % Declare Alice's SJR at Willie
Nsym = 8192                         % Number of symbols Alice generates

% Detector Settings
N = 4096                            % Number of samples Willie observes
time_vec = (0:N-1)*Ts
max_iterations                      % Declare max number of iterations
A                                   % Construct A Matrix

rng(0,'twister')                    % Initialize random generator seed
for ii=1:max_iterations
   % Generate signals
   g         % Generate jammer's bpsk pulse shaped signal with
             % a fixed timing offset
   x         % Generate Alice's bpsk pulse shaped signal
             % and vary timing offset if needed
   noise     % Generate and normalize AWGN power
   % Scale Alice's signal power and jammer's signal power
   % according to SNR and SJR

   % Match filter
   mf_h0 = conv(g+noise,pulse)        % Willie's H0 observation
   mf_h1 = conv(x+g+noise,pulse)      % Willie's H1 observation
```

```
    % Implement standard power detector here, then the
    % timing based detector

    % Downsample observation on jammer's branch using τ̂_j
    m         % Find when first symbol should arrive if there
              % are no timing offsets
    rj_h0 = downsample(mf_h0(m+τ̂_j:end),Tb)
    rj_h1 = downsample(mf_h1(m+τ̂_j:end),Tb)

    % Downsample Willie's observation with different
    % estimated timing offsets for Alice

    % Example code below for τ̂_a=10
    ra_h0_10 = downsample(mf_h0(m+10-1:end),Tb)
    ra_h1_10 = downsample(mf_h1(m+10-1:end),Tb)

    % Calculate power in estimates and repeat process
    % for τ̂_a=20,30,40
    ba_h0_10 = A^[ra_h0_10(1:N),rj_h0_10(1:N)]
    % measure power in ba_h0_10
    ba_h1_10 = A^[ra_h1_10(1:N),rj_h1_10(1:N)]
    % measure power in ba_h1_10
end for loop

