ON COVERTNESS IN COMMUNICATIONS SCHEMES

SUR LE CARACTÈRE SECRET DES SYSTÈMES DE COMMUNICATION

A Thesis Submitted to the Division of Graduate Studies
of the Royal Military College of Canada
by

Lucas Rooyakkers

In Partial Fulfillment of the Requirements for the Degree of
Master of Applied Science in Electrical and Computer Engineering

March 2025

Dr. Francois Chan, Dr. Tricia Willink

Abstract

Covert communications is a field that investigates if and how information can be transmitted over a channel with a low probability of being detected or intercepted. A covert communications scheme is usually designed by randomizing properties of the signal, so the signal "blends in" with background noise better. Several metrics for quantifying the covertness of a communications scheme exist, but are specific to detector or modulation type. This work compares the relative covertness of a wide variety of communications schemes against the main classes of signal detectors over an additive white Gaussian noise (AWGN) channel. Radiometric and cyclostationarity detectors are found to be effective for reliably detecting weak signals of all modulations, although the radiometer had an entirely predictable performance, regardless of modulation scheme, while the performance of the cyclostationarity detectors varied with modulation scheme. The communications schemes tested range from traditional modulations to spread spectrum techniques and chaotic modulations. The tradeoff between detectability and error rate is also considered in this work, and code division multiplex access (CDMA) was found to have the best overall balance between error rate and covertness.

Résumé

Les communications secrètes représentent un domaine qui étudie si et comment l'information peut être transmise sur un canal avec une faible probabilité d'être détectée ou interceptée. Une technique de communications secrètes est généralement conçue en randomisant certaines propriétés du signal afin que celui-ci se "fonde" mieux dans le bruit de fond. Plusieurs métriques existent pour quantifier le caractère secret d'une technique de communications, mais elles sont spécifiques au type de détecteur ou de modulation utilisé. Ce travail compare le degré relatif de secret d'une large gamme de techniques de communications face aux principales classes de détecteurs de signal sur un canal AWGN. Il est démontré que les détecteurs radiométriques et de cyclostationnarité sont efficaces pour détecter de manière fiable des signaux faibles, quelle que soit la modulation utilisée. Toutefois, la performance du radiomètre est entièrement prévisible, indépendamment du type de modulation, tandis que celle des détecteurs de cyclostationnarité varie en fonction du type de modulation. Les techniques de communications testées couvrent un large éventail, allant des modulations traditionnelles aux techniques d'étalement de spectre et aux modulations chaotiques. Ce travail examine également le compromis entre détectabilité et taux d'erreur, et le CDMA s'est avéré avoir le meilleur équilibre global entre taux d'erreur et secret.

Table of Contents

List of Abbreviations

M-PSK	M-Ary phase shift keying
ADC	Analog digital converter
ASK	Amplitude shift keying
AUC	Area under the curve
AWGN	Additive white Gaussian noise
BCE	Before common era
BER	Bit error rate
BFSK	Binary frequency shift keying
BMWD	Binary moving window detector
BPF	Band pass filter
BPSK	Binary phase shift keying
BSC	Binary symmetric channel
CAF	Cyclic autocorrelation function
CCS	Covert communications system
CD	Correlation detection
CDMA	Code division multiplex access
CEVR	Circular equivalent vulnerable radius
CFAR	Constant false alarm rate
COTS	Commercial off-the-shelf
CP	Cyclic prefix
CSI	Channel state information
CSK	Chaos shift keying
CSS	Chirp spread spectrum
DCS	Degree of cyclostationarity
DCSK	Differential chaos shift keying
DMC	Discrete memoryless channel
DSSS	Direct sequence spread spectrum
EM	Electro-magnetic
FAM	Frequency accumulation method
FFT	Fast-Fourier transform
FH	Frequency hopping
FH-CSS	Frequency-hopped chirp spread spectrum
FH-OFDM-DCSK	Frequency-hopped OFDM-DCSK
FHSS	Frequency-hopping spread spectrum
FPR	False positive rate

FSF	Frequency-selective fading
FSK	Frequency shift keying
I/Q	In-phase and quadrature
IF	Intermediate frequency
IFFT	Inverse fast-Fourier transform
IRS	Intelligent reflecting surface
ISI	Inter-symbol interference
LPD	Low probability of detection
LPF	Low pass filter
LPI	Low probability of interception
LRT	Likelihood ratio test
LS	Loosely-synchronous
MIMO	Multiple input multiple output
OFDM	Orthogonal frequency division multiplexing
PDF	Probability distribution function
PSD	Power spectral density
PSK	Phase shift keying
QAM	Quadrature amplitude modulation
QCSK	Quadrature chaos shift keying
QKD	Quantum key distribution
QPSK	Quadrature phase shift keying
RF	Radio frequency
ROC	Receiver operating characteristic
SCF	Spectral correlation function
SEVR	Spherical equivalent vulnerable radius
SISO	Single input single output
SNR	Signal-to-noise ratio
SOI	Signal of interest
SRL	Square-root law
SS	Spread spectrum
SSCA	Strip spectral correlation algorithm
TPR	True positive rate
TSM	Time smoothing method
UMP	Uniformly most powerful
WC	Walsh code

List of Symbols

$\mathbb{P}_{\mathbf{D}}$	Probability of detection
$\mathbb{P}_{\mathbf{FA}}$	Probability of false alarm
$\mathbb{P}_{\mathbf{MD}}$	Probability of missed detection
λ_0	Detector threshold
λ	Detector output
$\mathcal{D}(\cdot)$	Detector function
N_0	Awgn noise variance
T	Integration period
W	Bandwidth
f_s	Sampling rate/frequency
f_c	Carrier frequency
A	Sine wave amplitude
ϕ	Sine wave phase
k	Chip rate
$d(t)$	Data signal
d_i	Data signal sample at interval i
$s(t)$	Transmitted signal
s_i	Transmitted signal sample at interval i
$r(t)$	Received signal (with noise)
r_i	Received signal sample at interval i (with noise)
$\frac{E_b}{N_0}$	Signal-to-noise ratio per bit
H_0	The null hypothesis (noise only)
H_1	The transmit hypothesis (signal + noise)
$\mathcal{F}$	Fourier transform
$\operatorname{erf}(\cdot)$	Gauss error function: $\operatorname{erf}(z) = \frac{2}{\pi}\int_0^z e^{-t^2}dt$
$\operatorname{erfc}(\cdot)$	Complementary error function: $\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$
$\operatorname{ierfc}(\cdot)$	Inverse complementary error function: $\operatorname{ierfc}(\operatorname{erfc}(z)) = z$
$R_r^\alpha(t)$	Cyclic autocorrelation function of $r(t)$ at cycle frequency α
$S_r^\alpha(f)$	Spectral correlation function of $r(t)$ at cycle frequency α
$A_{\mathbf{det}}$	Area of detection
$V_{\mathbf{det}}$	Volume of detection

List of Figures

1 Introduction

In wireless transmission environments, it can be desirable to send messages reliably (i.e., that can be received by intended recipients without error), but deniably, (i.e., illegitimate users are unable to obtain evidence of message transmission). A communications scheme is undetectable, or covert, when it can transmit messages that have these properties.

This idea of undetectable communications has historical precedent. In 5^{th} century Ancient Greece before common era (BCE), Herodotus notes that Histiaeus tattooed a missive onto the shaved scalp of a messenger [1, §5.35.3]. The messenger grew back his hair, only to shave it again upon arrival at his destination to reveal the hidden message. During the 5^{th} century BCE in Ancient Greece, Aeneas Tacticus mentions many methods of hiding the existence of a message that remain largely unchanged to this day [2, Ch. 31]. These include the first known mention of invisible ink, and a plethora of ideas for smuggling written messages on hard to search areas of the body, or things like writing tablets. Tacticus also mentions steganographic techniques, like writing an ordinary letter about some topic that is sufficiently long, then embedding the hidden text by marking the letters of the hidden text with small marks, so the recipient deciphers the hidden text by only considering the marked letters. This demonstrates humanity has been thinking about this issue for millennia.

The desire to send messages without anyone else knowing the communication took place has not faded from the human psyche. While there have been many advances in undetectable messaging since then [3], the practice of modern steganography [4] concerns itself primarily with embedding hidden messages in digital files and computer systems. This thesis restricts its investigation to undetectable messages within the realm of wireless communications systems.

1.1 Covert Communications

Covert communications describe a situation that differs fundamentally from the traditional physical-layer security situation, largely in terms of the objective of the malicious user. In both scenarios, the legitimate user Alice wants to communicate with the legitimate user Bob. In the standard physical-layer security setup [5], the malicious user (usually an eavesdropper Eve or a decrypter Carol) seeks to discover *what* it is that Alice transmitted. Under the covert scenario, however, a warden Willie[1] seeks instead to find out *whether* Alice transmitted (or not). The warden Willie in this scenario has no regard for the contents of the message that Alice is sending, but only tries to establish evidence of transmission.

Theoretical fundamental limits have been established regarding the maximum information rate that can be achieved via covert communications, but little work has been done

to characterize and quantify the "detectability" or "covertness" of existing communications schemes by different types of detectors.

This thesis asks two questions about covert communications:

Q.1 Which communications protocols and modulation techniques can Alice employ to best evade detection by Willie?

Q.2 What detection method(s) can Willie employ to best detect signals transmitted by Alice?

These questions are intrinsically interrelated—they are two opposing perspectives of the same fundamental problem.

1.2 LPD & LPI

The terms low probability of detection (LPD) and low probability of interception (LPI) are both used to describe waveforms that attempt to be undetectable to illegitimate users. While low probability of interception (LPI) usually refers to properties of the waveforms, the phrase low probability of detection (LPD) communications describes the same overall situation as the term *covert communications*, which includes Alice's waveform and Willie's detector, i.e., both the transmitter and the detector of the illegitimate observer.

The terms LPD and LPI are often used interchangeably in the literature, and are treated interchangeably throughout this thesis. The primary difference between the two terms is that LPD refers to the signals that obey a specific mathematical formalism [6] with Alice, Bob, and Willie, whereas the term LPI is often used in the literature by waveform designers to mean "this signal is practically hard to intercept or detect", without formally quantifying exactly how or what it means for a waveform to be "difficult" to detect.

1.3 Contributions

This work measures the relative detectability of a wide variety of communications schemes against a variety of "blind-parameter" signal detectors, i.e., detectors that have no knowledge of the composition of the signal they are trying to detect. The deniability of a message, as measured by its probability of detection, is calculated as a function of signal-to-noise ratio (SNR) for each combination of detector and transmission scheme. The deniability of each transmission scheme is compared alongside its reliability, measured in terms of bit error rate (BER), to give an overall picture of covert performance. Properties that make communications schemes more covert are discussed. The question of which detectors are the most powerful at positively identifying transmissions is also addressed.

This work builds upon previous works that measure covertness as a function of gain difference between the intended user and the illegitimate observer [9], where only the radiometer is considered. It expands the literature by considering more detector types, like cyclostationarity detectors, as well as a plethora of different transmission schemes that have a variety of covertness properties.

[1] Although the first paper [6] and most papers on covert communications since then feature a warden Willie, early papers instead have a detector Dave [7, 8] who occupies the same role.

1.4 Thesis Structure

This thesis formally describes the covert communications scenario fully in Chapter 2, with the fundamental limits of covert communications discussed in Section 2.1 and a literature review of extensions and modifications to basic covert communications are discussed in Section 2.4 (e.g., covert communications under jamming, using relays, using multiple input multiple output (MIMO), etc.).

Next, the transmissions schemes that Alice and Bob use are described in Chapter 3, starting with basic modulations in Section 3.2, spread spectrum (SS) technologies in Section 3.3, and several more exotic chaotic communications schemes in Section 3.4.

Following this description of communications schemes, detector theory is explained in Chapter 4, where Willie applies the statistics of hypothesis testing to create effective detectors, alongside the mathematics of all the different types of detectors that Willie could employ.

After the basic strategies that Alice and Willie can employ are laid out in the sections above, we move on to a literature review of prior works that provide a metric to evaluate the covertness of transmission schemes in Chapter 5.

The model that underpins the simulations in this thesis is described in Chapter 6. Herein, the channel model and implementation details of the detectors are explained thoroughly. The results of simulations using this model are shown in Chapter 7. The detectors are compared in Section 7.1, and transmission schemes are compared in Section 7.2, while the overall implications for the covert communications scenario are considered in Section 7.3.

The thesis concludes with a summary of the contributions in Chapter 8, which addresses several open questions and ideas for future work in Section 8.2. The key take-away points of my investigation into quantifying covert communications are given in Section 8.3.

2 Covert Communications

The fundamental limits on how much information can be sent "reliably" to Bob but "deniably" to Willie have been found and precisely characterized for a wide variety of situations. These limits come in the form of achievability bounds, most often proved using either Kullback–Leibler divergence [6,10–12] (also known as relative entropy) or variational distance [13–15] as a metric for quantifying covertness.

These achievability arguments are based on measuring how similar the observations of the channel look when Alice transmits compared to the case when she is not transmitting. Although theorems have been discovered that prove that a certain covert information capacity is achievable, no algorithm has been developed to build a communications scheme that achieves that capacity[1].

This chapter discusses fundamental limits around the throughput rate of information that can be transmitted covertly in a variety of different scenarios, as well as practical techniques to increase the "covertness" of a transmission.

2.1 The Square-Root Law

In general, it can be shown that $\mathcal{O}(\sqrt{n})$ bits[2] of information can be transmitted covertly in n channel uses[3] [6–8, 10, 11, 17–23]. The square-root law (SRL) applies to many versions of the problem, and the exact scaling constant has been found for many scenarios, including single input single output (SISO) AWGN channels [6, 10, 11], binary symmetric channels (BSCs) [20, 24], multiple input multiple output (MIMO) Rayleigh-fade channels [8], and a broad class of discrete memoryless channels (DMCs) [10, 19, 20].

In these scenarios, there are several common assumptions. Alice needs to have a lower bound on the noise level observed by Willie[4] [6, 7]. It is also assumed that Willie knows the channel state information (CSI), background noise level, and all the details about the communications protocol used by Alice and Bob, exempting a shared secret that Alice and Bob may have prearranged. No pre-shared secret key is required if Alice and Bob know that Willie observes less signal energy than Bob [10,17] (e.g., Alice is using a directional antenna pointed at Bob). If Willie and Bob see the same noise power and channel distortion, then a

[1] Under AWGN only channels it can be shown that Gaussian signaling is the optimal form of LPI/LPD communication [12, 16] for minimizing the bit error rate (BER) at Bob, but not for maximizing covertness at Willie [17].

[2] For an explanation of the big-$\mathcal{O}(\cdot)$, small-$o(\cdot)$, and $\Omega(\cdot)$ notation, see Appendix A.1.

[3] A *channel use* in the context of DMCs (see Section 2.3) is a single message that Alice sends within a discrete time block.

secret key of length $\mathcal{O}(\sqrt{n})$ needs to be shared between Alice and Bob beforehand. The pre-shared key could allow Alice and Bob to coordinate the transmission signal characteristics in several ways. They key could pre-arrange transmission times to evade detection by Willie, or may be a spreading sequence, or frequency hopping pattern.

The square-root scaling comes from the mathematics of binary hypothesis testing [11]; the SRL in covert communications mirrors a similar SRL first discovered for steganography [4, 25], where $\mathcal{O}(\sqrt{n})$ bits can be covertly hidden in data (e.g., an image file) of size n bits.

2.1.1 Exceptions to the Square-Root Law

There are some cases where $\Omega(\sqrt{n})$ bits can be sent in n channel uses. $\mathcal{O}(n)$ bits can be transmitted in several situations:

- when Willie does not have a good estimate of his channel [26–29],
- when Alice is aware that Willie observes less signal energy than Bob [10],
- when Alice can predict the instantaneous background noise and channel state [30,31][5],
- when a jammer helps (see Section 2.4.2),
- when Alice hides her message among public messages, or within a public message (see Section 2.4.5).

The $\mathcal{O}(n)$ scenarios listed above allow for a positive (non-zero) covert rate (see Section 2.2). This means that given infinte time, Alice can send an unbounded amount of information without being detected.

If Willie's receiver observes noise, but Bob's receiver is perfectly noiseless, then Alice can transmit $\mathcal{O}(\sqrt{n}log(n))$ bits in n channel uses [24,32]. If Willie does not know in which time slots Alice transmits, her covert capacity becomes $\mathcal{O}(\sqrt{n\log(T(n))})$, where Alice transmits in one randomly selected symbol period every $T(n)$ symbol periods (see Section 2.4.1).

2.2 Covert Capacity

The covert capacity of a channel is how much data can be sent covertly within n channel uses in the asymptotic limit where $n \to \infty$. Scenarios characterized by the SRL have zero covert capacity because $\lim_{n\to\infty} \frac{\mathcal{O}(\sqrt{n})}{n} = 0$. This implies Alice cannot reliably transmit (with arbitrarily low error rate) to Bob forever without Willie detecting her.

Although the covert capacity is zero, this result only applies in the limit as $n \to \infty$. For any finite period, Alice can still reliably and covertly transmit a finite number of bits [19, 33–35] with the SRL scaling. Thus, in LPI/LPD communications we are more concerned with the number of reliably and deniably transmittable bits within n channel uses with a finite n, as opposed to the number of bits per channel use as $n \to \infty$.

A positive covert capacity, i.e., transmitting $\Omega(n)$ bits covertly, is achievable via every $\mathcal{O}(n)$ situation described in Section 2.1.1. A positive covert capacity implies that Alice

[4]Indeed, if there is no noise, distortion, fade, or path loss at all at Willie's receiver, then Alice cannot transmit without being detected.

[5]The situation where Alice knows the instantaneous background noise at every moment at Willie's receiver requires her to essentially predict the future. This is considered impossible, yet the mathematics have been worked out [30, 31] should this situation should ever arise.

can transmit an unbounded quantity of information to Bob, eternally, without Willie ever acquiring solid evidence of her having used the channel. Moreover, in most practical environments, there will be public messages on the channel, complex channel distortion, and Willie will not know the transmission time. A positive covert capacity should therefore be practically achievable, enabled by multiple factors.

2.3 Discrete Memoryless Channels

Wireless communications involve using antennas to measure and change the electromagnetic spectrum to receive and send messages. Although electromagnetic fields are treated mathematically as continuous, when antenna input signal is sampled by the analog digital converter (ADC) of a digital receiver, they are discretized in time by sampling rate and voltage. The output of an ADC is a sequence of numbers representing the voltage over time. This is essentially the receiver side of a DMC. The input side of the DMC is the information bits the sender transmits, and in-between the channel is modelled as the function that maps what symbols were transmitted to what symbols were received.

Formally, a DMC is an input alphabet, $X^n = \{x_1, x_2, \ldots, x_n\}$, which maps to an output alphabet $Y^n = \{y_1, y_2, \ldots, y_n\}$ via a random conditional transition probability law, $\mathcal{P}_{Y^n|X^n}$. An example is seen in Fig. 2.1, where $\mathcal{P}_{Y^n|X^n}$ is the transition probabilities for Bob, and $\mathcal{P}_{W^n|X^n}$ is the transition probabilities for Willie's channel and Willie's output alphabet. Fig. 2.1 presents a depiction of such a DMC.

The random transition law models a signal being corrupted by random noise, and abstracts away all the physical details of the problem to just the output symbols of the detector. A DMC is "memoryless" because the transition probabilities do not depend on prior channel usage.

My simulation for this thesis [36] has an alphabet consisting of tuples of IEEE-754 [37] 64-bit floating point numbers representing the sampling of real values of an antenna by an ADC.

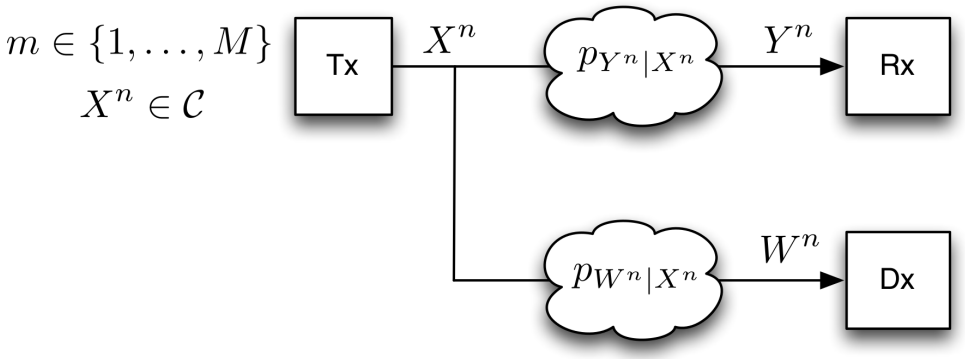


Figure 2.1: Model of a covert communications channel as a DMC. (Source: Letzepis [19])

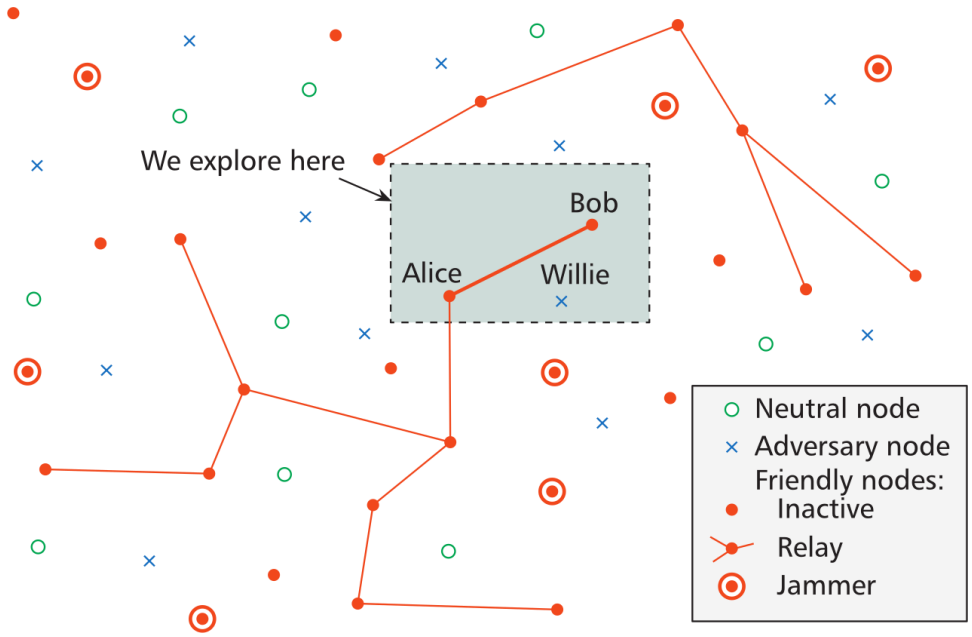


Figure 2.2: A depiction of a shadow network. The general problem can account for multiple discreet transceivers and multiple wardens, as well as jammers, relays, and neutral transmitters. In this work, the view is restricted to just the scenario involving Alice, Bob, and Willie. (Source: Bash *et al.* [23])

2.4 Further Scenarios

The literature has characterized many other variations of this problem. This section discusses some of these other scenarios and their implications. Fig. 2.2 shows the small subset of the generalized shadow network problem considered in this thesis. All the extensions in this section are beyond the bounds of this work, as they do not cut to the core of the covert communications problem.

2.4.1 Pre-Arranged Transmission Time

If Alice and Bob arrange before to only transmit during one random slot out of every $T(n)$ slots, then Alice can send $\mathcal{O}(\min\{\sqrt{n\log(T(n))}, n\})$ bits in n channel uses. Since Willie does not know when Alice transmits, he has to observe the channel for a longer period than Bob does, collecting more background noise, thus increasing Alice's covert capacity [38]. The only additional cost to Alice and Bob is the necessity of a length $\mathcal{O}(log(T(n)))$ pre-shared secret key between them.

2.4.2 Transmitting Noise

Artificial noise received by Willie can deceive him about the true channel statistics, which allows Alice to transmit more bits covertly. Alice can even achieve a positive covert capacity and transmit $\mathcal{O}(n)$ bits when a jammer is present [39–43] (represented by a circle in a red

dot () in Fig. 2.2). This is true even if Alice cannot coordinate with or control the jammer [40–43]. Alice and Bob can also transmit artificial noise themselves to increase their channel capacity [44, 45].

2.4.3 Shadow Networks

The minimal covert communications scenario just concerns Alice, Bob, and Willie, but this situation can be extended to include multiple discreet transceiving parties, multiple wardens [46], neutral parties, jammers, and more, as illustrated in Fig. 2.2. The goal of the covert parties is to form a "shadow network" [23], where communications are transmitted reliably amongst a network of covert transceivers while not being detectable to a group of wardens.

Situations with multiple covert parties transmitting to either a single receiver [21, 47] or multiple receivers [22, 48] both follow the SRL. Section 2.4.7 discusses how beamforming to steer more energy towards Bob and away from Willie can allow for non-zero covert capacity. This is a simple way to create a usable shadow network (provided Willie is not on the beampath).

2.4.4 Relay Networks

Relays repeat transmissions that they receive, transponding messages to endpoints, which are otherwise out of reach. Use of relays (a red dot with multiple connective edges () in Fig. 2.2) in covert communications allows a greater amount of data to be transmitted [39, 49–52], as Alice can lower her transmission power—she only has to reach the relay instead of transmitting all the way to Bob.

2.4.5 Using Public Messages

Most electromagnetic spectrum users are *not* concerned with transmitting discreetly[6], and their transmissions have predictable structure and occupy predictable frequencies and bandwidths. These non-discreet transmitters are said to be broadcasting "public" messages, which can be used to hide covert messages [53–56], usually with positive covert capacity.

Any detector that simply measures the power level of the spectrum and compares it to an estimate of noise power (see Section 4.2) is rendered useless in this scenario, as it does not discriminate between signal types whatsoever. If Willie knows the structure of the public message types and filters them out, this may not be enough—for Alice can embed her message *inside* the public message if she knows its structure as well [55, 56].

For Alice, this is an easy way to avoid low transmit power and low transmit rates, as transmissions of covert users become undetectable when the number of non-covert users increases [53], and a positive covert capacity can be achieved[7].

[6] As far as we can tell.

[7] This also depends on Alice having an accurate assumption of what public messages Willie can see. As in the "hidden node" problem [57, 58], Alice may see public messages from a transmitter due north of her, so if Willie is due south then he will receive the public messages at a much lower signal-to-noise ratio (SNR), leading Alice to overestimate the public message power.

2.4.6 LPI/LPD Radar

There is a very active literature that focuses upon LPI/LPD radar [59–63]. While important, LPI/LPD radar poses a slightly different problem than LPI/LPD *communications*. LPI/LPD radar waveforms need to successfully reflect electromagnetic energy off their target to detect it, which brings in a host of other problems and considerations. Fortunately Alice has no such requirement to willfully send any energy towards Willie in the covert communications scenario.

2.4.7 Beamforming & MIMO

Whenever Alice has more antennas, she can transmit more information covertly [7,15,64–67]. With either a directional antenna or multiple antennas in an array, combined with knowledge of Bob's location, Alice can steer her beam so that Bob receives more power than Willie. This allows a positive covert capacity when it results in Bob having a higher SNR than Willie.

If Alice has knowledge of Willie's location, she can use beamforming to aim a null at him for increased covertness. In the massive MIMO case (or if using a laser beam [68]) the sidelobes disappear, and Alice achieves the regular MIMO channel capacity limit with Bob, as Willie is unable to observe *any* energy from Alice unless he lies on the beampath. This "wiretap" scenario falls under the SRL regime.

If Alice has location data of Bob and/or Willie, in conjunction with control over one or more intelligent reflecting surfaces (IRSs) [69], she can use their beamforming abilities to reflect more energy away from Willie towards Bob.

2.4.8 Quantum Mechanics

The SRL has been shown to apply to bosonic[8] channels [70, 71]. A bosonic channel consists of a beam (potentially a laser beam) and a beamsplitter where Bob and Willie each receive one half of the split beam. AWGN is replaced by quantum thermal noise in this model.

Other research has shown that repurposing algorithms for quantum key distribution (QKD) to make use of anti-eavesdropper properties [72] demonstrates that positive rate covert communication is possible on a quantum internet with both passive and active eavesdroppers.

[8] A boson is a particle that has spin-$\frac{1}{2}$, which means an unlimited number of them may occupy the same quantum state. Bosons include photons, which make up the electromagnetic field, as well as all the other so-called "force-carrying" particles.

3 Transmission Schemes

Alice and Bob will be testing the efficacy of a variety of spread spectrum (SS) techniques and other LPI/LPD methods, alongside several traditional forms of communications modulation to serve as a baseline of performance.

The papers in the previous section all take an information-theoretic approach to derive the fundamental limits of undetectable communications using random coding arguments. This means that any arguments for how to construct covert codes only apply to the abstract symbol alphabet of a DMC, and are divorced from the physical constraints of a wireless transmission environment. Thus, we do not have an algorithm for generating covert communications systems (CCSs), as information symbols must be modulated onto electromagnetic waves; symbols do not simply teleport from a transmitter to a receiver.

This section describes several techniques to modulate electromagnetic waveforms in a way that decreases detectability. The techniques listed below can be (and often are) combined [73–75] to acquire the benefits of their differing properties.

Most modulation schemes encode data onto a sinusoidal wave at a particular frequency in the electro-magnetic (EM) field, and all the parameters of this sinusoidal waveform are mathematically represented in (3.1):

$$A\cos(2\pi f_c t + \phi). \tag{3.1}$$

3.1 Baseband & Passband

Disturbances in the EM field can be measured by an antenna and converted to a digital signal by an ADC, described in Section 2.3. As per the Shannon-Nyquist theorem [76], digital signals cannot have frequency components ($f_{\max}$) greater than twice the sampling rate, f_s in order to avoid aliasing:

$$f_{\max} = \frac{f_s}{2}. \tag{3.2}$$

This means that monitoring a wider band requires a higher sample rate.

3.1.1 Complex Baseband

The simulation was implemented using signals represented in complex baseband, which decreased the number of samples required to produce results. Baseband, also referred to as in-phase and quadrature (I/Q) data, allows the entire sine wave of (3.1) to be described as a single point in the 2D complex plane. Given a complex point $z = x + iy = Ae^{i\theta}$, the

amplitude A of the sine wave is $|z| = |Ae^{i\phi}| = A$. The phase is $\arg(z) = \phi$. Thus, an entire sine wave is perfectly represented as a single complex number, instead of as a discrete list of real numbers representing a sine wave being sampled at a rate f_s. The baseband representation notably also allows for a description of a modulation without reference to a specific carrier frequency f_c, and without needing to specify a bandwidth. Baseband is so frequently employed in modern radio frequency (RF) technology because of the reduction in sample rate required to faithfully represent the signal.

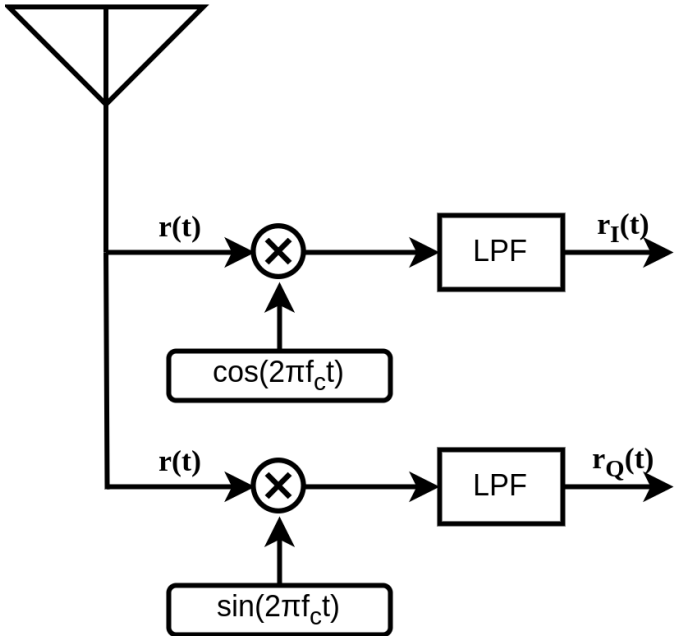


Figure 3.1: Conversion of a passband signal to baseband I/Q symbols. The received signal $r(t)$ is multiplied by a sine and a cosine wave at the carrier frequency, f_c. This reproduces the signal at zero Hertz (or baseband), where it can be isolated by the low pass filter. $r_I(t)$ represents the in-phase (real) part, while $r_Q(t)$ represents the quadrature (imaginary) part.

This ability to represent a signal independent of frequency requires the narrowband assumption; i.e., the carrier frequency f_c must be larger than the bandwidth W, or $f_c > W = \frac{f_s}{2}$. Thus the baseband model is equivalent to multiplying the antenna signal by the carrier frequency f_c, then running it through a lowpass filter of size W, as in Fig. 3.1. The real component, or the I portion, comes from multiplying the received signal by a cosine wave of frequency f_c. The imaginary, or Q component, comes from multiplying the signal by a cosine wave of frequency f_c that is phase shifted by $\frac{\pi}{2}$.

3.2 Basic Modulations

The basic transmission schemes described in this section serve as a baseline to compare against more "covert" waveforms designs described later. These work by modulating data

onto a set of complex symbols, called a constellation. Several constellations for basic modulations are depicted in Fig. 3.2. The constellation points consist of different values for the amplitude A, frequency f_c, and phase ϕ [77, Ch. 3], and form the basis for understanding more advanced transmission schemes.

Fortunately, as a result of using the complex baseband representation of a signal described in Section 3.1, any variation in the amplitude A and phase ϕ can be represented as a point in the complex plane, which disregards the carrier frequency f_c.

