UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN ANALYSIS OF THE INFORMATION CONTENT OF RADAR DETECTION

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

TREY T. CRUMP

Norman, Oklahoma

2022

AN ANALYSIS OF THE INFORMATION CONTENT OF RADAR DETECTION

A THESIS APPROVED FOR THE

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Justin Metcalf, Chair

Dr. Joseph P. Havlicek

Dr. Nathan Goodman

Abstract

The availability of the electromagnetic spectrum (EMS) was an unseen issue in the past, as there was sufficient spectrum access to suit the needs of its' consumers. Today, the use of the EMS has been become integrated within our daily lives. Applications varying from civil infrastructure to automotive radar has readily consumed the spectrum to communicate, sense, and interpret information [1]. Given the inflation of spectrum use, it is important that we investigate the amount of information and bandwidth that different spectrum-based applications are using and how different parameters can impact spectrum use. Previous work has identified the fundamental decision bound of pulse-Doppler radar, defined by the Rayleigh range-Doppler resolution, of 1 decision per second per Hz of transmitted bandwidth and has identified a Bayesian detection capacity expression [1].

In this thesis, a closed form expression is derived for detection capacity of radar, which does not require *a priori* probabilities. Furthermore, detection capacity expressions for multiple receivers and the use of M of N integration are also derived. These expressions are used to analyze the information content of different commonly used radar detectors for different assumptions. Another primary focus of this thesis was the analysis of the information content of both fluctuating targets and cell averaging constant false alarm rate (CA-CFAR). The novel analysis of these different radar/target assumptions show the effect that radar detection has on the spectrum. Finally, detection bandwidth and transmit range are connected to determine how signal-to-noise ratio can impact transmit range and spectral usage.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Within the last generation, the use of the electromagnetic spectrum (EMS) has become integrated within our daily lives. In the past, the availability of the spectrum was an unseen issue as there was sufficient spectrum access to suit the needs of its' consumers. Today, applications varying from civil infrastructure to radar has readily consumed the spectrum to communicate, sense, and interpret information [1]. Given the recent inflation of spectrum use, it is important that we have a means to quantify the amount of information and bandwidth that is being used for different spectrum-based applications.

To quantify the information/bandwidth used over a network, information theoretics have been widely used to characterize communication systems and their capacity, benefiting from their uniform nature [1], [2]. There have been some successful radar detection channel models to find the information theoretics of a system [1], [3], although they require *a priori* probabilities, which are often very difficult or even impossible to determine, especially in radar.

Different parameters of radar and target characteristics have yet to be analyzed from an information theoretic standpoint. It is important to find the information from these assumptions and characteristics in order to properly investigate spectral management from a radar perspective. The spectral bound of information that a pulse-Doppler can make is

1 decision per second per Hz of transmitted bandwidth [1]. To best manage this resource, we must determine how different processors use the transmitted bandwidth, such as cell average constant false alarm rate (CA-CFAR) processors. The fundamental goal is optimize and regulate our use of the spectrum from a radar standpoint given our spectrally cluttered environment.

1.2 Thesis Objective

The primary aim of this thesis is to examine the information capacity from a radar detection perspective for different target models. Furthermore, this thesis aims to expand on previously defined Bayesian detection capacity to develop a model that does not rely on *a priori* probabilities. Another aim of this thesis is to examine the information of common hypothesis distributions for varying sensor and variance values. Lastly, this thesis aims to find the effect of using different detection processors such as CA-CFAR.

1.3 Thesis Layout

This thesis contains five chapters which detail the aspects of the application of information theoretics to radar detection for different radar and target assumptions.

Chapter 2 will be focused on background and related work. Section 2.1 discusses radar fundamentals, focusing on the principle and operation of radar. Key concepts such as maximum unambiguous range, range resolution, the Doppler effect, and the radar equation are covered. Section 2.2 focuses on information theory, mainly the fundamental ideas of information. Concepts such as entropy, differential entropy, joint and conditional entropy, relative entropy and mutual information, Fisher information, and channel capacity are discussed. Section 2.3 discusses the theory behind detection from both a general and radar focused perspective and outlines important results and detectors. Bayesian detection, the Neyman-Pearson test, detection considerations within the confines of radar, distributed detection, and Bayesian distributed detection without fusion are summarized.

Chapter 3 provides differential entropy expressions/information content expressions for commonly used hypothesis distributions. Differential entropy expressions for Gaussian, Rayleigh, and Erlang distributions are provided.

Chapter 4 reviews previous models while providing a closed-form expression for detection capacity that does not rely on *a priori* probabilities. Furthermore, the detection capacity of different system and target assumptions are analyzed. Section 4.2 covers detection capacity expressions for the cases of single/multiple receivers and no integration/integration. Section 4.3 shows examples of different detectors for the case of hard decision detection with no integration. Section 4.4 shows examples of different detectors for the case of hard decision detection with integration. Section 4.5 gives expressions for the detection capacity of CFAR. Furthermore, fluctuating targets are also covered from a non-CFAR and CFAR perspective within the section to best show the change in detection capacity for different target assumptions.

Finally the conclusions of the thesis are presented in Chapter 5 including a discussion of possible extensions to the work.

Chapter 2

Background and Related Work

2.1 Radar Processing

RADAR, or RAdio Detection And Ranging, is a system that emits electromagnetic energy from an antenna (or antennas) for the detection and location of reflecting objects such as aircraft, ships, spacecraft, vehicles, and people [4]. Operation entails radiating energy into space and detecting the echo signal reflected from the object/target. The energy reflected is returned to the radar which indicates the presence (or absence) of a target. The basic principle of radar is shown in Figure 2.1.

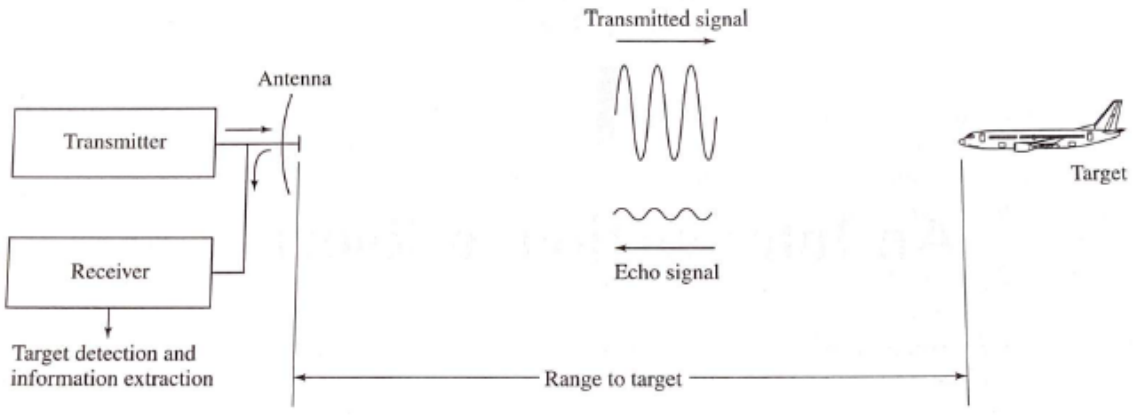

Figure 2.1: Basic radar operating principles [4].

2.1.1 Operating Principles

The radar transmitter transmits an electromagnetic pulse through an antenna and the electromagnetic energy is echoed back from the target/object to be received by the radar receiver. The range to a target can be found by measuring the time it takes for the radar signal to travel to the target and back to the radar. The most common radar waveform is a series of short-duration, rectangular-shaped pulses (known as a pulse train). Electromagnetic energy in free space travels at the speed of light c, so intuitively the range R with respect to target at time T_r is defined as [4]

$$R = \frac{cT_r}{2}\,. \tag{2.1}$$

2.1.2 Maximum Unambiguous Range

Once a radar transmits a signal, sufficient time must be allowed for all of the echo signals to return to the radar before the next pulse is transmitted to avoid destructive interference between the signals. If the time between pulses T_p is too short, an echo signal from a long-range target could arrive after the transmission of the following pulse and could be mistakenly associated with the pulse. The range beyond which targets appear as second-time-around echos is the maximum unambiguous range, R_{un}, given by [4]

$$R_{un} = \frac{cT_p}{2}\,, \tag{2.2}$$

where T_p is defined as the pulse repetition interval (sometimes denoted as PRI), or the amount of time between radar pulses. The inverse of the pulse repetition period is the pulse repetition frequency F_p (sometimes denoted as PRF), varying from several hundred pulses per second and several tens of thousands of pulses per second.

The range resolution, ΔR, is the minimum range between two targets that can be distinguished. In general, the radar range resolution is inversely proportional to signal bandwidth

β, defined as [5]

$$\Delta R \approx \frac{c}{2\beta}\,. \tag{2.3}$$

2.1.3 Doppler Effect

If a radar and a scatterer/target are not at rest with respect to each other, the frequency of the received echo will differ from the transmitted frequency due to the Doppler effect. Properly accounted for Doppler shifts can be used to detect echos from moving targets in the presence of much stronger echos from clutter or can be used to improve cross-range resolution. Uncompensated or improperly sensed Doppler shifts can cause a loss of sensitivity for some types of waveforms. Doppler shift can be expressed as the change in frequency between the transmitted signal, F_t, and the modified received signal, $\alpha_v F_t$. Given a target moving at velocity v, define $\beta_v \equiv v/c$. Let $\alpha_v = \frac{1+\beta_v}{1-\beta_v}$ so the Doppler shift F_d can be expressed as

$$F_d = \alpha_v F_t - F_t = \frac{2v}{(1-\beta_v)\lambda}\,, \tag{2.4}$$

where $\lambda = c/F_t$, or the wavelength of the transmitted sinusoidal waveform signal [5].

2.1.4 The Radar Equation

The radar equation is the fundamental connection between the range of a radar and the physical characteristics of the radar/target. The characteristics of the radar include the transmitter, receiver, antenna, target, and the environment. This equation allows for us to determine the maximum range that a radar can detect a target and how the different characteristics can affect the overall radar performance. The simplest form of the radar range equation can be expressed as

$$R_{max} = \left[\frac{P_t G A_e \sigma}{(4\pi)^2 S_{min}}\right]^{\frac{1}{4}}\,, \tag{2.5}$$

where P_t is the transmitted power (W), G is the antenna gain, $A_e = G\lambda^2/4\pi$ is the antenna effective aperture (m^2), σ is the radar cross section of the target (m^2), and S_{min} is the minimum detectable signal (W) [4]. The radar cross section of a target is defined as

$$\sigma = \frac{\text{power reflected toward source/unit solid angle}}{\text{incident power density}/4\pi} = 4\pi R^2 \frac{|E_r|^2}{|E_i|^2}\,, \tag{2.6}$$

where R is the range to the target, E_r is the electric field strength of the echo signal back to the radar, and E_i is the electric field strength incident on the target [4]. Change in viewing aspect of a radar can result in major changes in the radar cross section.

2.1.5 Detection in Noise

Even if a radar were to be operating in a completely noise free environment, there would still be noise generated by the thermal agitation of the receiver. The resistance of the receiver has an effective temperature T_0 (K) which causes thermal noise with spectral density N_0, defined as $N_0 = kT_0$ where k is Boltzmann's constant and is defined as $k = 1.38 * 10^{-23}$J/deg. Assuming an standard temperature $T_0 = 290$K modified by a noise factor F_n and a half-power bandwidth B with signal-to-noise ratio χ, the minimum detectable signal is defined as [4]

$$S_{min} = kT_0BF_n\chi\,. \tag{2.7}$$

Furthermore, by incorporating losses $L = L_T L_R L_{other}$, where L_T is the transmit loss, L_R is the receive loss, and L_{other} is random loss, we can express the radar range equation with noise and losses as [4]

$$R_{max} = \left[\frac{P_t G A_e \sigma}{(4\pi)^2 kT_0 BF_n \chi L}\right]^{1/4}\,. \tag{2.8}$$

For an example of the radar range equation, let us calculate the range for the radar characteristics found in Table 2.1 for χ values of 13dB, 10dB, and 3dB.

Table 2.1: Radar characteristics for range plots.

Radar Range Equation Parameter	Value
Frequency, f	8 GHz
Wavelength, $\lambda = c/f$	0.0375 m
Transmit Power, P_T	60 dBW
Directive Transmit Gain, G_T	38 dB
Directive Receive Gain, G_R	38 dB
Bandwidth, $B = 1/\tau_p$	2.5 MHz
Noise Figure, F_n	8 dB
Loss, $L = L_T L_R L_{other}$	7 dB
T_0	290 K
Radar Cross Section, σ	3.98 m^2

The results of plugging the radar characteristics found in Table 2.1 into Equation (2.25), are found in Table 2.2.

Table 2.2: R_{max} for varying χ values.

Signal-to-Noise Ratio, χ	Radar Transmit Range, R_{max}
13dB	75.1 km
10dB	77.2 km
3dB	92.9 km

Lower SNR values correspond with higher transmit range as the minimum detectable signal threshold is reduced as SNR decreases.

2.1.6 Radar Signal Processing

From our monostatic, pulsed operation of a radar system previously described, we now have the means to process the received signal. Furthermore, we need to process the received

signal to extract measurement data such as target range and radial velocity. An individually transmitted radar pulse with radar carrier frequency F_t can be described as

$$\bar{x}(t) = a(t)\sin(2\pi F_t t + \theta(t))\,, \tag{2.9}$$

where $a(t)$ is the constant amplitude pulse envelope and $\theta(t)$ is the phase modulation of the pulse [5]. $a(t)$ is an ideal square pulse envelope of amplitude A and duration τ seconds. The received signal will be be a combination of echoes of $\bar{x}(t)$ from targets and noise, expressed as

$$\bar{y}(t) = ka(t - t_0)\exp\left(j[2\pi F_t(t - t_0) + \theta(t - t_0) + \phi(t)]\right) + n(t)\,, \tag{2.10}$$

where $n(t)$ is the receiver noise, k is the echo amplitude factor due to propagation losses, and $\phi(t)$ is the echo phase modulation due to target interaction [5]. The most important parameters of the received signal are the delay time t_0, the echo component amplitude $k|a(t)|$ and its relative power to the noise component, and the echo phase modulation function $\theta(t - t_0) + \phi(t)$ [5]. These parameters allow for us to estimate important values such as target range, radial velocity, scattering strength, and range resolution.

2.2 Information Theory

2.2.1 Entropy

Entropy is the measure of uncertainty of a single random variable [2]. It is the average amount of information intrinsic to the given random variable's outcomes. The entropy of a random variable $\mathcal{X}$ with a probability mass function (PMF) of $p(x)$ is given by

$$H(\mathcal{X}) = -\sum_{x\in\mathcal{X}} p(x)\log_2 p(x)\,, \tag{2.11}$$

where logarithms are to base 2 due to entropy being measured in bits [2]. Thus, entropy is the average number of bits required to describe the random variable. Suppose we have an

unbiased game of heads and tails. Since the game is unbiased, the probabilities of winning are $(\frac{1}{2}, \frac{1}{2})$. The entropy of the unbiased game of heads and tails can be calculated as

$$H(\mathcal{X}) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1 \text{ bit.} \tag{2.12}$$

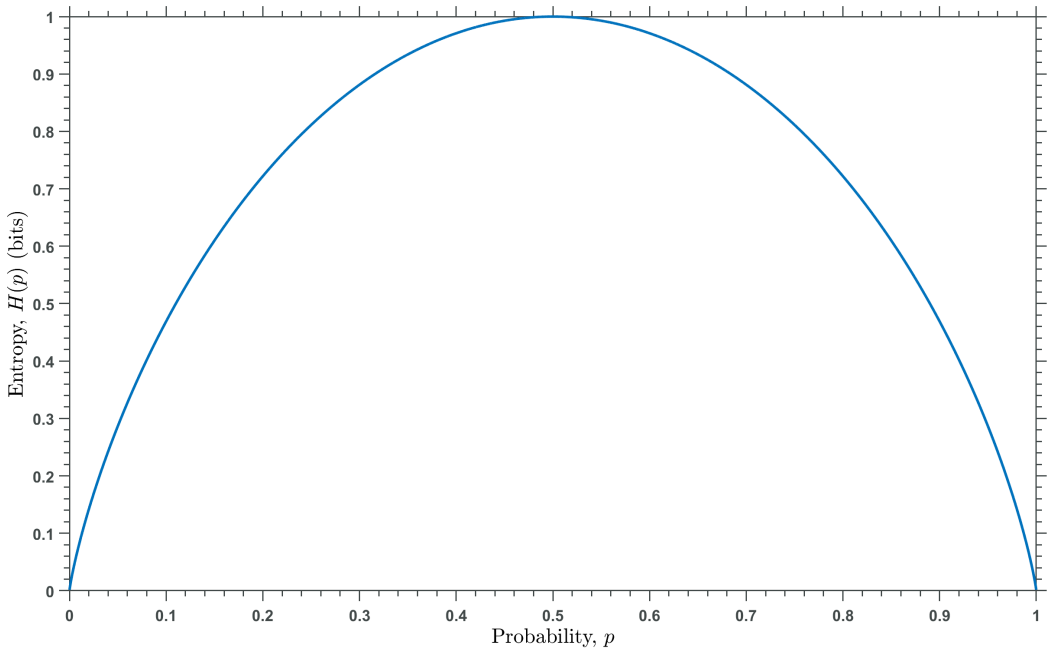

Figure 2.2: Binary entropy $H(p)$ v. p [2].