% Generate ROC results for power detector results
% Generate ROC results for ba_h0_10 and ba_h1_10
% Generate ROC results for ba_h0_20 and ba_h1_20
% Generate ROC results for ba_h0_30 and ba_h1_30
% Generate ROC results for ba_h0_40 and ba_h1_40
```

BIBLIOGRAPHY

[1] B. A. Bash, *Fundamental Limits of Covert Communication*. PhD thesis, University of Massachusetts Amherst, May 2015.

[2] S. Singh, *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*. Anchor Books, 2000.

[3] E. Tews, R.-P. Weinmann, and A. Pyshkin, "Breaking 104 bit wep in less than 60 seconds," 2007.

[4] M. Curtin, *Brute Force: Cracking the Data Encryption Standard*. Copernicus, 2005.

[5] P. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proc. 35th Annu. Symp. on Foundations of Comput. Sci.*, pp. 124–134, Nov. 1994.

[6] T. Spiller, "Quantum information processing: cryptography, computation, and teleportation," *Proc. of the IEEE*, vol. 84, pp. 1719–1746, Dec. 1996.

[7] D. Bacon and W. van Dam, "Recent progress in quantum algorithms," *Commun. ACM*, vol. 53, pp. 84–93, Feb. 2010.

[8] M. Sharbaf, "Quantum cryptography: An emerging technology in network security," in *IEEE Int. Conf. on Technologies for Homeland Security (HST)*, pp. 13–19, Nov. 2011.

[9] D. Kahn, *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet*. Scribner, 1996.

[10] J. Ball, "Nsa stores metadata of millions of web users for up to a year, secret files show." The Guardian, Sept. 2013. http://www.theguardian.com/world/2013/sep/30/nsa-americans-metadata-year-documents.

[11] B. Bash, D. Goeckel, and D. Towsley, "Square root law for communication with low probability of detection on awgn channels," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, pp. 448–452, July 2012.

[12] B. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," *IEEE J. on Select. Areas in Commun.*, vol. 31, pp. 1921–1930, Sept. 2013.

[13] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, "Quantum-secure covert communication on bosonic channels," *Nature Commun.*, vol. 6, Oct. 2015.

[14] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, pp. 2945–2949, July 2013.

[15] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. IEEE Inform. Theory Workshop*, pp. 30–34, Nov. 2014.

[16] L. Wang, G. Wornell, and L. Zheng, "Limits of low-probability-of-detection communication over a discrete memoryless channel," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, pp. 2525–2529, June 2015.

[17] L. Wang, G. Wornell, and L. Zheng, "Undetected communication over a dmc," in *Inform. Theory and Applicat. Workshop*, 2015.

[18] M. Bloch, "A channel resolvability perspective on stealth communications," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, pp. 2535–2539, June 2015. arXiv:1503.08778 [cs.IT].

[19] B. Bash, D. Goeckel, and D. Towsley, "LPD communication when the warden does not know when," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, pp. 606–610, June 2014.

[20] S. Lee and R. Baxley, "Achieving positive rate with undetectable communication over awgn and rayleigh channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 780–785, June 2014.

[21] S. Lee, R. Baxley, J. McMahon, and R. Frazier, "Achieving positive rate with undetectable communication over mimo rayleigh channels," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, pp. 257–260, June 2014.

[22] S. Lee, R. Baxley, M. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE J. Select. Topics Signal Process.*, vol. 9, pp. 1195–1205, Oct. 2015.

[23] D. Goeckel, B. Bash, S. Guha, and D. Towsley, "Covert communications when the warden does not know the background noise power," *IEEE Commun. Lett.*, vol. 20, pp. 236–239, Feb. 2016.

[24] W. Gardner, *Cyclostationarity in Communications and Signal Processing*. New Jersey: IEEE Press, 1993.

[25] W. A. Gardner, "Signal interception: a unifying theoretical framework for feature detection," *IEEE Trans. on Commun.*, vol. 36, pp. 897–906, Aug. 1988.

[26] W. Gardner, "Spectral correlation of modulated signals: Part I - analog modulation," *IEEE Trans. on Commun.*, vol. 35, pp. 584–594, June 1987.

[27] W. Gardner, W. Brown, and C.-K. Chen, "Spectral correlation of modulated signals: Part II - digital modulation," *IEEE Trans. on Commun.*, vol. 35, pp. 595–601, June 1987.

[28] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. of the IEEE*, vol. 55, pp. 523–531, Apr. 1967.

[29] B. Carrara and C. Adams, "Estimating the steganographic capacity of band-limited channels," in *IEEE Canadian Conf. on Electrical and Comput. Eng. (CCECE)*, pp. 1–5, May 2016.

[30] B. Sklar, *Digital communications: fundamentals and applications*. Prentice Hall Communications Engineering and Emerging Technologies Series, Prentice-Hall PTR, 2001.

[31] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, pp. 2334–2354, May 2016.

[32] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inf. Theory*, vol. 62, pp. 3493–3503, June 2016.

[33] A. Sheikholeslami, B. Bash, D. Towsley, D. Goeckel, and S. Guha, "Covert communication over classical-quantum channels," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, July 2016. arXiv:1601.06826[cs.IT].

[34] R. Soltani, B. Bash, D. Goeckel, S. Guha, and D. Towsley, "Covert single-hop communication in a wireless network with distributed artificial noise generation," in *Proc. Conf. Commun., Control Comput. (Allerton)*, pp. 1078–1085, Sept. 2014.

[35] R. Soltani, D. Goeckel, D. Towsley, and A. Houmansadr, "Covert communications on poisson packet channels," in *Proc. Conf. Commun., Control Comput. (Allerton)*, pp. 1046–1052, Sept. 2015.

[36] R. Soltani, D. Goeckel, D. Towsley, and A. Houmansadr, "Covert communications on renewal packet channels," in *Proc. Conf. Commun., Control Comput. (Allerton)*, pp. 548–555, Sept. 2016.

[37] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge University Press, 2005.

[38] B. Bash, D. Goeckel, and D. Towsley, "Covert communication gains from adversary's ignorance of transmission time," *IEEE Trans. on Wireless Commun.*, vol. 15, pp. 8394–8405, Dec. 2016.

[39] S. Kay, *Fundamentals of Statistical Signal Processing: Detection theory*. Fundamentals of Statistical Signal Processing, PTR Prentice-Hall, 1998.

[40] M. Shaked and J. Shanthikumar, *Stochastic Orders and their Applications*. Academic Press, 1994.

[41] T. Sobers, B. Bash, D. Goeckel, S. Guha, and D. Towsley, "Covert communication with the help of an uninformed jammer achieves positive rate," in *Proc. Asilomar Conf. on Signals, Systems, and Comput.*, Nov. 2015.

[42] S. M. Kay, *Fundamentals of statistical signal processing, volume I: estimation theory*. Prentice Hall, 1993.

[43] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. New York: Cambridge University Press, 1st ed., 2009.

[44] V. Korzhik, G. Morales-Luna, and M. H. Lee, "On the existence of perfect stegosystems," in *Proc. 4th Int. Workshop Digital Watermarking (IWDW)*, (Siena, Italy), pp. 30–38, Sept. 2005.

[45] S. Craver and J. Yu, "Subset selection circumvents the square root law," in *Proc. SPIE Media Forensics Security*, vol. 7541, (San Jose, CA), pp. 754103–1–754103–6, 2010.

[46] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2006.

[47] J. Proakis, *Digital Communications*. McGraw-Hill, 2001.

[48] Y. Zeng and Y. C. Liang, "Robustness of the cyclostationary detection to cyclic frequency mismatch," in *21st Annu. IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, pp. 2704–2709, Sept. 2010.

[49] D. L. Goeckel and Q. Zhang, "Slightly frequency-shifted reference ultra-wideband (uwb) radio," *IEEE Trans. on Commun.*, vol. 55, pp. 508–519, Mar. 2007.

[50] S. Miller and D. Childers, *Probability and Random Processes: With Applications to Signal Processing and Communications*. Elsevier Academic Press, 2004.

[51] D. Torrieri, *Principles of Spread-spectrum Communication Systems*. Boston, MA, USA: Springer: Springer, 2005.

[52] S. Kullback, *Information theory and statistics*. New York: Wiley, 1959.

[53] P. Tobin, *PSpice for digital communications engineering*. Morgan and Claypool Publishers, 2007.

[54] S. Yan, B. He, Y. Cong, and X. Zhou, "Covert communication with finite blocklength in AWGN channels," *Computing Research Repository (CoRR)*, vol. abs/1701.08891, 2017.

[55] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. McGraw-Hill Book Company, Inc., 1961.