3.2.1 Phase Shift Keying

In phase shift keying (PSK), data is modulated onto a carrier wave by changing its phase. The phase is represented by ϕ in (3.1). The simplest variant of phase shift keying (PSK) is binary phase shift keying (BPSK), which has two phase offsets to encode the "1" and "0" bits: ϕ is either 0 or π. The I/Q constellation for binary phase shift keying (BPSK) has points at $1 + 0i$ and $-1 + 0i$, as seen in Fig. 3.2a. The receiver for BPSK checks which constellation point the received symbol is closest to, which amounts to checking if the real part of the symbol is greater than or less than zero.

Putting two BPSK systems together—one on the I-axis, and another rotated 90° on the Q-axis—produces quadrature phase shift keying (QPSK), which has 4 constellation points, as in Fig. 3.2b. QPSK is often used as it has twice the spectral efficiency of BPSK. That is, the BER of QPSK and BPSK is identical under an AWGN channel (see Appendix A.2), and QPSK transmits twice as many bits per symbol, so it makes better use of the spectrum.

Generalizing the modulation further, M-ary phase shift keying (M-PSK) is a modulation with M distinct phases. The phase values of the M-PSK constellation are equidistant points of a circle around the origin—the roots of unity—with the n-th phase being given by the following equation:

$$\phi_n = e^{2\pi i \frac{n}{M}}, \tag{3.3}$$

where $n \in 1, 2, \ldots, M$. The receiver here is once again just checking which constellation point the received symbol is closest to.

3.2.2 Quadrature Amplitude Modulation

Quadrature amplitude modulation (QAM) is best understood by first looking at the constellation diagram (Figs. 3.2d, 3.2e). While M-PSK bunches points around a circle that become closer and closer together as M increases, QAM seeks to spread out the constellation points in a denser pattern, to maximize[1] the inter-symbol distance for a given area. 16-QAM and 64-QAM constellations can be seen in Fig. 3.2d and Fig. 3.2e.

QAM constellations form grids of squares, and every M-QAM constellation transmits $\sqrt{M}$ bits per symbol. The receiver is once again just finding the symbol that has the lowest Euclidian distance to the received symbol (as is the case with PSK and amplitude shift keying (ASK)). A 4-QAM constellation (see Fig. 3.2b) is identical to the QPSK constellation, and has the same performance and error characteristics.

[1] Although regular rectangular M-QAM is used throughout this thesis, QAM that uses a hexagonal grid [77, Ch. 4.7] instead of a rectangular grid is the densest pattern and maximizes inter-symbol distance.

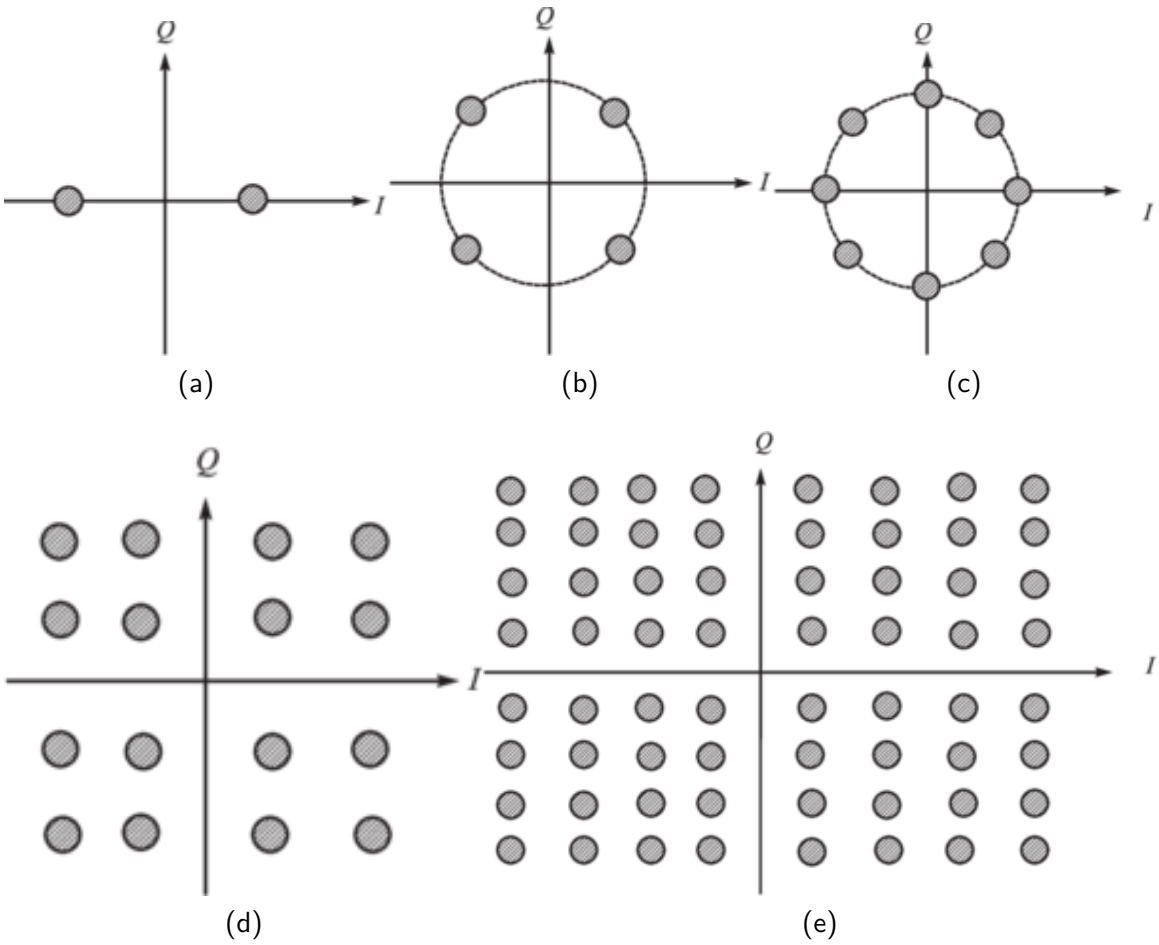


Figure 3.2: The constellations of several common modulations including BPSK (3.2a), QPSK/4-QAM (3.2b), 8-PSK (3.2c), 16-QAM (3.2d), and 64-QAM (3.2e). (Source: Singh *et al.* [78])

3.2.3 Frequency Shift Keying

Frequency shift keying (FSK) modulates data by transmitting sine waves of one or more frequencies (i.e., modulating data via f_c in (3.1). The simplest FSK modulation is to transmit one frequency (f_1) for a "1" bit and a second frequency (f_2) for a "0" bit. This is called binary frequency shift keying (BFSK), and is represented mathematically thusly:

$$s(t) = \begin{cases} 1: & A\cos(2\pi f_1 t) \\ 0: & A\cos(2\pi f_2 t) \end{cases}. \tag{3.4}$$

The passband representation of M-FSK with M evenly spaced frequencies of spacing Δf, where f_c is the lowest frequency, is given by:

$$s(t) = A\cos^{2\pi(f_c+m\Delta f)t}. \tag{3.5}$$

where $m \in \{1, 2, \ldots, M\}$.

A single complex number only represents a particular phase and amplitude—the complex baseband representation of changes to frequency must also be a function of time. To imagine what BFSK symbols might look like in complex baseband, translate the signal to the origin by using middle frequency, $f_{IF} = \frac{1}{2}(f_1 + f_2)$, such that the resulting frequencies are $\pm\Delta f = \pm\frac{1}{2}|f_1 - f_2|$, centered around the zero frequency.

This results in the baseband M-FSK data following the function

$$s(t) = Ae^{2\pi i m\Delta f t}, \tag{3.6}$$

where $m \in \{-\frac{M}{2}, -\frac{M-1}{2}, \ldots, \frac{M-1}{2}, \frac{M}{2}\}$. The resulting I/Q points rotate around the origin, with positive frequencies rotating counter-clockwise and negative frequencies rotating clockwise. Doubling the frequency means doubling the rotation rate.

The modulations we have seen thus far each have no time dependency in their complex baseband representation, and thus each symbol can be represented as a single sample (i.e., a single complex number). This is because complex baseband is often used to describe communications schemes without reference to a carrier frequency, while FSK uses several carrier frequencies. As (3.6) has time dependence, in a discrete-time simulation multiple samples of $s(t)$ are required to faithfully represent the signal, as each symbol is no longer a static complex number, but a function of time.

A baseband BFSK receiver measures the angular velocity from sample to sample and declares a "1" bit if the overall angular rotation was positive (or counter-clockwise), and a "0" bit when the overall rotation was negative (or clockwise).

3.2.4 Orthogonal Frequency Division Multiplexing

More complex than the previous modulations we have looked at, orthogonal frequency division multiplexing (OFDM) splits, or multiplexes, a high bandwidth signal onto several lower bandwidth subcarrier frequencies [79–81]. This spreads out the frequency spectrum of the signal and increases resistance to multipath interference and frequency-selective fading (FSF) [82]. The usage of orthogonal frequencies in OFDM means there is no inter-symbol interference (ISI) between subcarriers. Orthogonal frequencies are found automatically when the OFDM symbol is generated by an inverse fast-Fourier transform (IFFT).

OFDM can be transmitted by taking groups of M complex data symbols and passing these signals through an IFFT with M bins, as in Fig. 3.3. Some of these M bins may not contain data bits in order to act as pilot signals, which increase synchronization at the cost of lower data throughput [82]. The receiver similarly groups received complex symbols into groups of M before passing them through a fast-Fourier transform (FFT) with M bins to recover the transmitted data stream.

The modulated data that is the input in Fig. 3.3 can be any modulation with any amplitude and/or phase-modulated signal, like ASK, PSK, and QAM.

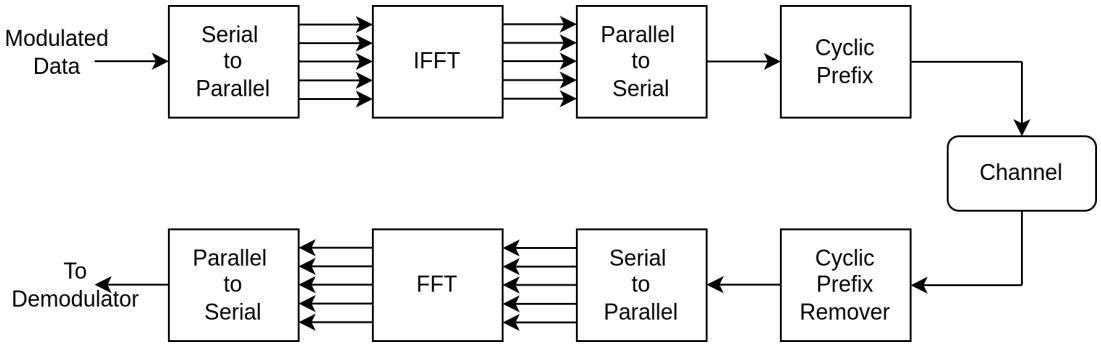


Figure 3.3: A block diagram of an OFDM transmitter and receiver. (Source: Kaur & Kansal [83])

A suitable demodulator needs to be added to the end of the block diagram in order to recover the original bitstream. Thus, OFDM must always be chained with another modulation scheme. The simplest OFDM scheme is to have the "1" and "0" bits be $\pm 1 + 0i$, which would be OFDM-BPSK.

OFDM introduces three new parameters to define the modulation. The first is the number of subcarriers, which determines the size of the FFT needed. The second parameter is how many of the subcarriers are "pilot" signals. Pilot signals do not transmit any data, and instead are used to help with channel estimation and synchronization. For example, a OFDM scheme with 64 subcarriers may use anywhere from 0 to 63 of them as pilot signals. The latter case would be equivalent to a low rate version of the base modulation, *sans* OFDM. The third parameter introduced by OFDM is the size of the cyclic prefix (CP). The CP is a prefix added to help with multipath propagation.

3.3 Spread Spectrum

Section 3.2 described the basic digital modulations widely used in communications. While they do not possess notable covertness properties, measuring the relative "covertness" of different modulation schemes is challenging (as discussed in Chapter 5). It is intuitive that the more the signal looks like the background noise, and the lower the signal energy relative to the noise floor across the signal bandwidth, the harder it is to be confident of the existence of the transmitted signal.

In this section we consider spread spectrum (SS) techniques used to "flatten" a signal out in the frequency domain [74], as measured by the power spectral density (PSD) of a SS signal. Traditionally, SS techniques provide anti-jamming and anti-interference properties [74], as jamming and interference are usually bandlimited. This helps with the covertness goal; if Willie monitors a limited bandwidth, the SNR will be lower for him.

3.3.1 Direct Sequence Spread Spectrum & Code Division Multiplex Access

In direct sequence spread spectrum (DSSS) systems, a signal is multiplied by a spreading sequence that has a higher rate—the *chip rate*, k—than the bit rate of information signal. The higher the chip rate, the more spread out the PSD becomes in the frequency domain. This lowers the signal energy at Willie if he is using a bandlimited detector. To recover the data signal, Bob needs to multiply his received signal by the spreading sequence.

Multiple users can occupy the same carrier frequency and bandwidth, transmitting reliably at the same time, if their spreading sequences are orthogonal. This is called code division multiplex access (CDMA). Orthogonal spreading sequences suitable for CDMA can be constructed by using the rows of a Hadamard matrix, which are known as Walsh codes (WCs). As CDMA is more widely used in the literature, I use CDMA to refer to both CDMA and direct sequence spread spectrum (DSSS) generally throughout this thesis.

3.3.2 Frequency-Hopping Spread Spectrum

The frequency-hopping spread spectrum (FHSS) technique spreads out the frequency spectrum by rapidly changing the carrier frequency over time (or *hopping*). This makes the signal harder to intercept unless the hopping pattern is known. Multiple symbols transmitted within a single hop is called slow-hopping. Fast-hopping occurs when the hopping rate exceeds the symbol rate (i.e., each symbol is split over multiple frequencies). A wideband receiver could listen over the whole spectrum to hear the message at any frequency, but knowing the hopping pattern allows a receiver to focus on that narrow bandwidth of the spectrum and sample the signal at a higher fidelity. The frequency hopping table is derived from the key that Alice and Bob share, as described in Section 2.1.

3.3.3 Chirp Spread Spectrum

A chirp is a sinusoidal signal whose frequency either monotonically increases (an upchirp) or decreases (a downchirp). They are difficult to detect due to the fact they continuously change in frequency and can cover a wide band. Chirp spread spectrum (CSS) occurs when wideband chirps[2] are used to transmit information. Owing to their nature, chirps are resistant to Doppler effects, and have good performance under multipath fading [84].

Frequency-hopped chirp spread spectrum (FH-CSS) [75] combines frequency hopping with chirp spread spectrum to prevent signal detection. Bits are represented as upchirps and downchirps. The chirps occur within a fixed bandwidth, but at randomly selected frequencies. For Bob to receive the chirps, the series of frequency hops must be derivable from a shared secret with Alice.

3.4 Chaotic Communications

A chaotic system is a dynamical system that produces a drastically different output given a small perturbation of the input [85,86]. Signals from such systems are “irregular, aperiodic,

[2]Not to be confused with the *chips* of DSSS (Section 3.3.1).

uncorrelated, broad-band, and impossible to predict over longer times" [87]. Chaotic signals have zero cross-correlation and zero auto-correlation everywhere except with zero delay (i.e. $t = 0$) [88], and lowered cyclostationarity [89], combined with the spreading characteristic. This makes chaotic signals very difficult to detect.

Modulating an information signal by a chaotic one works effectively as a communications scheme [90–93], but creates a new problem of synchronizing the chaos generators between Alice and Bob [94].

3.4.1 Chaotic Shift Keying

Chaos shift keying (CSK) was the first chaotic communications scheme described [87], and requires N chaos generators to be synchronized (where $N \geq 2$). It was first introduced in 1993 [95]. The chaos generators generally have similar attractors, so their values can be kept in the same broad range. CSK allows a total of N message symbols by simply assigning the output of each of the N chaos generators, $g_i(t)$, $t = 1, \ldots, N$, to a message symbol:

$$s_{CSK}(t) = \begin{cases} g_1(t), & \text{if } m(t) = m_1 \\ g_2(t), & \text{if } m(t) = m_2 \\ \vdots & \vdots \\ g_N(t), & \text{if } m(t) = m_N \end{cases}. \tag{3.7}$$

3.4.2 Differential Chaotic Shift Keying

Differential chaos shift keying (DCSK) is a classic chaotic communications protocol [90, 92, 96–99] that solves the synchronization problem of CSK noted in Section 3.4.1 by transmitting the chaotic sequence alongside the signal. Two time slots are needed for transmitting each bit. During the first slot an element of a chaotic sequence from a chaos generator is transmitted. During the second time slot, the chaotic reference is transmitted again for a "1" bit, and an inverted copy of the reference is transmitted for a "0" bit.

To receive DCSK, wideband delay lines are used to check the correlation of adjacent time slots. If two time slots are correlated then a "1" bit is output, and if they're anti-correlated then a "0" bit is output. Thus DCSK does not employ a shared secret between Alice and Bob, and—while solving the chaotic synchronization problem—leaves DCSK symbols recoverable by anyone.

3.4.3 Quadrature Chaos Shift Keying

Quadrature chaos shift keying (QCSK) is to DCSK as QPSK is to BPSK, in the sense that QCSK is composed of two orthogonal DCSK systems. DCSK only modulates data onto only the real part of the complex I/Q plane, so QCSK exploits this inefficiency by adding another DCSK modulator onto the complex plane to double the number of bits per symbol. These two orthogonal DCSK modulators are, of course, using two separate chaotic sequences.

3.4.4 Frequency-Hopped Orthogonal Frequency Division Multiplexing Differential Chaos Shift Keying

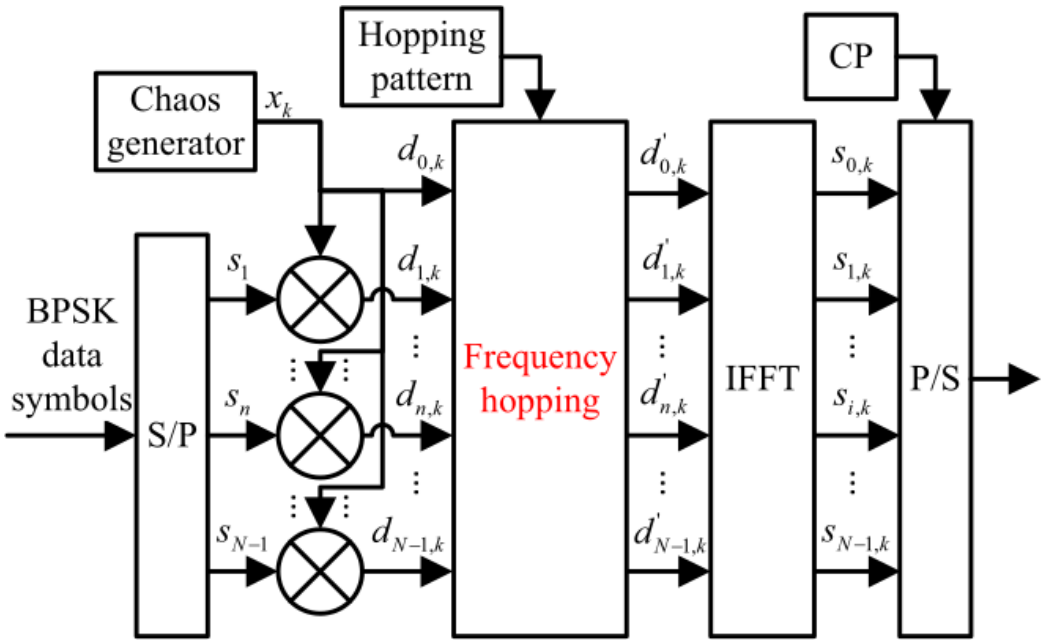


Figure 3.4: A block diagram of a FH-OFDM-DCSK transmitter. (Source: Lie *et al.* [80])

Frequency-hopped OFDM-DCSK (FH-OFDM-DCSK) [80,100] combines frequency hopping (FH), OFDM, and chaos with DCSK to make a more covert waveform.

A block diagram of FH-OFDM-DCSK is shown in Fig. 3.4. For FH-OFDM-DCSK with N OFDM subcarriers, $N-1$ symbols are split into parallel and multiplied by the current output of the chaos generator (x_k in Fig. 3.4). $N-1$ of the subcarriers have data, but one ($d_{0,k}$) is just the chaotic symbol x_k, left as a pilot symbol to aid in demodulation. The next block, the "frequency hopping" block of Fig. 3.4, just rearranges all the input symbols randomly (according to a pre-shared secret key). The remainder of the transmitter is identical to OFDM (an IFFT, parallel-to-serial conversion, with addition of the CP). The primary difference to regular OFDM in Section 3.2.4 is that the input symbols are modulated by a chaotic generator, and the inputs to the IFFT are shuffled for every symbol.

4 Warden Detection Schemes

How does our warden Willie determine whether Alice has transmitted or not? This is a binary yes/no question that asks whether Alice's signal is present, given Willie's (inherently noisy) observation of the EM spectrum. Answering this question requires delving into the mathematics that underlies signal detection: hypothesis testing.

4.1 Hypothesis Testing

Given his observation of the channel, i.e., the received signal $r(t)$, Willie has to decide between two hypotheses to determine whether Alice transmitted: H_0, that he only observed channel noise $n(t)$, and H_1, that he observed both noise and Alice's signal, $s(t)$:

$$\begin{array}{ll} H_0: & r(t) = n(t) \\ H_1: & r(t) = s(t) + n(t) \end{array}. \tag{4.1}$$

Accurately determining which of these hypotheses is true is the crux of detection.

4.1.1 Detector Theory

A detector function takes a received signal $r(t)$, and outputs a value, λ, that is compared to a threshold. The threshold determines which of the two hypotheses in (4.1) is selected. $\mathcal{D}(\cdot)$ denotes the detector function, whose domain is the signal space, and whose output is the test statistic λ. It is useful to think of a detector as a black box that takes a signal and outputs a single number λ. This output test statistic, λ, can then be compared to a predetermined threshold value λ_0. If the threshold is exceeded, then the H_1 hypothesis is selected, otherwise the null hypothesis H_0 is chosen. This is represented mathematically as:

$$\mathcal{D}(r(t)) = \lambda \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_0. \tag{4.2}$$

There are four cases that can occur with this setup:

- *detection*: positively identifying a signal when one was transmitted, with probability $\mathbb{P}_{\mathrm{D}}$,
- *false alarm*: believing there is a signal when none was transmitted, with probability $\mathbb{P}_{\mathrm{FA}}$,
- *missed detection*: believing that no signal was transmitted when one was transmitted, with probability $\mathbb{P}_{\mathrm{MD}}$.

- *true negative*: believing that no signal was transmitted when one was transmitted, with probability $\mathbb{P}_{\text{TN}}$.

These can be defined as conditional probabilities:

$$\mathbb{P}_{\text{D}} = \mathbb{P}(\lambda \geq \lambda_0 | H_1), \tag{4.3}$$

$$\mathbb{P}_{\text{FA}} = \mathbb{P}(\lambda \geq \lambda_0 | H_0), \tag{4.4}$$

$$\mathbb{P}_{\text{MD}} = \mathbb{P}(\lambda < \lambda_0 | H_1) = 1 - \mathbb{P}_{\text{D}}, \tag{4.5}$$

$$\mathbb{P}_{\text{TN}} = \mathbb{P}(\lambda < \lambda_0 | H_0) = 1 - \mathbb{P}_{\text{FA}}. \tag{4.6}$$

4.1.2 Neyman-Pearson Lemma

There is a mathematical way to maximize detections (and ergo $\mathbb{P}_{\text{D}}$) for a fixed false alarm rate $\mathbb{P}_{\text{FA}}$. This is the Neyman-Pearson lemma which provides with the uniformly most powerful (UMP) likelihood ratio test (LRT) [101]. It works by taking the test statistic λ that is output by the detector $\mathcal{D}(\cdot)$, and assessing whether that output from the detector is more likely when Alice transmitted ($\mathbb{P}(\lambda|H_1)$) or more likely when she did not ($\mathbb{P}(\lambda|H_0)$):

$$\mathcal{D}(r(t)) = \lambda \Rightarrow \frac{\mathbb{P}(\lambda|H_1)}{\mathbb{P}(\lambda|H_0)} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_0. \tag{4.7}$$

Our assessment of whether the test statistic λ in (4.7) is more likely to appear in the H_0 case or the H_1 case relies on our prior assumptions about the channel and the signal of interest (SOI). The technique that I used to determine the optimal threshold λ_0 is discussed in Section 6.2.

4.2 Radiometric Detectors

The radiometer—also known as the energy detector, total power detector, or quadratic detector—is a quintessential tool for any sort of spectrum sensing. By simply summing the signal power over time, energy detectors are agnostic to all other underlying signal parameters. If the only thing Willie knows is that Alice's signal, $s(t)$, is a stationary Gaussian stochastic process, and the noise, $n(t)$, is AWGN, then the optimal detector as per the Neyman-Pearson lemma above in Section 4.1.2, (4.7), is the following equation that calculates the total energy of the signal.:

$$\mathcal{D}(r(t)) = \int_0^T |r(t)|^2 dt \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_0. \tag{4.8}$$

A generic radiometer is depicted in Fig. 4.1, and consists of a bandpass or lowpass filter of bandwidth W, followed by a squared magnitude detector and an integrator with integration period T that feeds into a threshold detector. Willie turns a radiometer into a detector by measuring the total energy of the bandwidth he is monitoring and comparing this energy against the power he expects to see from AWGN with variance N_0.

The optimal threshold for the radiometer in stationary AWGN has been analytically determined [102, 103]. With the threshold value equal to N_0, Willie's estimate of the noise variance, such that $\lambda_0 = N_0$, the detector $\mathcal{D}(\cdot)$, specified in (4.8) is:

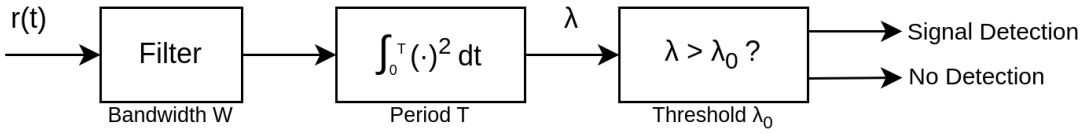


Figure 4.1: Block diagram of a bandlimited radiometric detector. The received signal is passed through a filter of bandwidth W, then squared and integrated over a period T. The total energy is output and compared to the threshold λ_0. If the threshold is exceeded, then a detection event is recorded.

$$\mathcal{D}(r(t)) = \sqrt{\frac{1}{n}\sum_{t=1}^{n} r(t)^2} \underset{H_1}{\overset{H_0}{\lessgtr}} N_0. \tag{4.9}$$

This thesis work uses the log-normalized version of the metric in (4.8) to stay consistent with other research [104]:

$$\lambda_{\text{Energy}} = 10\log_{10}\left[\sum_{r_i \in r(t)} |r_i|^2\right]. \tag{4.10}$$

How I calculated the optimal threshold, λ_0, is discussed later on in Section 6.2.

The radiometer of (4.8) is the most powerful detector of an unknown deterministic signal under AWGN [105]. As this method relies on an accurate estimate of N_0 (and a constant N_0), it does not perform well when the noise power level changes over time, or under narrowband interference.

A radiometer can be baseband (via intermediate frequency (IF) multiplication plus low pass filter (LPF), then square and integrate), or bandpass (via band pass filter (BPF)), then square and sum) [106, Ch. 10]. It is very easy for any other signal to interfere and trigger a false alarm, as this detector does not discriminate whether the energy source is background noise, Alice's signal, or an interfering signal. Using a Chebyshev filter has been shown to be better than using a Butterworth or a Bessel filter [107] in the filter step of Fig. 4.1.

The probability of signal detection, $\mathbb{P}_{\text{D}}$, and the false alarm probability, $\mathbb{P}_{\text{FA}}$, both have closed form solutions under constant variance AWGN. These depend on the bandwidth W of Willie's radiometer, as well as how long he observes the channel (i.e., his integration period T). Multiplied together these form the TW product. Alongside selecting a threshold value λ_0, all the parameters of the radiometer in Fig. 4.1 have been used to determine $\mathbb{P}_{\text{D}}$ and $\mathbb{P}_{\text{FA}}$ for *any* deterministic signal using just the energyE_{signal} [102, 105]:

$$\mathbb{P}_{\text{FA}} = \frac{1}{2}\operatorname{erfc}\left[\frac{\lambda_0 - N_0 TW}{\sqrt{2N_0^2 TW}}\right], \quad TW \gg 1, \tag{4.11}$$

$$\mathbb{P}_{\text{D}} = \frac{1}{2}\operatorname{erfc}\left[\frac{\lambda_0 - N_0 TW - E_{\text{signal}}}{\sqrt{2N_0^2 TW}}\right], \quad TW \gg 1. \tag{4.12}$$

Here, erfc(·) is the complementary error function, equal to $1 - \text{erf}(\cdot)$, where erf(·) is the error function. Equation (4.11) and (4.12) only account for the case where Alice is either transmitting during the *whole* integration period T, and over the *whole* bandwidth W, or that she is not transmitting at all. However, analytical equations exist for the cases where Willie only sees part of Alice's signal [108, 109]; either because her signal is shorter than Willie's integration period T, or because of frequency hopping (if Willie is using a channelized radiometer, see Section 4.2.1).

Equation (4.11) can be re-arranged to calculate the optimal constant false alarm rate (CFAR) threshold, λ_0, using only the TW product, noise variance, and by choosing an acceptable false alarm rate, $\mathbb{P}_{\text{FA}}$ [105]:

$$\lambda_0 = \sqrt{2N_0^2 TW}\,\text{ierfc}\left[1 - 2\mathbb{P}_{\text{FA}}\right] + N_0 TW. \tag{4.13}$$

Here, ierfc(·) is the inverse complementary error function, defined such that $\text{ierfc}(\text{erfc}(x)) = x$.

4.2.1 Channelized Radiometers

In the sections above we assumed that the bandwidth W is monitored entirely by a single radiometer. Having a single radiometer is also known as a *wideband* radiometer.

Creating a channelized radiometer works by dividing the total monitored bandwidth W into multiple discrete frequency bins that are each monitored by their own radiometer. If any of the radiometers in the subchannels detects a signal, then a detection event is registered. This approach of allowing any individual radiometer to unilaterally trigger the detector is known as a binary moving window detector (BMWD) [110].

By using a channelized radiometer, Willie can filter out public messages more readily [111]. However, the channelized radiometer has worse performance than the wideband one with frequency hopping (FH), because when the hop rate increases, each channel has less signal energy that can be integrated [109,111,112]. Generalized analytical expressions for $\mathbb{P}_{\text{D}}$ have been found for channelized radiometers [113], which take into account circumstances like the radiometer only observing part of a signal transmission due to frequency hopping.

A channelized radiometer may be sweeping [106] (i.e., it is a single radiometer that monitors the bandwidth around different frequencies), or may consist of a bank of parallel detectors that monitor a fixed frequency bin. Using a bank of parallel detectors has better performance than using a frequency-sweeping detector, as they are less likely to miss a frequency-hopped modulation [108]. However, a bank of parallel detectors is more complex and expensive.

This work only concerns a one-channel wideband radiometer instead of a channelized version.

4.3 Matched Filters

Willie can attain better performance by incorporating all the information he knows about Alice's signal (e.g., for DSSS-BPSK, the code-rate, symbol rate, and carrier frequency), since "[i]t is very difficult to design a signal that is not vulnerable to a dedicated detector" [9].

The optimal detector LRT for a known signal waveform is a matched filter [108, 114–117]. Matched filters have better performance than a radiometer, but only for the waveform they are matched to.

A detector that consists of a bank of matched filters being run in parallel on the received signal can be a highly-effective method of detecting those signals whose waveforms are present in the filter bank. Matched filters perform far better than other detectors when there is low SNR, interfering signals, or unknown noise power.

4.4 Cyclostationarity Analysis

The radiometer of Section 4.2 ignores signal composition entirely, relying only on the total energy, while matched filters, discussed in Section 4.3, *only* examine how closely a received signal matches a reference signal. Cyclostationarity analysis instead seeks to distinguish signal from noise by looking for periodic components [118]. Essentially all signals originating from human sources are based off some modulation of the parameters of the sine wave from (3.1), so any signal that Alice transmits should have periodic components.

The cyclostationarity of a signal is characterized by its spectral correlation function (SCF), which is the cyclostationary equivalent of the PSD (see Section 4.4.2). Cyclostationarity analysis is the best method of detecting weak signals with poor SNR [119].

4.4.1 The Cyclic Autocorrelation Function

In order to calculate the spectral correlation function (SCF), one must first know the cyclic autocorrelation function (CAF). It measures the autocorrelation response of a signal at given cycle frequency α, when the signal is at different delays τ. The CAF function, $R_r^\alpha(\tau)$, for received signal $r(t)$, evaluated at cycle frequency α, and delay τ is:

$$R_r^\alpha(\tau) = \int_{-\infty}^{\infty} r(t - \frac{\tau}{2}) r^*(t + \frac{\tau}{2}) e^{-2\pi i \alpha t} dt. \tag{4.14}$$

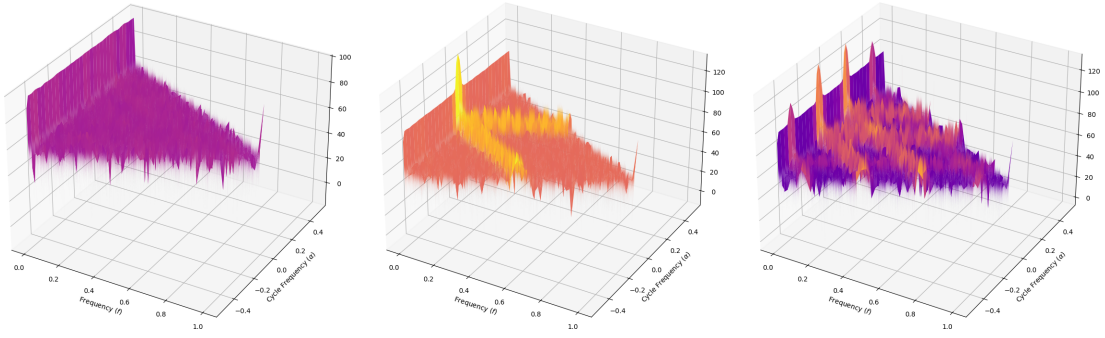
Equation (4.14) has a dependence on the delay of the autocorrelation signals, τ. To remove the dependence on τ, one integrates over all possible delays to obtain the SCF proper, as shown in the next section.