Figure 2.2 shows the relationship of $H(p)$ and p for a binary entropy distribution. This allows us to see that the entropy of a distribution is a concave function and is equal to 0 when $p = 0$ and $p = 1$.

2.2.2 Differential Entropy

Differential entropy provides a measure of the uncertainty, or information, of a continuous random variable. For random variable $\mathcal{X}$ and probability density function $f(x)$, the differen-

tial entropy is defined as [2]

$$H(\mathcal{X}) = -\int_{\mathcal{X}} f(x) \log f(x) dx \,. \tag{2.13}$$

2.2.3 Joint and Conditional Entropy

From the definition of entropy, it follows that the joint entropy is the measure of uncertainty between two random variables [6]. The joint entropy $H(\mathcal{X}, \mathcal{Y})$ of a pair of discrete random variables $(\mathcal{X}, \mathcal{Y})$ with joint distribution $p(x, y)$ is defined as [2]

$$H(\mathcal{X}, \mathcal{Y}) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y) \,. \tag{2.14}$$

Additionally, the conditional entropy of two random variables is the average amount of information needed to encode the outcome of one random variable given the value of the other random variable is known. If $H(\mathcal{X}, \mathcal{Y}) \sim p(x, y)$, the conditional entropy $H(\mathcal{X}, \mathcal{Y})$ is defined as

$$\begin{aligned} H(\mathcal{Y}|\mathcal{X}) &= -\sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \\ &= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \\ &= -E \log p(\mathcal{Y}|\mathcal{X}) \,, \end{aligned} \tag{2.15}$$

where E denotes the expected value operator [2]. One important theorem tells us that conditioning reduces entropy, meaning more trials reduces the average amount of surprisal between two random variables

$$H(\mathcal{X}|\mathcal{Y}) \leq H(\mathcal{X}) \,, \tag{2.16}$$

with equality if and only if $\mathcal{X}$ and $\mathcal{Y}$ are independent [2].

2.2.4 Relative Entropy and Mutual Information

The relative entropy, or Kullback–Leibler divergence, is the measure of statistical difference between two statistical objects (ex. random variables or probability distributions). For Kullback-Leiblber divergence $D_{KL}(p||q)$, this would be the measure of how different distribution q is from distribution p. It can be used to quantify the information gain between different states if the state of the environment can be described statistically. The relative entropy $D_{KL}(p||q)$ of the probability mass function p with respect to the probability mass function q is defined as [2]

$$D_{KL}(p||q) = \sum_{x} p(x) \log \frac{p(x)}{q(x)} \,. \tag{2.17}$$

Moreover, the mutual information between two random variables is the amount of information obtained from one random variable by observing the other random variable. The mutual information between two random variables $\mathcal{X}$ and $\mathcal{Y}$ is defined as [2]

$$\begin{aligned} I(\mathcal{X};\mathcal{Y}) &= \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) \,. \end{aligned} \tag{2.18}$$

Due to mutual information being the amount of information shared between two random variables, we can also use it to quantify information gain between different states. Figure 2.3 shows the relation between entropy and mutual information [2].

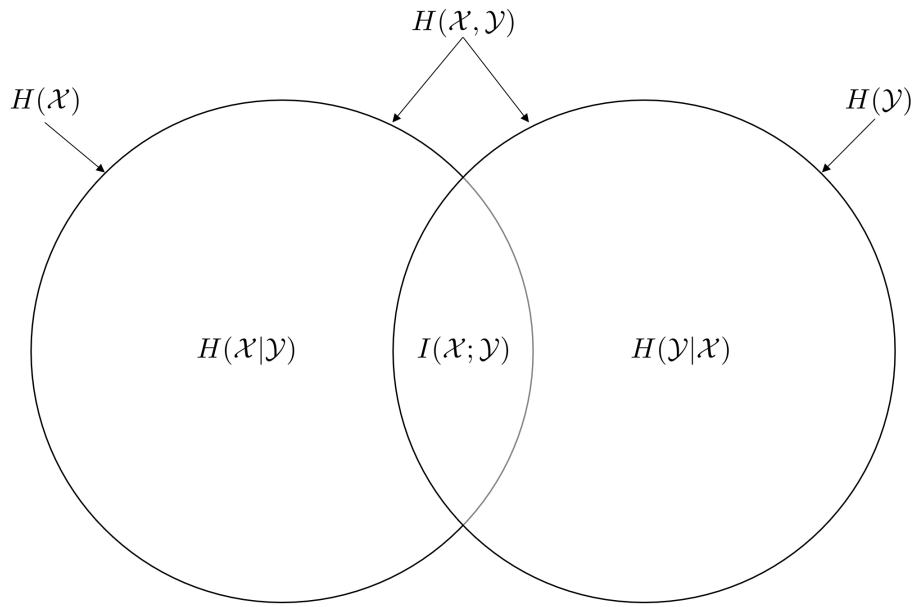

Figure 2.3: Relationship between mutual information and entropy [2].

2.2.5 Fisher Information

Fisher information, J, quantifies the amount of information that can be extracted from a measurement/score function [7]. The score, V, of distribution $f(x;\theta)$ is $V = \frac{\partial}{\partial\theta}\ln f(\mathcal{X};\theta)$ where $\mathcal{X} \sim f(x;\theta)$. The Fisher information is the variance of the score, defined as [2]

$$J(\theta) = E_\theta\left[\frac{\partial}{\partial\theta}\ln f(\mathcal{X};\theta)\right]^2 . \tag{2.19}$$

Per the Cramer-Rao inequality, the inverse of the Fisher Information Matrix is bounded by the Cramer-Rao lower bound which bounds the variance of the unbiased maximum likelihood estimates $\hat{x}$ of x, for example giving us the track accuracy [7]

$$E[[\hat{x}(\mathcal{X}) - x]^2] \geq J^{-1} . \tag{2.20}$$

2.2.6 Channel Capacity

The basic idea of communication is the transmission of information from point A to B through a channel/medium μ_{AB}. This transfer of information through a communication channel is much like a wave propagating through a dielectric material. The wave/information will propagate or be sent from point A to point B but will suffer from dielectric losses/noise. These losses in a communication channel come from noise and imperfections within the process of sending information. Within our communication channel, the highest information rate that can be achieved with an arbitrarily small error probability is known as the channel capacity. The information channel capacity of a discrete memoryless channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ and a probability transition matrix $p(y|x)$, shown in Figure 2.4 is defined as [2]

$$\mathcal{C} = \max_{p(x)} I(\mathcal{X};\mathcal{Y})\,. \tag{2.21}$$

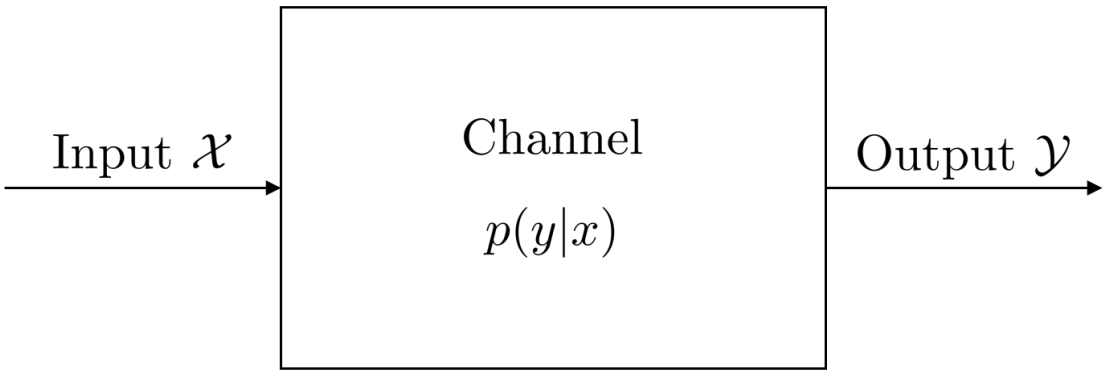

Figure 2.4: Basic information channel.

For an example of channel capacity, let's examine the binary symmetric channel (BSC), shown in Figure 2.5. Binary channels have two inputs/outputs, with input alphabet $\mathcal{X}$ with elements $\{0,1\} \in \mathcal{X}$ and output alphabet $\mathcal{Y}$ with elements $\{0,1\} \in \mathcal{Y}$ with an error probability

of p. An error occurs when a 0 is received as a 1, or vice versa. Since we do not know where the error occurs, given that none of the received bits are reliable (or certain), they are random variables.

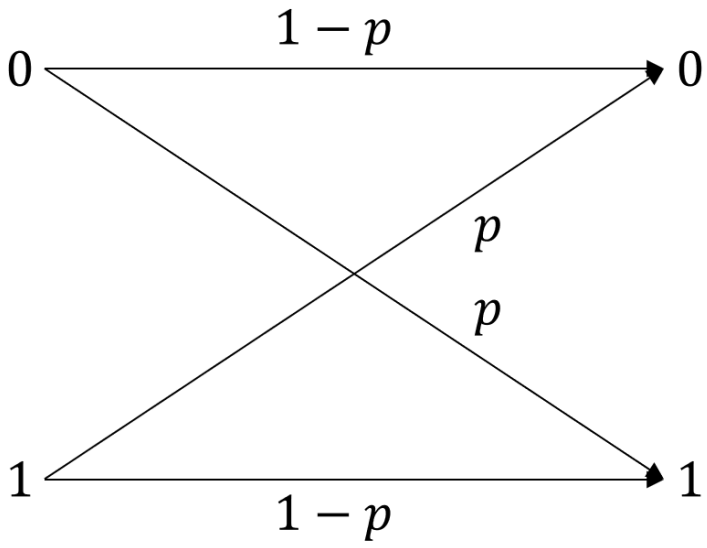

Figure 2.5: Binary symmetric channel [2].

The mutual information between the input alphabet and output alphabet is found by using Equation (2.18) [2]

$$\begin{aligned} I(\mathcal{X};\mathcal{Y}) &= H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) \\ &= H(\mathcal{Y}) - \sum p(x) H(\mathcal{Y}|\mathcal{X}=x) \\ &= H(\mathcal{Y}) - H(p) \\ &\leq 1 - H(p)\,, \end{aligned} \tag{2.22}$$

where $H(p)$ is the binary entropy function, defined as

$$H(p) = -p\log(p) - (1-p)\log(1-p)\,. \tag{2.23}$$

The inequality, $I(\mathcal{X};\mathcal{Y}) \leq 1 - H(p)$, follows due to $\mathcal{Y}$ being a binary random variable. Equality

is achieved when the input distribution is uniform [2]. Since we are looking at the maximization of mutual information between the two alphabets, the channel/information capacity is achieved by the equality (i.e. uniform input distribution)

$$\mathcal{C} = 1 - H(p) \text{ bits.} \tag{2.24}$$

The channel capacity of the binary symmetric channel is shown in Figure 2.6. We can see that the channel capacity of the binary symmetric channel is convex, unlike concave nature of binary entropy (shown in Figure 2.2). This is due to the channel capacity being the complement of the binary entropy function (for the binary symmetric channel).

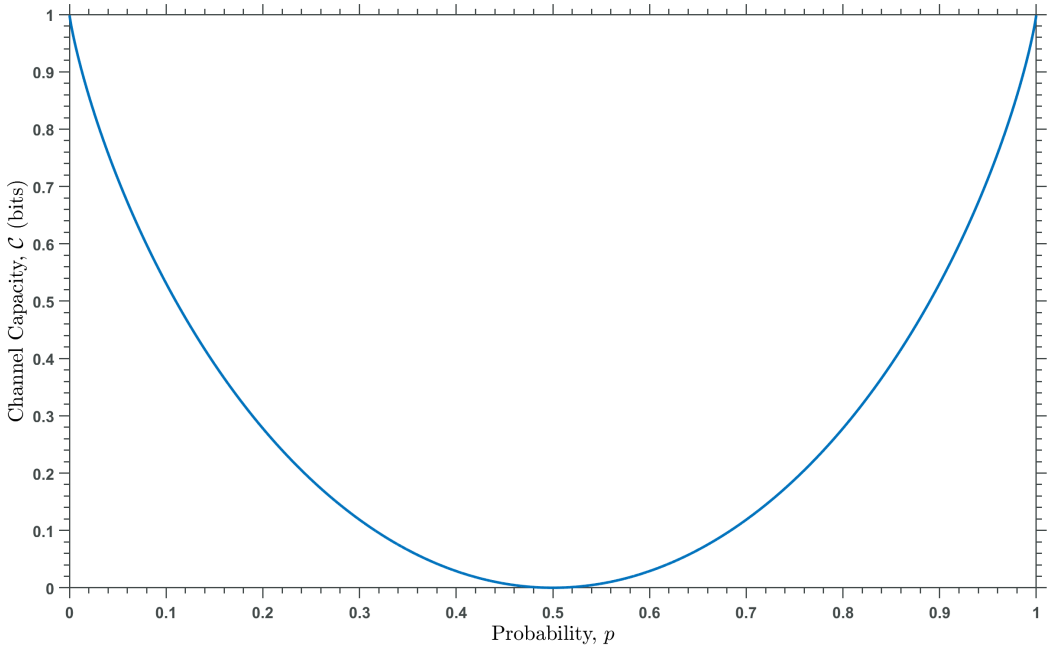

Figure 2.6: Channel capacity of binary symmetric channel.

2.3 Detection Theory

There are many situations in which we are presented with a decision-making problem, or the problem of making a choice with a variety of different possibilities. For example, in radar a decision must be made as to whether or not a target is present or absent when the signal

comes back to the receiver. In communications, a waveform is transmitted over a channel and based on the received waveform, we need to determine the symbol that was transmitted. From these applications, we can see that the basic problem is making a decision from many possible choices with noisy observations.

In a decision making problem, we are faced with a multitude of different hypotheses. The simplest choice is that of binary decisions. Binary decisions are decisions that are either a hard 'yes' or a hard 'no' (sometimes referred to as hard decisions), typically encoded with either a '1' or a '0'. An example of a binary decision process is the choice between two hypotheses, typically denoted $\mathcal{H}_0$ and $\mathcal{H}_1$ for target absence and target presence, respectively. Another decision process is that of soft decisions. It is essentially a 'maybe', typically taking a value between 0 and 1.

The formation of a hypothesis problem starts with a source, which outputs one of the two hypotheses. The source is not known as there would be no decision problem. After the source, there is some probabilistic transition mechanism that generates a point in the observation space by some probability law. A decision rule partitions the observation space into different decision regions according to different hypotheses. The hypothesis for which the decision rule maps the observation to is then marked as 'true' or '1'. In this chapter, we will be outlining different detection methods relevant to the problem that we are studying. The derivations in the following section largely follow from [8].