4.4.2 The Spectral Correlation Function

By the Wiener-Khinchin theorem, the SCF for cycle frequency α is simply a Fourier transform of the CAF:

$$S_r^\alpha(f) = \mathcal{F}\{R_r^\alpha(\tau)\} = \int_{-\infty}^{\infty} R_r^\alpha(\tau) e^{-2\pi i f \tau} d\tau. \tag{4.15}$$

The SCF, $S_r^\alpha(f)$, is the input and basis of all cyclostationarity detectors. The SCF is equivalent to the PSD at the cycle frequency $\alpha = 0$. Fig. 4.2 depicts the output of the SCF for an AWGN signal (Fig. 4.2a), a BPSK signal (Fig. 4.2b), and a CDMA signal (Fig. 4.2c). The AWGN signal has very little autocorrelation anywhere, except at $\alpha = 0$, where the SCF equals the PSD. The BPSK signal shows strong autocorrelation lines around the carrier frequency. The CDMA shows further autocorrelation spikes when the cycle frequency equals the chip rate as well.



(a) (b) (c)

Figure 4.2: The SCF of AWGN noise (4.2a), BPSK (4.2b), and CDMA (4.2c). The carrier frequency is 0.05Hz, and the bit rate is $\frac{1}{10}$. The noise constant for all three plots is $N_0 = \frac{1}{2}$. The CDMA signal has a chip rate of 64. The sample frequency, f_s, is 2560. The frequency axis, f, and is normalized to range from 0 to 1 (unmapped, it runs from 0 to f_s). The cycle frequency axis, α, is normalized to range from 0 to 1.
The SCF here is calculated using my implementation [36] of the strip spectral correlation algorithm (SSCA) (see Section 6.3), with $N = 2^{15}$ and $N_p = 2^7$.

4.5 Cyclostationarity Detectors

The optimal test statistic of the LRT for a weak cyclostationary signal in AWGN [119, 120] once again depends on Willie's assumptions about the structure of Alice's signal, $s(t)$. By generating the SCF of Alice's expected waveform, $S_s^\alpha(f)$, Willie can compute the SCF of his received signal $r(t)$ to get $S_r^\alpha(f)$. The optimal test statistic is thus:

$$\lambda = \frac{1}{N_0^2} \sum_\alpha \int_{-\infty}^{\infty} S_s^\alpha(f)^* S_r^\alpha(f) df. \tag{4.16}$$

Equation (4.16) depends on knowledge of $s(t)$, and incorporating this knowledge serves to create a matched filter (discussed in Section 4.3).

Nevertheless, we can proceed by finding other ways to reduce the SCF to a single number, λ, when Willie has no knowledge of Alice's signal. One way of doing this to detect a signal with an unknown composition under AWGN is by taking the period-normalized sum of the SCF:

$$\lambda(\alpha) = \frac{1}{T} \sum_f |S_r^\alpha(f)_{\Delta t}|. \tag{4.17}$$

This produces a radiometer, equivalent to (4.8), with the same inherent advantages and disadvantages detailed in Section 4.2. The metric presented in (4.17) does not exploit any of the unique properties of the SCF, so several other methods that *do* are presented in this section.

4.5.1 Degree of Cyclostationarity Detector

There are two main metrics in the literature that reduce the SCF to a single output statistic λ. The first and most widely employed way to characterize the cyclostationarity of signals is the degree of cyclostationarity (DCS) [104, 120–122].

The DCS^{α} for a particular cycle frequency α is calculated by

$$\mathrm{DCS}^{\alpha} = \frac{\sum_f |S_r^{\alpha}(f)_{\Delta t}|}{\sum_f |S_r^{0}(f)_{\Delta t}|}. \tag{4.18}$$

The total DCS of the signal is calculated by summing over all non-zero cycle frequencies α:

$$\mathrm{DCS} = \sum_{\alpha \neq 0} \sum_f |S_r^{\alpha}(f)_{\Delta t}|. \tag{4.19}$$

The $\alpha = 0$ cycle frequency is ignored, which is that part of the SCF that is equal to the PSD. The particular metric I am using is a log-normalized version of (4.19):

$$\lambda_{\mathrm{DCS}}(\alpha) = 10 \log_{10} \left[\sum_{\alpha \neq 0} \sum_f |S_r^{\alpha}(f)_{\Delta t}| \right]. \tag{4.20}$$

4.5.2 Max Cut Detector

The second metric that uses the SCF employed in this work is the maximum cut method [120], which picks out the largest squared peak for each cycle frequency. It is given by:

$$\lambda_{\mathrm{MAX}}(\alpha) = 10 \log_{10} \left[\max_f (|S_r^{\alpha}(f)_{\Delta t}|^2) \right]. \tag{4.21}$$

The metric provided in (4.21) is quite similar to DCS, except it notably includes the PSD of the signal found at $\alpha = 0$.

4.6 Other Detector Methods

There exist several other methodologies for creating detectors, including correlation detection (CD) [123], cepstrum analysis [124], and Eigenvalue detectors [125]. For a more comprehensive taxonomy of detector types (as well as their relative merits and disadvantages) see Ali *et al.* [125].

4.6.1 Normal-Distribution Test

Also known as D'Agostino and Pearson's omnibus test of normality [126, 127], the normal-distribution test checks how similar a received signal is to a Gaussian distribution by examining the skew and kurtosis of the received samples. This is sort of like creating a matched filter for the channel noise, instead of for Alice's signal.

This test is only expected to work on AWGN channels, as other channel distortion types will result in non-Gaussian noise. It takes advantage of more information from the signal than the radiometer does. This is the only correlation detection (CD) detector I am employing.

5 Metrics for Covertness

Several metrics for quantifying covertness have been presented in the literature. Some take into account specific geometric and physical information, like position of the agents, and transmitter antenna pattern. Other metrics are more generalized and abstract away the physical layer.

5.1 Energy Based Metrics

Since radiometers are agnostic to the underlying signal structure, much research has been conducted to determine the relationship between $\mathbb{P}_{\mathrm{D}}$, $\mathbb{P}_{\mathrm{FA}}$, and the SNR for these systems. Urkowitz [105] first established analytical bounds for $\mathbb{P}_{\mathrm{D}}$ and $\mathbb{P}_{\mathrm{FA}}$ for wideband radiometers at different SNRs, signal bandwidths, and observation times. These are presented in (4.11) and (4.12). Expressions for $\mathbb{P}_{\mathrm{D}}$ and $\mathbb{P}_{\mathrm{FA}}$ have also been found for channelized radiometers [113].

These analyses assume that Willie knows the noise variance, N_0, and this noise variance does not change. But how does Willie come to learn the noise variance? When Willie has to estimate N_0 himself, it can be shown that even a small amount of uncertainty in the noise power estimate can lead to markedly reduced detector performance [128, 129] compared to theoretical perfect knowledge case. Throughout this thesis I assume that Willie has perfect knowledge of N_0, (and Alice's SNR) however. Some more concrete metrics of covertness are discussed below.

5.1.1 Detectability Distance

In the classic paper that created a metric to quantify communications covertness, Weeks *et al.* [130] define the "detectability distance". Assume that Alice and Bob are close together while Willie is a distance r from Alice and a distance d from Bob, with $r > d$. Place Willie collinear to Alice and Bob, with Bob sitting in between Alice and Willie (such that Alice is a distance $r - d$ from Bob), as in Fig. 5.1. Assume that Alice and Bob are transmitting using the minimum required SNR[1] to achieve their desired BER. The metric is based off the ratio $\frac{r}{d}$ of Willie to Alice and Bob: "As the detectability distance gets smaller, [Willie] must get closer to the transmitter to detect a transmission and the system becomes more covert." [130] The detectability distance is thus the distance ratio $\frac{r}{d}$ where $\mathbb{P}_{\mathrm{D}}$ has some acceptable value[2].

[1] Weeks *et al* [130] actually use $\frac{E_b}{N_0} = 10$dB to evaluate the covertness systems in their paper.

[2] Weeks *et al* [130] use $\mathbb{P}_{\mathrm{D}} = \frac{1}{2}$.

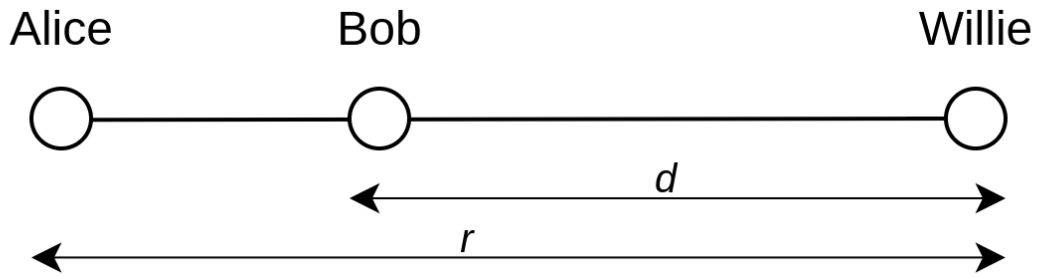


Figure 5.1: The setup for the detectability distance metric [130]. Alice, Willie, and Bob are all collinear, while Alice and Bob use the minimum required SNR to communicate.

This geometric model thus accounts for path loss (assuming unity gain omnidirectional antennas for Alice and Bob). Weeks *et al.* [130] tested several commercial off-the-shelf (COTS) modulations with detectability distance: GSM, IS-54, IS-95 and wideband CDMA. These cellular modulations leave Alice playing the role of the base station, while Bob acts as a mobile subscriber.

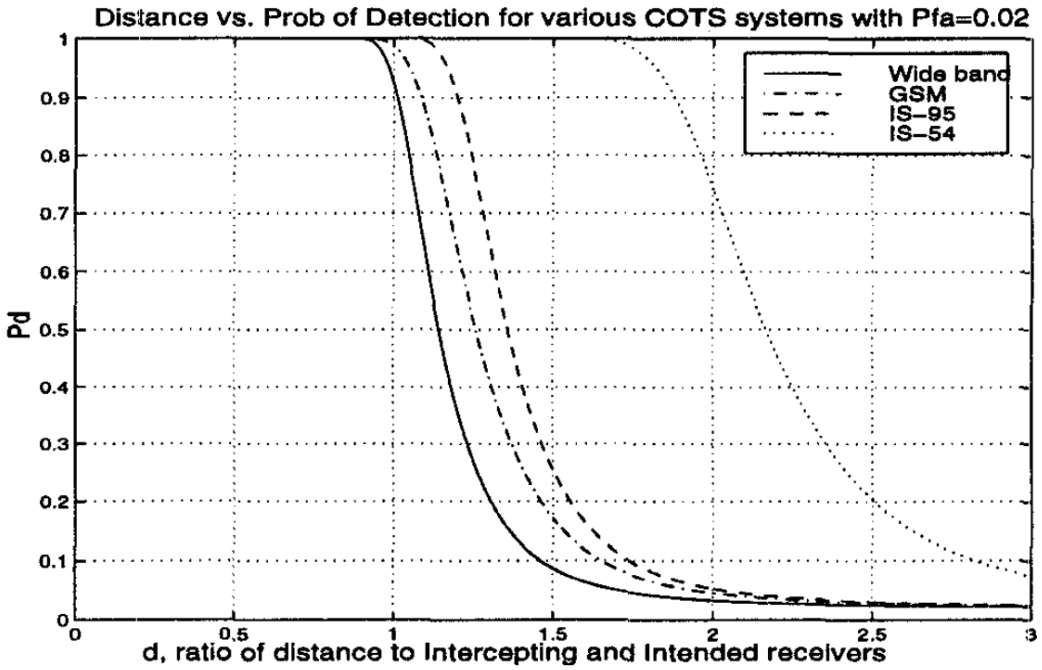


Figure 5.2: The probability of detection $\mathbb{P}_{\mathrm{D}}$ versus *detectability distance* for several COTS modulations. The curves are step-function-like, and can be reduced to the single point where $\mathbb{P}_{\mathrm{D}} = 0.5$ to compare between modulations. (Source: Weeks *et al.* [130])

Further research has shown that having multiple Alices and Bobs in a peer-to-peer shadow network drastically increases detectability by Willie [131], as both the average and

the minimum detectability distances decreases when there are more covert transceivers.

5.1.2 CEVR & SEVR

Circular equivalent vulnerable radius (CEVR) [132,133] and spherical equivalent vulnerable radius (SEVR) [134] are two energy-based detection metrics that apply not only to covert communications, but to LPI/LPD radar and RF aircraft stealth as well. This implies they only look at how detectable Alice is to Willie, ignoring the BER for Bob. Both circular equivalent vulnerable radius (CEVR) and spherical equivalent vulnerable radius (SEVR) focus on only three parameters:

1. the performance of Willie's detector,
2. the channel between Alice and Willie,
3. Alice's antenna pattern.

The primary difference between the two metrics is that CEVR is circular and 2-dimensional, while SEVR is spherical and 3-dimensional, taking into account the full 3-D antenna pattern at Alice.

They are both evaluated by selecting the maximum $\mathbb{P}_{\text{FA}}$ allowed and the minimum $\mathbb{P}_{\text{D}}$ required by Willie. To evaluate the CEVR, integrate the total area from each viewing angle of the antenna where the SNR is sufficient to attain the specified $\mathbb{P}_{\text{D}}$, called the area of detection, or A_{det} in (5.1). A circle with an equivalent area defines the CEVR, or area of probable detection. Note that this smooths out the effects of lobes.

$$\text{CEVR} = \sqrt{\frac{A_{\text{det}}}{\pi}}. \tag{5.1}$$

SEVR, as the three-dimensional extension of CEVR, defines a *detection volume*, V_{det}, wherein the SNR is sufficient to achieve the defined $\mathbb{P}_{\text{D}}$ and $\mathbb{P}_{\text{FA}}$. The SEVR is defined below in (5.2) as the radius of a sphere with volume V_{det}:

$$\text{SEVR} = \sqrt[3]{\frac{3V_{\text{det}}}{4\pi}}. \tag{5.2}$$

As these metrics rely on physical and geometric parameters like antenna pattern and distance, they are not considered in this work, as introducing these physical layer characteristics reduces the generalizability of the results.

5.1.3 Detectability Gain

Detectability gain [9] generalizes detectability distance by considering the gain (or SNR) difference between Willie and Bob. The gain difference between Willie and Bob is

$$G_{B,W} = \text{SNR}_W - \text{SNR}_B, \tag{5.3}$$

which accounts for all path losses, antenna gains, and system losses. If Willie has a path loss Pl_W, antenna gain a_W, and other "system" losses, given by L_W, and if, correspondingly, Bob has path loss Pl_B, antenna gain a_B, and system loss L_B, one can expand (5.3):

$$G_{B,W} = Pl_B - Pl_W - a_B + a_W + L_B - L_W. \tag{5.4}$$

Here "losses" (in dB) are defined to be positive numbers that are subtracted, and "gains" are added, in order to better see the effect on covertness. Note that the gain difference, $G_{B,W}$, has no dependence on Alice's transmit power. If (5.4) is evaluated only in terms of the path loss, then the detectability distance metric from Section 5.1.1 is derived as a special case of detectability gain,

The full detectability gain, G, also accounts for the loss that Willie suffers from not integrating Alice's entire signal, L_E, and the loss incurred by Willie integrating excess noise, L_N:

$$G = G_{B,W} - L_E - L_N. \tag{5.5}$$

The procedure to calculate the detectability gain is to first assume that Bob's SNR is sufficient to achieve a specified BER. Willie's probability of false alarm and a TW product are also specified. Next, the probability of detection, $\mathbb{P}_{\mathrm{D}}$, is calculated as a function of G. The detectability gain is thus the value of G where $\mathbb{P}_{\mathrm{D}}$ reaches a specified value.

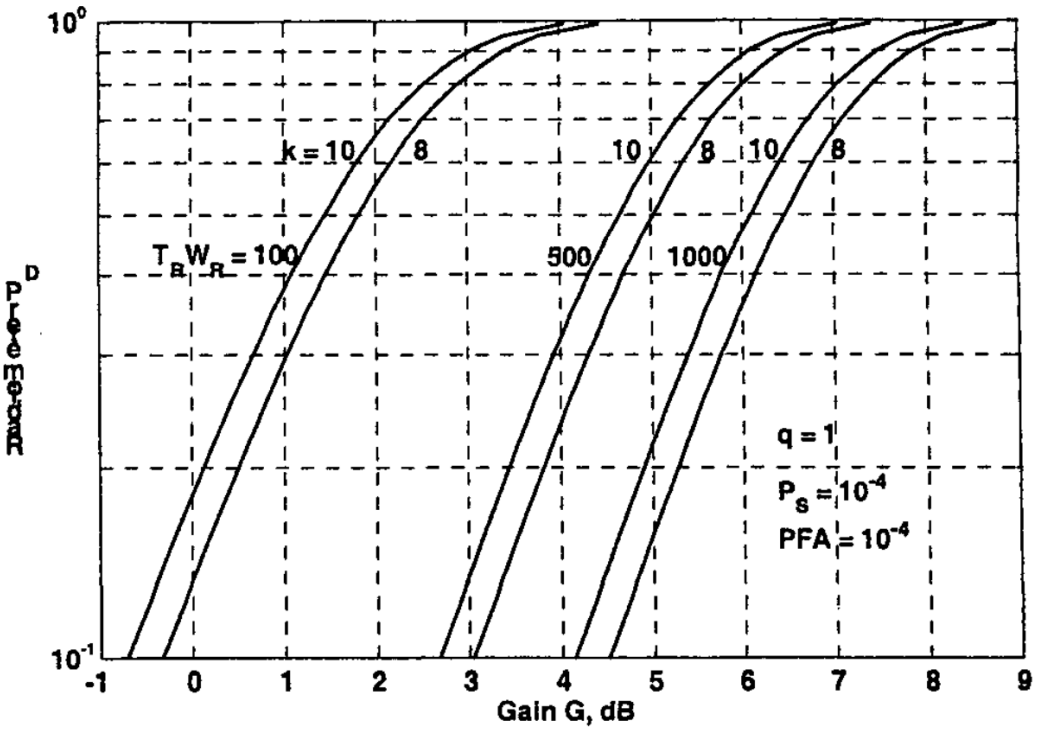


Figure 5.3: The probability of detection $\mathbb{P}_{\mathrm{D}}$ versus the gain difference G in dB for a single symbol sent by Alice. G is the SNR gain difference between Willie and Bob after accounting for all system and path losses/gains. These results are for a single symbol ($q = 1$) containing either $k = 8$ or $k = 10$ bits with $T_R W_R$ product values of 100, 500, and 1000. The probability of symbol error at Bob, P_s, and the probability of false alarm for Willie, PFA, are both 10^{-4}. (Source: Dillard and Dillard [9])

Fig. 5.3 plots the probability of detection $\mathbb{P}_{\mathrm{D}}$ versus the gain difference, G, between Willie and Bob. Fig. 5.3 shows that for a single symbol ($q = 1$) containing $k = 10$ or $k = 8$ bits, Willie must increase his gain G relative to Bob's to achieve an increase in $\mathbb{P}_{\mathrm{D}}$

for a fixed symbol error rate P_s and $\mathbb{P}_{\text{FA}}$. If Willie increases his time-bandwidth product ($T_R W_R$ in Fig. 5.3), he must achieve a larger gain G to maintain the same $\mathbb{P}_{\text{D}}$. Since Alice is transmitting a single symbol in this example, Willie integrates unnecessary noise when he increases his TW product, which reduces his detection ability, necessitating a higher gain difference, G. If Alice was transmitting symbols to occupy Willie's entire TW product, his gain difference would be lower.

Dillard and Dillard [9], when introducing the concept of detectability gain, only considered a radiometer under an AWGN channel. Notwithstanding, the framework they presented is actually generalizable to other types of detectors. Detectability gain depends on only $\mathbb{P}_{\text{D}}$, $\mathbb{P}_{\text{FA}}$, and the gain difference between Willie and Bob; there is no specific dependency to any kind of detector. However, analytic solutions may not be available. By generalizing beyond the radiometer one introduces several new dimensions to the problem—mainly the explosion in parameter space from the combination of each detector type with each transmission scheme. With the radiometer, this was not a problem as it is intrinsically agnostic to the signal structure.

5.2 Cyclostationarity Metrics

Multiple studies look at the probability of detection with cyclostationary detectors [89, 120, 135, 136], most of which use either the max cut metric ((4.21), Section 4.5.2), DCS ((4.20), Section 4.5.1), or simply act as a radiometer as in (4.17).

These detectors can all be combined so that a detection event occurs whenever any of the sub-detectors detect a signal. Combining cyclostationarity detectors has been shown to be more effective than relying on any lone detector [137].

The probability of false alarm, $\mathbb{P}_{\text{FA}}$, can be analytically determined for cyclostationarity detectors because calculating it only depends on the structure of the noise—not the structure of the signal being detected [137].

5.2.1 DCS Ratio

Many papers are concerned with lowering the degree of cyclostationarity (DCS) of their signal [120–122, 138], instead of trying to evaluate $\mathbb{P}_{\text{D}}$ and $\mathbb{P}_{\text{FA}}$. The desire to reduce cyclostationary properties of a signal comes not from detector designers, but from signal waveform designers, who are trying to engineer better transmission schemes for Alice. The DCS Ratio [122] metric is usually used to quantify an "improvement" to covertness compared to a some reference signal:

$$DCS \text{ ratio} = \frac{DCS \text{ of selected signal}}{DCS \text{ of reference signal}}. \tag{5.6}$$

DCS ratio allows waveform designers to test and compare how different techniques affect the detectability of signals by cyclostationarity detectors. This metric only concerns transmit signal structure, totally ignoring the BER for Bob and the $\mathbb{P}_{\text{D}}/\mathbb{P}_{\text{FA}}$ for Willie for different SNRs. DCS ratio essentially assumes that Willie is using a DCS detector as described in Section 4.5.1, such that when Alice lowers the DCS of her transmit signal it should also result in a decrease to Willie's $\mathbb{P}_{\text{D}}$. It is assumed that given two signals, the

one with the lower DCS ratio is "more covert", supposing they are measured relative to the same base signal.

This metric only concerns one aspect of cyclostationarity, so I will be ignoring it in this work insofar as it does not account for any channel conditions or alternate detectors that Willie may be using. Additionally, there is no standard term for the denominator in (5.6) to set as a base signal.

6 Methods

I sought to answer two questions relating to covert communications:

- **Q.1**: Which covert communications schemes are best for Alice to use?
- **Q.2**: What detectors are best for Willie to use?

Answering these questions necessitated building a communications simulation setup as in Fig. 2.1, with a channel between Alice and Bob, and a channel between Alice and Willie.

I published the simulation code with an open-source licence, and it can be found in a public repository [36]. This enables others to easily reproduce, modify, and extend the results of this work, by adding additional modulation schemes, detectors, or channel conditions.

This section details the overall structure of the communications simulation, as well as all the statistical methods and techniques needed to reproduce the results and figures.

6.1 Simulation Model

Analyzing the tradeoff between the covertness of a transmission scheme and the error rate requires calculating both the probability of detection, $\mathbb{P}_{\mathrm{D}}$, and the bit error rate (BER) over an AWGN channel. Finding the BER entails measuring the number of bit errors Bob experiences at different SNRs. Further details and plots of the BERs for all modulations are located in Appendix. A.2.

Finding the probability of detection, $\mathbb{P}_{\mathrm{D}}$, is less straightforward than calculating the BER. Fig. 6.1 portrays the four of states from (4.3)–(4.6) that Willie can experience. His detector either receives a signal that is pure noise, as in Fig. 6.1a, or Alice's signal plus noise, as in Fig. 6.1b. In this work, in the H_0 case, Alice does not transmit at all, and in the H_1 case, Alice transmits during the *entirety* of Willie's observation period, T.

The first step towards finding the probability of detection is to run the detector on many received signals that represent both the H_0 and H_1 cases at many different SNRs, in order to estimate the probability distribution functions (PDFs) of the detector output in either case. The detector function, $\mathcal{D}(\cdot)$, discussed in Section 4.1.1, outputs a value λ. This detector *output* is not the same as the detector *threshold*, λ_0, which is found using the process described in Section 6.2.2.

6.1.1 Detectors Available to Willie

All the detectors I have made available to Willie operate without any knowledge of the signal of interest (SOI), making them "blind-parameter" detectors, and are described in

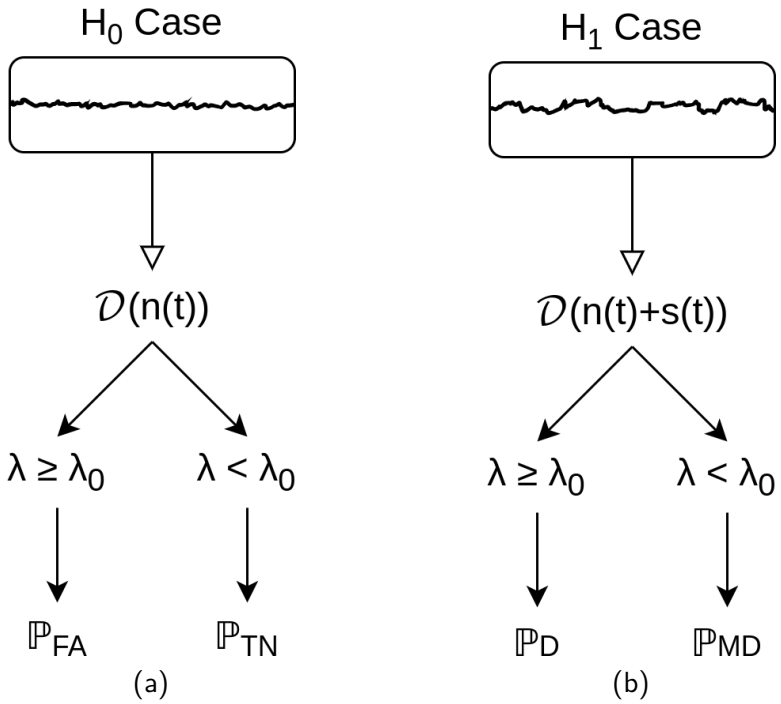


Figure 6.1: This shows the possible events for Willie. For the H_0 case (6.1a), if Willie outputs $\lambda \geq \lambda_0$ he makes a false detection (which has probability $\mathbb{P}_{\text{FA}}$). When Alice does transmit, (6.1b), Willie has either $\lambda \geq \lambda_0$, where he makes a true detection with probability $\mathbb{P}_{\text{D}}$, or $\lambda < \lambda_0$, where he misses the detection of Alice with probability $\mathbb{P}_{\text{MD}}$.

detail in Chapter 4. This excludes any sort of matched filter (Section 4.3), as matched filters require specific knowledge of parameters of the SOI.

The detectors available to Willie are:

- The Radiometer (Equation 4.10)
- The Max Cut Detector (Equation 4.21)
- The DCS Detector (Equation 4.20)
- The Normal-Distribution Test (Section 4.6.1)

6.1.2 Transmission Schemes Available to Alice

Here is the list of the transmission schemes that I implemented for Alice and Bob in the simulation:

- BPSK (Section 3.2.1),
- QPSK (Section 3.2.1),
- CDMA-BPSK, 16 and 64 bit chips (Section 3.3.1),
- CDMA-QPSK, 16 and 64 bit chips (Section 3.3.1),
- QAM, 16-QAM and 64-QAM (Section 3.2.2),
- BFSK (Section 3.2.3),

- OFDM—modulated by BPSK and QPSK with 16 and 64 subcarriers (Section 3.2.4),
- Chirp Spread Spectrum (CSS) (Section 3.3.3),
- CSK (Section 3.4.1),
- DCSK (Section 3.4.2),
- QCSK (Section 3.4.3), and
- FH-OFDM-DCSK (Section 3.4.4).

6.2 Calibrating the Detector

A detector takes a received signal, then outputs a single real number[1], λ, as a test statistic. When λ is greater than the detector threshold value, λ_0, the detector "goes off", and Willie notes a detection event of Alice's transmission as in Fig. 6.1.

How do we know which λ_0 is optimal for a specified detector? λ is easy to find; it is simply the number output by the detector. Although we have analytically determined λ_0 for the radiometer in (4.9), Section 4.2, there is not a method to find it generally for an arbitrary detector.

The detector outputs for the H_0 and H_1 cases will produce two PDFs[2], as shown in Fig. 6.2. The more that the two PDFs overlap, the less Willie's detector is able to distinguish between noise and Alice's signal. When the means differ, the H_0 case becomes clearly distinguishable from the H_1 case, and a threshold can be found to minimize classification error. Finding λ_0 thus becomes a binary classification problem of whether λ is more likely in the H_0 case or the H_1 case.

6.2.1 Receiver Operating Characteristic

With estimates of the PDFs of the detector output λ for the H_0 and the H_1 cases from the previous section, we can evaluate the performance of detectors for different threshold values λ_0. To measure the performance of a detector with various thresholds, we can use a receiver operating characteristic (ROC) plot. The ROC is a graph that shows how effective different threshold values are at performing binary classification. The abscissa[3] is the false positive rate (FPR):

$$\mathrm{FPR} = \frac{\mathbb{P}_{\mathrm{FA}}}{\mathbb{P}(H_1)} = \frac{\mathbb{P}_{\mathrm{FA}}}{\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{TN}}} \tag{6.1}$$

and the ordinate[4] is the true positive rate (TPR):

$$\mathrm{TPR} = \frac{\mathbb{P}_{\mathrm{D}}}{\mathbb{P}(H_0)} = \frac{\mathbb{P}_{\mathrm{D}}}{\mathbb{P}_{\mathrm{D}} + \mathbb{P}_{\mathrm{MD}}}. \tag{6.2}$$

There are two main extremes that can arise in a ROC curve. If the binary classifier is useless (i.e., the true positive rate (TPR) equals the false positive rate (FPR)), then the

[1] In general, the detector output λ does not strictly need to be a real number, but can be any totally ordered number.

[2] Under AWGN noise.

[3] Or x-axis.

[4] Or y-axis.

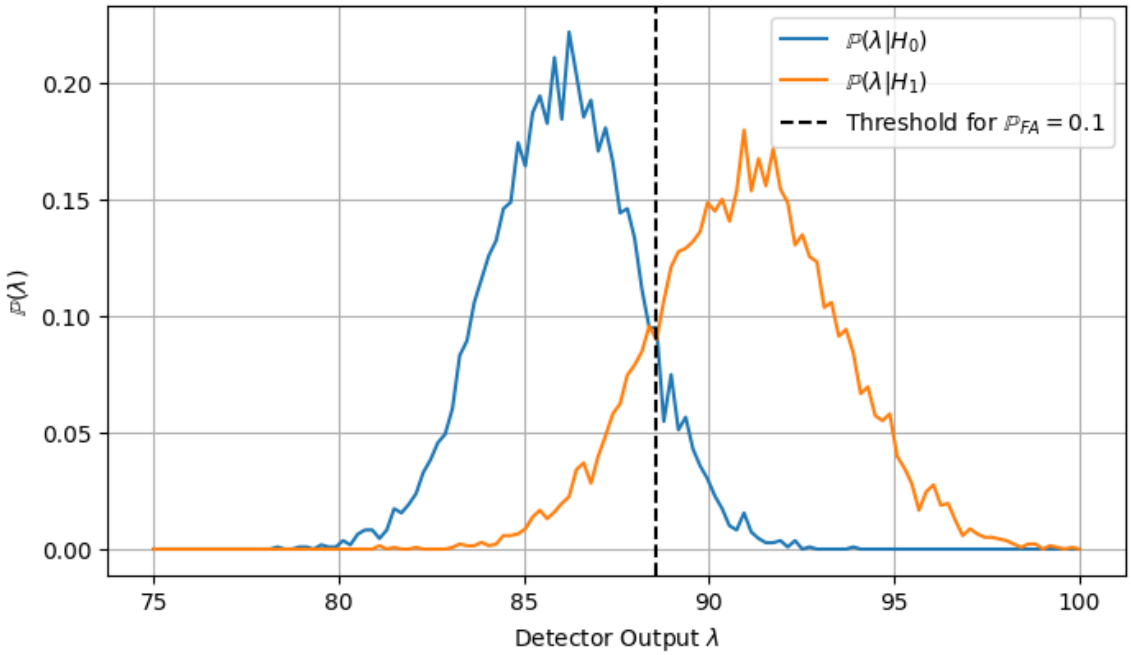


Figure 6.2: The output of the detector, λ, for both the H_0 (noise only) and H_1 (noise plus signal) cases. Although there is some overlap between the PDFs, they are clearly bimodal and the optimal threshold between them is demarcated with a dotted line at $\lambda = 88$, where the 10% of detections are false alarms.

ROC curve will be the diagonal $y = x$ line spanning from $(0, 0)$ to $(1, 1)$[5]. On the other hand, a classifier that that is accurate 100% of the time will be represented by a point at $(0, 1)$, where the TPR is 1 and the FPR is 0. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve can also be used to judge the effectiveness of a detector. An accurate detector has the area under the curve (AUC) approach one, while the AUC for a useless detector is $\frac{1}{2}$.

An ROC curve of this type is generated by taking the H_0 and H_1 PDFs and sliding the threshold λ_0 across the whole range of potential λ values. A non-perfect yet not-useless classifier will have points on the TPR-FPR ROC graph somewhere between the $y = x$ line and the point $(0, 1)$. Fig. 6.3 depicts the ROC plot of the PDFs from Fig. 6.2, where we can see that our optimal threshold, λ_0, occurs when $\mathbb{P}_{\text{FA}} \approx 0.1$.

6.2.2 Constant False Alarm Rate

Now, we can assess the performance of a detector given the ROC curve in the previous section, but we have not yet found a way to pick the ideal threshold λ_0. This work only tests "blind-parameter" detectors for Willie, which stipulates that he does not have any information about the PDF of the H_1 case, where Alice transmits. This restriction precludes

[5]The ROC curve can never be "below" the $x = y$ line where TPR = FPR. If the ROC curve goes below this line the detector needs to relabel the H_0 and H_0 cases, and now performs better than chance.

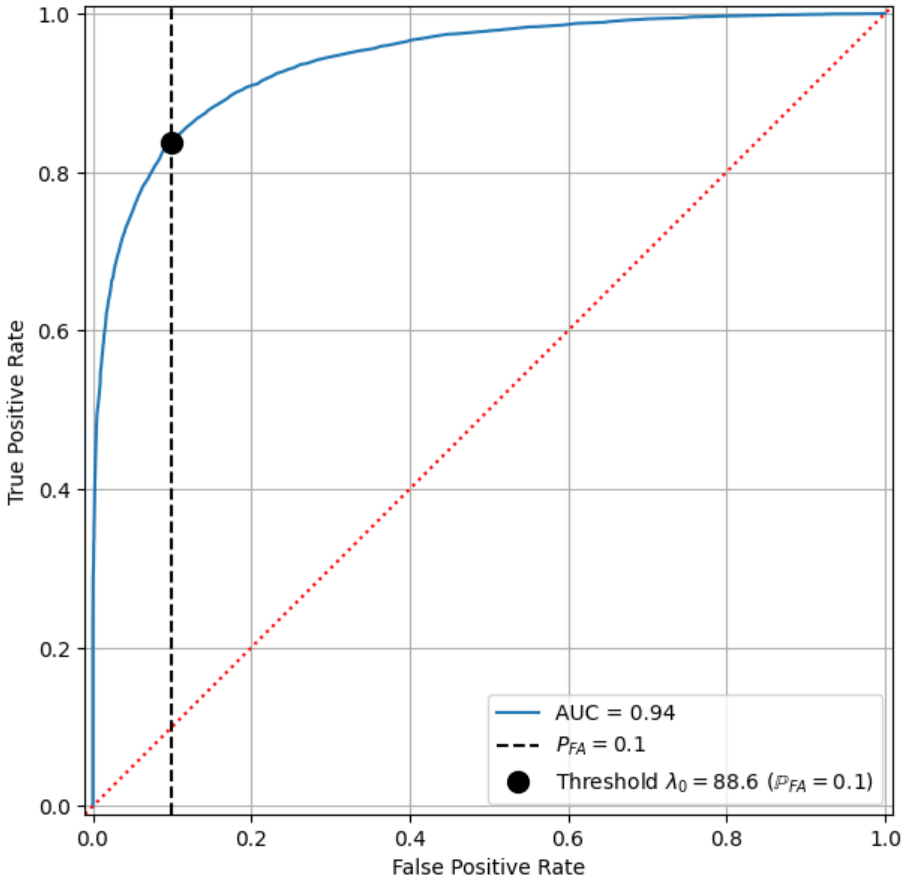


Figure 6.3: The plot shows the TPR-FPR ROC curve of Fig. 6.2. The optimal threshold, λ_0, (depicted by the large black dot) is around 89, as in Fig. 6.2.

other methods of finding that ideal threshold, like Youden's J-index [139], that incorporate both PDFs and thus achieve better performance.

Indeed, for Willie to estimate the PDF of the H_1 case requires that he has a made many valid detections of Alice's signal, and that he knows Alice's SNR (which, under AWGN, is the combination of Alice's transmit power and Willie's noise variance, N_0). Building the large dataset required to accurately depict the PDF of the H_1 case while also knowing the SNR of the SOI is unlikely in reality, so Willie must choose the detector threshold using the output of the detector on a channel consisting solely of noise.

The simplest way to do this, which is ubiquitous in the literature, is the constant false alarm rate (CFAR) method [9, 102, 108, 113, 120, 131, 137, 140]. With this technique, one first chooses an acceptable false alarm probability, $\mathbb{P}_{\text{FA}}$, and then finds the threshold, λ_0, that achieves the desired $\mathbb{P}_{\text{FA}}$. This can be visualized as choosing λ_0 such that $\mathbb{P}_{\text{FA}}$ the area under the PDF of the H_0 case in Fig. 6.2 to the right of λ_0 is equal to $\mathbb{P}_{\text{FA}}$, or

$$\int_{\lambda_0}^{\infty} \mathbb{P}(\lambda|H_0)d\lambda = \mathbb{P}_{\text{FA}}. \tag{6.3}$$

This is also equivalent to choosing the threshold in the ROC plot in Fig 6.3 by selecting the λ_0 value that is one the line above the desired $\mathbb{P}_{\text{FA}}$. In Fig. 6.3, $\mathbb{P}_{\text{FA}} = 0.1$, which happens to be the optimal threshold for these PDFs.

Thus, with the CFAR method, tuning the threshold only requires choosing $\mathbb{P}_{\text{FA}}$, which in Fig. 6.1a only depends on the PDF of the H_0 case. Under AWGN, this means that Willie must have accurate knowledge of his noise variance, N_0, at every point in time to get the PDFs for the null hypothesis at different SNRs.

6.3 The Strip Spectral Correlation Algorithm

The cyclostationarity detectors discussed in Section 4.4 rely on calculating the spectral correlation function (SCF) of the received signal $r(t)$, or $S_r^\alpha(f)$. The strip spectral correlation algorithm (SSCA) is a computationally efficient algorithm for estimating the SCF in (4.15) at *all* values of the cycle frequency, α [104,120,141,142]. It is more computationally efficient than alternative algorithms [104, 141, 142] that estimate the SCF, like the time smoothing method (TSM) and frequency accumulation method (FAM).

This is the algorithm I have implemented for Willie to estimate the SCF and perform cyclostationarity detection. An outline of the algorithm is depicted in Fig. 6.4. First, $N + N_p$ samples[6] are taken from the received signal, where $N > N_p$, and put into windows of size N_p. These blocks are $xt(1, k), \ldots, xt(N, l)$ in Fig. 6.4, and are then subject to Hadamard multiplication by a Hamming window function, $a(k)$, and the resulting blocks ($axt(1, k), \ldots, axt(N, k)$ in Fig. 6.4) are then put through an N_p-point FFT. After this FFT, the result is multiplied by its complex conjugate and by another Hamming window, $g(k)$, which is rotated 90°, and are multiplied by exponential terms

$$e^{-\frac{2\pi i(n-1)k}{N_p}} \tag{6.4}$$

for $k = -\frac{N_p}{2}, \ldots, \frac{N_p}{2} - 1$ and $n = 1, \ldots, N$. This is all put through an N-point FFT and the output is rotated 45° in the final mapping step of Fig. 6.4 to make it align properly with the frequency f and cycle frequency α domains.

6.4 The Frequency Accumulation Method

I also implemented the frequency accumulation method (FAM) algorithm to estimate the SCF. Although it has a higher computational cost compared to the strip spectral correlation algorithm (SSCA), it is not a one-shot algorithm, and the frequency resolution is tuneable. This means that a higher resolution SCF can be calculated, but it takes far longer than the SSCA. It is also a conceptually simpler algorithm to implement.

The frequency accumulation method (FAM) allows you to estimate the SCF for any arbitrary cycle frequency α, compared to the one-shot SSCA, which has the set of cycle frequencies computed built into the algorithm. To get a picture of the overall SCF, one can run the FAM in a "for"-loop over the desired cycle frequencies to get arbitrary resolution on any particular section of the SCF.

[6] Where, for the FFTs to be more computationally efficient, N and N_p are both powers of 2.

To perform the FAM algorithm, first the entire signal is Fourier transformed. Next, for each cycle frequency α, the Fourier transformed signal is circularly shifted left by α, and multiplied by the complex conjugate of the same Fourier transformed signal that is shifted right by α. The output of that multiplication is then convolved with a Hamming window of size N_p, and the result of this provides a single slice of the SCF at that cycle frequency α. The steps above are completed for each α of the SCF that is desired.

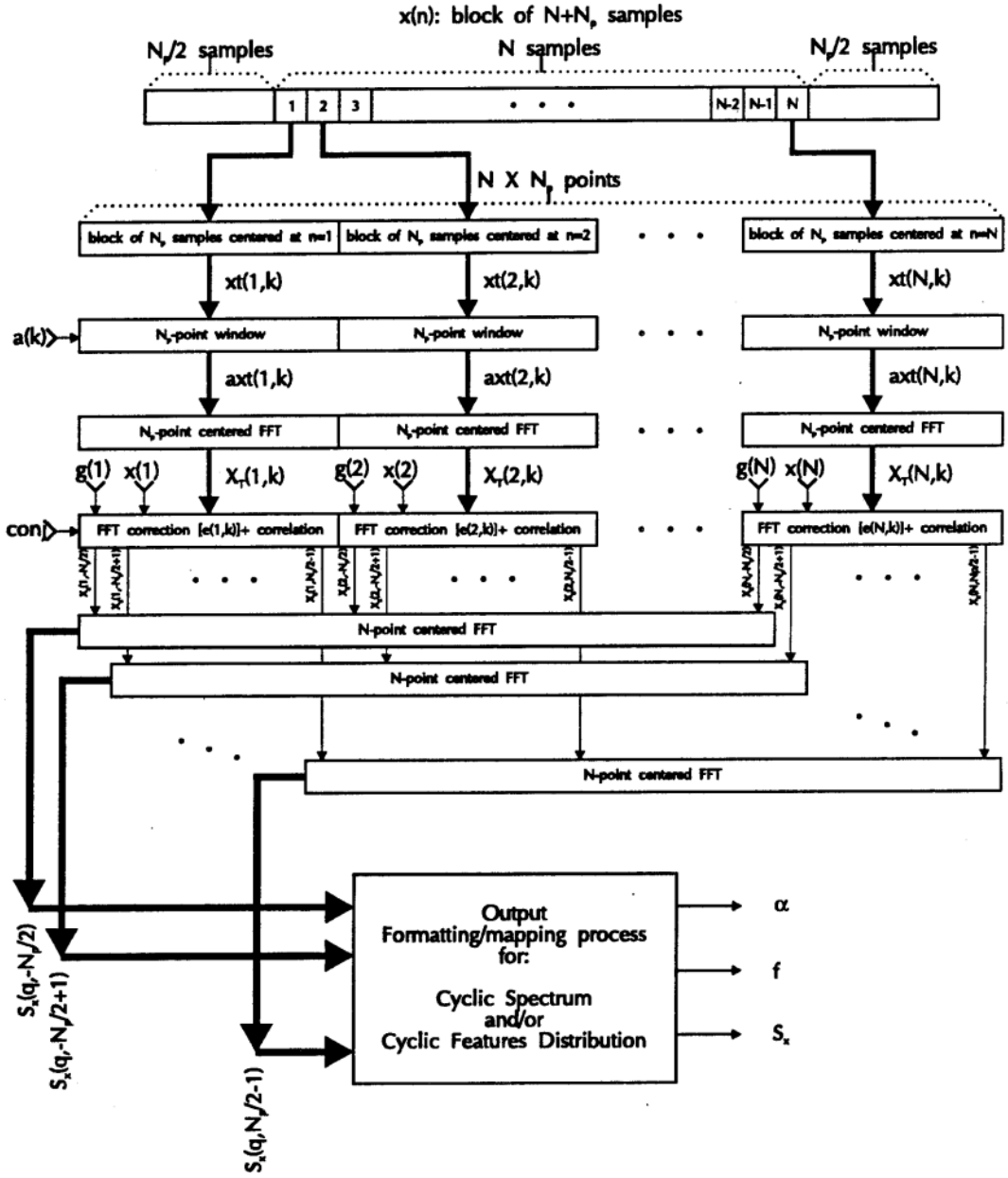


Figure 6.4: The structure of the strip spectral correlation algorithm (SSCA). Samples are windowed into blocks of size N_p and Fourier transformed before being rotated into blocks of size N and being Fourier transformed again. Lastly, an output mapping algorithm makes the data suitable for display and plotting. (Source: April [104])

7 Results & Discussion

Using the methods described above in Chapter 6, the performance of the four detectors listed in Section 6.1.1 are compared against the covertness of 21 transmission schemes, which are listed in Section 6.1.2.

The SNRs tested were 150 logarithmically spaced SNRs ranging from -45dB to 12dB (i.e., they were linearly spaced between the decibel range $[-45, 12]$). For each combination of transmission scheme, detector, and SNR, $10,000$ trials were conducted. That is, for each SNR, the detector was given $10,000$ baseband signals containing only AWGN (to represent the H_0 case) and $10,000$ baseband signals that consist of AWGN plus the transmitted signal. Each trial signal is $2^6 + 2^{12} = 64 + 4096 = 4160$ complex baseband samples. Two powers of two are required for the SSCA algorithm that powers the cyclostationarity detectors as discussed in Section 6.3. Thus, the overall number of samples is a sum of powers of two, and not a power of two itself. The results were analyzed using the CFAR method with a false alarm rate of $\mathbb{P}_{\text{FA}} = 0.01$.

To make the figures easier to read, the modulations are plotted in specific groups. The groups are listed below for reference.

Group 1

- BPSK.
- QPSK (4-QAM).
- 16-QAM.
- 64-QAM.
- CDMA-BPSK with a 16 key.
- CDMA-QPSK with a 16 key.
- CDMA-QPSK with a 32 key.
- CDMA-QPSK with a 64 key.

Group 2

- BFSK with 16 samples per symbol.
- BFSK with 32 samples per symbol.
- BFSK with 64 samples per symbol.
- OFDM-BPSK with 16 subcarriers.
- OFDM-QPSK with 16 subcarriers.
- OFDM-BPSK with 64 subcarriers.
- OFDM-QPSK with 64 subcarriers.

Group 3

- CSS with 16 samples per symbol.
- CSS with 64 samples per symbol.
- CSK.
- DCSK.
- QCSK.
- FH-OFDM-DCSK.

7.1 Detector Comparison

In this section, the relative performance of the four detectors that Willie is using are discussed. The radiometer, max cut, and DCS detectors were found to be the most useful for detecting a wide variety of modulations, with the cyclostationarity detectors performing both better and worse than the radiometer, depending on the modulation. The normal-distribution detector has significantly worse performance than the other detectors.

7.1.1 Probability of Detection Versus SNR

In order to directly compare the performance of Willie's detectors, we can make plots like those in Figs. 7.1–7.10. These plots show how the probability of detection, $\mathbb{P}_{\mathrm{D}}$, varies with SNR for various detectors with a fixed TW product. For all detectors, the TW product (see Section 7.1.4) is formed from $2^6 + 2^{12} = 4160$ complex baseband samples. The abscissa displays the SNR and the ordinate shows the probability of detection, $\mathbb{P}_{\mathrm{D}}$. With the CFAR method, $\mathbb{P}_{\mathrm{D}} \to \mathbb{P}_{\mathrm{FA}}$ as the SNR decreases, and $\mathbb{P}_{\mathrm{D}} \to 1$ as the SNR increases. A more effective detector will have this transition from $\mathbb{P}_{\mathrm{D}} = \mathbb{P}_{\mathrm{FA}}$ to $\mathbb{P}_{\mathrm{D}} = 1$ occur at a lower SNR than a worse detector.

Radiometer

As the radiometer is signal agnostic, it only depends on the SNR. Fig. 7.1 shows that every modulation reaches $\mathbb{P}_{\mathrm{D}} \approx 1$ by the time that the SNR is -10dB, and that every modulation reaches $\mathbb{P}_{\mathrm{D}} = \frac{1}{2}$ when the SNR is around -14.4dB for this TW product. Increasing the TW product increases the performance of the detector, which is discussed in Section 7.1.4.

The radiometer detectability curve thus sits as a benchmark of comparison for the other detector types.

Max Cut Detector

The max cut detector is the detector with the most varied and interesting results. It essentially looks for the "biggest spike" in the SCF of the signal. Figs. 7.2–7.4 show that some modulations are more easily detected with the max cut detector than with the radiometer in Fig. 7.1. Several modulations have the same detectability with the max cut method and the radiometer. These are DCSK, QCSK, and CDMA with a 16-bit chip rate. In these figures the SSCA was used to estimate the SCF. Similar plots of $\mathbb{P}_{\mathrm{D}}$ versus SNR for the max cut detector using the FAM approach to estimate the SCF are presented in Section 7.1.2.

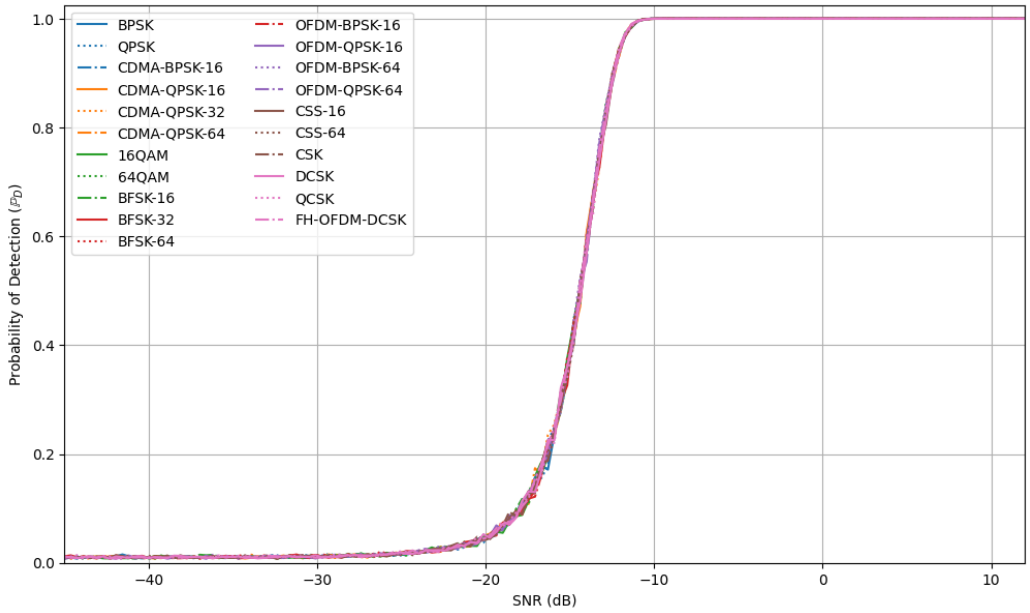


Figure 7.1: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the radiometer. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

Several modulations are *more* detectable with max cut than with the radiometer. The most detectable modulation is CSK, which is somewhat unexpected given that CSK is a synchronized chaotic method. CDMA also appears to become more and more detectable as the chip-rate increases. The max cut detector had worse performance than the radiometer for PSK, QAM, OFDM, and CSS.

All other modulations were detected with $\mathbb{P}_{\mathrm{D}} \approx 1$ when the SNR is -4dB. And every modulation had $\mathbb{P}_{\mathrm{D}} \geq \frac{1}{2}$ when the SNR is -7.25dB. A bunch of the modulations cluster together with $\mathbb{P}_{\mathrm{D}} = \frac{1}{2}$ at this -7.25dB gain point, providing a limit to how much a communications scheme can avoid the max cut detector.

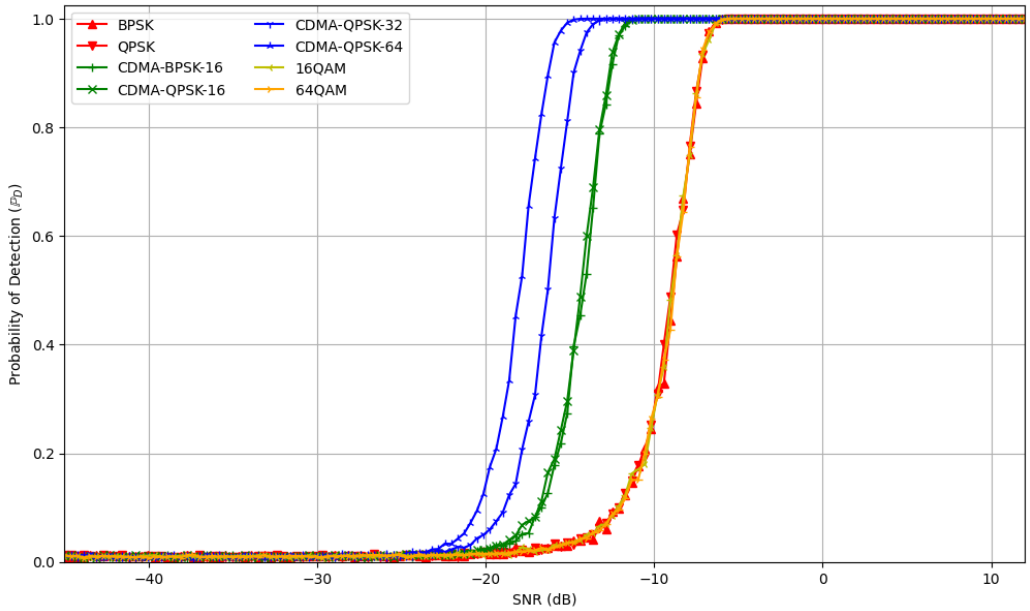


Figure 7.2: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the max cut detector with Group 1 with the SSCA algorithm. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

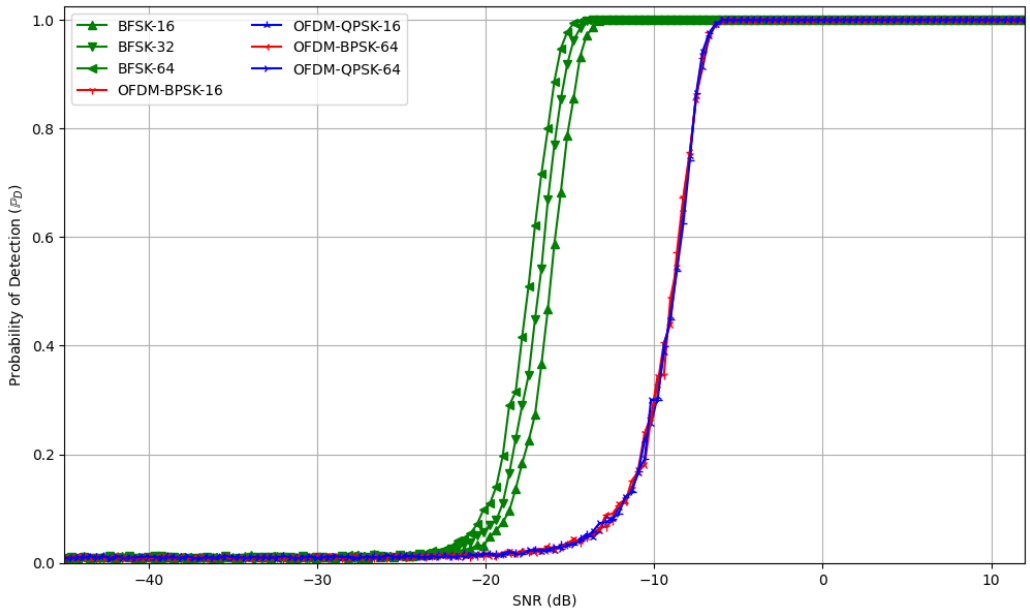


Figure 7.3: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the max cut detector with Group 2 with the SSCA algorithm. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

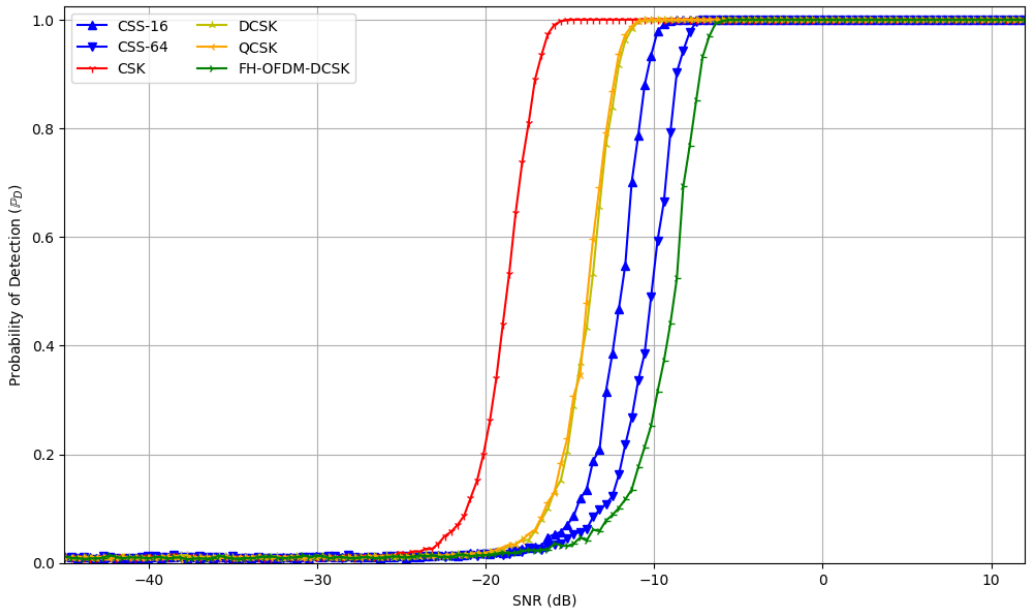


Figure 7.4: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the max cut detector with Group 3 with the SSCA algorithm. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

DCS Detector

The performance of the degree of cyclostationarity (DCS) detector has much less variance than the max cut detector. The $\mathbb{P}_D$ curves in Figs. 7.5–7.7 are bifurcated into two branches; one branch was detectable with $\mathbb{P}_D = \frac{1}{2}$ at an SNR of -15dB, while the other branch reached the same detectability with the SNR at -13dB. The modulations in the "more-detectable" group at -15dB were BFSK, CDMA, and the chaotic modulation, CSK. The "less-detectable" -13dB group included all the other modulations. One of these branches performs slightly better, and one slightly worse than the radiometer, which had $\mathbb{P}_D = \frac{1}{2}$ when the SNR is -14.4dB.

Note that Figs. 7.5–7.7 depict the detectability with SCF estimated by the FAM method. Using the SSCA produced worse results than the FAM—more of the modulations were in the -15dB branch, and all modulations had worse performance than with the radiometer. Further discussion of the differences between the SSCA and the FAM is found in Section 7.1.2.

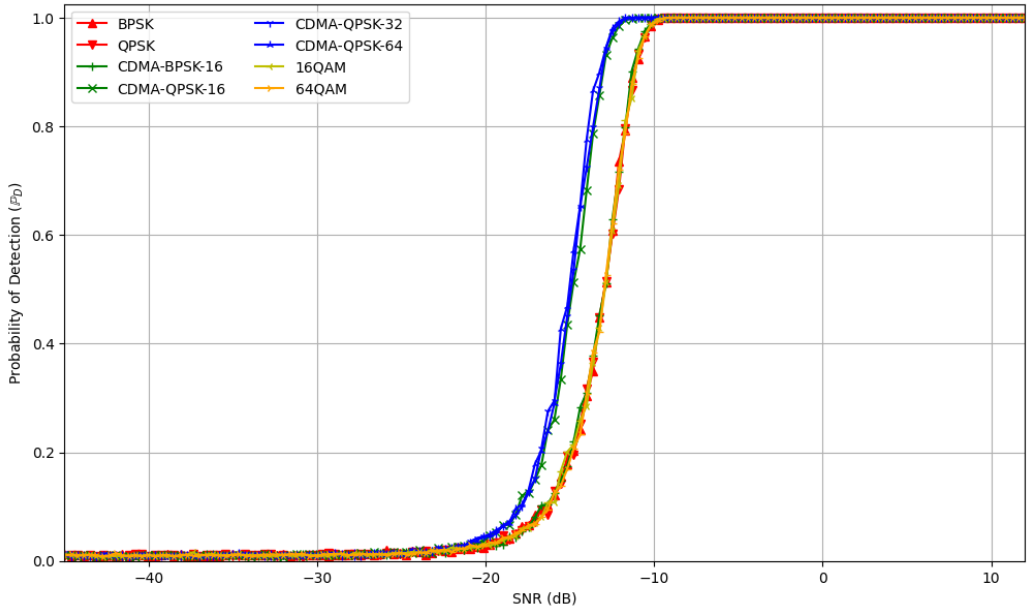


Figure 7.5: $\mathbb{P}_D$ versus SNR for the DCS detector with Group 1 with the FAM algorithm. $\mathbb{P}_{FA} = 0.01$.

As discussed in Section 7.1.2, when the same metric is run on the SCF estimate produced by the SSCA the detection power decreases, and the $\mathbb{P}_D$ curves shift to the right by about 1dB to 3dB. So using the SSCA with the DCS metric is almost strictly worse than the radiometer. As the FAM algorithm has a tuneable parameter to increase its resolution of the SCF, it may be that increasing the SCF resolution further would continue to increase the power of this detector. The results in Figs. 7.5–7.7 estimated 500 equally spaced cycle frequencies α. Doing more, however, would increase the computation time required to produce results, and so was not pursued further in this thesis.

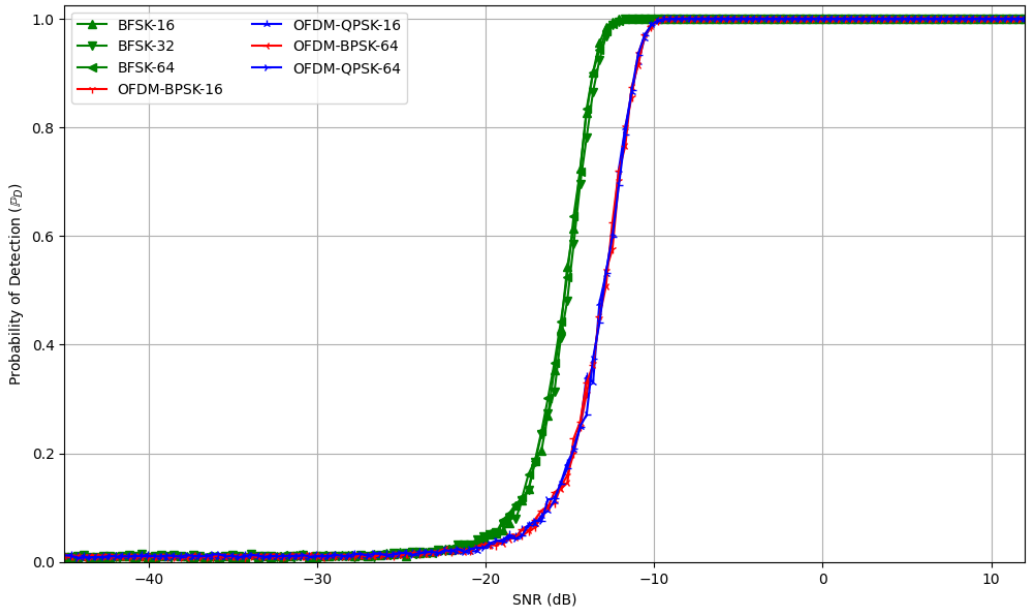


Figure 7.6: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the DCS detector with Group 2 with the FAM algorithm. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

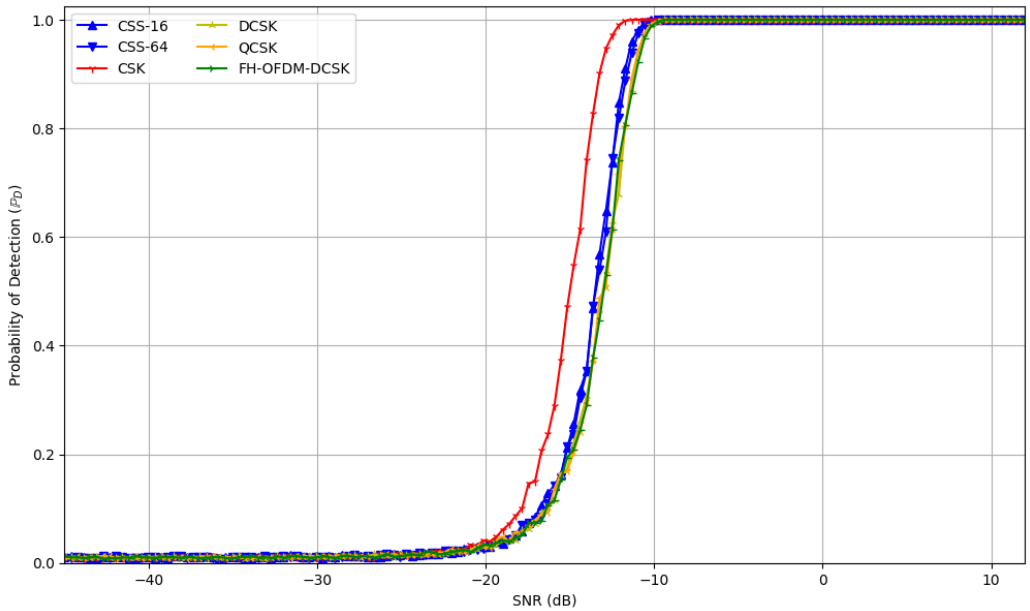


Figure 7.7: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the DCS detector with Group 3 with the FAM algorithm. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

Normal-Distribution Detector

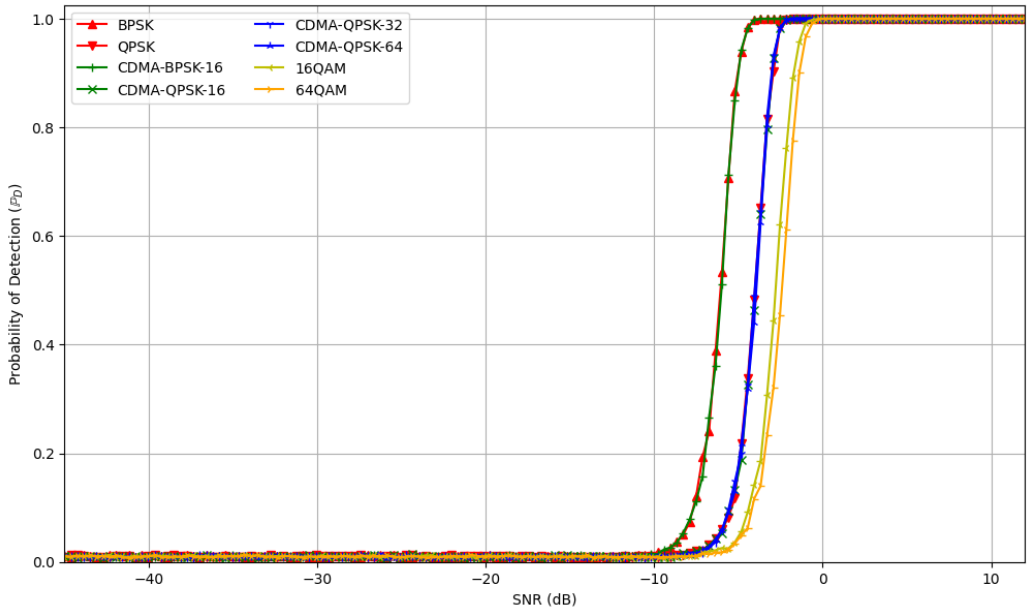


Figure 7.8: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the normal-distribution detector with Group 1. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

This detector had the worst performance of all tested. It could detect BPSK with $\mathbb{P}_{\mathrm{D}} = \frac{1}{2}$ when the SNR was -6dB, which was the best case of any transmission scheme (shared with DCSK), as seen in Figs. 7.8–7.10. The best case scenario for the normal-distribution detector was far worse than the worst case of any other detector by about 10dB.