2.3.1 Bayesian Detection

Let us assume a simple binary hypothesis testing problem, where the hypotheses are either $\mathcal{H}_0$ or $\mathcal{H}_1$. The observation will be denoted as $'y'$ so that the conditional densities under the two different hypotheses are $p(y|\mathcal{H}_0)$ and $p(y|\mathcal{H}_1)$. In this section, the two conditional densities are assumed as known. This means that the observation space is filled by points under the two hypotheses. The Bayesian problem formulation denotes the two *a priori* probabilities of the two hypotheses as P_0 and P_1 respectively. Within a binary detection

problem, four outcomes can occur. Two outcomes are deciding the correct hypotheses and two outcomes are incorrectly deciding the hypotheses. In this scenario, costs are assigned to each possibility, denoted as C_{ij} for $i = 0, 1$ and $j = 0, 1$. This represents the cost of declaring $\mathcal{H}_i$ true while $\mathcal{H}_j$ is present. For the Bayesian problem formulation, the decision rule that minimizes the average cost must be determined. The Bayes risk function $\mathcal{R}$ is given as

$$\mathcal{R} = \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} P_j \int_{Z_i} p(y|\mathcal{H}_j) dy \,, \tag{2.25}$$

where Z_i is the decision region corresponding to hypothesis $\mathcal{H}_i$ [8]. Equation (2.25) can be rewritten as [8]

$$\mathcal{R} = P_0 C_{00} + P_1 C_{11} + \int_{Z_0} [P_1(C_{01} - C_{11}) p(y|\mathcal{H}_1) - P_0(C_{10} - C_{00}) p(y|\mathcal{H}_0)] dy \,. \tag{2.26}$$

The likelihood ratio test (LRT) can be expressed as [8]

$$\frac{p(y|\mathcal{H}_1)}{p(y|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} \,. \tag{2.27}$$

The left hand side is known as the likelihood ratio and the right hand side is denoted as the threshold. The likelihood ratio is tested against the threshold to determine which hypothesis should be determined.

When $C_{00} = C_{11} = 0$ and $C_{10} = C_{01} = 1$, the cost of a correct decision is set to zero and the cost of an error is set to one. The Bayes risk function is now

$$\mathcal{R} = P_0 \int_{Z_1} p(y|\mathcal{H}_0) dy + P_1 \int_{Z_0} p(y|\mathcal{H}_1) dy \,, \tag{2.28}$$

which is the average probability of error [8]. The Bayes test in this case simply minimizes the average probability of error, the threshold η is given by $\frac{P_0}{P_1}$ in this case. In digital communication systems, when $\eta = 1$, the resulting receivers are called minimum probability

of error receivers. The error decisions used in radar are summarized in Table 2.3.

Table 2.3: Different detection probabilities.

Probability	Description	Equation
Probability of Detection, P_D	The probability that a target **is declared** when a target **is present**	$\int_{Z_1} p(y\|\mathcal{H}_1)dy$
Probability of False Alarm, P_{FA}	The probability that a target **is declared** when a target is **not present**	$\int_{Z_1} p(y\|\mathcal{H}_0)dy$
Probability of Miss, P_M	The probability that is a target is **not declared** when a target **is present**	$\int_{Z_0} p(y\|\mathcal{H}_1)dy$

The average probability of error in decision making within a Bayesian framework is given by [8]

$$P_{error} = P_0 P_{FA} + P_1 P_M \,. \tag{2.29}$$

We can rewrite the Bayesian risk function in terms of parameters P_{FA} and P_M, giving $\mathcal{R}$ as [8]

$$\begin{aligned} \mathcal{R} &= C_{00}(1 - P_{FA}) + C_{10}P_{FA} + \\ & P_1[(C_{11} - C_{00}) + (C_{01} - C_{11})P_M - (C_{10} - C_{00})P_{FA}] \,. \end{aligned} \tag{2.30}$$

2.3.2 Neyman-Pearson Test

In most practical applications, the *a priori* probabilities are usually unknown and the cost assignments are difficult to determine. The cost of missing a target cannot be calculated easily since the target's presence may not be known. In radar, we use a special case of the Bayes criterion called the Neyman-Pearson criterion. Within this criterion, the rule/decision process is designed to maximize P_D while keeping P_{FA} under a set value. The radar system designer will designate the acceptable P_{FA} based on the operation of the radar and its sen-

sitivity to higher/lower P_{FA} values. Given a fixed system, increasing P_D has the implication of increasing P_{FA} as well. Let α denote the acceptable value of P_{FA}. The function that maximizes P_D while keeping P_{FA} is defined as follows [8]

$$F = \lambda(1-\alpha) + \int_{Z_0} [p(y|\mathcal{H}_1) - \lambda p(y|\mathcal{H}_0)]\, dy\,, \tag{2.31}$$

for Lagrange multiplier $\lambda \geq 0$, different from our previously defined λ, which was defined as wavelength. The LRT is given as [8]

$$\Lambda(y) = \frac{p(y|\mathcal{H}_1)}{p(y|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda\,. \tag{2.32}$$

The threshold of the test is λ that satisfies the constraint [8]

$$P_{FA} = \int_{\lambda}^{\infty} p(\Lambda|\mathcal{H}_0) d\Lambda = \alpha\,. \tag{2.33}$$

2.3.3 Radar Detection

For a given radar measurement testing for the presence of a target, one of two hypotheses can be true:

$$\mathcal{H}_0 = \text{Target is not present}$$

$$\mathcal{H}_1 = \text{Target is present}$$

where $\mathcal{H}_0$ is denoted as the null hypothesis while $\mathcal{H}_1$ is denoted as the alternative hypothesis. Due the statistic nature of the signals, the analysis starts with a conditional PDF of each of the two hypotheses:

$$p_y(y|\mathcal{H}_0) = \text{PDF of y given that a target was } \textbf{not} \text{ present}$$

$$p_y(y|\mathcal{H}_1) = \text{PDF of y given that a target } \textbf{was} \text{ present}$$

One of the most important parts of the detection problem is developing these two PDFs to achieve detection performance as close to optimal as possible. However, this thesis is primarily focused on the information of these hypotheses, so the design these PDFs will not be covered. Generally, detection will be based on N samples of data y_n forming a column vector $\mathbf{y}$, where $\mathbf{y}$ is

$$\mathbf{y} \equiv [y_0 ... y_{N-1}]^T . \tag{2.34}$$

The simplest detection example is the detection of the presence or absence of a constant in zero-mean Gaussian noise of variance σ_w^2, denoted $N(\mu, \sigma^2) = N(0, \sigma_w^2)$. Let $\mathbf{w}$ be a vector of independent identically distributed (i.i.d.) zero mean Gaussian random variables. When the hypothesis is $\mathcal{H}_0$, meaning the constant is absent, observation vector $\mathbf{y} = \mathbf{w}$ will follow an N-dimensional normal distribution with a scaled identity covariance matrix. When the constant is detected (hypothesis $\mathcal{H}_1$), $\mathbf{y} = \mathbf{m} + \mathbf{w} = m\mathbf{1}_N + w$ and the distribution is shifted to a nonzero mean of $m\mathbf{1}_N$

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{y} \sim N(\mathbf{0}_N, \sigma_w^2 \mathbf{I}_N) \\ \mathcal{H}_1 &: \mathbf{y} \sim N(m\mathbf{1}_N, \sigma_w^2 \mathbf{I}_N) , \end{aligned} \tag{2.35}$$

where $m > 0$ and $\mathbf{0}_N$, $\mathbf{1}_N$, and $\mathbf{I}_N$ are vectors of N zeros, N ones, and the identity matrix of order N respectively [5]. The required PDFs are [5]

$$\begin{aligned} p(\mathbf{y}|\mathcal{H}_0) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{1}{2}\left(\frac{y_n}{\sigma_w}\right)^2\right) \\ p(\mathbf{y}|\mathcal{H}_1) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{1}{2}\left(\frac{y_n - m}{\sigma_w}\right)^2\right) . \end{aligned} \tag{2.36}$$

After modeling the PDFs, the likelihood ratio and the log-likelihood ratio can be calculated

from Equation (2.36) as [5]

$$\Lambda(\mathbf{y}) = \frac{\prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{1}{2}\left(\frac{y_n - m}{\sigma_w}\right)^2\right)}{\prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{1}{2}\left(\frac{y_n}{\sigma_w}\right)^2\right)}, \tag{2.37}$$

$$\begin{aligned} \ln \Lambda \mathbf{y} &= \sum_{n=0}^{N-1} \left(-\frac{1}{2}\left(\frac{y_n - m}{\sigma_w}\right)^2 + \frac{1}{2}\left(\frac{y_n}{\sigma_w}\right)^2\right) \\ &= \frac{1}{\sigma_w^2} \sum_{n=0}^{N-1} m y_n - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} m^2 \,. \end{aligned} \tag{2.38}$$

Substituting Equation (2.38) into the log-likelihood ratio gives the decision rule

$$\sum_{n=0}^{N-1} y_n \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{\sigma_w^2}{m} \ln(-\lambda) + \frac{Nm}{2} \equiv T \,, \tag{2.39}$$

where all of the constraints have been combined into a single constant T [5].

The left hand term, $\sum_{n=0}^{N-1} y_n$ is denoted as the sufficient statistic for this problem. The sufficient statistic is a function of the data $\mathbf{y}$ that allows for the data to appear in the likelihood ratio only through a single function, $\Upsilon(\mathbf{y})$. Thus when making a decision that is optimal under the Neyman-Pearson criterion, knowing the sufficient statistic is as good as knowing the actual data [5]. It can be interpreted as a geometric coordinate transformation chosen to place all of the useful information in the first coordinate [5], [9].

2.3.4 Constant False Alarm Rate Detection

Constant false alarm rate (CFAR) detection sets a constant false alarm rate via an adapative detector threshold. Cell-averaging CFAR tests each available data sample for the presence or absence of a target, the current cell under test (CUT), x_i, is compared against the threshold determined by interference power, which is estimated from the data. If the data in the CUT exceeds the threshold set by the interference power, the detection processor declares a target present within the range-Doppler bin. The detection processing is based on two major

assumptions [5]:

- The neighboring cells contain interference with the same statistics as the CUT (homogeneous interference), so that they are representative of the interference that is competing with the potential target
- The neighboring cells do not contain any targets; purely interference

Zero mean, complex interference statistics can be estimated from the measured samples in the nearby cells. For Gaussian interference and linear or square law detectors, the interference will be Rayleigh or exponential distributed, respectively. In either case, the interference PDF has only one free parameter, the mean interference power. Thus, the CFAR processor must estimate the mean interference power in the CUT by using the measured data in the adjoining cells [5].

Consider the optimal detector, the square law case. The PDF of a cell x_i, assuming the interference is independent and identically distributed (i.i.d.) white Gaussian noise (WGN) in the I and Q signals with power $\sigma_w^2/2$ in each, is [5]

$$p_{x_i}(x_i) = \frac{1}{\sigma_w^2} \exp(-x_i/\sigma_w^2)\,. \tag{2.40}$$

Knowledge of σ_w^2 is needed to set the threshold. When exact knowledge is not available, it must be estimated. Assume that N cells in the vicinity of the cell under test are used to estimate σ_w^2, and that the interference in each is i.i.d. WGN. The joint PDF of a vector $\mathbf{x}$ of N such samples is [5]

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\sigma_w^{2N}} \prod_{i=1}^{N} \exp(-x_i/\sigma_w^2)\,. \tag{2.41}$$

For the case of a square law detector, the expression for the estimated threshold is [5]

$$\hat{T} = \frac{\alpha}{N} \sum_{i=1}^{N} x_i\,. \tag{2.42}$$

Defining $z_i = (\alpha/N)x_i$ gives $\hat{T} = \sum_{i=1}^{N} z_i$ with the PDF of z_i being

$$p_{z_i}(z_i) = \frac{N}{\alpha\sigma_w^2} e^{-Nz_i/\alpha\sigma_w^2}, \tag{2.43}$$

where α is the acceptable value of P_{FA} provided by Equation (2.33) [5]. The PDF of $\hat{T}$ is the Erlang density [5]

$$p_{\hat{T}}(\hat{T}) = \begin{cases} \left(\frac{N}{\alpha\sigma_w^2}\right)^N \frac{\hat{T}^{N-1}}{(N-1)!} e^{-N\hat{T}/\alpha\sigma_w^2}, & \hat{T} > 0 \\ 0, & \hat{T} < 0 \end{cases}. \tag{2.44}$$

Assuming that we have a single sample of a Swerling I or II target with threshold $\hat{T}$, we have a probability of detection and probability of false alarm equal to [5]

$$\begin{aligned} P_D &= \exp[-\hat{T}/(1+\bar{\chi})] \\ P_{FA} &= \exp[-\hat{T}/\sigma_w^2]\,. \end{aligned} \tag{2.45}$$

2.3.5 Distributed Detection

Detection networks can be single or multi-nodal networks with varying topologies. An example of a distributed detection network is in MIMO radar, where an array of receivers each make local detection decisions based on local thresholds/decision rules. Non-fused distributed detection focuses on making local sensor decisions based on local decision rules without performing fusion to obtain a global decision. Fused distributed detection makes local sensor decisions based on local decision rules and fuses those decisions to obtain a global decision.

2.3.6 Bayesian Distributed Detection

Consider a binary hypotheses problem with $\mathcal{H}_0$ and $\mathcal{H}_1$ being the two hypotheses with *a priori* probabilities P_0 and P_1, respectively. Let two detectors, DM_1 and DM_2, collect

observations y_1 and y_2 and make decisions μ_1 and μ_2 respectively. The joint conditional density functions under the two hypotheses are $p(y_1, y_2|\mathcal{H}_i)$, $i = 0, 1$. The decisions μ_i are given by [8]

$$\mu_i = \begin{cases} 0, & \mathcal{H}_0 \text{ is declared present} \\ 1, & \mathcal{H}_1 \text{ is declared present}\,. \end{cases} \tag{2.46}$$

Local decisions μ_i are only based on the local observations of their respective detector/sensor. The costs are given as C_{ijk} with $i, j, k = \{0, 1\}$, where C_{ijk} represents the cost of DM_1 deciding H_i, DM_2 deciding $\mathcal{H}_j$ when $\mathcal{H}_k$ is present.

The goal is to obtain decision rules at both of the detectors that minimize the average cost. Like the Bayesian example in Section 2.3.1, the Bayes risk function is given by [8]

$$\mathcal{R} = \sum_{i,j,k} \int_{y_1,y_2} P_k C_{ijk} p(\mu_1, \mu_2|y_1, y_2, \mathcal{H}_k) p(y_1, y_2|\mathcal{H}_k)\,. \tag{2.47}$$

Due to the local decisions being independent and not depending on the hypothesis present, Equation (2.47) can be rewritten as [8]

$$\mathcal{R} = \sum_{i,j,k} \int_{y_1,y_2} P_k C_{ijk} p(\mu_1|y_1) p(\mu_2|y_2) p(y_1, y_2|\mathcal{H}_k)\,. \tag{2.48}$$

Expanding and simplifying Equation (2.48), $\mathcal{R}$ becomes [8]

$$\mathcal{R} = \int_{y_1} p(\mu_1 = 0|y_1) \sum_{j,k} \int_{y_2} P_k p(\mu_2|y_2) p(y_1, y_2|\mathcal{H}_k) \left[C_{0jk} - C_{1jk}\right]\,. \tag{2.49}$$

Minimizing $\mathcal{R}$ allows us to create the decision rule and making the assumption that $C_{0j0} < C_{1j0}$, meaning that the cost of DM_1 making an error when $\mathcal{H}_0$ is present is more than the cost of it being right regardless of the decision of DM_2 [8]

$$\frac{P_1 p(y_1|\mathcal{H}_1)}{P_0 p(y_1|\mathcal{H}_0)} \underset{\mu_1=1}{\overset{\mu_1=0}{\gtrless}} \frac{\sum_j \int_{y_2} p(\mu_2|y_2) p(y_2|y_1, \mathcal{H}_0)[C_{1j0} - C_{0j0}]}{\sum_j \int_{y_2} p(\mu_2|y_2) p(y_2|y_1, \mathcal{H}_1)[C_{0j1} - C_{1j1}]}\,. \tag{2.50}$$

Assuming that observations y_1 and y_2 are conditionally independent given any hypothesis, the right hand side of Equation (2.50) becomes a threshold t_1 [8]

$$t_1 = \frac{P_0 \int_{y_2} p(y_2|\mathcal{H}_0) \left([C_{110} - C_{010}] + p(\mu_2 = 0|y_2)[C_{100} - C_{000} + C_{010} - C_{110}]\right)}{P_1 \int_{y_2} p(y_2|\mathcal{H}_1) \left([C_{011} - C_{111}] + p(\mu_2 = 0|y_2)[C_{001} - C_{101} + C_{111} - C_{011}]\right)} . \tag{2.51}$$

This threshold t_1 is a function of $p(\mu_2 = 0|y_2)$ which decides the decision rule at DM_2, giving us a function of t_2. Therefore these thresholds can be expressed as functions where [8]

$$\begin{aligned} t_1 &= f_1(t_2) \\ t_2 &= f_2(t_1) . \end{aligned} \tag{2.52}$$

2.4 Related Work

Information theoretic measures have been the fundamental measure of defining spectral utility in communications, with information capacity/channel capacity being the fundamental bound for the utility of traditional communication systems [2], [10]. While the uniform nature of communication systems allows for convenient use of information theoretic analyses, in radar it is more challenging to establish a model.