Several schemes were almost undetectable with the normal-distribution detector—namely any transmission scheme that used OFDM, including FH-OFDM-DCSK. Increasing the number of subcarriers in the OFDM scheme increased covertness for Alice with this detector. This test evaluates whether the received signal deviates from a Gaussian distribution. As the modulation gets more complex, it will tend to look more Gaussian by the central limit theorem, and therefore be harder to detect using this method.

PSK, FSK, CDMA, and QAM were all detected with $\mathbb{P}_{\mathrm{D}} \approx 1$ by the time that the SNR was 0dB, so while the normal-distribution detector *does* function properly as a detector, it is not a very useful one.

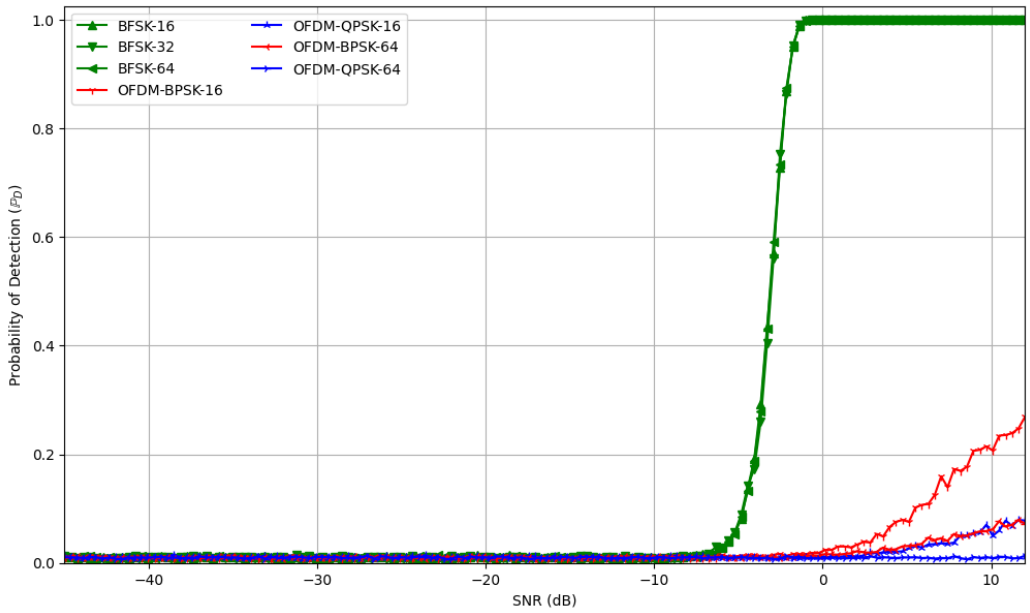


Figure 7.9: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the normal-distribution detector with Group 2. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

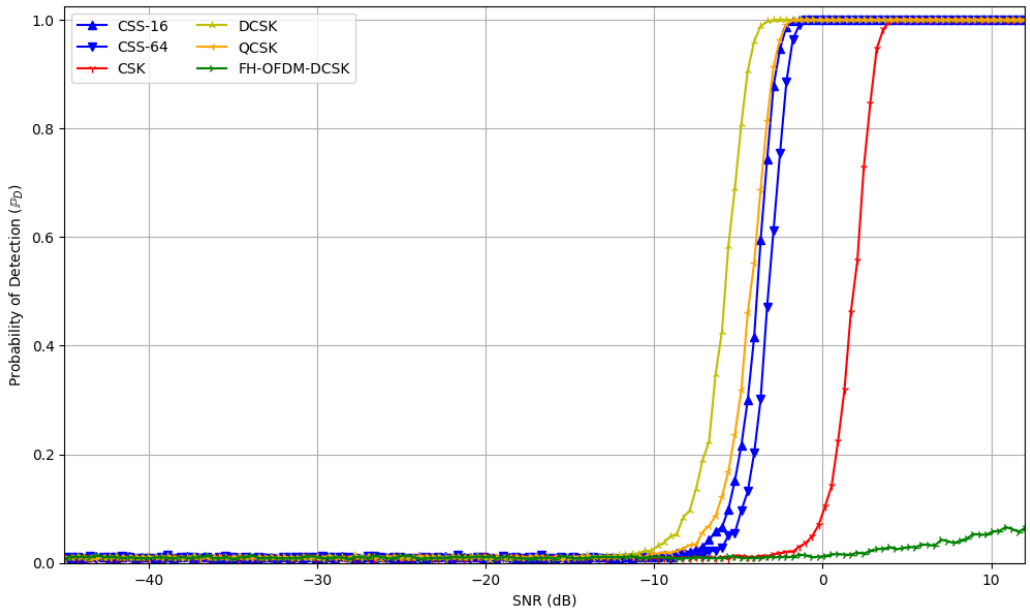


Figure 7.10: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the normal-distribution detector with Group 3. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

7.1.2 SSCA Versus FAM

Cyclostationarity detectors rely on calculating the spectral correlation function (SCF) of the received signal. To this end, I implemented both the strip spectral correlation algorithm (SSCA), described in Section 6.3, and the frequency accumulation method (FAM), described in Section 6.4. The SSCA has much lower computational complexity and is thus more efficient than the FAM. The SSCA is also a one-shot algorithm where the selected cycle frequencies are baked in to the algorithm, thus the spectral resolution is fixed. The FAM, however, has a tuneable resolution that makes it superior at truly estimating the SCF, as one can simply run the FAM in a "for-loop" over the desired cycle frequencies α. This work used 500 equally spaced cycle frequencies, whereas the SSCA was only estimating 64 cycle frequencies[1].

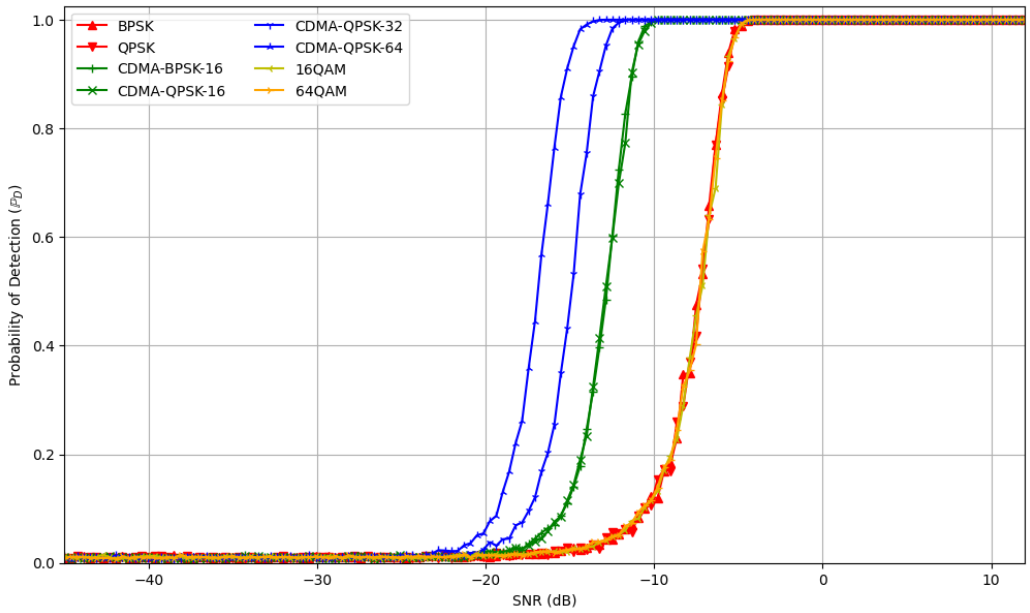


Figure 7.11: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the max cut detector with Group 1 with the FAM technique. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

Regarding the max cut detector, the SSCA had less variance overall in its ability to detect transmission schemes compared to the FAM. Figs. 7.11–7.13 show $\mathbb{P}_{\mathrm{D}}$ as a function of SNR for the max cut detector when the SCF is estimated using the FAM. Comparing these figures with the corresponding graphs in Fig. 7.2–7.4 (where the SSCA algorithm is used), shows the SSCA allowed the max cut detector to detect every tested modulation at an SNR of 5dB lower than with the FAM. The SSCA also detected CDMA at a lower SNR than with the FAM, as seen in Fig. 7.11. The FAM, was better at detecting CSK than the SSCA by 3dB, as in Fig. 7.13. On average, however, the SSCA had better performance than the FAM with the max cut detector.

[1]The SSCA algorithm does not benefit much from increasing the number of cycle frequencies beyond 64 [104].

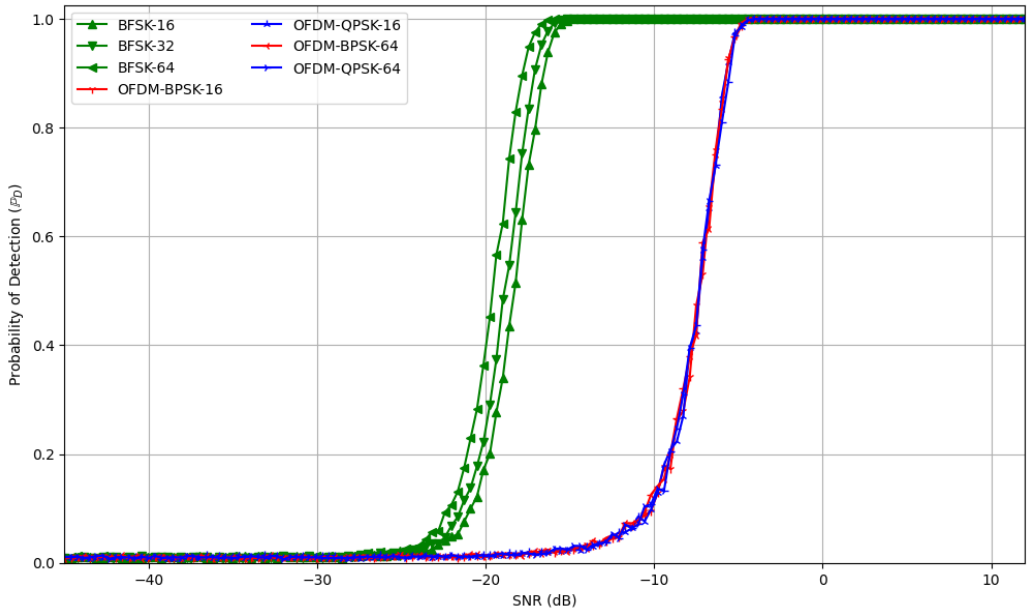


Figure 7.12: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the max cut detector with Group 2 with the FAM technique. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

For the DCS detector in Section 7.1.1, there was a clear bifurcation of the modulations into two branches. In that section, in Figs. 7.5–7.7, the DCS detector results used the FAM. When the DCS detector is fed by the SSCA instead, the detector performs worse, and more of the modulations cluster towards the lower branch (where $\mathbb{P}_{\mathrm{D}} = \frac{1}{2}$ when the SNR is -15dB). As a result, using the DCS detector with the SSCA was strictly worse than the radiometer in this model. When the DCS detector used the SCF estimate provided by the FAM, some of the modulations were detected at a lower SNR than with the radiometer, and some were detected at a higher SNR than the radiometer. The detectability for the DCS detector with the SSCA is plotted for select modulations in Figs. 7.20–7.25 in Section 7.2.1.

These results highlight the sensitivity that detectors have to the estimate of the SCF. Although the FAM took much longer to run than the SSCA, it had superior performance for the DCS detector, and for several modulations with the max cut detector. This highlights a tradeoff between the power of a detector and its computational costs; if Willie is computationally constrained he may consider using the SSCA, as it is much more efficient, and sometimes performs better than the FAM method.

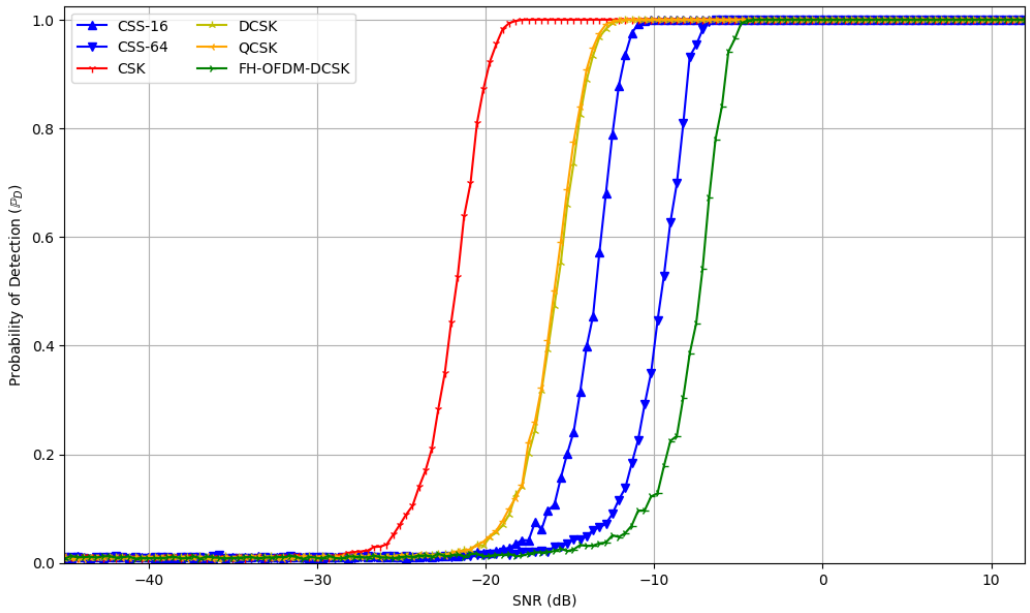


Figure 7.13: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the max cut detector with Group 3 with the FAM technique. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

7.1.3 Effect of the False Alarm Rate

This work uses the constant false alarm rate (CFAR) method (Section 6.2.2) to determine the detector threshold and to calculate $\mathbb{P}_{\mathrm{D}}$. The CFAR technique introduces a new free parameter—the false alarm rate, $\mathbb{P}_{\mathrm{FA}}$. Naturally, the subsequent question becomes determining which value of $\mathbb{P}_{\mathrm{FA}}$ is most suitable. When Willie chooses a higher alarm rate, the total number of detections increases, but a larger portion of these "detections" are actually false positives. As $\mathbb{P}_{\mathrm{FA}}$ decreases, $\mathbb{P}_{\mathrm{D}}$ also decreases, because the detector threshold is higher, leading to a reduction in both true detections and false alarms.

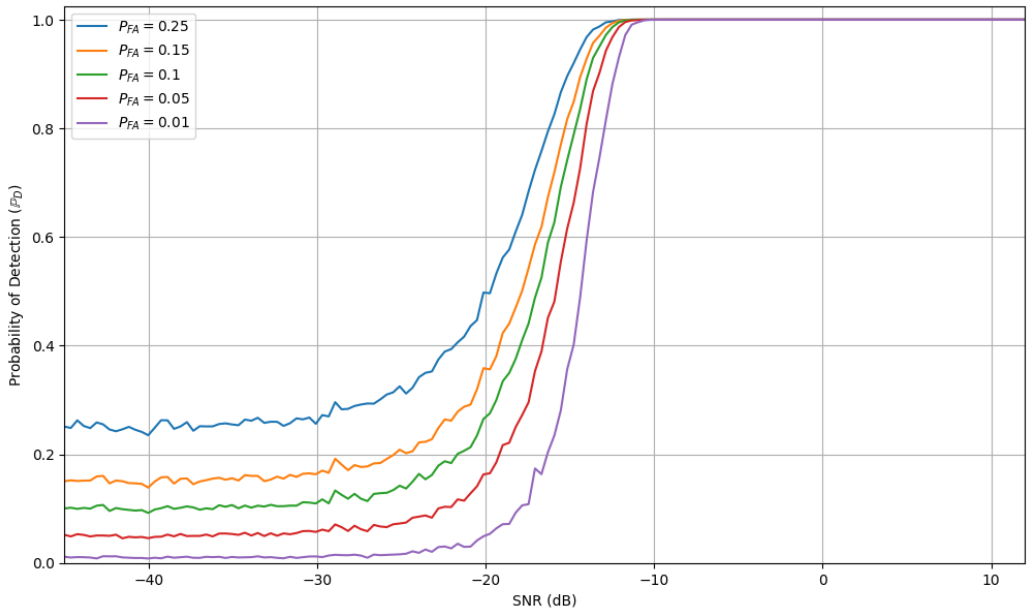


Figure 7.14: The probability of detection as a function of SNR for the radiometer with different false alarm rates. The values of $\mathbb{P}_{\mathrm{FA}}$ shown are 0.25, 0.15, 0.1, 0.05, and 0.01.

Fig. 7.14 shows the $\mathbb{P}_{\mathrm{D}}$ versus SNR plot for the radiometer with different values of $\mathbb{P}_{\mathrm{FA}}$ using CFAR. CDMA-QPSK with a 64-bit spreading sequence is depicted, but the signal agnostic nature of the radiometer detector means that the curve shown in Fig. 7.14 is the same for every modulation. When the SNR decreases, $\mathbb{P}_{\mathrm{D}}$ settles to $\mathbb{P}_{\mathrm{FA}}$. Also note that when the false alarm rate is lower, the detector requires a higher SNR to achieve $\mathbb{P}_{\mathrm{D}} = 1$, as a result of the higher threshold used. This highlights the importance that Willie choose an acceptably small $\mathbb{P}_{\mathrm{FA}}$ to increase his true positive rate (TPR).

7.1.4 Effect of TW Product

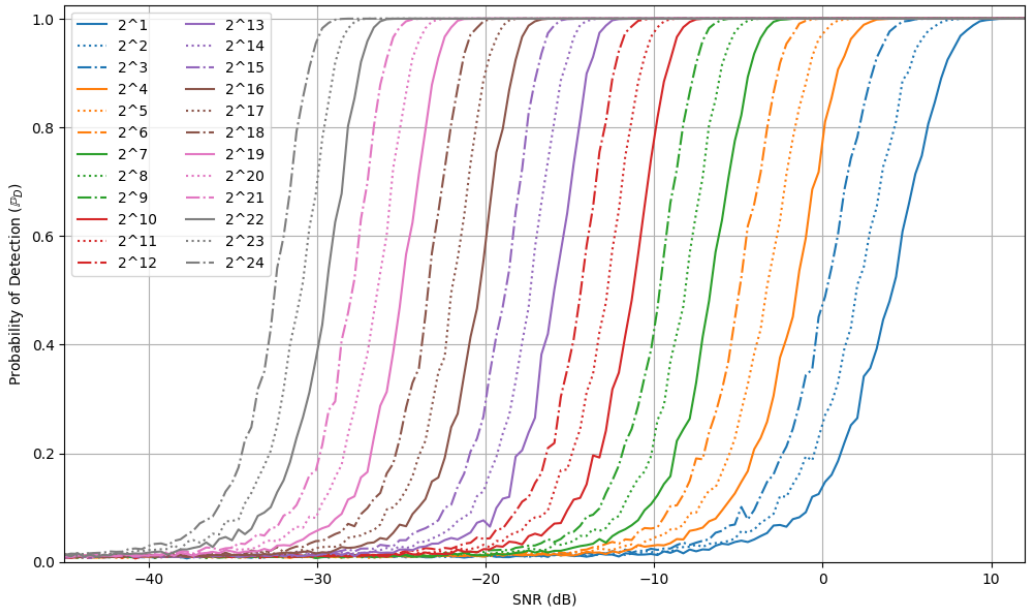


Figure 7.15: $\mathbb{P}_{\mathrm{D}}$ versus SNR for the radiometer with different TW products (made by changing the integration time). $\mathbb{P}_{\mathrm{FA}} = 0.01$. Doubling the TW product has the effect of increasing detectability by about 1.6dB.

Section 4.2 introduced the TW product as a factor in the performance of the radiometer. This is the product of two properties of Willie's detector: the bandwidth W and integration period T. Section 5.1.3 explains that if Willie increases his TW product while Alice holds hers constant, Willie incurs a performance loss due to integrating additional noise. Fig. 5.3 shows the consequences for the probability of detection in this case. A larger TW product for the warden also increases the odds of signals from other users, degrading his performance further.

This work assumes that Alice's transmission either occupies the whole bandwidth and integration time (H_1), or none of it (H_0), as discussed in Section 6.1. Increasing the TW products of Alice and Willie in lockstep also increases Willie's probability of detection, as he has more information upon which to base his decision. Fig. 7.15 illustrates that doubling the TW product[2] has the effect of increasing the "detectability" of the transmission by about 1.6dB. This can be seen in Fig. 7.15 as all points of the $\mathbb{P}_{\mathrm{D}}$ versus SNR curve move -1.6dB (or leftwards). This is expected as per (4.12). In all other sections, the sampling time consists of $2^6 + 2^{12} = 4160$ complex baseband samples between all detectors in order to keep the TW product constant and to make comparisons between detectors fair.

While the TW *product* is the only thing that matters for assessing $\mathbb{P}_{\mathrm{D}}$, cyclostationarity detectors are not dependent on the TW product overall, but are rather sensitive to the particular values of T and W themselves. Increasing the TW product increases the proba-

[2]Doubling the TW product is done in the simulation by doubling the integration time.

bility of detection (again, assuming that Alice's signal still occupies the entire bandwidth W and observation period T). While this is simple to do for the radiometer, it introduces significant computational complexity for cyclostationarity detectors, as larger and larger FFTs are required to estimate the SCF.

7.1.5 PDFs of the H_0 & H_1 Cases

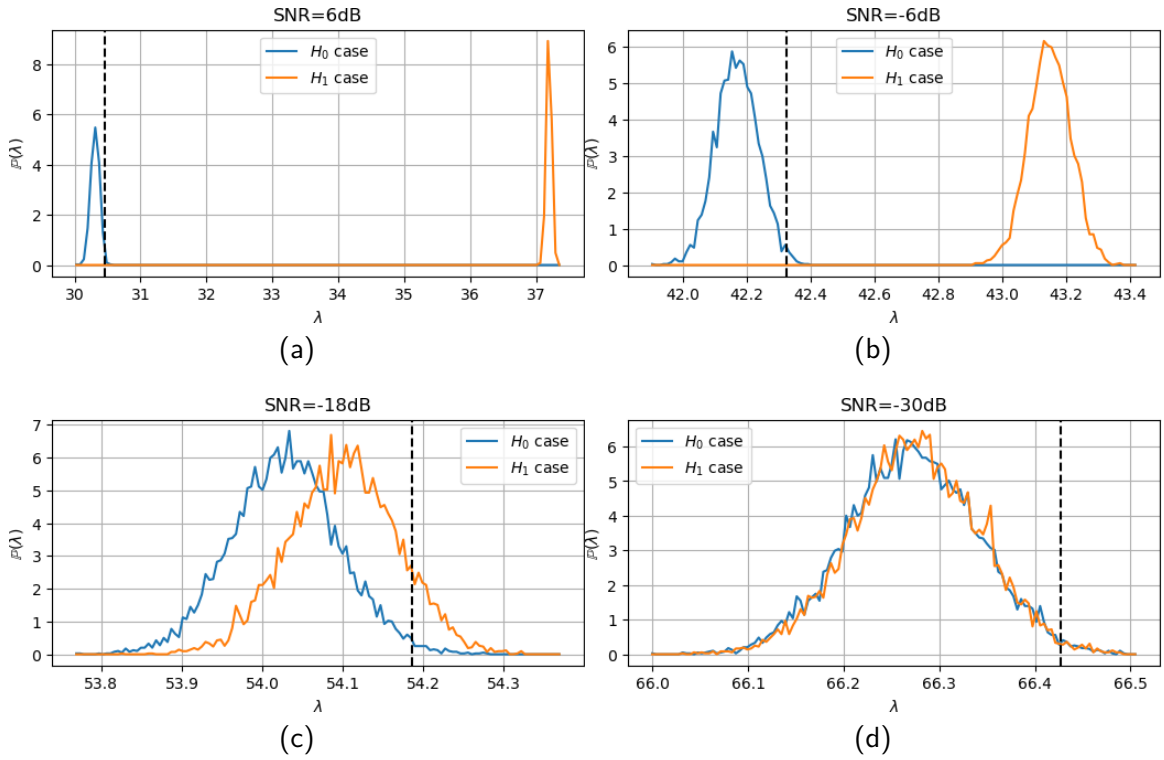


Figure 7.16: The PDFs of the detectors outputs (λ) for the H_0 and H_1 cases for BPSK under the radiometer. The SNRs are 6dB (7.16a), −6dB (7.16b), −18dB (7.16c), and −30dB (7.16d). $\mathbb{P}_{\text{FA}} = 0.01$. The vertical dashed line indicates the detector threshold λ_0 that was calculated using the CFAR technique for $\mathbb{P}_{\text{FA}} = 0.01$.

Fig. 7.16 shows the PDFs of the H_0 and H_1 cases for radiometer at different SNRs. When the SNR is sufficiently high, the detector has no problem differentiating H_0 and H_1. This is evident in Fig. 7.16a and Fig. 7.16b, where the probability of detection is $\mathbb{P}_{\text{D}} \approx 1$. As the SNR decreases, the two PDFs begin to overlap. Alice has only a 10% chance of being detected (i.e., $\mathbb{P}_{\text{D}} \approx 0.1$) when the SNR is −18dB, as in Fig. 7.16c. When the SNR drops sufficiently, as in Fig. 7.16d, the PDFs overlap, and the transmission scheme becomes undetectable, leading to $\mathbb{P}_{\text{D}} = \mathbb{P}_{\text{FA}}$.

7.1.6 Threshold λ_0 Versus SNR

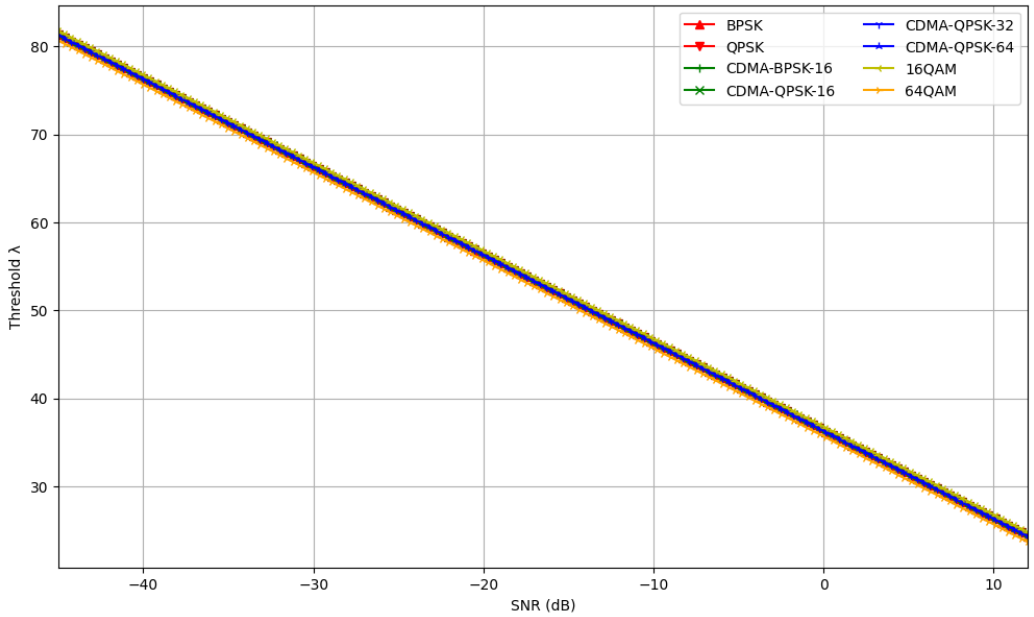


Figure 7.17: The calculated best threshold, λ_0, as a function of SNR for the radiometer with Group 1. $\mathbb{P}_{\text{FA}} = 0.01$.

The optimal detector threshold changes with channel conditions. In both Figs. 7.17–7.18, the optimal threshold λ_0 decreases as the SNR increases. This makes sense for a fixed integration time, as the signal easily cuts through the noise at higher SNRs and is readily detectable. The line in Fig. 7.17 matches up with the theoretical curve for the radiometer with this TW product and $\mathbb{P}_{\text{FA}}$ [105].

The results for the max cut detector in Fig. 7.18 mirror the results of Fig. 7.2, discussed in Section 7.1.1. Modulations in groups 2 & 3 are similar to those depicted here already in this section, and are omitted for the sake of brevity,

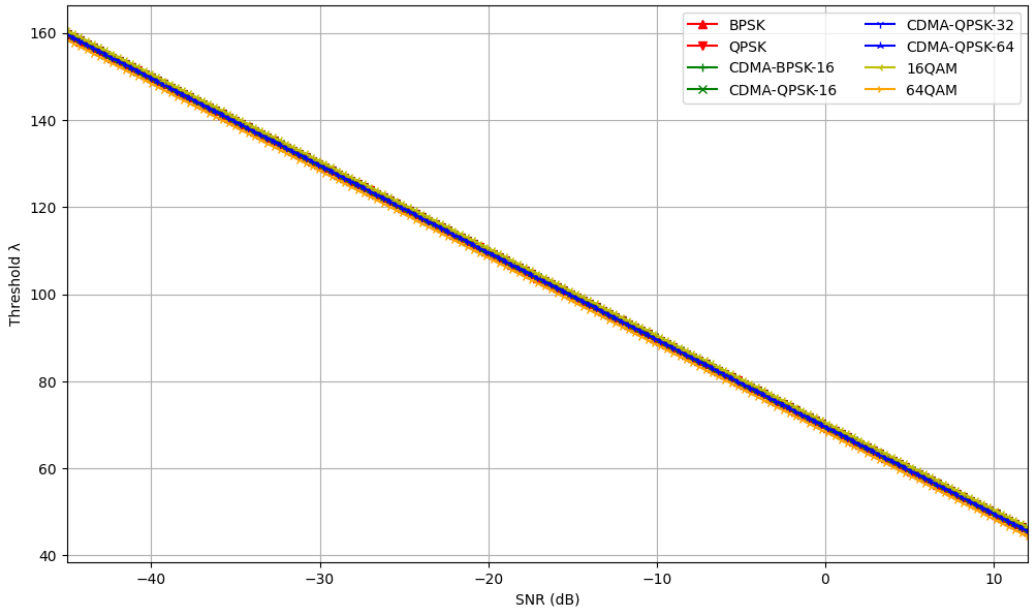


Figure 7.18: The calculated best threshold, λ_0, as a function of SNR for the max cut detector with Group 1. $\mathbb{P}_{\text{FA}} = 0.01$.

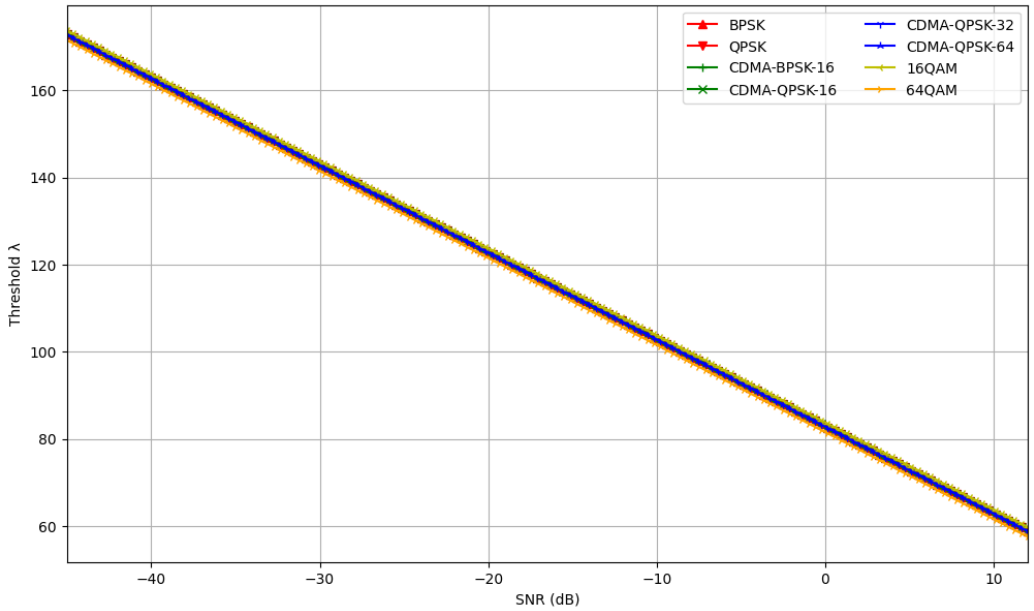


Figure 7.19: The calculated best threshold, λ_0, as a function of SNR for the DCS detector with Group 1. $\mathbb{P}_{\text{FA}} = 0.01$.

7.2 Transmission Scheme Performance

This section details which transmission schemes are best for Alice and Bob to employ, by taking into account both the detectability *and* the BER of the protocols. A list of all the transmission schemes that were tested is in Section 6.1.2, and they are plotted in the groups listed at the start of Chapter 7.

7.2.1 Probability of Detection and BER

Plots like those in Figs. 7.20–7.25 can help to visualize the performance of a transmission scheme in terms of both BER and the detectability across multiple detectors. The abscissa for these plots is the SNR in dB, and there are two different ordinates overlaid on the same plot. The BER is shown in red on the left-hand side, whilst the $\mathbb{P}_{\mathrm{D}}$, as measured by CFAR method, is in blue on the right-hand side, with different line styles for the different detectors. The BER always starts at 0.5 when the SNR is low[3], then approaches zero as the SNR increases. The probability of detection, $\mathbb{P}_{\mathrm{D}}$, on the other hand, settles to $\mathbb{P}_{\mathrm{FA}}$ for low SNRs where the signal is buried in the noise before approaching one (for a functioning detector) as the signal energy increases.

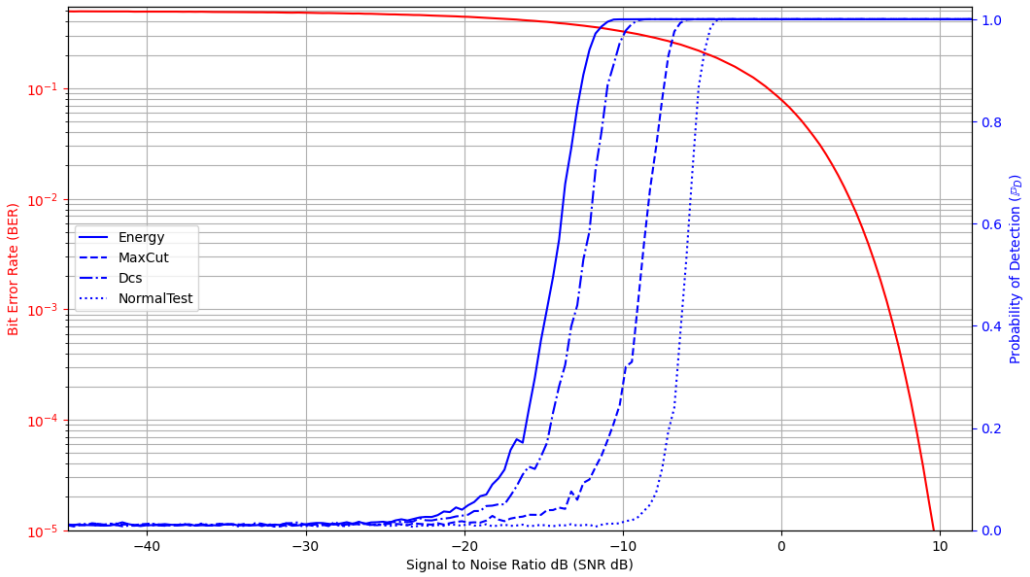


Figure 7.20: Plot of BER (red, left-hand side), and $\mathbb{P}_{\mathrm{D}}$ (blue lines, right-hand side) for BPSK. The SSCA was used to generate the SCF for the cyclostationarity detectors, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

The plots in Figs. 7.20–7.25 allow for visually determining if Alice can achieve her desired BER before Willie reaches a given $\mathbb{P}_{\mathrm{D}}$ for a given transmission scheme. This allows Alice to accurately assess the overall covertness of her situation and the probability of her being

[3]The BER can never be higher than 0.5 (i.e., a random coin toss). If the BER is calculated to be 1, then the *true* BER is zero, as one needs to merely swap the labels on the "0" and "1" bits.

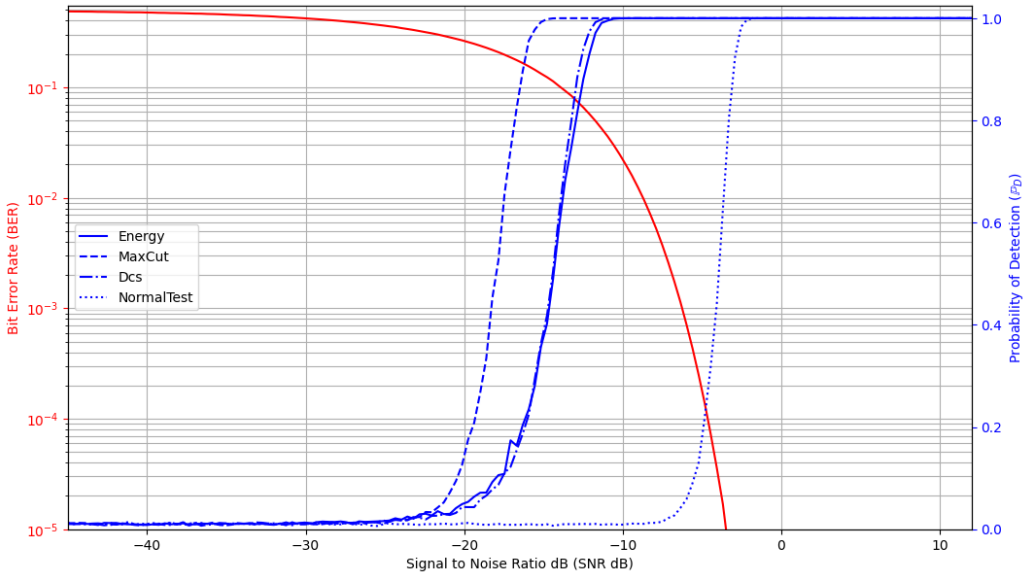


Figure 7.21: Plot of BER (red, left-hand side), and $\mathbb{P}_{\mathrm{D}}$ (blue lines, right-hand side) for CDMA-QPSK with a 64-bit chip rate. The SSCA was used to generate the SCF for the cyclostationarity detectors, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

detected, as well as the expected error rate. BPSK, depicted in Fig. 7.20 is harder to detect than CDMA-QPSK, depicted in Fig. 7.21. The $\mathbb{P}_{\mathrm{D}}$ curves for BPSK in Fig. 7.20 require higher SNRs to reach $\mathbb{P}_{\mathrm{D}} \approx 1$ than for CDMA-QPSK in Fig. 7.21. However, by examining the BER, we see that the CDMA-QPSK has a lower BER than BPSK in these regions, indicating that Alice can transmit more bits reliably to Bob. If Alice knows that both Bob and Willie have an SNR of at most -20dB, then she may choose CDMA-QPSK to transmit information.

All the plots in Figs. 7.20–7.25 have the radiometer line in the same spot, which serves as a visual benchmark to see the relative shifts in performance of the other detector types. The results of Section 7.1 are reaffirmed here; Willie can use the radiometer and DCS detector to reliably detect Alice with $\mathbb{P}_{\mathrm{D}} \approx 1$ when the SNR is -10dB. The normal-distribution detector is consistently the worst performing of all; in Fig. 7.25, we see that FH-OFDM-DCSK was never reliably detected by it, even up to $+12$dB. The cyclostationarity detectors usually perform slightly better or worse than the radiometer.

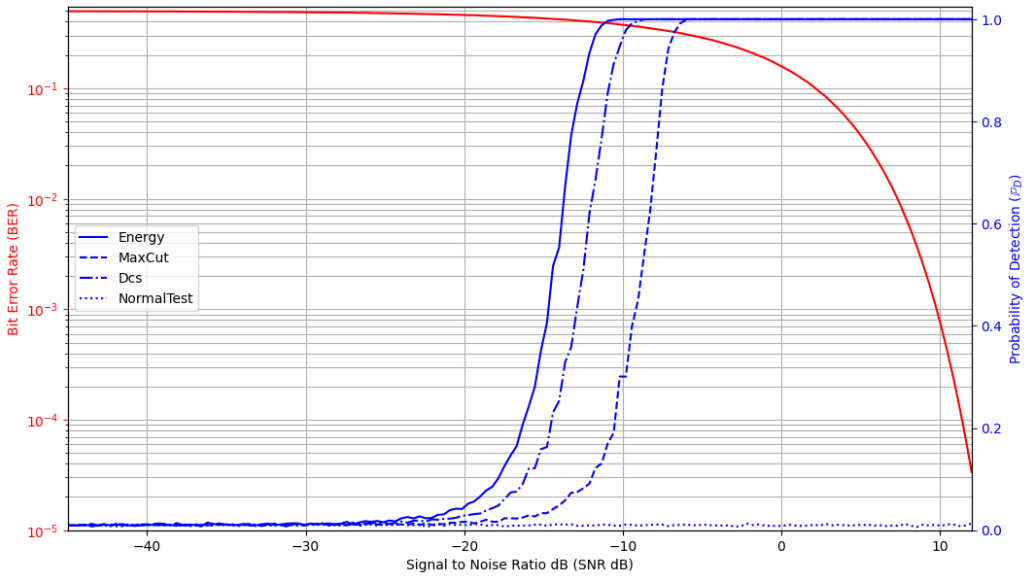


Figure 7.22: Plot of BER (red, left-hand side), and $\mathbb{P}_{\mathrm{D}}$ (blue lines, right-hand side) for OFDM-QPSK with 64 subcarriers. The SSCA was used to generate the SCF for the cyclostationarity detectors, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

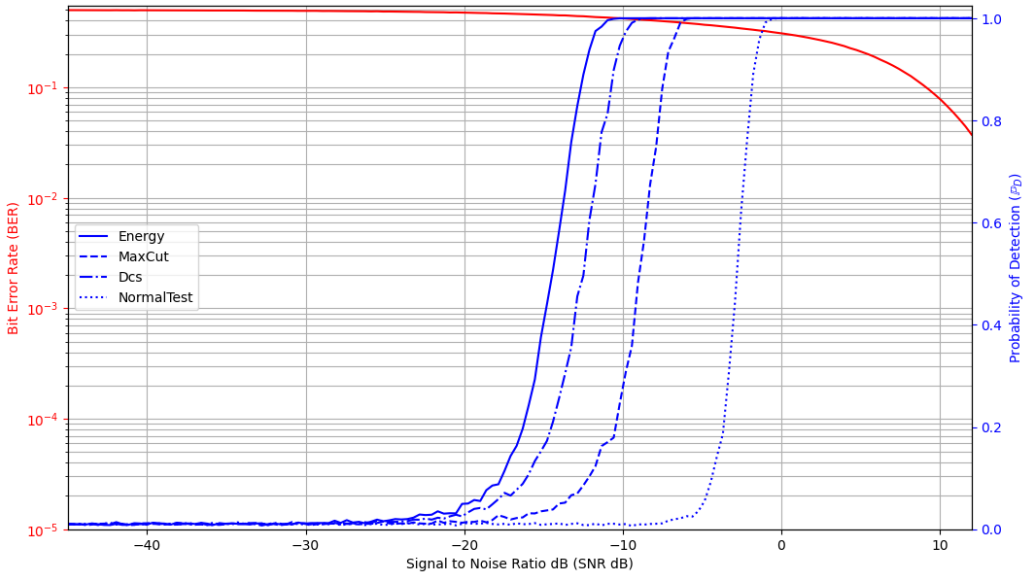


Figure 7.23: Plot of BER (red, left-hand side), and $\mathbb{P}_{\mathrm{D}}$ (blue lines, right-hand side) for 16-QAM. The SSCA was used to generate the SCF for the cyclostationarity detectors, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

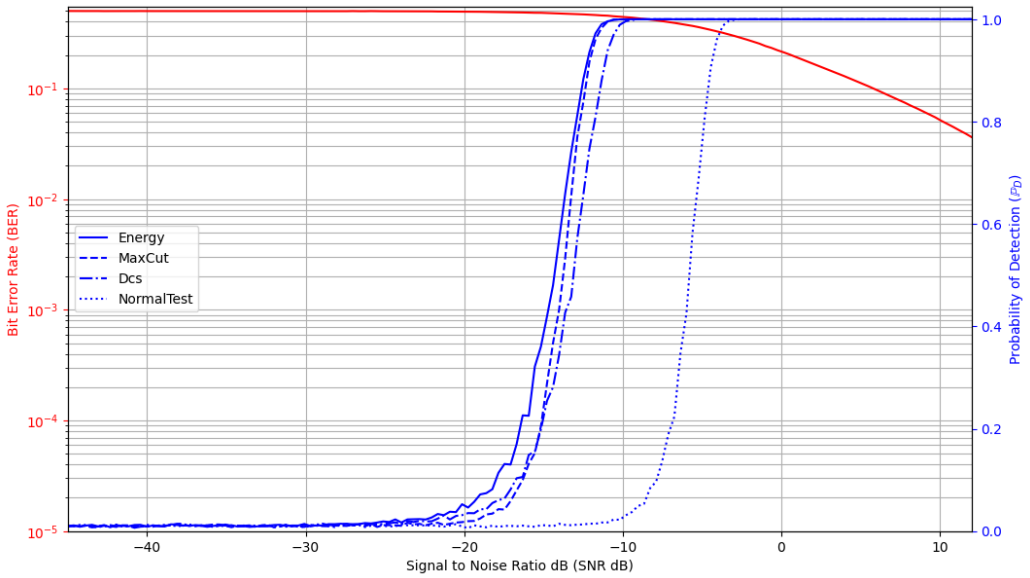


Figure 7.24: Plot of BER (red, left-hand side), and $\mathbb{P}_{\mathrm{D}}$ (blue lines, right-hand side) for DCSK. The SSCA was used to generate the SCF for the cyclostationarity detectors, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

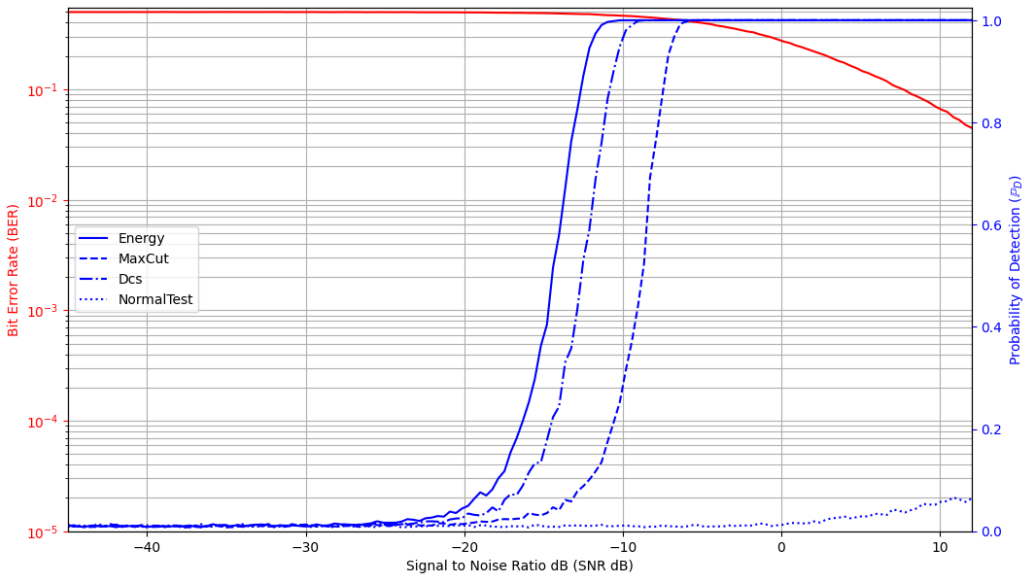


Figure 7.25: Plot of BER (red, left-hand side), and $\mathbb{P}_{\mathrm{D}}$ (blue lines, right-hand side) for FH-OFDM-DCSK. The SSCA was used to generate the SCF for the cyclostationarity detectors, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

7.2.2 $\mathbb{P}_{\mathrm{D}}$ Versus BER ROC Plot

If one knows the SNR of both the Alice–Bob channel and the Alice–Willie channel, one can examine the tradeoff between reliability and deniability by generating ROC plots that compare the BER with the probability of detection. Each point on the ROC plot corresponds to the BER and probability of detection at a specific SNR. A point near the origin at $(0, 0)$ indicates that the transmission scheme has no errors and is undetectable at some SNR. A point near $(1, 1)$ in this curve indicates the transmission scheme is completely unreliable and always detectable at a given SNR for the specified detector and modulation.

Plotting these ROC curves requires setting the gain difference between Bob and Willie, and in Figs. 7.26–7.31 the SNRs for Bob and Willie are set to be equal (i.e., the gain difference is zero). This is not a realistic assumption, as it is likely in practice that Bob and Willie have different SNRs. A plot of all the BERs of all the modulations can be found in Appendix A.2.

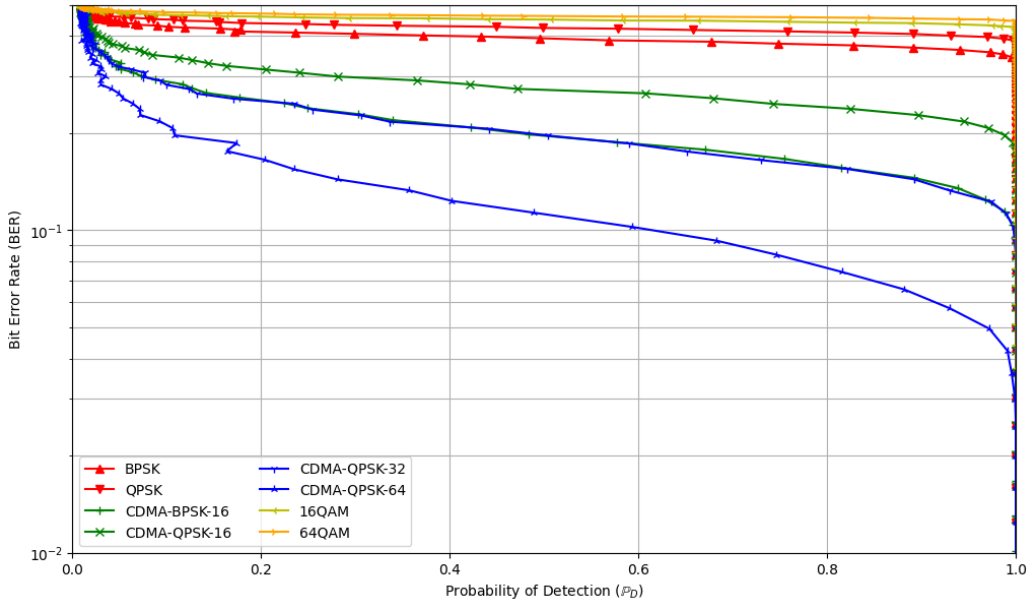


Figure 7.26: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the radiometer with Group 1 when Bob and Willie have the same SNR. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

Recurring patterns emerge when accounting for the BER/$\mathbb{P}_{\mathrm{D}}$ tradeoff in Figs. 7.26–7.34. CDMA appears as the winner across multiple detectors; i.e., CDMA curves always come closest to the origin point, $(0, 0)$ (where both the BER and probability of detection are zero), compared to other detectors. Specifically, CDMA-QPSK with a 64-bit spreading sequence has the best performance for the radiometer and DCS detector, as seen in Fig. 7.26 and Fig. 7.32. However, Fig. 7.29 shows that CDMA-BPSK with a 16 bit spreading sequence was the best modulation under max cut detector. Increasing the size of the spreading sequence increased covertness.

Figs. 7.26–7.34 all display a somewhat consistent ordering between classes of transmission schemes. CDMA has the best BER/$\mathbb{P}_{\mathrm{D}}$ tradeoff, followed by PSK and OFDM. BPSK

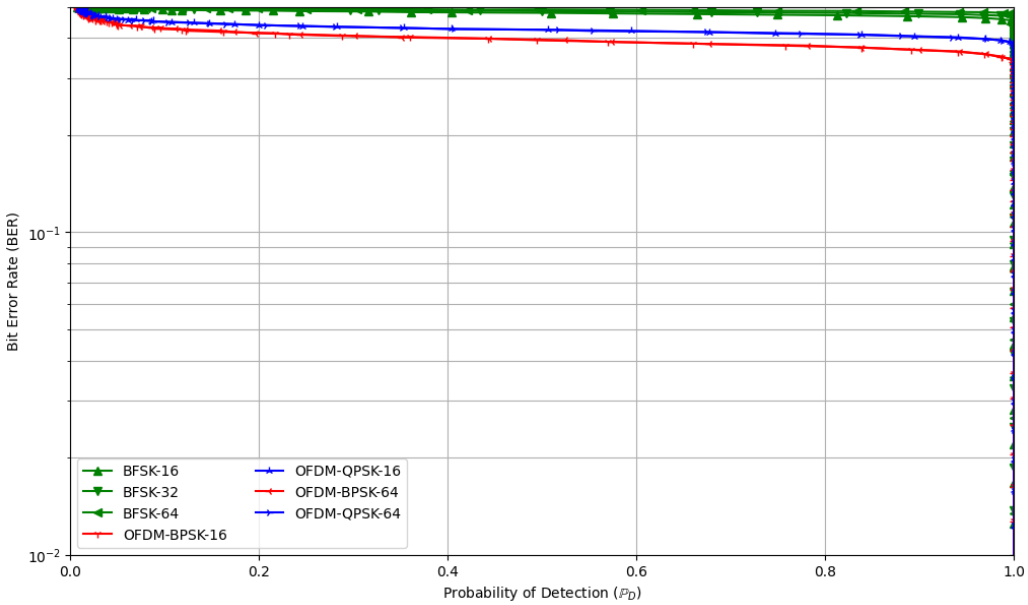


Figure 7.27: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the radiometer with Group 2 when Bob and Willie have the same SNR. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

and OFDM-BPSK have overlapping curves in all three plots. QPSK and OFDM-QPSK have worse performance than BPSK and OFDM-BPSK. After PSK and OFDM, is QAM. The higher the order of the QAM modulation, the worse the performance.

The remaining modulations (FSK, CSS, and the chaotic modulations) all performed the worst throughout Figs. 7.26–7.34. They each were detected with $\mathbb{P}_{\mathrm{D}} \approx 1$ before the BER had a chance to drop below 0.4.

Figs. 7.35–7.37 highlight the low effectiveness of the normal-distribution detector first discussed in Section 7.1.1. Accounting for BER does not change this fact. Figs. 7.35–7.37 show that OFDM is basically undetectable for the normal-distribution detector across a wide range of SNRs, shown by the lines of the curve for all OFDM modulations being close to the origin, $(0, 0)$, indicating there are SNRs where Alice and Bob have a positive covert capacity.

In Fig. 7.37, we see that the line for FH-OFDM-DCSK does not make contact with any edge of the graph. It is undetectable in the given SNR range, as the line does not come close to $\mathbb{P}_{\mathrm{D}} = 1$, but is also an unreliable transmission scheme, as the BER did not even reach 10^{-2} in the simulation SNR range.

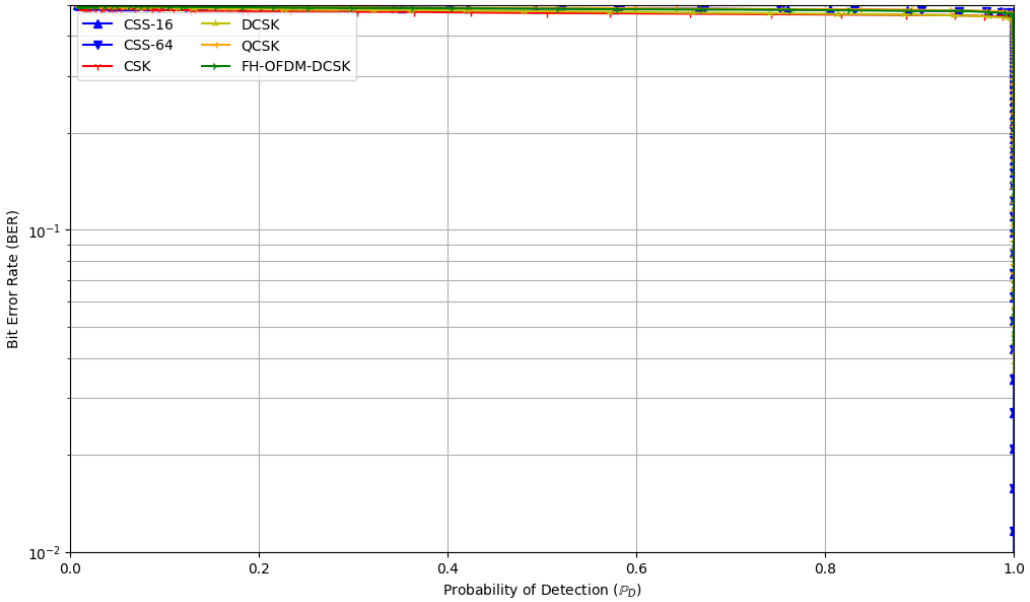


Figure 7.28: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the radiometer with Group 3 when Bob and Willie have the same SNR. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

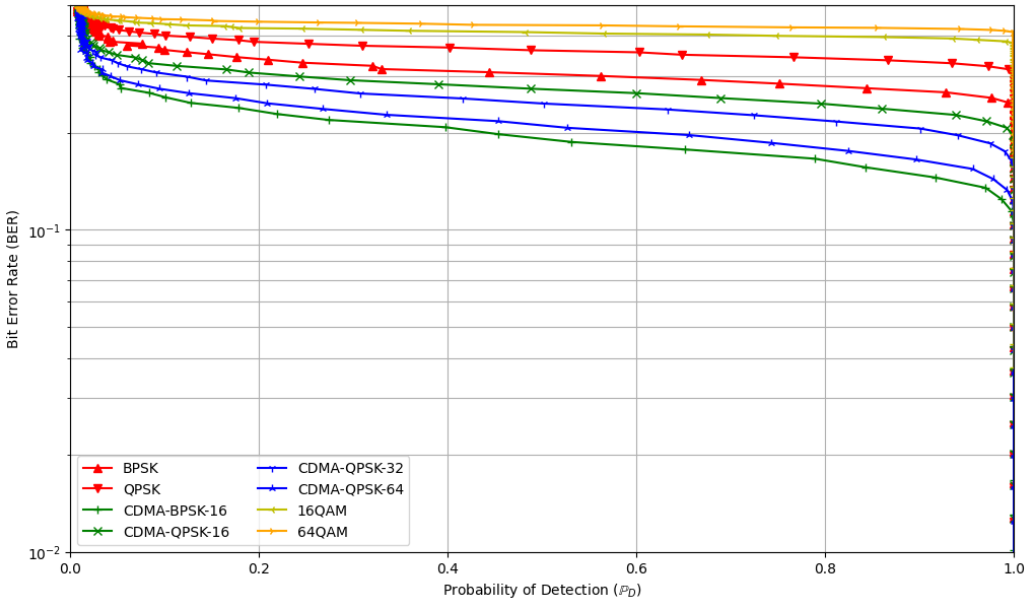


Figure 7.29: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the max cut detector with Group 1 when Bob and Willie have the same SNR. The SSCA was used to estimate the SCF, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

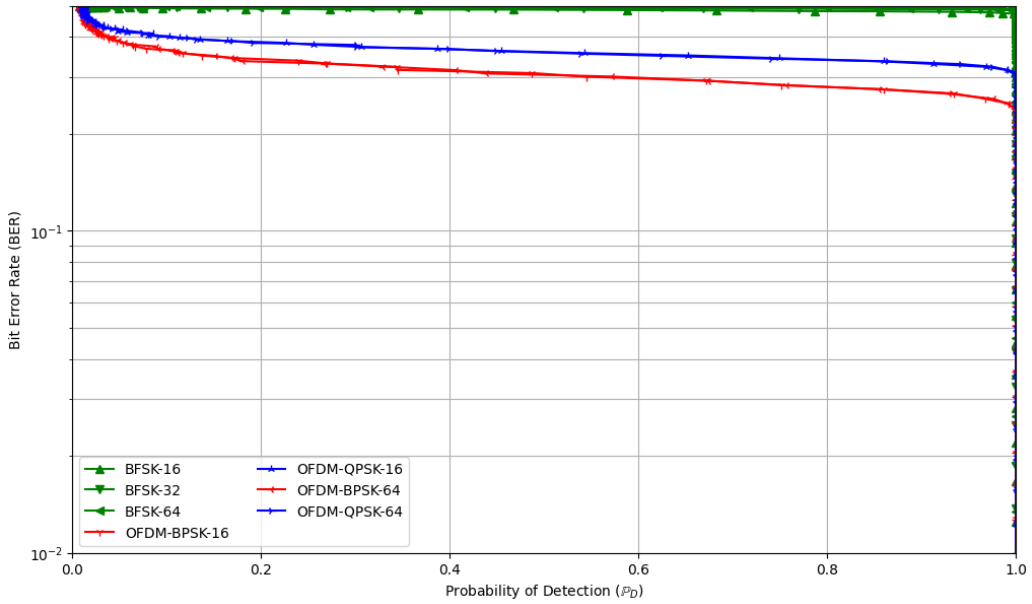


Figure 7.30: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the max cut detector with Group 2 when Bob and Willie have the same SNR. The SSCA was used to estimate the SCF, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

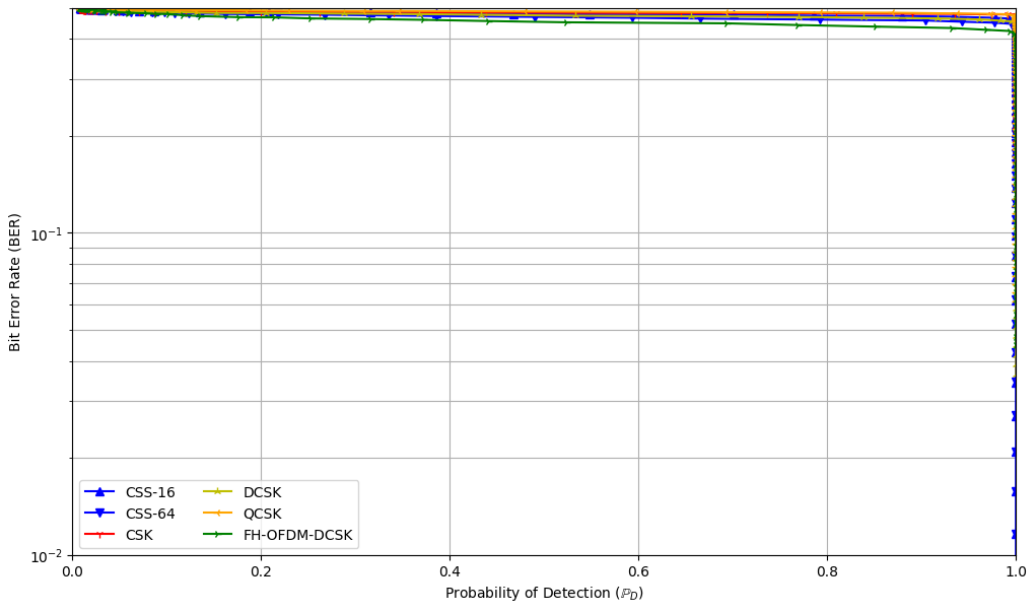


Figure 7.31: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the max cut detector with Group 3 when Bob and Willie have the same SNR. The SSCA was used to estimate the SCF, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