2.4.1 Detection Rate

The detection rate, or the fundamental bound of decisions per second per Hz of transmitted bandwidth, has been derived in [1] where a pulse-Doppler radar was assumed. From Section 2.1, we know that a pulsed radar unambiguously measures $R_{ua} = \frac{cT_{PRI}}{2}$ meters per PRI. From the pulsed nature of the radar, there are

$$N_{r,ua} = \frac{R_{ua}}{\Delta_r} = \frac{cT_{PRI}}{2}\frac{2B}{c} = BT_{PRI} \tag{2.53}$$

range bins sampled every PRI [1]. Collecting M pulses yields a Doppler space with M bins associated with each N range bin. So each CPI gives

$$N_{RD} = BT_{PRI}M \tag{2.54}$$

total range-Doppler bins [1]. Provided that

$$N_{CPI} = \frac{1}{T_{CPI}} = \frac{1}{T_{PRI}M} \tag{2.55}$$

CPIs are measured each second [1]. The highest number of range-Doppler bins that a radar can generate per second is then equal to

$$N_{bins} = N_{RD}N_{CPI} = BT_{PRI}M\frac{1}{T_{PRI}M} = B\,, \tag{2.56}$$

giving us the fundamental bound on the number of unique decisions a pulse-Doppler radar may make per second, defined by the Rayleigh range-Doppler resolution [1]. This means that the radar is able to make 1 decision per second per Hz of transmitted bandwidth, independent of center frequency and type of waveform transmitted [1].

Chapter 3

Information of Common Hypothesis Distributions

In this chapter we will be analyzing the differential entropy of different observation distributions for 1 to M sensors with 1 to N observations.

3.1 Gaussian Distribution

Due to the Gaussian distribution's entropy properties (i.e. Gaussian distributions have the largest entropy, or information) we will first be analyzing this distribution for the information content for 1 to M sensors with 1 to N observations. Firstly, we assume that we know the noise distribution and the target (i.e. a constant in Gaussian noise). From our previous example in Section 2.3.3, we know that Gaussian decision distributions can be modeled as

$$\begin{aligned} P(y_{ij}|\mathcal{H}_0) &= \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{1}{2}\left(\frac{y_{ij}}{\sigma_w}\right)^2\right) \\ P(y_{ij}|\mathcal{H}_1) &= \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{1}{2}\left(\frac{y_{ij}-s_j}{\sigma_w}\right)^2\right), \end{aligned} \tag{3.1}$$

for s_j individual, independently spaced sensors, where $j = 1, ..., M$. Furthermore, there are $i = 1, ..., N$ observations y, so we have 1 to N observations for each 1 to M sensors. The

notation for observation i on sensor j is denoted y_{ij}. To illustrate how the sensor number impacts the PDF $P(y_{ij}|\mathcal{H}_1)$, Figure 3.1 shows how the PDF is shifted for 0 to 3 sensors.

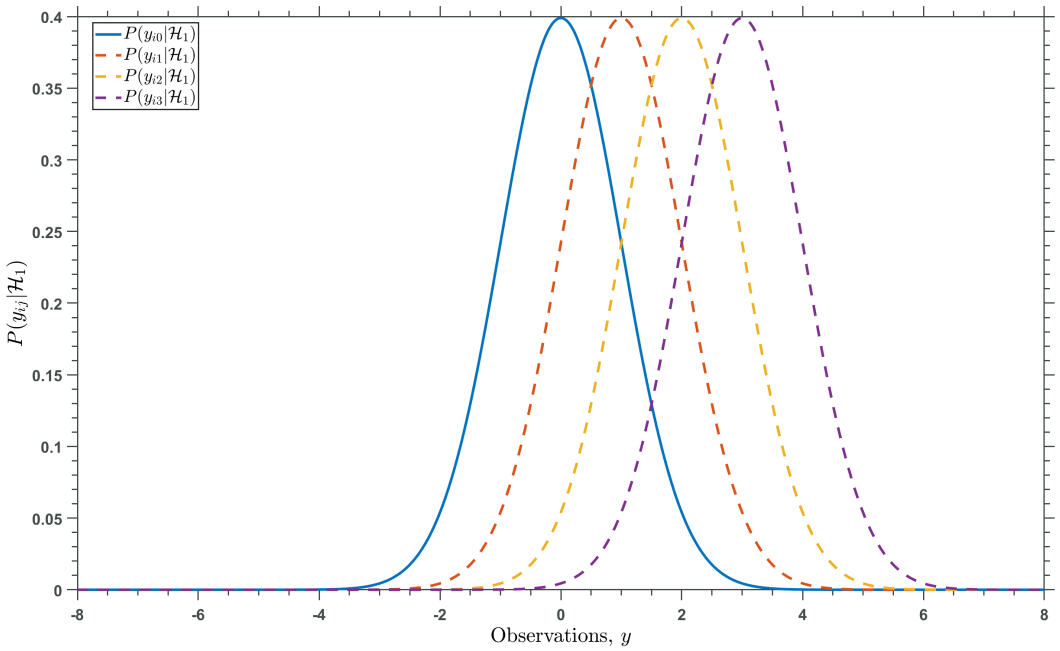

Figure 3.1: $P(y_{ij}|\mathcal{H}_1)$ for j independently spaced sensors.

Due to Gaussian entropy not relying on mean, the entropy of both hypothesis distributions will be the same, regardless of sensor number. For a Gaussian distribution, the differential entropy is [11]

$$h(y_{ij}^{\mathcal{H}_0}) = h(y_{ij}^{\mathcal{H}_1}) = \frac{1}{2}\log(2\pi e\sigma^2)\,. \tag{3.2}$$

The relationship between differential entropy for different variances is seen in Figure 3.2.

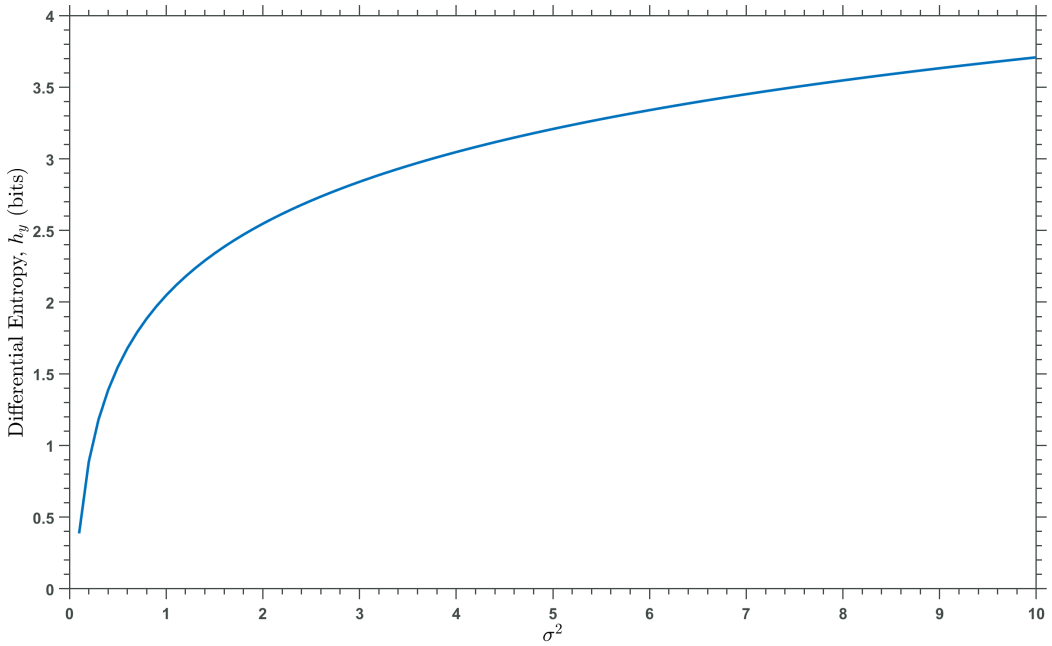

Figure 3.2: Differential entropy of a Gaussian with varying σ^2.

3.2 Rayleigh Fluctuation Environment

A common distribution for the problem of radar detection is the Rayleigh fluctuation environment. Under hypothesis $\mathcal{H}_0$, the PDF of $z_n = |y_n|$ is Rayleigh [5]

$$p_{z_n}(z_n|\mathcal{H}_0) = \begin{cases} \frac{2z_n}{\sigma_w^2} e^{-z_n^2/\sigma_w^2} & z_n \geq 0\,, \\ 0 & z_n < 0\,. \end{cases} \tag{3.3}$$

Under hypothesis $\mathcal{H}_1$, z_n is a Rician voltage density [5]

$$p_{z_n}(z_n|\mathcal{H}_1) = \begin{cases} \frac{2z_n}{\sigma_w^2} e^{-(z_n^2+\tilde{m}^2)/\beta^2} I_0\left(\frac{2\tilde{m}z_n}{\sigma_w^2}\right) & z_n \geq 0\,, \\ 0 & z_n < 0\,, \end{cases} \tag{3.4}$$

where $I_0(\bullet)$ is the modified Bessel function

$$I_0(x) = \sum_{m=0}^{\infty} \frac{1}{m!\Gamma(m+1)} \left(\frac{x}{2}\right)^{2m}. \tag{3.5}$$

The differential entropy under hypothesis $\mathcal{H}_0$ is [11]

$$h(z_n^{\mathcal{H}_0}) = 1 + \log \frac{\sigma_w}{2\sqrt{2}} + \frac{\gamma}{2}, \tag{3.6}$$

where γ is the Euler-Mascheroni constant given by

$$\gamma = -\int_0^{\infty} e^{-x} \log x dx \approx 0.57721566490153286060651209008. \tag{3.7}$$

The differential entropy under hypothesis $\mathcal{H}_1$ can only be found numerically by using the formula found in Equation (2.13).

3.3 Erlang Distribution

The differential entropy of the estimated threshold $\hat{T}$ from Section 2.3.4 can be expressed as [11]

$$h(\hat{T}) = \log \left(\frac{\Gamma(N)\alpha\sigma_w^2}{N} e^{N+(1-N)\psi(N)} \right), \tag{3.8}$$

where $\Gamma(x)$ is the gamma function and $\Psi(y)$ is the digamma function, respectively provided as

$$\begin{aligned} \Gamma(x) &= (x-1)! \\ \Psi(y) &= \frac{d}{dy} \ln(\Gamma(y)) \sim \ln y - \frac{1}{2y}. \end{aligned} \tag{3.9}$$

Chapter 4

Detection Capacity

4.1 Introduction

From our elementary understanding of detection from the radar detection section, we know that one way to decide if something is detected or not is encoding a decision 0 for $\mathcal{H}_0$ and a decision 1 for $\mathcal{H}_1$, which is called a hard decision. While this relatively simplistic view of encoding either a 0 or 1 is adequate in theory, actual detection scenarios rely on soft decisions. A soft decision is essentially a 'maybe' instead of a hard 'yes' or 'no', where instead of deciding either a 0 or 1, it takes the value of somewhere in-between. Soft decisions allow for more accurate detection scenarios at the cost of computational resources. Instead of sending a simple one-bit decision to the sensor fusion center, it sends multi-bit decisions.

There are three main methods for data fusion in these multi-sensor distributed binary detection systems (i.e. MIMO radar), that vary in performance and resource based trade-offs. The best approach theoretically is centralized detection, where all of the sensor observations are sent directly to a central detector where the decision processing is performed. This has optimal performance, as all of the detection data is sent to the fusion center, at the cost of communication bandwidth and computational resources. Another method that is used is binary decision fusion. Decisions are made at the local sensors, sending the fusion center either a 0 or 1 instead of soft decisions that could be multiple bits. This method has the

advantages of optimal communication bandwidth and computational cost, although at the cost of performance. This cost in performance comes from the fact that the fusion center only receives binary partial information.

In Section 4.2, the derivation of detection capacity found in [1] will be expanded upon to find a closed form expression of channel capacity. The key benefit of having this closed form expression is that this expression does not depend on *a priori* probabilities. Subsection 4.2.1 shows the derivation of a closed form expression for detection capacity of hard decisions and shows a plot of channel capacity for various probabilities of detection and probabilities of false alarm. Subsection 4.2.3 expands on the channel capacity expression shown in Section 4.2.1 to add integration over multiple trials. Binary, or M of N integration is implemented, with plots showing the effect of integration on channel capacity in a hard detection system. Section 4.3 shows the effect that different detectors (constant, known RCS in known noise with known/unknown phase) have on detection capacity with no integration. The results match up with the results provided in [1], which confirms that our novel aspect of finding an expression that does not depend on *a priori* probabilities is correct. Section 4.4 shows the effect that different detectors have on detection capacity with integration. Section 4.5 shows the effect of using different processors such as CA-CFAR as well as the effect of having target fluctuation and clutter.

4.2 Detection Capacity of Hard Decisions

4.2.1 Single Receiver

Assume the case of one decision maker, DM_1, and that this decision maker has only makes hard decisions, the case that $\mathcal{H}_0 \in \{0, 1\}$ or $\mathcal{H}_1 \in \{0, 1\}$. Furthermore, assume that there is only a single trial, so there is no integration within the fusion center. We can model this detection problem as a binary channel model as we are only deciding between encoding a 0 or 1. For the case of $P_{FA} = P_D$, we have a completely noisy channel of zero capacity. If we have that $P_D = 1 - P_{FA}$, this case reduces to the case of a binary symmetric channel, where

the channel capacity is $\mathcal{C}_{HD} = 1 - H_b(P_D)$ bits [2]. Given the case of $P_{FA} = 0$, it reduces to a Z-channel. If P_{FA} and P_D do not match the first three cases, more specifically the channel cross-over probability is not the same, we can model this as a binary asymmetric channel, shown in Figure 4.1.

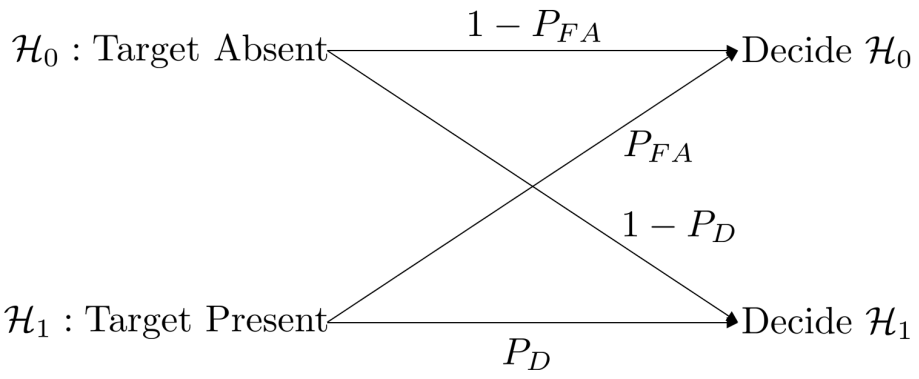

Figure 4.1: Binary asymmetric channel.

For symmetry, the values of P_{FA} and $1 - P_D$ are restricted to the following values [12]

$$0 \leq P_{FA} \leq 1 - P_D \leq 1 \tag{4.1}$$

$$P_{FA} \leq 1 - P_{FA} \tag{4.2}$$

$$P_{FA} \leq P_D \,. \tag{4.3}$$

Figure 4.2 shows the different regions of interest/soft constraints for our model. The region Ω are the bounds provided by Equations (4.1) and (4.2). For the cases that violate the constraints, the model can simply be flipped (which will use the same detection capacity expression). For the case of $P_{FA} > 1 - P_{FA}$, we can simply flip all zeros to ones and vice-versa to get an equivalent channel with $P_{FA} \leq 1 - P_{FA}$. The case where $P_{FA} > P_D$, we can flip the output to get an equivalent channel with $P_{FA} \leq P_D$. Furthermore, for the case where

$P_{FA} > P_D$, we can flip the input to get an equivalent channel with $P_{FA} \leq P_D$ [12].