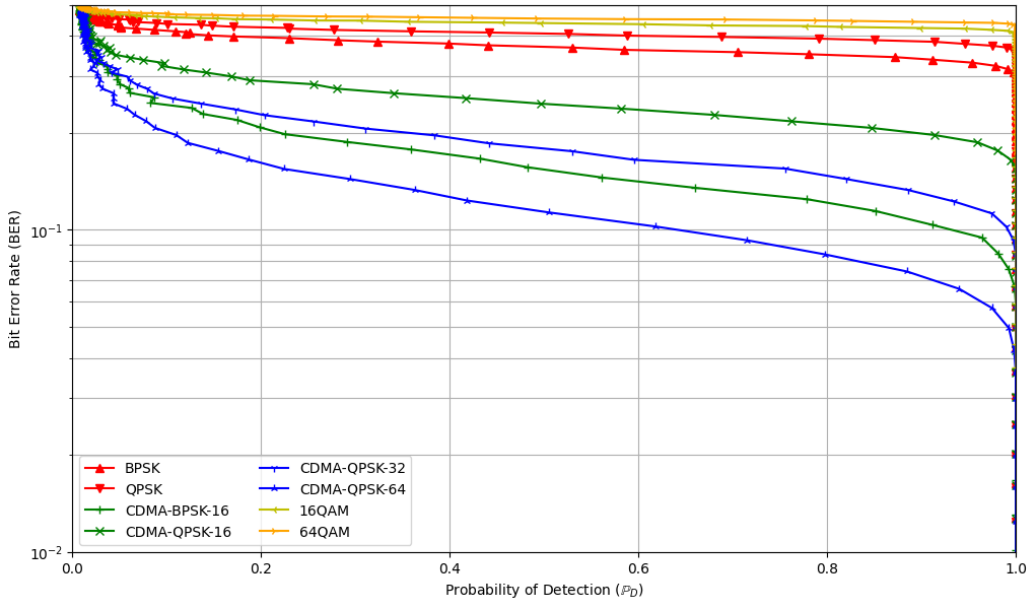


Figure 7.32: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the DCS detector with Group 1 when Bob and Willie have the same SNR. The SSCA was used to estimate the SCF, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

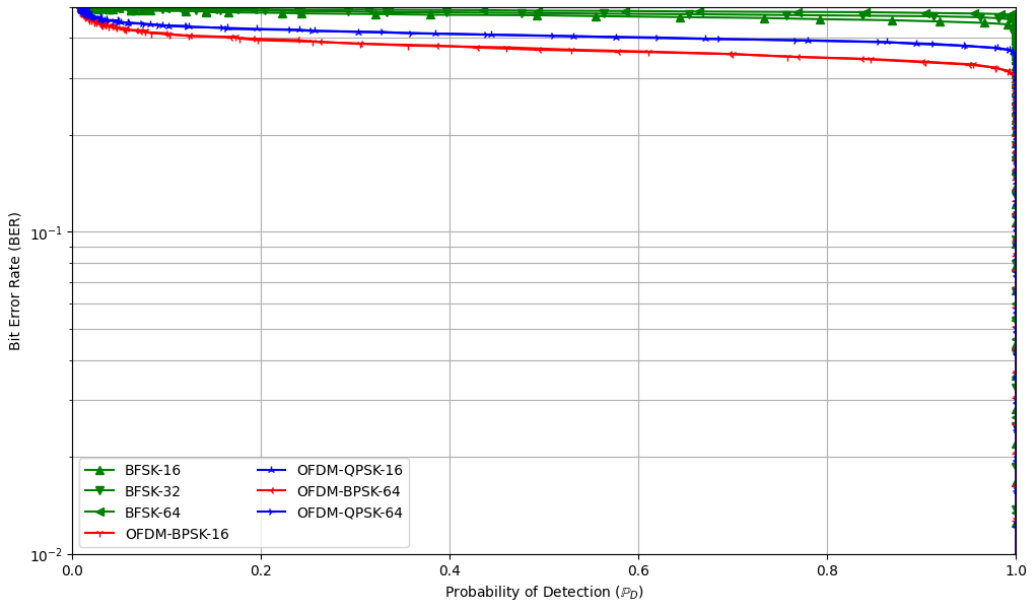


Figure 7.33: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the DCS detector with Group 2 when Bob and Willie have the same SNR. The SSCA was used to estimate the SCF, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

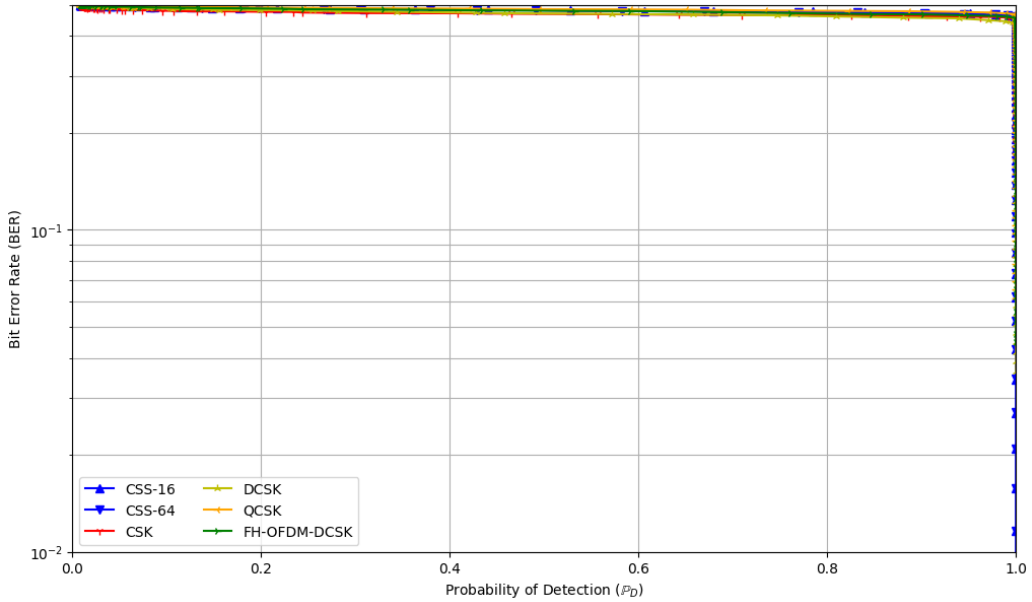


Figure 7.34: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the DCS detector with Group 3 when Bob and Willie have the same SNR. The SSCA was used to estimate the SCF, and $\mathbb{P}_{\mathrm{FA}} = 0.01$.

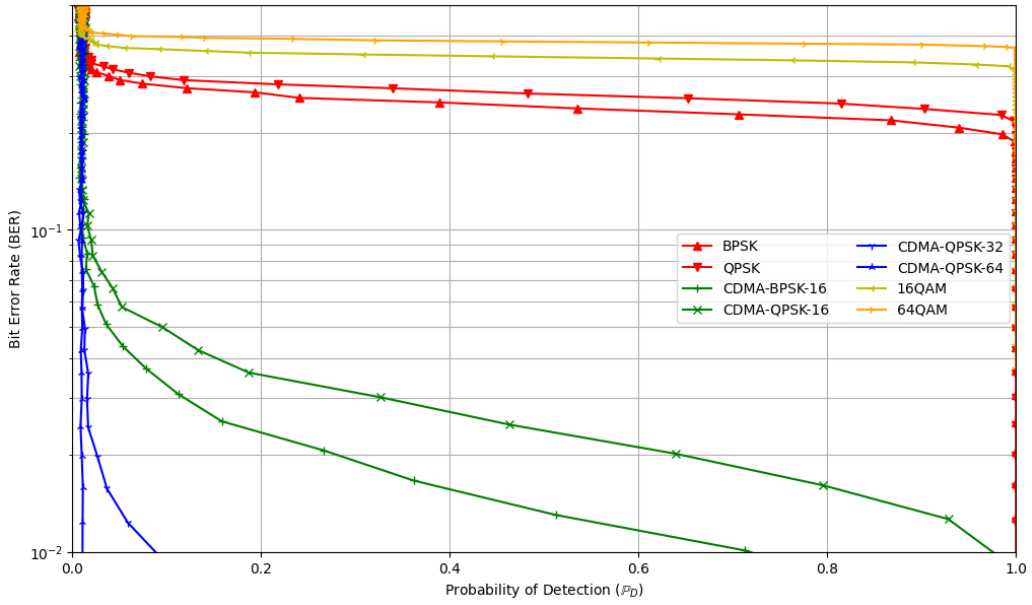


Figure 7.35: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the normal-distribution detector with Group 1 when Bob and Willie have the same SNR. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

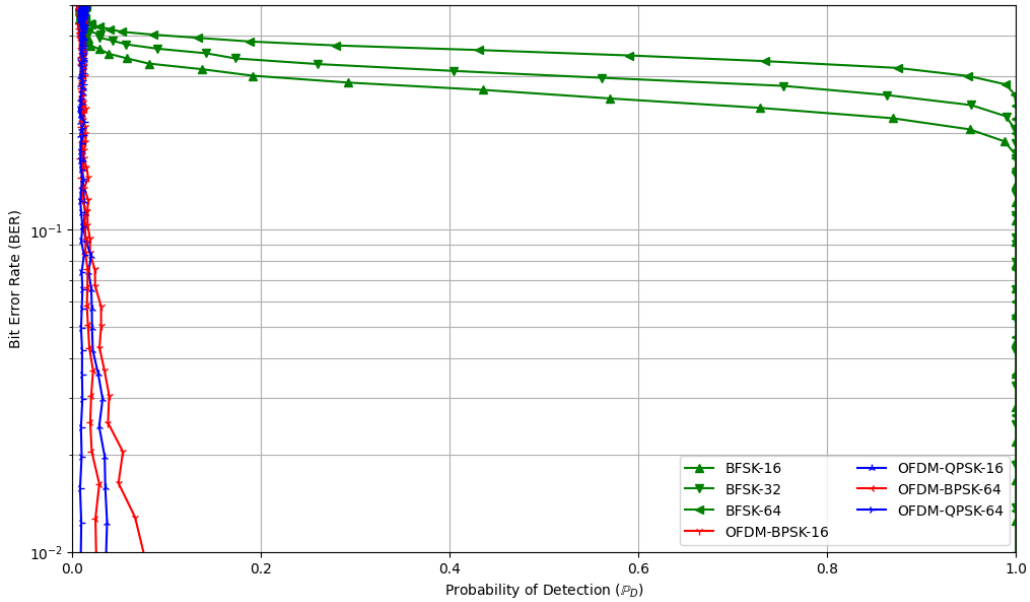


Figure 7.36: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the normal-distribution detector with Group 2 when Bob and Willie have the same SNR. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

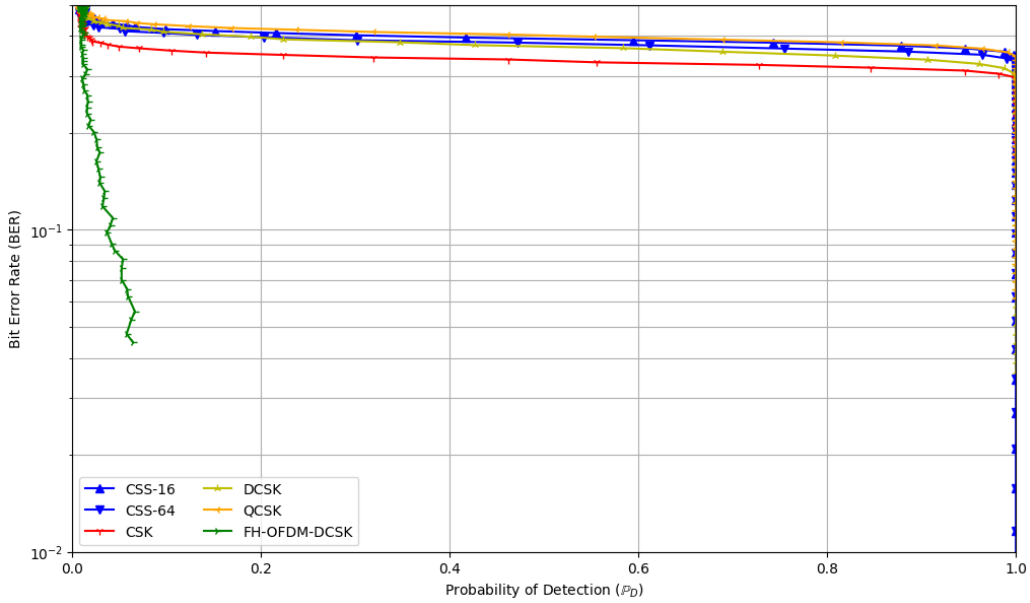


Figure 7.37: The ROC plot of $\mathbb{P}_{\mathrm{D}}$ versus BER for the normal-distribution detector with Group 3 when Bob and Willie have the same SNR. $\mathbb{P}_{\mathrm{FA}} = 0.01$.

7.3 Discussion

The results in Sections 7.1 and 7.2 demonstrate that not all detectors and transmission schemes are created equal. While we can draw conclusions about which detectors Willie ought to employ and which transmission schemes Alice and Bob should use, it is important to keep in mind the scope and limitations of the model used in this work, and how the implications of these results mesh with real-world transmission scenarios. The results of this investigation of covert communications are summarized here, and the implications for engineering practical systems are discussed.

7.3.1 Comparing Detector Performance

The radiometer, max cut, and DCS detectors each had good performance, and the normal-distribution detector was significantly worse than the others. The radiometer had the most consistent performance of any detector—the $\mathbb{P}_{\mathrm{D}}$ curves for all modulations are identical for the radiometer, regardless of any details of the transmission scheme. The max cut and DCS detectors had greater variance in their performance; they detected some modulations at a lower SNR than the radiometer, but were able to reliably detect other modulations at higher SNRs than the radiometer.

In this work Willie selects his detector threshold, λ_0, using the CFAR approach, which requires him to accurately know his noise variance, N_0, If Willie is not magically provided with the value of N_0, then he must estimate it himself to figure out which detector threshold should be used. Any error in Willie's estimate of the noise variance decreases detector performance [128, 129], and may not affect each detector equally. It may be that cyclostationarity detectors shine compared to the radiometer when this is taken into account. Choosing the best threshold based on a channel estimate is a difficult problem that inevitably affects detector design.

Overall, this work confirms that both cyclostationarity detectors (max cut, DCS) and energy detectors (the radiometer) are viable options for effectively detecting unknown transmissions in AWGN, while the normal-distribution detector should be avoided. The optimal detector depends on the signal that is being detected, but Willie can expect stable detection performance for *every* transmission scheme Alice might use when he deploys a radiometer.

7.3.2 Comparing Transmission Scheme Covertness

The transmission schemes in this work were assessed by using two primary properties: deniability (covertness), and reliability (BER). A total of 21 transmission schemes were tested, which varied wildly in terms of both these properties. The modulation technique that consistently increased covertness was DSSS, which was represented in this simulation by CDMA whose spreading sequences are Hadamard codes. This confirms DSSS as a sound method for generating covert signals—at the cost of a reduced data rate.

It is important to consider *both* the detectability *and* the BER in the covertness problem. If a modulation performs well by covertness metrics (i.e., it is hard for Willie to detect), but has an unacceptably high BER, then the messages are not readable by the intended recipient. FH-OFDM-DCSK is an example of this—Fig. 7.25 and Fig. 7.10 show that the normal-distribution detector was unable to reliably detect FH-OFDM-DCSK, even at an SNR of

+12dB. This was the best covertness performance of every combination of transmission scheme and detector tested in this work. The BER of FH-OFDM-DCSK, however, was one of the worst overall, which massively reduces its appeal.

The consequences of the tradeoff between the probability of detection, $\mathbb{P}_{\mathrm{D}}$, and the BER are perhaps best illustrated by the ROC plots shown in Section 7.2.2. To compare $\mathbb{P}_{\mathrm{D}}$ with the BER, one must first fix the SNRs of Bob and Willie, and Figs. 7.26–7.37 plot the ROC curve when Bob and Willie have the same SNR (which may be an unrealistic comparison). If we disregard the normal-distribution detector (which had the worst detector performance), then not a single modulation could be undetectable *and* have a BER less than 10^{-2} under these conditions. A communications scheme is often considered useable when the BER is less than 10^{-4}. This paints a grim picture for Alice and Bob's covertness prospects, and highlights the importance of Alice and Bob ensuring that their channel has a better SNR than Willie.

In this work, transmission schemes transmitted random bits that were completely uncoded, including Gray coding. The use of error correction techniques could be employed by Alice and Bob to reduce the error rate of their channel. A low BER is critical because errors will force Alice to retransmit dropped packets. Having to re-transmit data increases her total signal power and time spent using the channel, which consequently increases her probability of being detected by Willie.

The modulation with the best BER/$\mathbb{P}_{\mathrm{D}}$ tradeoff overall was CDMA. CDMA-QPSK was better than CDMA-BPSK because of the increased spectral efficiency of sending twice as many bits per symbol. Using a longer spreading sequence increased covertness with the DCS and normal-distribution detector, but *decreased* covertness with the max cut detector (spreading sequence had no effect on the radiometer). Thus, the best performing variation of CDMA was generally CDMA-QPSK with a 64 bit spreading sequence. However, the reduced data rate means that Alice must transmit longer to achieve the same total throughput, exposing her signal to Willie for a longer duration.

The only other techniques that have a mentionable performance while balancing covertness with BER are PSK and OFDM. Using BPSK is better than QPSK, and using OFDM-BPSK is better than OFDM-QPSK. Increasing the number of OFDM subcarriers reduced covertness and increased detectability. The chaotic modulation had a lackluster performance. I had expected that signals modulated with chaos generators would be more resistant to cyclostationarity detectors, but the chaotic modulations were often *more* detectable than other, more elementary, modulations.

8 Conclusion

This work outlined the fundamental limits of covert communications in Chapter 2, and explored different extensions to the base problem. Several communications schemes of various levels of proclaimed covertness were described in Chapter 3. The statistics of hypothesis testing and the mathematical theory behind several different classes of detectors were described in Chapter 4, while a review of existing metrics for quantifying covertness was conducted in Chapter 5.

An explanation of simulation methods used to generate results and plots was given in Chapter 6. The results of the simulation are shown in Chapter 7; the performance of detectors in Section 7.1, and the performance of the transmission schemes in Section 7.2, with a discussion of the overall results and implications of this experiment found in Section 7.3.

This section summarizes the results of previous sections and provides the key insights about covert communications that were learned through this work. It also discusses open questions and highlights promising avenues for further work.

8.1 Contributions of This Work

The approach in this thesis addresses a deficiency in the current literature by providing a comprehensive performance assessment of multiple transmission schemes with the primary classes of blind-parameter signal detectors. In addition, this work also considered the BER versus $\mathbb{P}_{\mathrm{D}}$ tradeoff, which is usually omitted in other published work. Judging a transmission scheme by its covertness properties alone does not provide a complete picture of its utility. This tradeoff is important because if the deniability of a transmission comes at the cost of reliability of Bob understanding the message then the modulation fails its primary role as a communications scheme.

I investigated and answered two questions in this work:

- **Q.1**: What are the most covert wireless transmission schemes?
- **Q.2**: What are the most powerful signal detectors?

Answering **Q.1** showed that Alice is best off using direct sequence spread spectrum (DSSS) techniques to maximally profit from the tradeoff between bit error rate (BER) and the probability of detection, $\mathbb{P}_{\mathrm{D}}$. The transmission scheme with the best performance tested in this work was CDMA-QPSK. The tradeoff between BER and $\mathbb{P}_{\mathrm{D}}$ is important, as Alice wants to maximize the BER and minimize the probability of detection. Having a high BER means that Alice will have to transmit for a longer duration because of retransmission of dropped packets. This reduces the overall information capacity of the channel, as more of the channel slots are used for packets with errors. This result validates the efficacy of DSSS

techniques as a reliable method of increasing signal covertness, at the cost of a reduced data rate.

Q.2 analyzes the problem of covert communications from the perspective of the illegitimate observer. Warden Willie always achieves better results when he has more information about the transmitted signal. The more characteristics of the transmitted signal that Willie knows, the better of a matched filter he can construct to detect it.

With knowledge of only the signal bandwidth, W, and the AWGN noise variance, N_0, Willie can employ a radiometer or cyclostationarity detector to reliably detect Alice at SNRs as low as -10dB. If Bob and Willie have the same SNR, then the BER for Bob is likely to be unacceptably high—not a single modulation that was tested achieved a BER less than 10^{-2} without also being detected with $\mathbb{P}_{\mathrm{D}} = 1$ (for the TW product used here). The results of this work make a strong case that it is much easier for Willie to achieve his goals than it is for Alice and Bob, who need to, by any means, ensure that Willie has a lower SNR than Bob. Otherwise, they are forced to drastically reduce the throughput of their channel to avoid detection.

The best detectors are the radiometer and the cyclostationarity detectors. The normal-distribution test detector had the worst performance, but was still effective for a variety of modulation schemes. The radiometer the same performance for every transmission scheme, as it only examines signal energy. The cyclostationarity detectors had a wider variance in their ability to detect different modulations, although the DCS detector had less variance than the max cut detector. Sometimes the cyclostationarity detectors performed better than the radiometer, and sometimes they performed worse. This shows that Willie stands to gain from employing multiple detectors simultaneously to increase his probability of detection.

8.2 Recommendations for Further Work

This work established a framework for covert communications and compared the major classes of conventional modulations and detectors. Further work could extend these ideas to consider additional systems and approaches, and the key ones are identified here.

8.2.1 Additional Modulations

This work tested a wide variety of communications schemes, covering all the elementary ones, as well as more exotic modulations designed with covertness in mind. There are other communications schemes that exist, some of which combine characteristics of the different core schemes. These modulations and their trade-offs in respect to complexity, BER performance, and covertness could be investigated using the framework herein.

Additionally, many communications schemes have tuneable parameters (e.g., the spreading sequence used for DSSS) that affect their performance. There are potentially infinite options for modulation parameters, so examining tradeoffs between the various parameters could be examined in further detail. Direct sequence spread spectrum (DSSS), also known as CDMA for a single user, was found to be the best performing modulation in this work, and Hadamard codes were used as the spreading sequence. Work needs to be done to deter-

mine which spreading sequence (e.g., pseudo-random sequences, loosely-synchronous (LS) codes [121]) is best, or if it matters.

Due to the open source nature of this work [36], anyone can easily extend the simulation by adding additional transmission schemes.

8.2.2 Parallel Detector Bank

Section 7.1.1 shows that some detectors have better performance with certain modulations than others. If Willie could employ multiple detectors running in parallel on the same received signal, he could potentially increase his overall detection performance by combining the outputs of all of them. The detector bank, however, incurs an increase in system complexity, computational requirements, and cost.

The simplest way that Willie could connect all the detectors into a single detector would be to use a binary moving window detector (BMWD) [110]. The output of each individual detector is a binary "1" if the detector believes hypothesis H_1 (there was a transmission), and "0" if the detector believes hypothesis H_0 (no transmission). The binary outputs of each detector are fed into the BMWD, which records a detection event (H_1) when one or more of the input detectors has registered a detection event within the period of the moving window. This setup could allow Willie to get the performance advantages of each detector, while minimizing the drawbacks. Research shows that parallel detector banks improve Willie's detection power [137].

The suggestion above with the BMWD was a simple first step. More advanced ways of integrating the information from the detectors in the bank could be used to improve Willie's detection capacity further, like Kalman filters, or machine learning techniques.

8.2.3 Frequency-Channelized Detectors

The parallel detector bank idea found above in Section 8.2.2 discusses a single bandwidth being monitored by multiple detectors. Another idea is to divide the monitored bandwidth into smaller frequency bins, where each subchannel is monitored by a detector.

The output of each detector for each frequency bin could then connect to a BMWD, or via another technique, make a decision from the output of each individual detector. Channelized detectors can be more effective against frequency-hopping spread spectrum (FHSS) and CSS, as mentioned in Section 4.2.1). These could also help Willie to filter out public users to further avoid false alarms.

Willie could also combine channelized detectors with a parallel bank of detectors operating on each subchannel. This mega detector bank could allow him to acquire the benefits of both methods. The tradeoff here, again, is an increase in system complexity and computational requirements.

8.2.4 Partial and Burst Transmissions

As described in Section 6.1, this work assumes that in the H_1 case, Alice is transmitting for Willie's entire integration period, and that in the H_0 case, she is not transmitting at all, so Willie receives only noise. In the real world it is likely that Willie's observation period does not always perfectly overlap Alice's transmission window. He might catch only

the beginning or the tail end of a transmission within his observation period. Alice could also transmit information in short bursts that are potentially much smaller than Willie's observation period. She could send one or more of these packets within this observation time, so further research should be done measuring the performance of detectors when the transmission does not occupy the entire bandwidth.

This is important, because it is a necessary consideration not only to build practical detectors, but also because it may reveal additional performance differences between detectors. Since the radiometer is a total power detector, its performance suffers increasingly as Alice's transmissions occupy less and less of Willie's observation period. The cyclostationarity detectors ought to fare better than the radiometer under these conditions, but further research is required to establish the relative performance degradation of detectors that only receive partial or burst transmissions.

8.2.5 Total Data Throughput

It has been seen that DSSS signals are the hardest to detect, and also suggested that error correction coding would improve Bob's BER performance. Both of these suggestions reduce the data rate, which means Alice must transmit for longer to achieve the same total data throughput. It would be useful to evaluate what the net effect is on the covertness of the signal when this larger timescale is taken into account. This would permit an understanding of how many "covert" bits the channel has with a given modulation scheme and detector.

8.3 Key Takeaways

The results of this work, in conjunction with the existing literature, provides several insights for the designers of both covert communications schemes and signal detectors. Many of the physical layer characteristics of the covert communications problem were abstracted away in this work in order to produce more general results. The high-level considerations extend beyond choosing a transmission scheme and detector type, and involve critical details for building these systems practically.

8.3.1 SNR (at Willie) Matters

The best thing that Alice and Bob can do to achieve a positive covert capacity is to ensure that Willie receives as little of Alice's signal energy as possible. Alice and Bob can achieve this by using a high gain antenna pointing at Bob while ensuring that Willie is not in the line of sight. Alice and Bob can also employ beamforming, MIMO, and IRSs to achieve this same end, as discussed in Section 2.4. Alternatively, Alice can lower her transmit power if she knows that Bob is closer to her than Willie, as in the setup for detectability gain found in Section 5.1.1.

Every detector fails when the SNR at Willie is sufficiently low, so it should be a high priority for Alice and Bob to prevent Willie from seeing as much of the transmitted message as possible. Likewise, if Willie knows the location of Alice, he can point an antenna at her to increase his gain.

8.3.2 Willie's Estimate of Channel Conditions Matters

While the radiometer has been proven to be the mathematically the theoretically optimal detector for an unknown deterministic signal under AWGN, the theory assumes that Willie at all times has a perfect estimate of the noise variance, N_0. As discussed in Section 5.1, when Willie has to estimate the noise variance himself, this gravely decreases detector performance [128, 129]. When Willie has a sufficiently bad estimate of N_0, then positive covert capacity is possible between Alice and Bob, as mentioned in Section 2.1.1.

This, of course, matters much less when the detector is a matched filter (Section 4.3) or a cyclostationarity detector (Section 4.4), because these detector types analyze the received signal structure beyond just the total energy, and thus perform better when the SNR is not known exactly.

8.3.3 Bandwidth W and Integration Period T

For the radiometer, only the overall TW product matters for detection. This work assumes that Alice's transmission occupies the entire bandwidth W and lasts for the entire integration period T (or, that she does not transmit for the entire integration period T, in the H_0 case). This is unrealistic, however, as there is no guarantee that Willie sees all of Alice's signal.

If the bandwidth, W, is too wide, or if the integration period, T, is longer than Alice's transmission, then Willie's detectors are integrating more noise power that they need to.

The values of T and W matter a *lot* for the cyclostationarity detectors. Whereas the radiometer cares merely for the TW *product*, there is far greater variation in detector performance when T and W are altered individually for the cyclostationarity detectors, as mentioned in Section 7.1.4.

A Appendices

A.1 Big-$\mathcal{O}$ Notation

Bachman-Landau notation [143, Ch. 3], colloquially known as big-$\mathcal{O}$ notation, is used to describe the asymptotic behaviour of functions.

To start, the notation $f(n) = \Theta(g(n))$ means that $f(n)$ grows exactly as fast as $g(n)$.

This implies that there exist positive constants c_1, c_2 and n_0 such that $c_1 g(n) \leq f(n) \leq c_2 g(n)$, $\forall n \geq n_0$. In Bachman-Landau notation, $\Theta(\cdot)$ is actually a *set* of functions, so $f(n) = \Theta(n)$ actually means that $f(n) \in \Theta(n)$. The above definitions imply for some positive constant k that:

$$f(n) \in \Theta(g(n)) \iff \lim_{n\to\infty} \frac{f(n)}{g(n)} = k.$$

Most often used in the literature is the big-$\mathcal{O}(\cdot)$. $f(n) = \mathcal{O}(g(n))$ signifies that the $f(n)$ grows *at most* as fast as $g(n)$ asymptotically. Formally, this means there exist positive constants c and n_0 such that $0 \leq f(n) \leq cg(n)$, $\forall n \geq n_0$.

The other similar terms used throughout are $\Omega(n)$ and $o(n)$. $f(n) = o(g(n))$ implies that $\forall c > 0$, $\exists n_0 > 0$ such that $0 \leq f(n) < cg(n)$, $\forall n \geq n_0$. This means that $f(n)$ strictly grows slower than $g(n)$. The term $f(n) = \Omega(g(n))$ means that $f(n)$ grows *at least* as fast as $g(n)$. Formally, there exist positive constants c and n_0 such that $0 \leq cg(n) \leq f(n)$, $\forall n \geq n_0$. For completeness, $f(n) = \omega(g(n))$ implies $f(n)$ grows strictly faster than $g(n)$. So $\forall c > 0$, there is an $n_0 > 0$ such that $0 \leq cg(n) < f(n)$.

Here is a rough chart of what the notations mean for asymptotic values of f and g:

$$\begin{aligned} f &= \mathcal{O}(g(n)) \approx f \leq g \\ &= \Omega(g(n)) \approx f \geq g \\ &= \Theta(g(n)) \approx f = g \\ &= o(g(n)) \approx f < g \\ &= \omega(g(n)) \approx f > g. \end{aligned}$$

A.2 Bit Error Rates of Modulations

The bit error rate (BER) of each modulation was calculated by counting the number of bit errors at the receiver at different SNRs under additive white Gaussian noise (AWGN). Throughout this work, the BER is shown in terms of the ratio of the total signal power to noise power:

$$\text{SNR} = \frac{\text{Total Signal Power}}{\text{Total Noise Power}}.$$

All the BERs, as a function of SNR, can be seen in Figs. A.1–A.3.

It is more conventional in the literature to measure BER as a function of the bit energy, $\frac{E_b}{N_0}$. This is ratio of the energy per bit, E_b, to the noise variance, and measures the spectral efficiency of the modulation. Figs. A.4–A.6 shows a plot of BER as a function of the bit-energy, $\frac{E_b}{N_0}$.

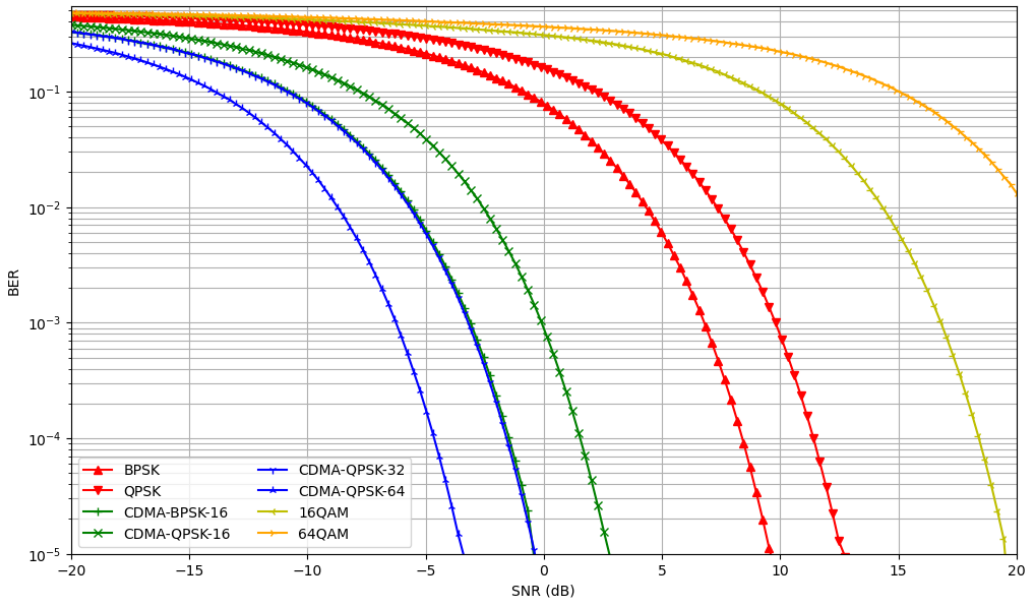


Figure A.1: The BER of Group 1 modulations as a function of the ratio of signal power to noise power (SNR).

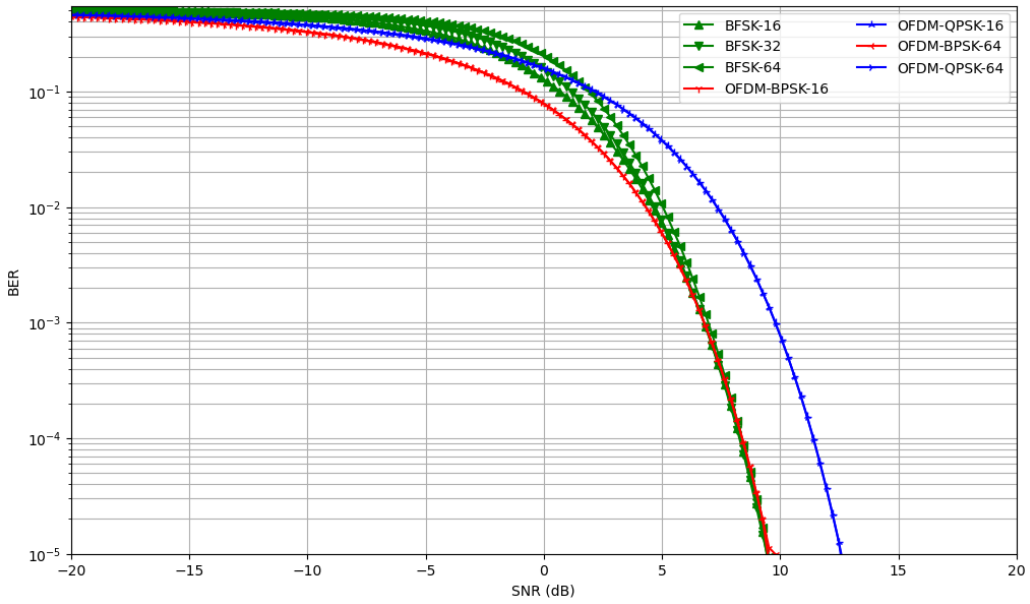


Figure A.2: The BER of Group 2 modulations as a function of the ratio of signal power to noise power (SNR).