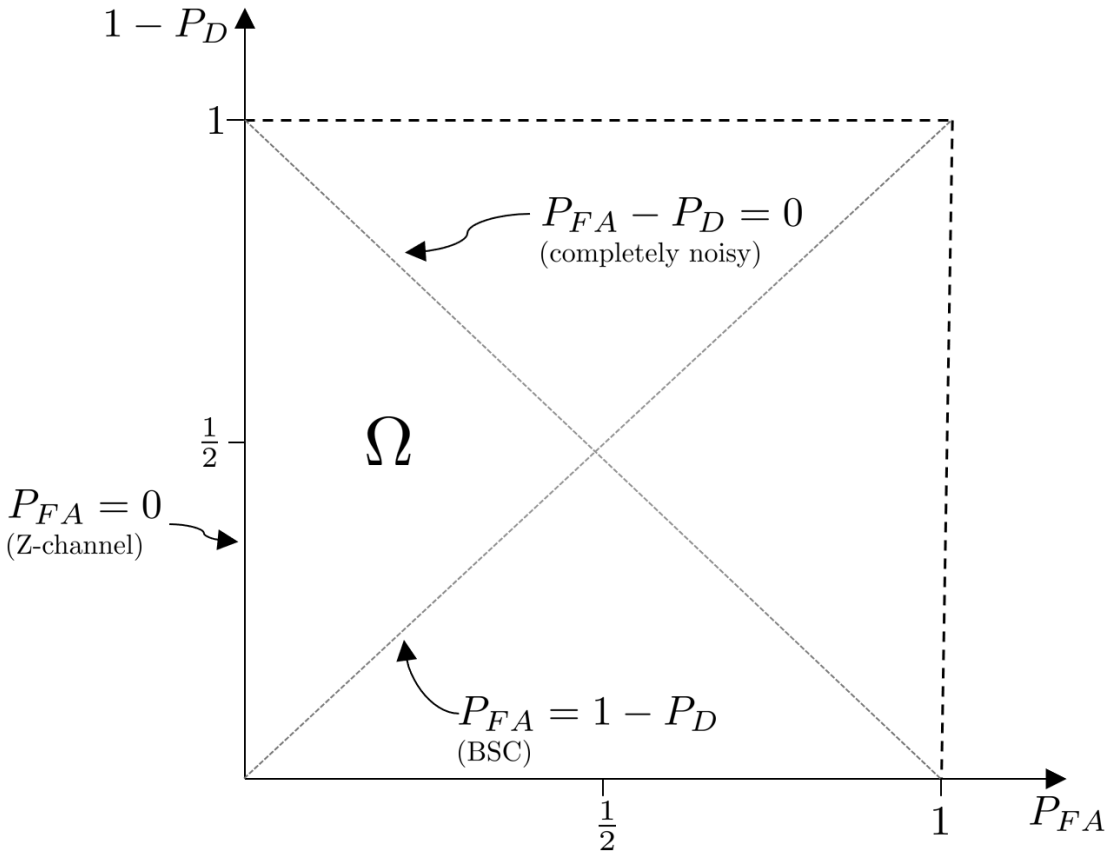

Figure 4.2: Binary channel bounds.

Radar detectors typically utilize a Neyman-Pearson test, where *a priori* probabilities do not play a part of the hypothesis procedure. Bayesian detection frameworks, which are primarily seen in communications and other applications, incorporate *a priori* probabilities into detection threshold estimates [3], [13]. From [1], we know that for random variable $\mathcal{H}$, where $\mathcal{H}$ denotes the probability of a target being present in a range-Doppler bin ($P(\mathcal{H}_0) = P_0$ and $P(\mathcal{H}_1) = 1 - P_0$), the probability of deciding that a target is not present is the marginalization

$$\begin{aligned} P(\mu_0) &= P(\mu = 0) = P(\mathcal{H}_0)P(\mu = 0|\mathcal{H}_0) + P(\mathcal{H}_1)P(\mu = 0|\mathcal{H}_1) \\ P(\mu_0) &= P_0(1 - P_{FA}) + (1 - P_0)(1 - P_D)\,. \end{aligned} \tag{4.4}$$

Similarly, the probability of deciding a target is present is the marginalization

$$\begin{aligned} P(\mu_1) &= P(\mu = 1) = P(\mathcal{H}_0)P(\mu = 1|\mathcal{H}_0) + P(\mathcal{H}_1)P(\mu = 1|\mathcal{H}_1) \\ P(\mu_1) &= P_0 P_{FA} + (1 - P_0)P_D \,. \end{aligned} \tag{4.5}$$

The mutual information of the input and output random variables of the hard decision detection problem can be expressed as [1], [3], [8]

$$\begin{aligned} I(\mathcal{H};\mu) &= \sum_T \sum_\mu P(\mathcal{H},\mu) \log \frac{P(\mathcal{H}|\mu)}{P(\mathcal{H})} \\ &= P_0(1 - P_{FA})[\log(1 - P_{FA}) - \log(P_{\mu_0})] \\ &+ (1 - P_0)(1 - P_D)[\log(1 - P_D) - \log(P_{\mu_0})] \\ &+ (1 - P_0)P_D[\log(P_D) - \log(P_{\mu_1})] \\ &+ P_0 P_{FA}[\log(P_{FA}) - \log(P_{\mu_1})] \,. \end{aligned} \tag{4.6}$$

Since we know that the channel capacity of a hard decision is the maximization of mutual information, we can express it as [12]

$$\mathcal{C}_{HD} = \frac{P_{FA}}{P_D - P_{FA}} H_b(1 - P_D) - \frac{P_D}{P_D - P_{FA}} H_b(P_{FA}) + \log\left[1 + 2^{\frac{H_b(P_{FA}) - H_b(1-P_D)}{P_D - P_{FA}}}\right] , \tag{4.7}$$

where $H_b(\bullet)$ is the binary entropy function, defined in Equation (2.23). The full derivation is provided in Appendix A. The input distribution $P(\mathcal{H})$ that achieves channel capacity is given by [12]

$$P(\mathcal{H}_0) = 1 - P(\mathcal{H}_1) = \frac{\frac{1}{2^x} + P_D - 1}{P_D - P_{FA}} , \tag{4.8}$$

where x is defined as

$$x = \frac{H_b(P_{FA}) - H_b(1 - P_D)}{P_D - P_{FA}} \,. \tag{4.9}$$

We can now plot the hard decision channel capacity for the case of one receiver, provided by Equation (4.7).

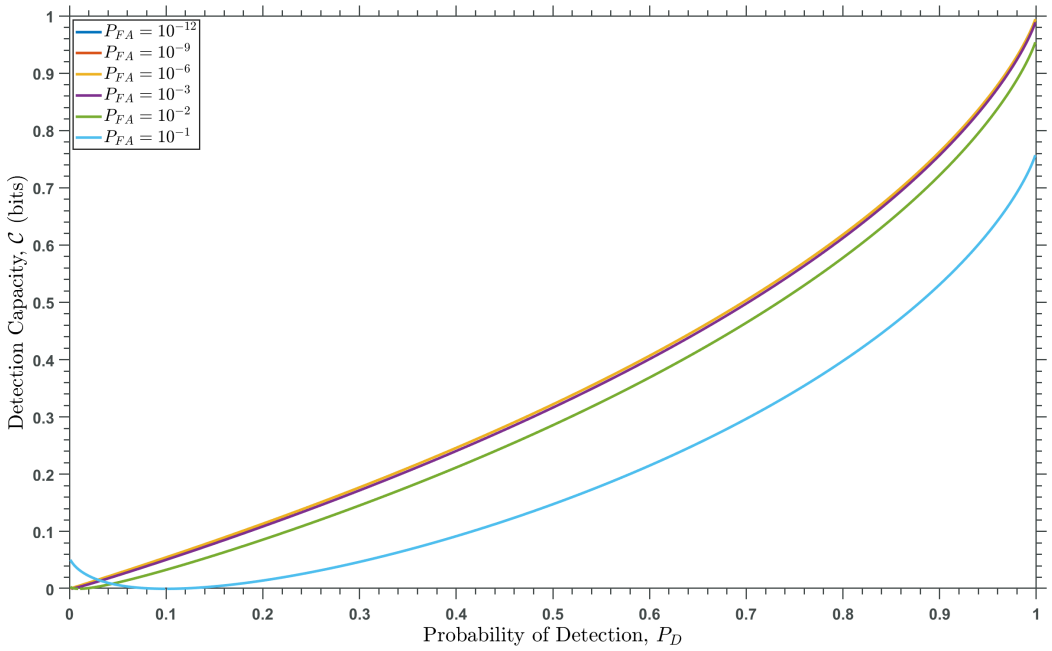

Figure 4.3: Hard decision detection capacity for various P_{FA} values.

Figure 4.3 shows the relationship between detection capacity and probability of false alarm. Lower P_{FA} values such as $P_{FA} = 10^{-12}$, $P_{FA} = 10^{-9}$, $P_{FA} = 10^{-6}$, and $P_{FA} = 10^{-3}$ have higher detection capacity rates as there is less uncertainty between the hypotheses. High P_{FA} values such as $P_{FA} = 10^{-2}$ and $P_{FA} = 10^{-1}$ have lower detection capacity rates as there is more uncertainty due to the probability of false alarm being high (i.e. there will be a higher chance that there is an incorrect hypothesis).

4.2.2 Multiple Receivers without Integration

We can expand our expression for channel capacity for N decision makers/sensors. To do this, we use the fact that our sensor array is independent. Given that our channels are independent, we can use the additive property of channel capacity $\mathcal{C}(c_1 \times c_2) = \mathcal{C}(c_1) + \mathcal{C}(c_2)$, giving us

$$\mathcal{C}_{HD_N} = \sum_{i=1}^{N} \mathcal{C}_{HD_i} \,. \tag{4.10}$$

This may seem to violate our fundamental bound, found in Equation (2.56), of 1 decision per second per Hz of transmitted bandwidth, but the overall value will be normalized by the fusion center. The normalization by the fusion center will be done by the fusion center's global threshold.

4.2.3 Binary Integration

To expand Section 4.2.1 to the case of multiple trials with integration, we can use the M of N decision rule [5]. After coherent or incoherent integration, the detection process is followed by comparing the data with a threshold test (LRT such as section 2.3.1 or 2.3.2). The result of the threshold test has two outcomes, the binary test of deciding the hypothesis of target absent or target present. Given N trials of the detection process, there will be N binary decisions available. Each decision of $\mathcal{H}_1$ will have some probability of P_D being correct and a probability of P_{FA} being incorrect. To improve the reliability of the binary detection decision, the decision rule can be modified such that there must be some number of M successful target detects for N detection trials. This process is also called *binary integration*, modeling our probabilities as binomial distributions. This allows for us to replace our P_D and P_{FA} in Equation (4.7) with [5]

$$\begin{aligned} P_{BD} &= \sum_{r=M}^{N} \binom{N}{r} p_D^r (1-p_D)^{N-r} \\ P_{BFA} &= \sum_{r=M}^{N} \binom{N}{r} p_{FA}^r (1-p_{FA})^{N-r} \,, \end{aligned} \tag{4.11}$$

where $\binom{N}{r}$ is equal to

$$\binom{N}{r} = \frac{N!}{(N-r)!r!} \,. \tag{4.12}$$

One example of the M of N decision rule is putting $N = 4$ and varying the M count. Table 4.1 shows the effect on probability for different numbers of success for the case of $N = 4$ trials [5]. Furthermore, Table 4.2 summarizes the effect of each of the different M cases for

the M of N decision rule.

Table 4.1: Binary integrated probabilities for M of 4 decision rule [5].

p	$P_{M=1}$	$P_{M=2}$	$P_{M=3}$	$P_{M=4}$
0.9	0.999	0.996	0.948	0.656
0.8	0.998	0.973	0.819	0.410
0.5	0.938	0.688	0.313	0.063
10^{-3}	0.004	5.992×10^{-6}	4.00×10^{-9}	1.0×10^{-12}
10^{-6}	4.0×10^{-6}	6.0×10^{-10}	4.00×10^{-18}	1.0×10^{-24}

Table 4.2: Summary of M of N decision rule for $N = 4$.

M of 4	Description
1 of 4	The "1 of 4" rule increases the effective P_D, reducing the SNR required to achieve the goal P_D, at the cost of also increasing P_{FA}
2 of 4	The "2 of 4" rule reduces P_{FA} and increases P_D so long as single-trial P_D is high ($P_D \geq 0.23$)
3 of 4	The "3 of 4" rule has the same benefits as the "2 of 4" rule but single-trial P_D must be high ($P_D \geq 0.75$)
4 of 4	The "4 of 4" rule is good for P_{FA} reduction but is not good for improving P_D

Figures 4.4, 4.5, 4.6, 4.7, and 4.8 show the non-normalized detection capacity for different values of M for $P_{FA} = 10^{-7}, 10^{-5}, 10^{-3}, 0.1,$ and 0.3. For the M of 4 rule, $M = 1$ has the greatest slope for the lower P_{FA} values such as $P_{FA} = 10^{-7}$ and $P_{FA} = 10^{-5}$. At $P_{FA} = 10^{-3}$, $M = 1$ starts to lower in the maximum detection capacity, peaking at approximately $\mathcal{C} = 0.98$. At $P_{FA} = 0.1$ and $P_{FA} = 0.3$, $M = 1$ continues to lower in detection capacity for the higher probabilities of detection, peaking at $\mathcal{C} = 0.46$ and $\mathcal{C} = 0.21$ respectively for

$P_D = 1$, while increasing the detection capacity at lower P_D, with values of $\mathcal{C} = 0.20$ and $\mathcal{C} = 0.55$ for $P_D = 0$. $M = 1$ has a detection capacity of zero at $P_D = 0.1$ and $P_D = 0.3$ for $P_{FA} = 0.1$ and $P_{FA} = 0.3$. For $P_{FA} = 10^{-7}$ and $P_{FA} = 10^{-5}$, $M = 2$ had the second greatest slope. Unlike the case of $M = 1$, $M = 2$ did not suffer from the lowering of maximal detection capacity for $P_{FA} = 10^{-3}$, achieving $\mathcal{C} = 1$ at $P_D = 1$.

For the cases of $P_{FA} = 0.1$ and $P_{FA} = 0.3$, the maximal detection capacity reduces to $\mathcal{C} = 0.85$ and $\mathcal{C} = 0.45$ at P_D while also achieving a detection capacity of $\mathcal{C} = 0.03$ and $\mathcal{C} = 0.21$ for $P_D = 0$. For the case of $M = 3$, maximal $\mathcal{C}$ at $P_D = 1$ is not reduced. $\mathcal{C}$ stays at around 0 for $P_{FA} = 0.1$ at $P_D = 0$ but increases to 0.05 for $P_{FA} = 0.3$. $\mathcal{C}$ is still high for $P_{FA} = 0.1$ and $P_{FA} = 0.3$ at $P_D = 1$, peaking at $\mathcal{C} = 0.98$ and $\mathcal{C} = 0.79$. $M = 4$ had the highest detection capacity peaks for higher P_{FA}, peaking at $\mathcal{C} = 0.98$ and $\mathcal{C} = 0.94$ at $P_D = 1$. Moreover, $M = 4$ never had a detection capacity above zero until the case of $P_D = P_{FA}$ for lower P_D.

Choosing the case of $M = 1$ maximizes detection capacity, or information rate, at lower P_D for higher P_{FA} such as $P_{FA} = 0.1$ and $P_{FA} = 0.3$. This is due to the "1 of N" achieving a high binary integrated P_D for relatively low single-trial P_D values, meaning that the information rate would be the most volatile at this point. Choosing a value of $M = 4$ minimizes the detection capacity at lower P_D for higher P_{FA} such as $P_{FA} = 0.1$ and $P_{FA} = 0.3$, but maximizes the detection capacity at $P_D \approx 1$. This is due to the "4 of 4" rule being good for false alarm reduction at the cost of being bad at improving detection, meaning it takes a very high P_D for the information rate to be maximized. The case of $M = 3$ has high detection capacity at higher P_{FA} at $P_D = 1$ and also has a non-zero detection capacity for $P_{FA} = 0.3$ at $P_D = 0$, albeit the lowest of the non-zero cases. This follows from our understanding of the $M = 3$ decision rule, as the "3 of 4" rule only increases probabilities equal to approximately 0.75 or higher. The case of $M = 2$ has the second highest detection capacity at higher P_{FA} values for $P_D = 0$ but also has the third highest detection capacity at $P_D = 1$. This makes sense as the "2 of 4" rule provides good false alarm reduction for small values of probability

while providing detection improvement for large values of probability, improving detection down to a probability value of approximately 0.25.

For smaller values of P_{FA} such as $P_{FA} = 10^{-7}$, $P_{FA} = 10^{-5}$, and $P_{FA} = 10^{-3}$, the detection capacity is maximized under all of the values of P_D for the case of $M = 1$. This is primarily due to the "1 of N" rule achieving a high binary integrated probability of detection, and since P_{FA} is very small, the simultaneous rise of P_{FA} is negligible.

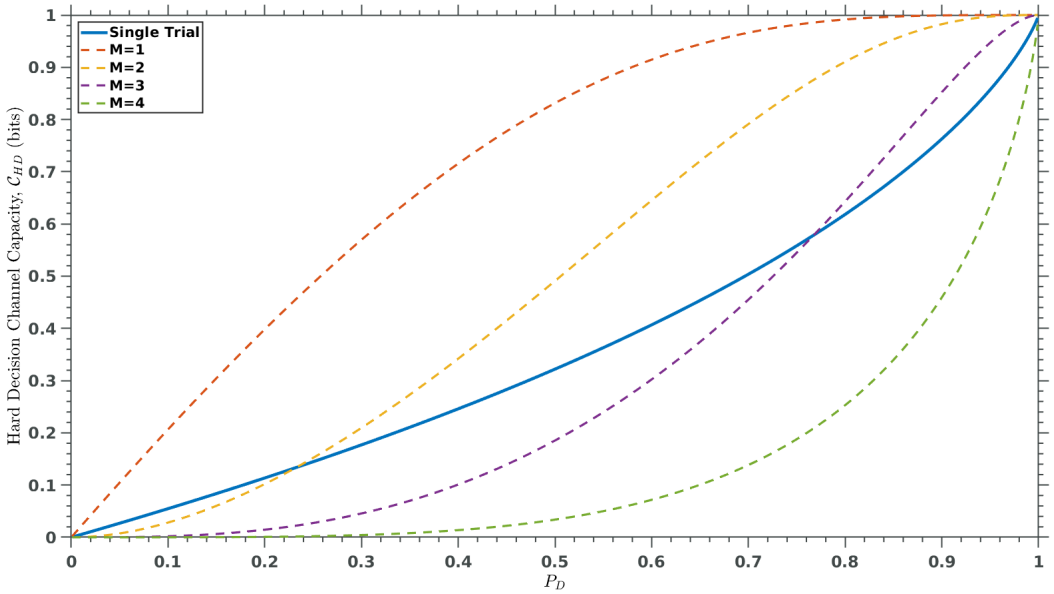

Figure 4.4: $\mathcal{C}_{HD}$ v. P_D for M of 4 Rule, $P_{FA} = 10^{-7}$

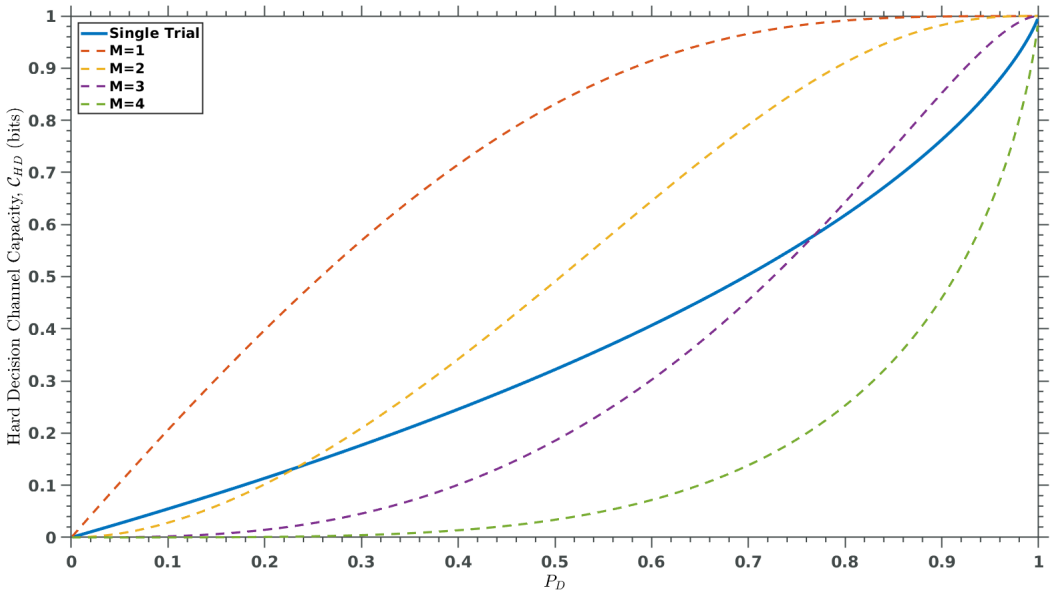

Figure 4.5: $\mathcal{C}_{HD}$ v. P_D for M of 4 Rule, $P_{FA} = 10^{-5}$

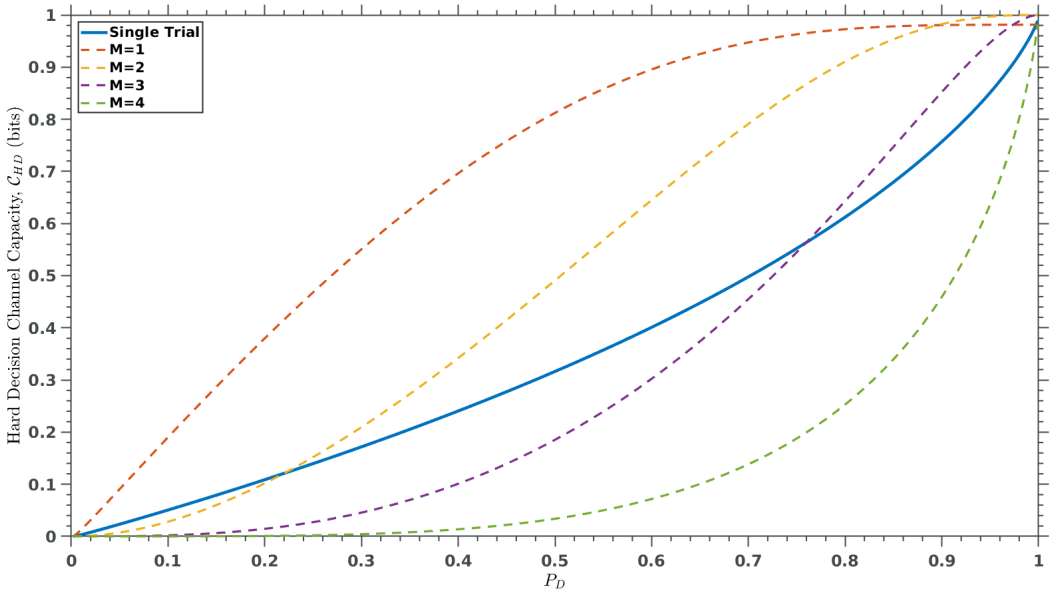

Figure 4.6: $\mathcal{C}_{HD}$ v. P_D for M of 4 Rule, $P_{FA} = 10^{-3}$

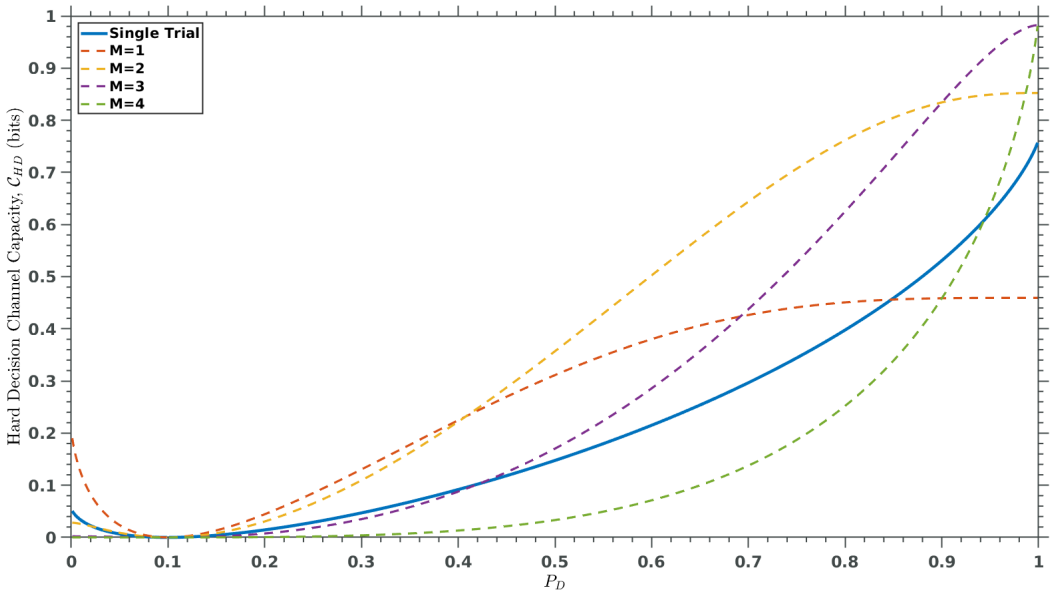

Figure 4.7: $\mathcal{C}_{HD}$ v. P_D for M of 4 Rule, $P_{FA} = 0.1$

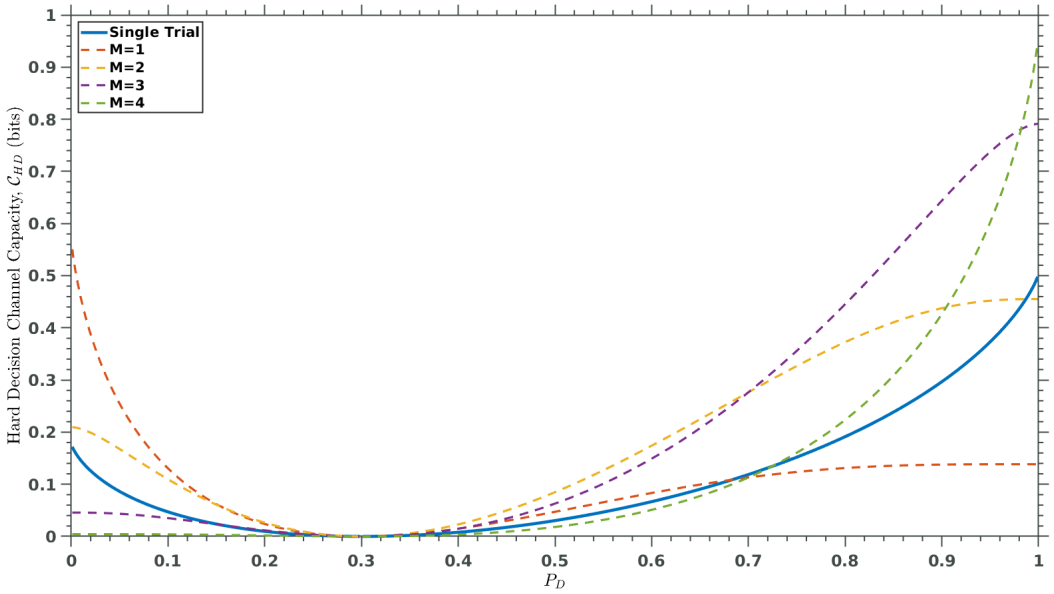

Figure 4.8: $\mathcal{C}_{HD}$ v. P_D for M of 4 Rule, $P_{FA} = 0.3$

4.3 Detection Capacity of Hard Decision Detectors without Integration

4.3.1 Constant, Known RCS in Known Noise with Known Phase

The simplest detector example is where the target reflection is a known, constant amplitude with known phase and known noise. Under the optimal likelihood ratio (LRT) test, the probability of detection is [5]

$$P_D = \frac{1}{2}\text{erfc}\left[\text{erfc}^{-1}(2P_{FA}) - \sqrt{\chi}\right] . \tag{4.13}$$

The detection capacity of a single receiver with no integration deciding only hard decisions is plotted in Figure 4.9 for SNR values of 13dB, 10dB, and 3dB.

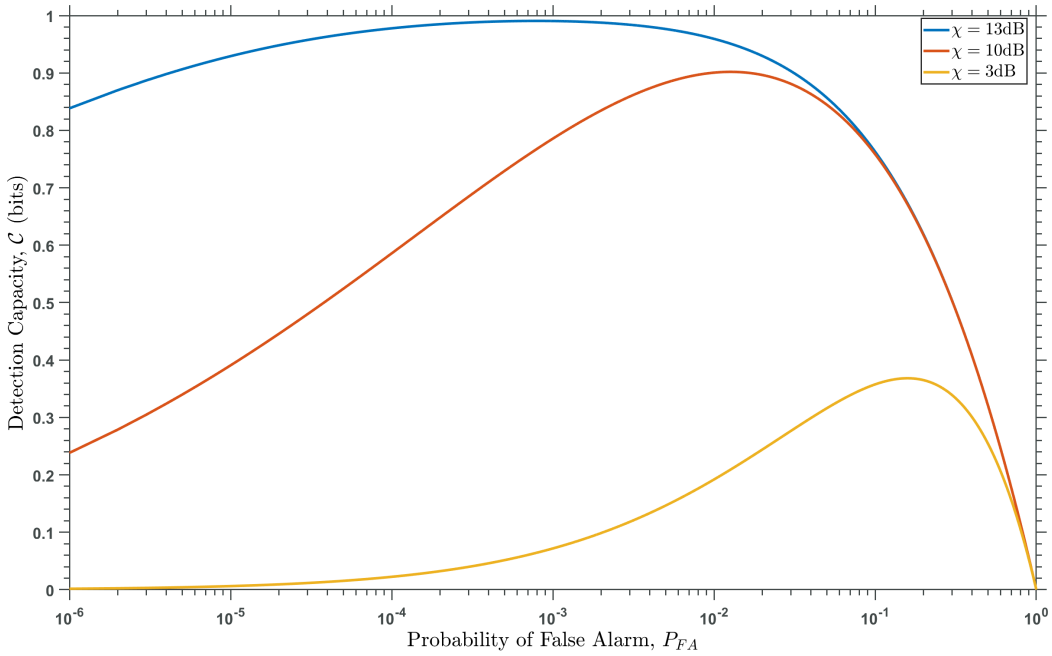

Figure 4.9: Detection capacity for constant, known RCS with known phase in known noise.

The detection capacity of a single receiver with no integration deciding only hard decisions is plotted against range in Figure 4.10 for P_{FA} values of 10^{-12}, 10^{-9}, and 10^{-6}. The transmit range is calculated with the radar characteristics provided in Table 2.1.