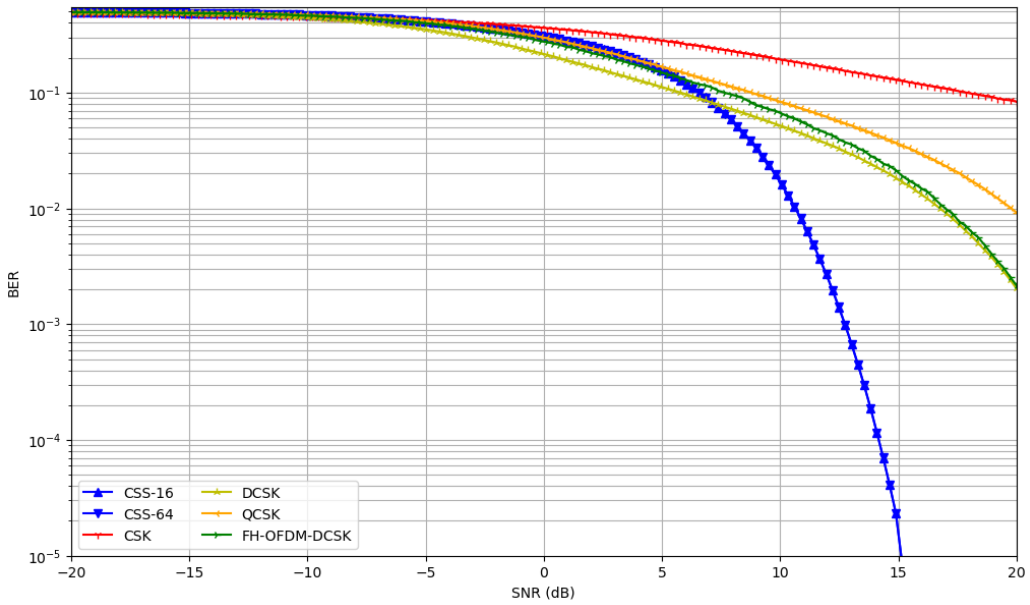


Figure A.3: The BER of Group 3 modulations as a function of the ratio of signal power to noise power (SNR).

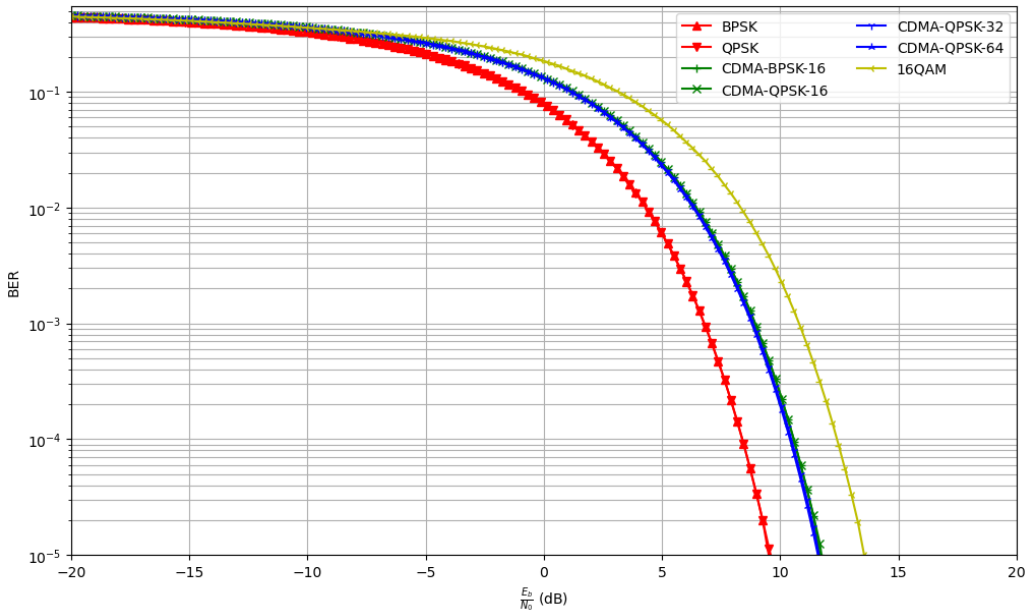


Figure A.4: The BERs of Group 1 modulations as a function of the ratio bit-energy to noise variance, $\frac{E_b}{N_0}$.

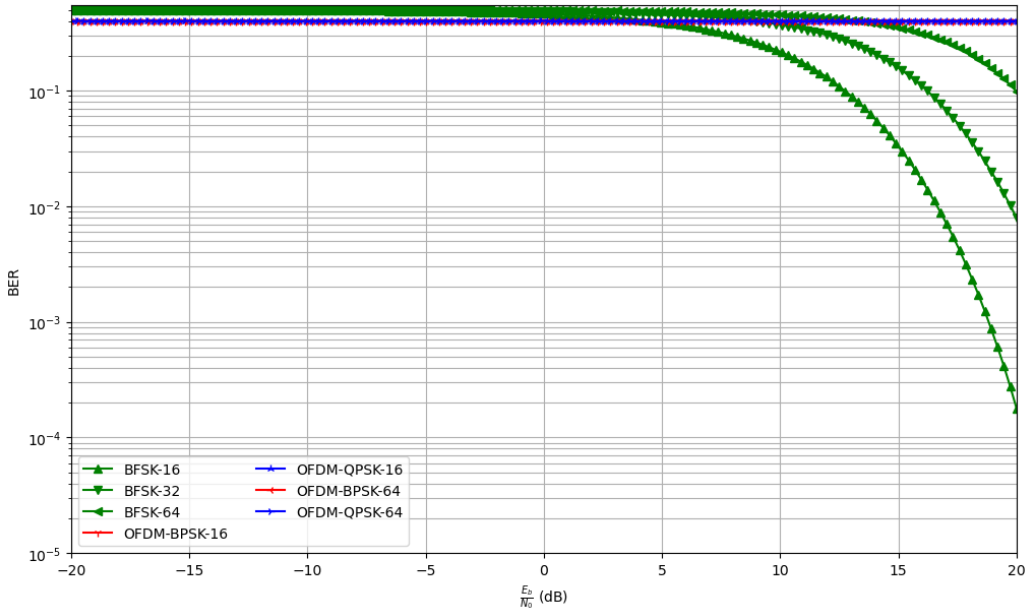


Figure A.5: The BERs of Group 2 modulations as a function of the ratio bit-energy to noise variance, $\frac{E_b}{N_0}$.

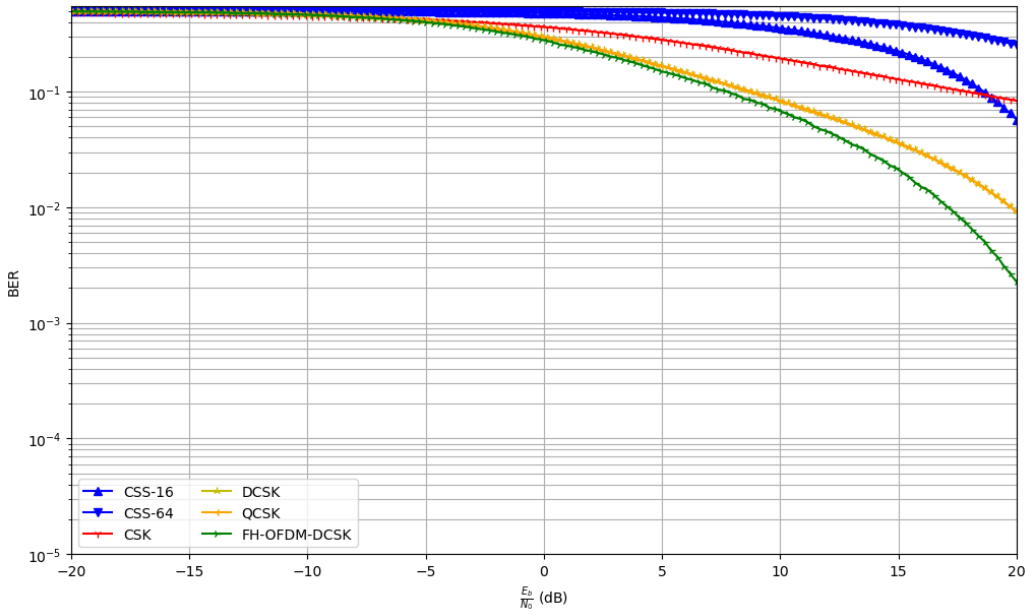


Figure A.6: The BERs of Group 3 modulations as a function of the ratio bit-energy to noise variance, $\frac{E_b}{N_0}$.

Bibliography

[1] Herodotus, *The histories of Herodotus, a translation by A. D. Godley.* Scribe Publishing, Oct. 1921. [Online]. Available: https://www.perseus.tufts.edu/hopper/text?doc=Hdt.+5.35.3

[2] A. Tacticus, *How to Survive under Siege*, ser. Loeb Classical Library. Illinois Greek Club, Jan. 1923. [Online]. Available: https://www.aeneastacticus.net/public_html/ab31.htm

[3] P. Wright, *Spycatcher.* Heinemann, Jul. 1987.

[4] A. D. Ker, T. Pevný, J. Kodovský, and J. Fridrich, "The square root law of steganographic capacity," in *Proceedings of the 10th ACM Workshop on Multimedia and Security*, ser. MM&Sec '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 107–116. [Online]. Available: https://doi.org/10.1145/1411328.1411349

[5] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[6] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, 2013.

[7] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1195–1205, 2015.

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

[9] G. Dillard and R. Dillard, "A metric for defining low probability of detection based on gain differences," in *Conference Record of Thirty-Fifth Asilomar Conference on Signals, Systems and Computers (Cat.No.01CH37256)*, vol. 2, 2001, pp. 1098–1102 vol.2.

[10] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, 2016.

[11] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, 2016.

[12] S. Yan, Y. Cong, S. V. Hanly, and X. Zhou, "Gaussian signalling for covert communications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3542–3553, 2019.

[13] M. Tahmasbi and M. R. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2190–2212, 2019.

[14] Q. E. Zhang, M. R. Bloch, M. Bakshi, and S. Jaggi, "Undetectable radios: Covert communication under spectral mask constraints," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 992–996.

[15] S.-Y. Wang and M. R. Bloch, "Covert MIMO communications under variational distance constraint," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4605–4620, 2021.

[16] Y. Katsuki, G. T. F. de Abreu, K. Ishibashi, and N. Ishikawa, "A new noncoherent Gaussian signaling scheme for low probability of detection communications," *IEEE Wireless Communications Letters*, vol. 12, no. 3, pp. 545–549, 2023.

[17] S. Yan, X. Zhou, J. Hu, and S. V. Hanly, "Low probability of detection communication: Opportunities and challenges," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 19–25, 2019.

[18] B. A. Bash, D. Goeckel, and D. Towsley, "Square root law for communication with low probability of detection on AWGN channels," in *2012 IEEE International Symposium on Information Theory Proceedings*, 2012, pp. 448–452.

[19] N. Letzepis, "A finite block length achievability bound for low probability of detection communication," in *2018 International Symposium on Information Theory and Its Applications (ISITA)*, 2018, pp. 752–756.

[20] L. Wang, G. W. Wornell, and L. Zheng, "Limits of low-probability-of-detection communication over a discrete memoryless channel," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 2525–2529.

[21] K. S. K. Arumugam and M. R. Bloch, "Covert communication over a K-user multiple-access channel," *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7020–7044, 2019.

[22] K.-H. Cho and S.-H. Lee, "Treating interference as noise is optimal for covert communication over interference channels," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 816–821.

[23] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha, "Hiding information in noise: fundamental limits of covert wireless communication," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 26–31, 2015.

[24] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *2013 IEEE International Symposium on Information Theory*, 2013, pp. 2945–2949.

[25] A. D. Ker, "The square root law of steganography: Bringing theory closer to practice," in *Proceedings of the 5th ACM Workshop on Information*

Hiding and Multimedia Security, ser. IH&MMSec '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 33–44. [Online]. Available: https://doi.org/10.1145/3082031.3083235

[26] D. Goeckel, B. Bash, S. Guha, and D. Towsley, "Covert communications when the warden does not know the background noise power," *IEEE Communications Letters*, vol. 20, no. 2, pp. 236–239, 2016.

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

[28] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *2014 IEEE Information Theory Workshop (ITW 2014)*, 2014, pp. 30–34.

[29] H. Q. Ta and S. W. Kim, "Covert communication under channel uncertainty and noise uncertainty," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[30] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with noncausal channel-state information at the transmitter," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 2830–2834.

[31] S.-H. Lee, L. Wang, A. Khisti, and G. Wornell, "Covert communication with channel-state information at the transmitter," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2310–2319, 2018.

[32] P. H. Che, S. Kadhe, M. Bakshi, C. Chan, S. Jaggi, and A. Sprintson, "Reliable, deniable and hidable communication: A quick survey," in *2014 IEEE Information Theory Workshop (ITW 2014)*, 2014, pp. 227–231.

[33] R. Xu, B. Zhang, D. Guo, H. Wang, and G. Ding, "Finite blocklength covert communications: When the warden wants to detect the communications quickly," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 10, pp. 11 278–11 283, 2022.

[34] S. Yan, B. He, Y. Cong, and X. Zhou, "Covert communication with finite blocklength in AWGN channels," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[35] H. Tang, J. Wang, and Y. R. Zheng, "Covert communications with extremely low power under finite block length over slow fading," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2018, pp. 657–661.

[36] L. Rooyakkers, "Bazbandilo," https://crates.io/crates/bazbandilo, 2024.

[37] "IEEE standard for floating-point arithmetic," *IEEE Std 754-2019 (Revision of IEEE 754-2008)*, pp. 1–84, 2019.

[38] B. A. Bash, D. Goeckel, and D. Towsley, "LPD communication when the warden does not know when," in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 606–610.

[39] H. Zhu, H. Wu, and X. Jiang, "Covert MIMO communication in two-hop relay systems," in *2021 International Conference on Networking and Network Applications (NaNA)*, 2021, pp. 63–68.

[40] T. V. Sobers, B. A. Bash, D. Goeckel, S. Guha, and D. Towsley, "Covert communication with the help of an uninformed jammer achieves positive rate," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, 2015, pp. 625–629.

[41] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6193–6206, 2017.

[42] H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Covert communication in the presence of an uninformed, informed, and coordinated jammer," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 306–311.

[43] R. Soltani, D. Goeckel, D. Towsley, B. A. Bash, and S. Guha, "Covert wireless communication with artificial noise generation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7252–7267, 2018.

[44] L. Yang, W. Yang, S. Xu, L. Tang, and Z. He, "Achieving covert wireless communications using a full-duplex multi-antenna receiver," in *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*, 2019, pp. 912–916.

[45] Y. Wang, S. Yan, W. Yang, and Y. Cai, "Covert communications with constrained age of information," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 368–372, 2021.

[46] R. Sun, S. Zeng, and X. Zhu, "Limits of covert communication over AWGN channels in the presence of multiple wardens," in *2018 International Conference on Networking and Network Applications (NaNA)*, 2018, pp. 143–146.

[47] J. Wang, P. Yu, S. Xiao, Y. Zhang, and W. Tang, "Achieving positive covert rate in distributed antenna system," in *2022 IEEE Globecom Workshops (GC Wkshps)*, 2022, pp. 625–630.

[48] V. Y. F. Tan and S.-H. Lee, "Time-division is optimal for covert communication over some broadcast channels," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1377–1389, 2019.

[49] J. Wang, W. Tang, Q. Zhu, X. Li, H. Rao, and S. Li, "Covert communication with the help of relay and channel uncertainty," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 317–320, 2019.

[50] Y. Su, H. Sun, Z. Zhang, Z. Lian, Z. Xie, and Y. Wang, "Covert communication with relay selection," *IEEE Wireless Communications Letters*, vol. 10, no. 2, pp. 421–425, 2021.

[51] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J. Wang, "Covert communication achieved by a greedy relay in wireless networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4766–4779, 2018.

[52] J. Bai, J. He, Y. Chen, Y. Shen, and X. Jiang, "On covert communication performance with outdated CSI in wireless greedy relay systems," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2920–2935, 2022.

[53] S. W. Kim and H. Q. Ta, "Covert communication by exploiting node multiplicity and channel variations," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[54] Y. Qian, W. Li, Y. Lin, L. Shi, X. Zhou, J. Li, and F. Shu, "Antenna coding and rate optimization for covert wireless communications," *IEEE Internet of Things Journal*, vol. 10, no. 3, pp. 2459–2472, 2023.

[55] K. S. Kumar Arumugam and M. R. Bloch, "Embedding covert information in broadcast communications," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2787–2801, 2019.

[56] A. Dutta, D. Saha, D. Grunwald, and D. Sicker, "Secret agent radio: Covert communication through dirty constellations," in *Information Hiding*, M. Kirchner and D. Ghosal, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 160–175.

[57] Tanu and D. Kaur, "A solution to the hidden node problem in cognitive radio networks," in *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, 2017, pp. 16–20.

[58] M. Carrick, "Cyclostationary methods for communication and signal detection under interference," Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, Aug. 2018.

[59] J. Jia, Z. Han, and L. Liu, "Review on low intercept radar signal design technology," in *2022 IEEE 4th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, 2022, pp. 434–437.

[60] H. Deng, "Waveform design for MIMO radar with low probability of intercept (LPI) property," in *2011 IEEE International Symposium on Antennas and Propagation (APSURSI)*, 2011, pp. 305–308.

[61] S. Yirong and C. Zengping, "Evaluation and simulation of LPI radar signals' low probability of exploitation," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, 2017, pp. 842–846.

[62] D. E. Lawrence, "Low probability of intercept antenna array beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 9, pp. 2858–2865, 2010.

[63] C. J. Pici and R. M. Narayanan, "Multifunctional radar and communications waveform using chaos," in *NAECON 2018 - IEEE National Aerospace and Electronics Conference*, 2018, pp. 568–572.

[64] A. Abdelaziz and C. E. Koksal, "Fundamental limits of covert communication over MIMO AWGN channel," in *2017 IEEE Conference on Communications and Network Security (CNS)*, 2017, pp. 1–9.

[65] A. Bendary, A. Abdelaziz, and C. E. Koksal, "Achieving positive covert capacity over MIMO AWGN channels," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 149–162, 2021.

[66] A. Hero, "Secure space-time communication," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3235–3249, 2003.

[67] W. Xiang, J. Wang, S. Xiao, and W. Tang, "Achieving constant rate covert communication via multiple antennas," in *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, 2022, pp. 1–6.

[68] J. Barry and G. Mecherle, "LPI optical communication system," in *MILCOM 1984 - IEEE Military Communications Conference*, vol. 2, 1984, pp. 259–262.

[69] X. Chen, T.-X. Zheng, L. Dong, M. Lin, and J. Yuan, "Enhancing MIMO covert communications via intelligent reflecting surface," *IEEE Wireless Communications Letters*, vol. 11, no. 1, pp. 33–37, 2022.

[70] M. S. Bullock, C. N. Gagatsos, S. Guha, and B. A. Bash, "Fundamental limits of quantum-secure covert communication over bosonic channels," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 3, pp. 471–482, 2020.

[71] L. Wang, "Optimal throughput for covert communication over a classical-quantum channel," in *2016 IEEE Information Theory Workshop (ITW)*, 2016, pp. 364–368.

[72] J. H. Shapiro, D. M. Boroson, P. B. Dixon, M. E. Grein, and S. A. Hamilton, "Quantum low probability of intercept," in *2019 Conference on Lasers and Electro-Optics (CLEO)*, 2019, pp. 1–2.

[73] S. Francis and G. Prescott, "Computer aided analysis of LPI signal detectability," in *MILCOM 91 - Conference record*, 1991, pp. 827–831 vol.2.

[74] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook (Revised Ed.)*. USA: McGraw-Hill, Inc., 1994.

[75] B. G. Mobasseri and K. D. Pham, "Chirp spread spectrum performance in low probability of intercept theater," in *MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM)*, 2018, pp. 329–335.

[76] C. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[77] J. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, NY: McGraw-Hill Professional, Nov. 2007.

[78] S. Singh, S. Sengar, R. Bajpai, and S. Iyer, "Next-generation variable-line-rate optical WDM networks: Issues and challenges," *Journal of Optical Communications*, vol. 34, pp. 331–350, 03 2014.

[79] J. Jung and J. Lim, "Chaotic standard map based frequency hopping OFDMA for low probability of intercept," *IEEE Communications Letters*, vol. 15, no. 9, pp. 1019–1021, 2011.

[80] Z. Liu, L. Zhang, Z. Wu, and J. Bian, "A secure and robust frequency and time diversity aided OFDM–DCSK modulation system not requiring channel state information," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1684–1697, 2020.

[81] B. Gao, Z. Fan, X. Liu, L. Zhang, and X. Ouyang, "OFDM covert communication system based on the QC-LDPC and symbol spread spectrum," in *2020 Cross Strait Radio Science & Wireless Technology Conference (CSRSWTC)*, 2020, pp. 1–3.

[82] Y. Liu, *Introduction to OFDM Receiver Design and Simulation*, 2019.

[83] M. El-Nabawy, M. Aboul-Dahab, and K. El-Barbary, "PAPR reduction of OFDM signal by using Walsh Hadamard transform with μ-law companding technique," *International Journal of Computer Networks & Communication*, vol. 6, 09 2014.

[84] X. Ouyang and J. Zhao, "Orthogonal chirp division multiplexing," *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 3946–3957, 2016.

[85] G. Kaddoum, "Wireless chaos-based communication systems: A comprehensive survey," *IEEE Access*, vol. 4, pp. 2621–2648, 2016.

[86] L. Kocarev, "Chaos-based cryptography: A brief overview," *Circuits and Systems Magazine, IEEE*, vol. 1, pp. 6 – 21, 09 2002.

[87] A. Abel and W. Schwarz, "Chaos communications-principles, schemes, and system analysis," *Proceedings of the IEEE*, vol. 90, no. 5, pp. 691–710, 2002.

[88] G. Heidari-Bateni and C. McGillem, "Chaotic sequences for spread spectrum: an alternative to PN -sequences," in *1992 IEEE International Conference on Selected Topics in Wireless Communications*, 1992, pp. 437–440.

[89] T. L. Carroll, "Chaos for low probability of detection communications," *Chaos, Solitons & Fractals*, vol. 103, pp. 238–245, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960077917302588

[90] M. P. Kennedy and G. Kolumbán, "Digital communications using chaos," *Signal Processing*, vol. 80, no. 7, pp. 1307–1320, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168400000384

[91] C. Williams, "Chaotic communications over radio channels," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 12, pp. 1394–1404, 2001.

[92] K. Busawon, P. Canyelles-Pericas, R. Binns, I. Elliot, and Z. Ghassemlooy, "A brief survey and some discussions on chaos-based communication schemes," in *2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*, 2018, pp. 1–5.

[93] N. J. Corron and J. N. Blakely, "Chaos in optimal communication waveforms," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 471, no. 2180, p. 20150222, Aug. 2015. [Online]. Available: https://doi.org/10.1098/rspa.2015.0222

[94] L. M. Pecora and T. L. Carroll, "Synchronization in chaotic systems," *Phys. Rev. Lett.*, vol. 64, pp. 821–824, Feb 1990. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.64.821

[95] H. Dedieu, M. Kennedy, and M. Hasler, "Chaos shift keying: modulation and demodulation of a chaotic carrier using self-synchronizing Chua's circuits," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 40, no. 10, pp. 634–642, 1993.

[96] G. Kolumbán, B. Vizári, W. Schwarz, and A. Abel, "Differential chaos shift keying : A robust coding for chaotic communication," *Proc. 6th Int. Specialist Workshop Nonlinear Dynamics Electronics Systems*, vol. 4, pp. 87–92, 1996.

[97] Y. Fang, G. Han, P. Chen, F. C. M. Lau, G. Chen, and L. Wang, "A survey on DCSK-based communication systems and their application to UWB scenarios," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1804–1837, 2016.

[98] M. Chen, W. Xu, D. Wang, and L. Wang, "Design of a multi-carrier different chaos shift keying communication system in doubly selective fading channels," in *2017 23rd Asia-Pacific Conference on Communications (APCC)*, 2017, pp. 1–6.

[99] W. Tam, C. Lau, and C. Tse, *Digital communications with chaos : multiple access techniques and performance*. Netherlands: Elsevier, 2007.

[100] Y. Mu, Z. Chen, L. Zhang, Y. Feng, and Z. Wu, "Demonstrating frequency hopping aided OFDM-DCSK and low-rank approximation of matrices based transmissions via USRP," in *2022 IEEE/CIC International Conference on Communications in China (ICCC)*, 2022, pp. 65–70.

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

[102] R. Mills and G. Prescott, "A comparison of various radiometer detection models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 1, pp. 467–473, 1996.

[103] D. Torrieri, "The radiometer and its practical implementation," in *2010 - MILCOM 2010 Military Communications Conference*, 2010, pp. 304–310.

[104] E. April, "On the implementation of the strip spectral correlation algorithm for cyclic spectrum estimation." Defence Research Establishment Ottawa, 1994.

[105] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[106] D. Torrieri, *Detection of Spread-Spectrum Signals*. Cham: Springer International Publishing, 2022, pp. 595–625. [Online]. Available: https://doi.org/10.1007/978-3-030-75343-6_10

[107] M. Wickert, K. Rhead, and D. Reed, "Practical limitations in limiting the rate-line detectability of spread spectrum LPI signals," in *IEEE Conference on Military Communications*, 1990, pp. 994–998 vol.3.

[108] R. A. Dillard, "Detectability of spread-spectrum signals," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-15, no. 4, pp. 526–537, 1979.

[109] J. Lehtomäk, "Analysis of energy based signal detection," *University of Oulu repository*, 12 2005.

[110] G. Dillard, "A moving-window detector for binary integration," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 2–6, 1967.

[111] R. Mills and G. Prescott, "Waveform design and analysis of frequency hopping LPI networks," in *Proceedings of MILCOM '95*, vol. 2, 1995, pp. 778–782 vol.2.

[112] R. F. Mills and G. E. Prescott, "Detectability models for multiple access low-probability-of-intercept networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 3, pp. 848–858, 2000.

[113] K. Watters and E. J. Coyle, "Expected probability of radiometric detection by channelized radiometer," in *MILCOM 2023 - 2023 IEEE Military Communications Conference (MILCOM)*, 2023, pp. 635–642.

[114] A. Polydoros and C. Weber, "Detection performance considerations for direct-sequence and time-hopping LPI waveforms," *IEEE Journal on Selected Areas in Communications*, vol. 3, no. 5, pp. 727–744, 1985.

[115] A. Polydoros and C. L. Weber, "Optimal detection considerations for low probability of intercept," in *MILCOM 1982 - IEEE Military Communications Conference - Progress in Spread Spectrum Communications*, vol. 1, 1982, pp. 2.1–1–2.1–5.

[116] N. Krasner, "Optimal detection of digitally modulated signals," *IEEE Transactions on Communications*, vol. 30, no. 5, pp. 885–895, 1982.

[117] C. L. Nikias and K. T. Woo, "Advanced LPI (Low-Probability-of-Intercept) intercept detector research," Final Report Axiomatix, Los Angeles, CA., Nov. 1985.

[118] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: half a century of research," *Signal Process.*, vol. 86, no. 4, p. 639–697, apr 2006. [Online]. Available: https://doi.org/10.1016/j.sigpro.2005.06.016

[119] W. Gardner, "Signal interception: a unifying theoretical framework for feature detection," *IEEE Transactions on Communications*, vol. 36, no. 8, pp. 897–906, 1988.

[120] A. M. Gillman, "Non co-operative detection of LPI/LPD signals via cyclic spectral analysis." Air Force Institute of Technology, 2012. [Online]. Available: https://scholar.afit.edu/etd/5253/

[121] C. Koumpouzi, P. Spasojevic, and F. T. Dagefu, "Low probability of detection QS-MC-DS-CDMA for low VHF„" in *2019 International Conference on Military Communications and Information Systems (ICMCIS)*, 2019, pp. 1–6.

[122] C. Koumpouzi, P. Spasojevic, and F. Dagefu, "Performance analysis of signal pattern reducing techniques for low probability of detection," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1–5.

[123] D. Nie, K. Xie, F. Zhou, and G. Qiao, "A correlation detection method of low SNR based on multi-channelization," *IEEE Signal Processing Letters*, vol. 27, pp. 1375–1379, 2020.

[124] X.-Y. Hu, C. Bai, and H.-P. Ren, "A chaotic pseudo orthogonal covert communication system," in *2022 6th International Conference on Communication and Information Systems (ICCIS)*, 2022, pp. 61–65.

[125] A. Ali and W. Hamouda, "Advances on spectrum sensing for cognitive radio networks: Theory and applications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1277–1304, 2017.

[126] R. B. D'Agostino, "An omnibus test of normality for moderate and large size samples," *Biometrika*, vol. 58, no. 2, pp. 341–348, 1971. [Online]. Available: http://www.jstor.org/stable/2334522

[127] R. D'Agostino and E. S. Pearson, "Tests for departure from normality. empirical results for the distributions of b_2 and $\sqrt{b_1}$," *Biometrika*, vol. 60, no. 3, pp. 613–622, 1973. [Online]. Available: http://www.jstor.org/stable/2335012

[128] A. Sonnenschein and P. Fishman, "Radiometric detection of spread-spectrum signals in noise of uncertain power," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 654–660, 1992.

[129] A. Sonnenschein and P. M. Fishman, "Limitations on the detectability of spread-spectrum signals," in *IEEE Military Communications Conference, 'Bridging the Gap. Interoperability, Survivability, Security'*, 1989, pp. 364–369 vol.2.

[130] G. Weeks, J. Townsend, and J. Freebersyer, "A method and metric for quantitatively defining low probability of detection," in *IEEE Military Communications Conference. Proceedings. MILCOM 98 (Cat. No.98CH36201)*, vol. 3, 1998, pp. 821–826 vol.3.

[131] Y. R. Zheng and L. L. Fan, "Performance metrics for low probability of detection in cooperative communication networks," in *OCEANS 2016 - Shanghai*, 2016, pp. 1–5.

[132] L. K. Nguyen, M. A. Blanco, and L. J. Sparace, "On the sensitivity of wideband radiometric detection for low probability of intercept and probability of detection (LPI/LPD) in frequency hopped systems," in *MILCOM 2013 - 2013 IEEE Military Communications Conference*, 2013, pp. 817–822.

[133] J. Oetting and J. East, "A comparison of the LPI performance of optical and MM-wave systems," in *MILCOM 1986 - IEEE Military Communications Conference: Communications-Computers: Teamed for the 90's*, vol. 1, 1986, pp. 10.6.1–10.6.5.

[134] J. F. Dishman and E. R. Beadle, "SEVR: a LPD metric for a 3-D battle space," in *MILCOM 2007 - IEEE Military Communications Conference*, 2007, pp. 1–5.

[135] L. Ma, C. Fan, W. Sun, and G. Qiao, "Comparison of detection methods for non-cooperative underwater acoustic DSSS signals," in *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2017, pp. 1–5.

[136] W. Gardner and C. Spooner, "Signal interception: performance advantages of cyclic-feature detectors," *IEEE Transactions on Communications*, vol. 40, no. 1, pp. 149–159, 1992.

[137] O. A. Yeste Ojeda and J. Grajal, "Detection of unknown signals based on spectral correlation measurements," in *2006 14th European Signal Processing Conference*, 2006, pp. 1–5.

[138] M. Kasher, F. T. Dagefu, J. Choi, C. Koumpouzi, and P. Spasojevic, "Low probability of detection communication via polarization diversity: An experimental study," in *2024 United States National Committee of URSI National Radio Science Meeting (USNC-URSI NRSM)*, 2024, pp. 200–201.

[139] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, p. 32–35, 1950. [Online]. Available: http://dx.doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

[140] S. Sedaghatnejad and M. Farhang, "Detectability of chaotic direct-sequence spread-spectrum signals," *IEEE Wireless Communications Letters*, vol. 4, no. 6, pp. 589–592, 2015.

[141] R. Roberts, W. Brown, and H. Loomis, "Computationally efficient algorithms for cyclic spectral analysis," *IEEE Signal Processing Magazine*, vol. 8, no. 2, pp. 38–49, 1991.

[142] N. J. Carter, "Implementation of cyclic spectral analysis methods," *Theory of Computing Systems Mathematical Systems Theory*, p. 4, 1992. [Online]. Available: https://api.semanticscholar.org/CorpusID:108966576

[143] T. Cormen, C. Leiserson, R. Rivest, and C. Stein, *Introduction To Algorithms*, ser. MIT Electrical Engineering and Computer Science. MIT Press, 2001.