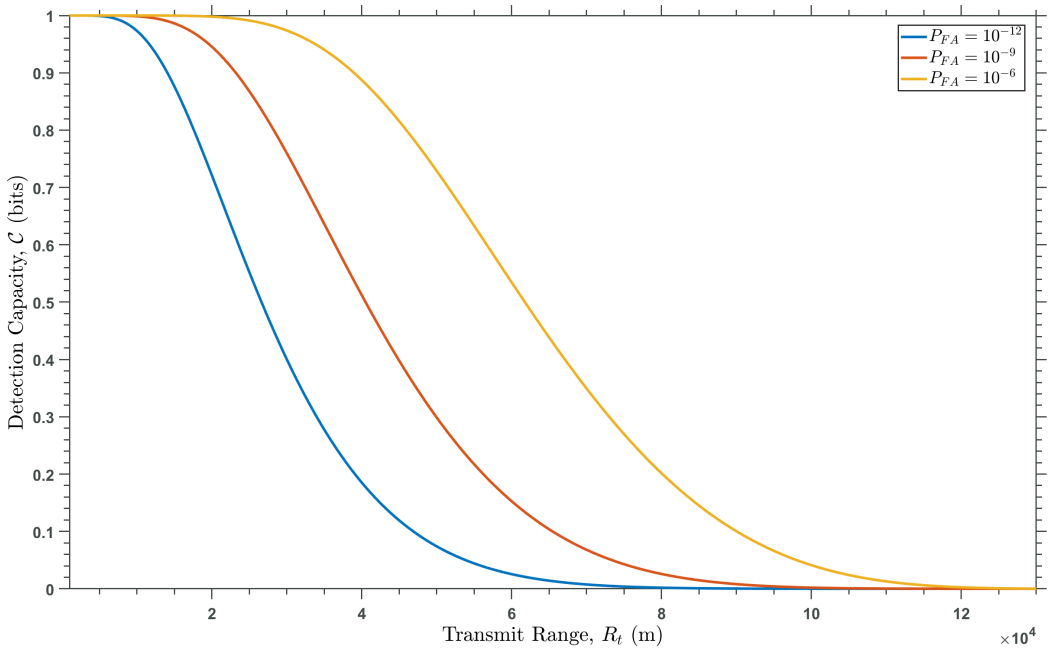

Figure 4.10: Detection capacity v. transmit range for constant, known RCS with known phase in known noise

We can see that as the transmit range increases, the detection capacity generally decreases as well. From Section 2.1, we know that as SNR decreases, transmit range is increased. It follows from Figure 4.9 that SNR decreasing aligns with lower detection capacity, which gives us an inverse relationship between transmit range and channel capacity.

4.3.2 Constant, Known RCS in Known Noise with Unknown Phase

To see the difference in detection capacity between detectors, we can find the channel capacity of a detector of having unknown or random phase. The optimum LRT envelope detector has a P_D of [5]

$$P_D = Q_M\left(\sqrt{2\chi}, \sqrt{-2\ln(P_{FA})}\right), \tag{4.14}$$

where $Q_M(\bullet)$ is known as Marcum's Q function, given by [5]

$$Q_M(\alpha, \gamma) = \int_{\gamma}^{+\infty} t\exp\left[-\frac{1}{2}(t^2+\alpha^2)\right] I_0(\alpha t)dt\,. \tag{4.15}$$

The difference between having known and unknown phase on P_D and P_{FA} for different SNR values can be seen in Figures 4.11, 4.12, and 4.13.

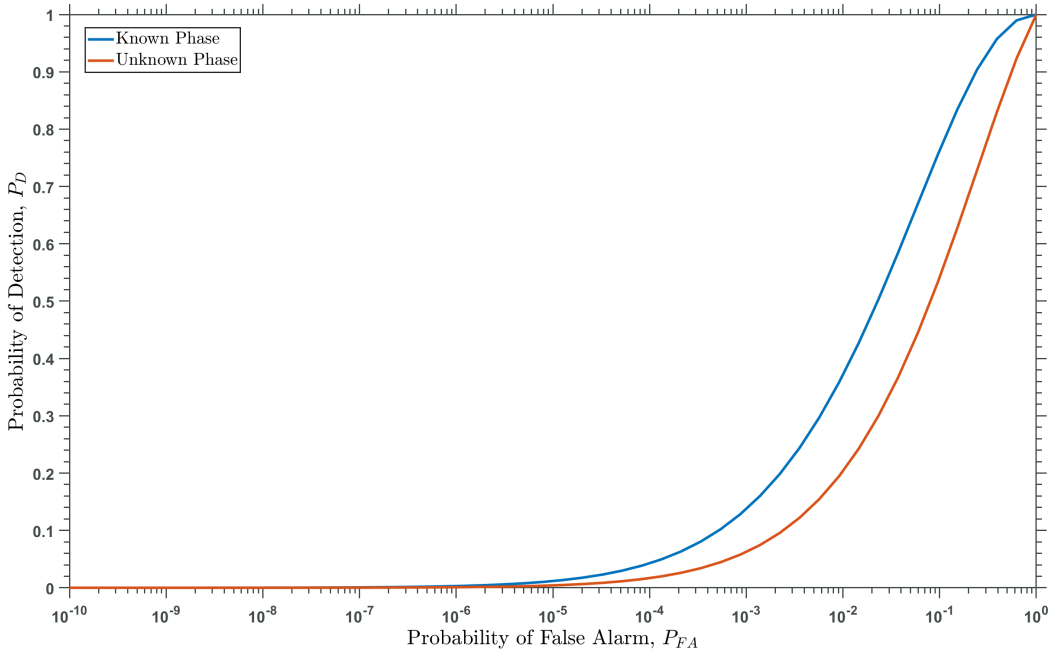

Figure 4.11: P_D v. P_{FA} for $\chi = 3$dB.

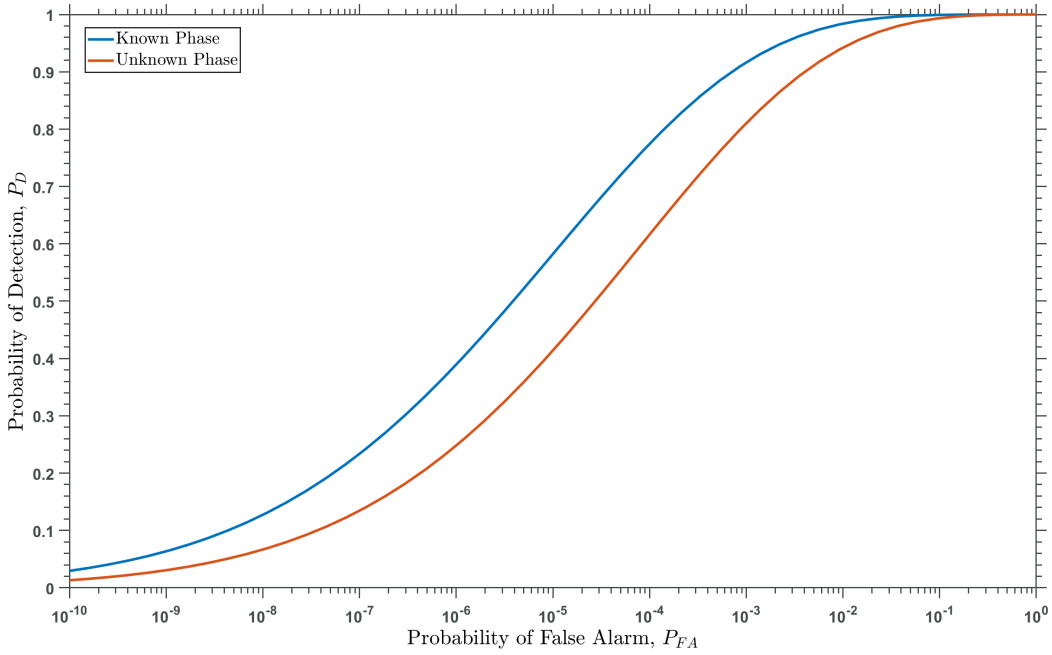

Figure 4.12: P_D v. P_{FA} for $\chi = 10$dB.

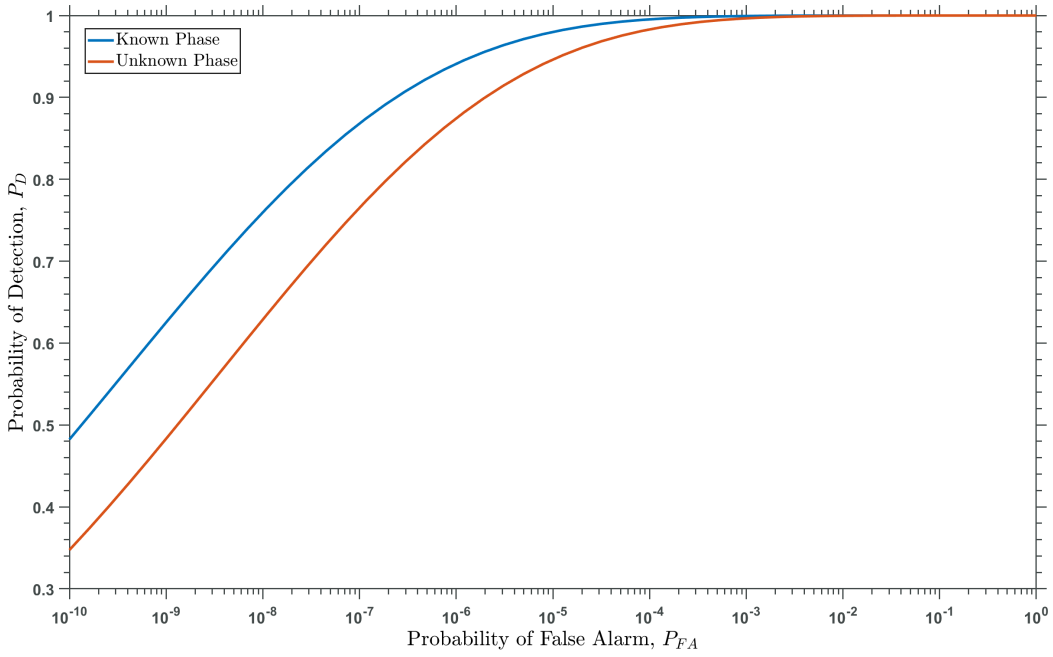

Figure 4.13: P_D v. P_{FA} for $\chi = 13$dB.

Furthermore, the SNR loss from having unknown phase is plotted for various P_{FA} values in Figures 4.14 and 4.15.

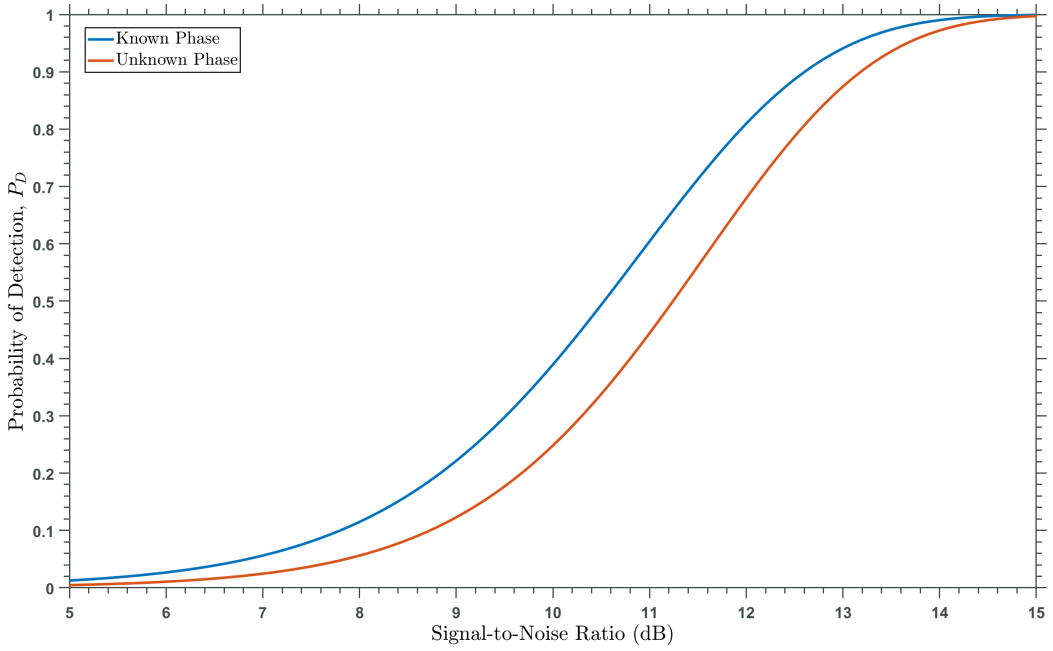

Figure 4.14: χ v. P_D for $P_{FA} = 10^{-6}$.

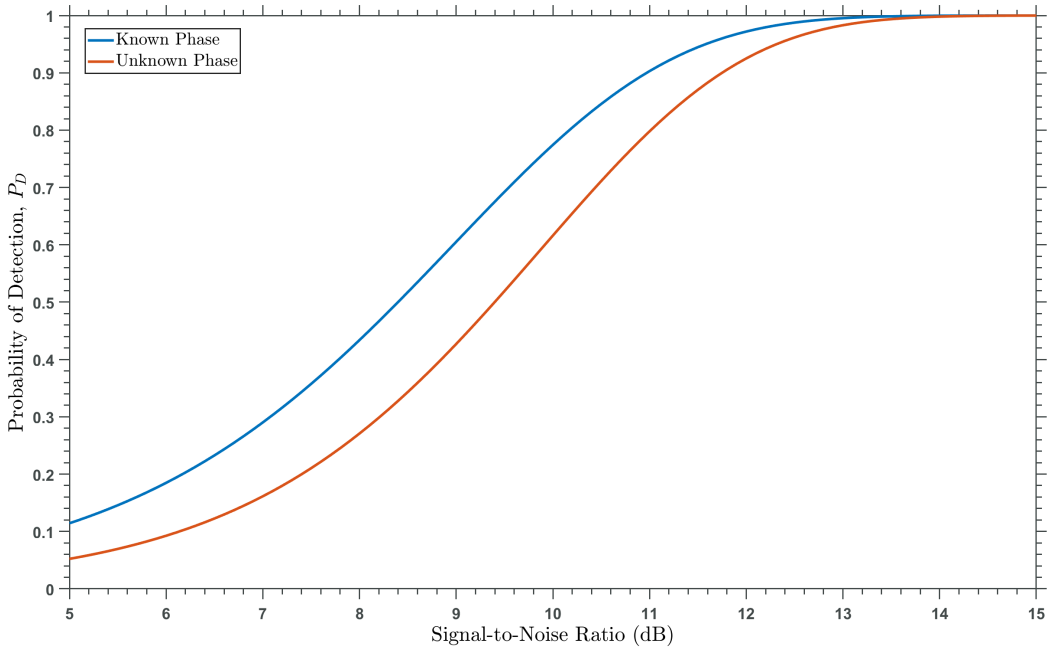

Figure 4.15: χ v. P_D for $P_{FA} = 10^{-4}$.

The detection capacity of a single receiver with no integration deciding only hard decisions is plotted in Figure 4.16.

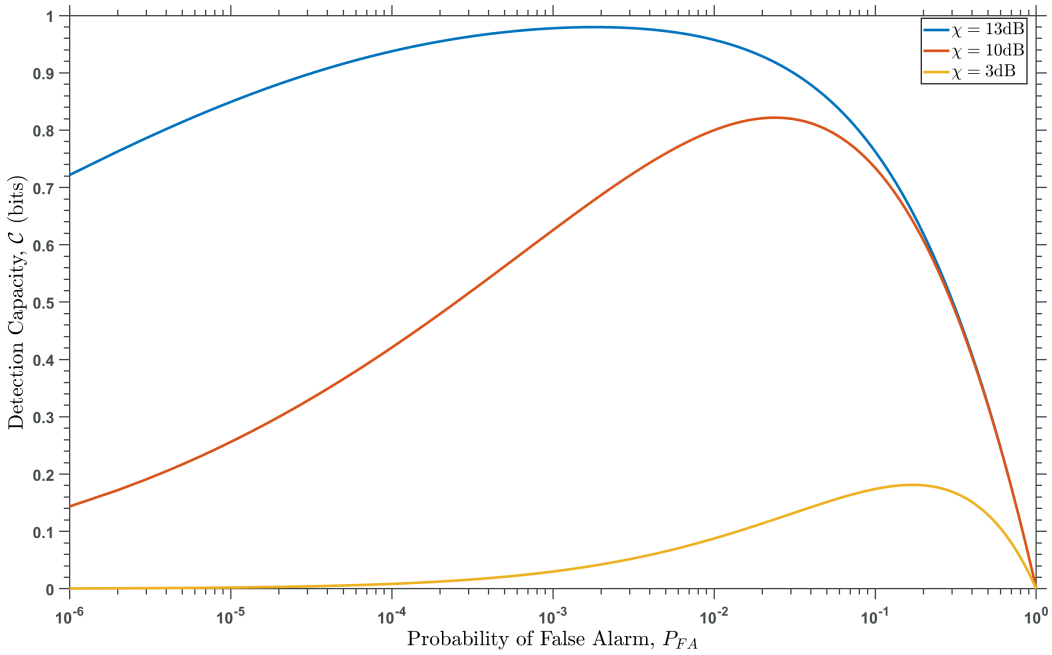

Figure 4.16: Detection capacity for constant, known RCS with unknown phase in known noise.

The detection capacity of a single receiver with no integration deciding only hard decisions is plotted against range in Figure 4.17 for P_{FA} values of 10^{-12}, 10^{-9}, and 10^{-6}.

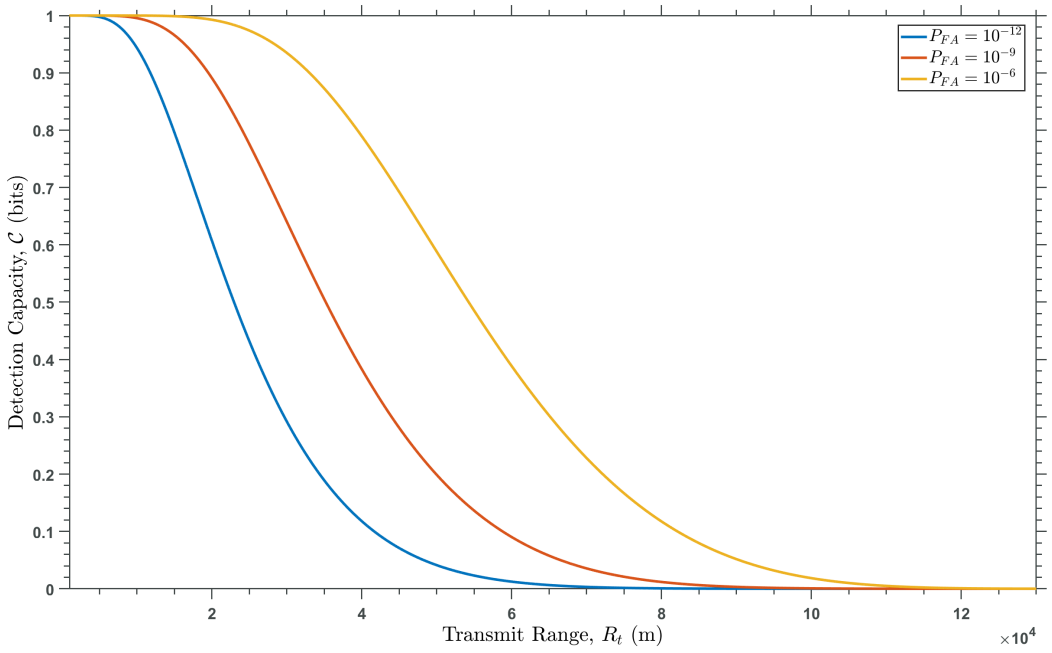

Figure 4.17: Detection capacity v. transmit range for constant, known RCS with unknown phase in known noise.

Comparing with Figure 4.9, we can see the y-intercept is lower than the case of having known phase. The introduction of a single random variable in detection process decreases the rate at which information can be transmitted without an arbitrarily low error. The difference in detection capacity between the two detectors is shown in Figure 4.18. We see that for higher SNR values, the information penalty is higher for near optimal detector performance ($P_{FA} \to 0$) while the information penalty is higher at worse detector performance ($P_{FA} \to 1$) for lower SNR values.

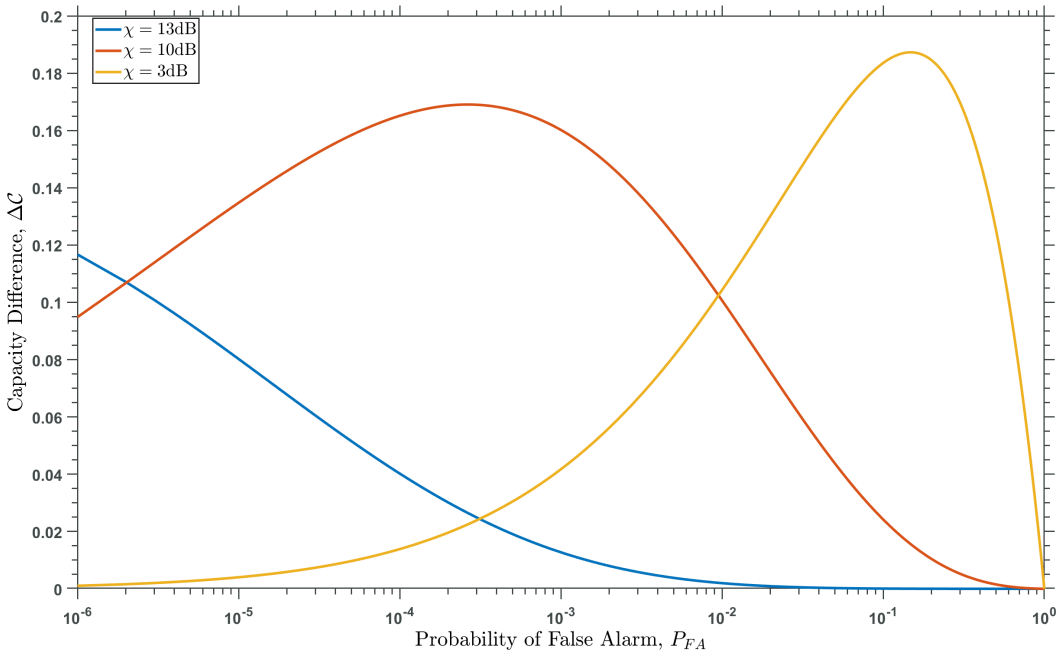

Figure 4.18: Difference in detection capacity, $\Delta\mathcal{C}$, for various SNR values.

4.4 Detection Capacity of Hard Decision Detectors with Integration

4.4.1 Constant, Known RCS in Known Noise with Known Phase

To find the detection capacity for a constant, known RCS in known noise detector using integration, we can plug the modified binary false alarm probability, Equation (4.11), into the expression for P_D in the constant, known RCS in known noise, Equation (4.13), to find the probability of detection with integration. Figures 4.19, 4.20, 4.21, and 4.22 show the effect of integration on detection capacity for M of N processing, plotting P_{FA} before integration against detection capacity.

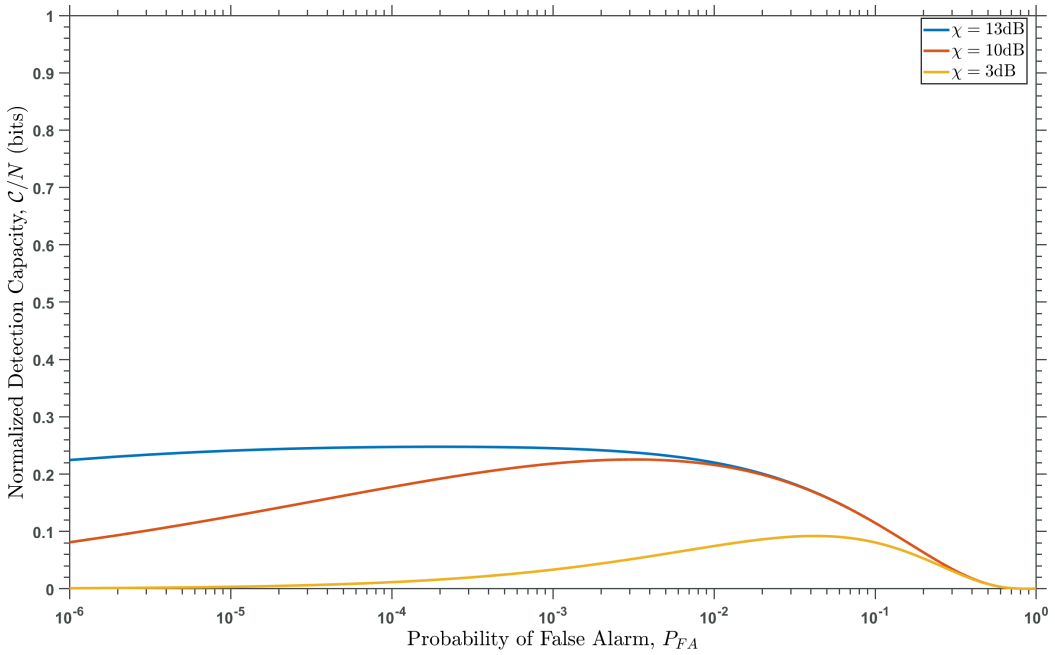

Figure 4.19: Detection capacity for constant, known RCS with known phase in known noise for 1 of N integration.

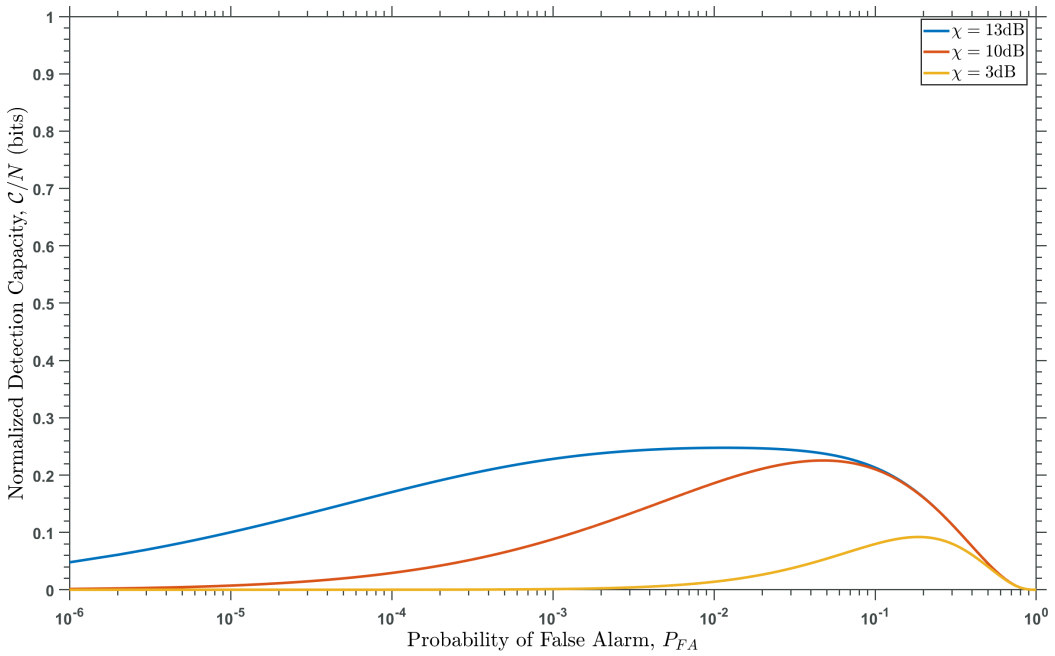

Figure 4.20: Detection capacity for constant, known RCS with known phase in known noise for 2 of N integration.

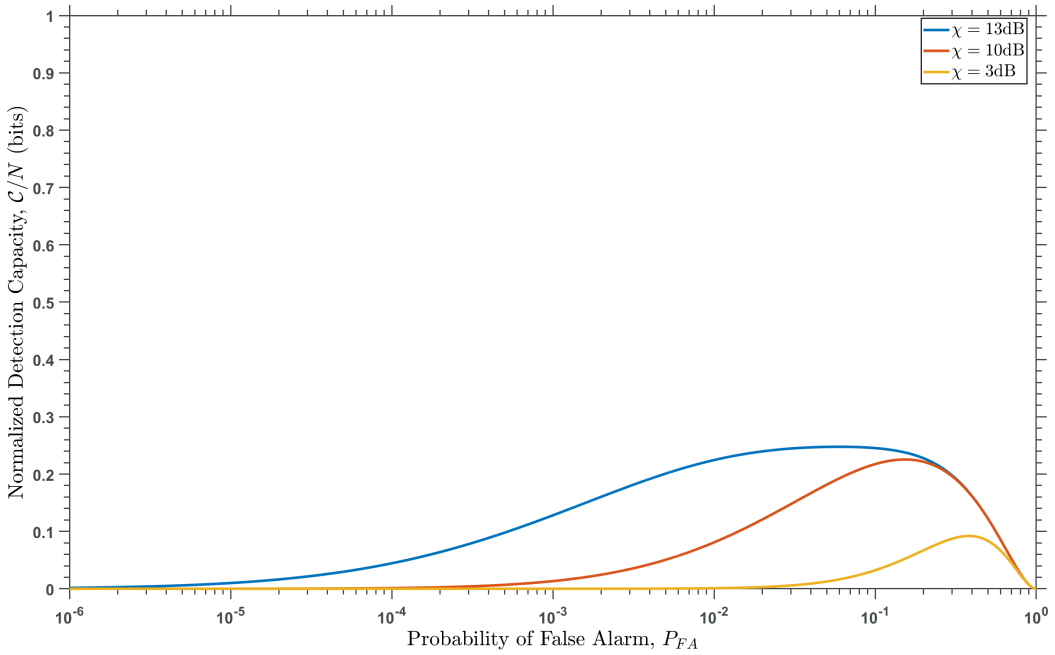

Figure 4.21: Detection capacity for constant, known RCS with known phase in known noise for 3 of N integration.

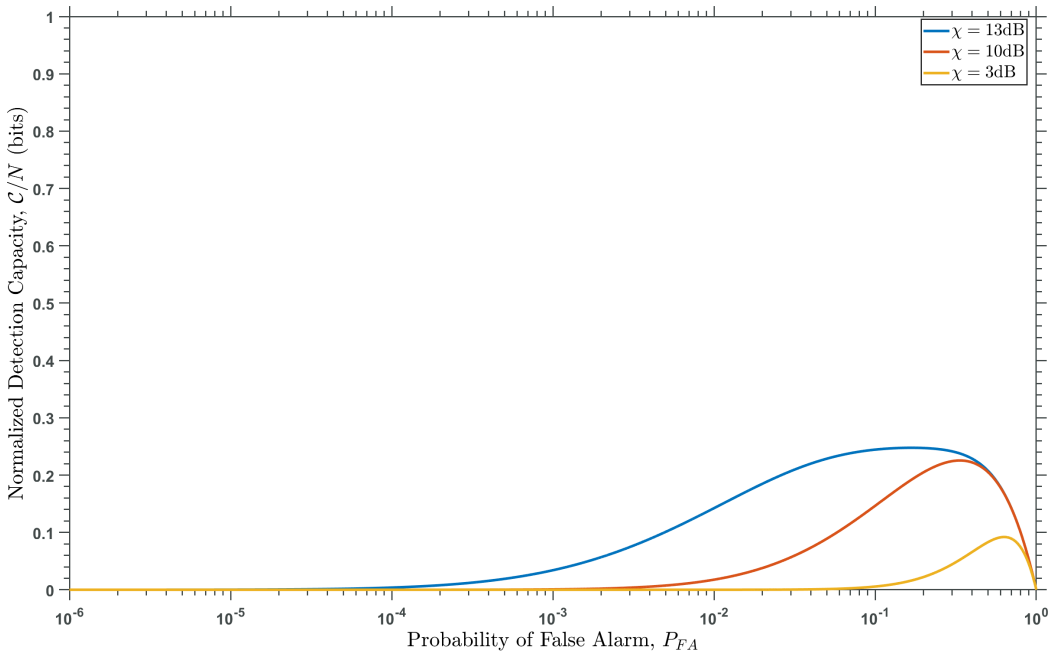

Figure 4.22: Detection capacity for constant, known RCS with known phase in known noise for 4 of N integration.

Plot 4.19 shows the 1 of 4 rule for a detector with constant, known RCS in known noise. We can see that the overall normalized channel capacity is high for smaller values of P_{FA}. This is due to the 1 of 4 rule increasing the effective P_D while also increasing P_{FA}, aligning with the fact that detection capacity depends primarily on P_D. Figures 4.20, 4.21, and 4.22 show the 2 of 4, 3 of 4, and 4 of 4 rules respectively. We can see that the channel capacity is significantly reduced at low P_{FA} as M increases. This is due to single trial P_D being relatively low for an actual detector, meaning P_D only increases at higher values of P_{FA}.

The detection capacity of a single receiver with integration deciding only hard decisions is plotted against range in Figures 4.23, 4.24, 4.25, and 4.26 for P_{FA} values of 10^{-12}, 10^{-9}, and 10^{-6}.

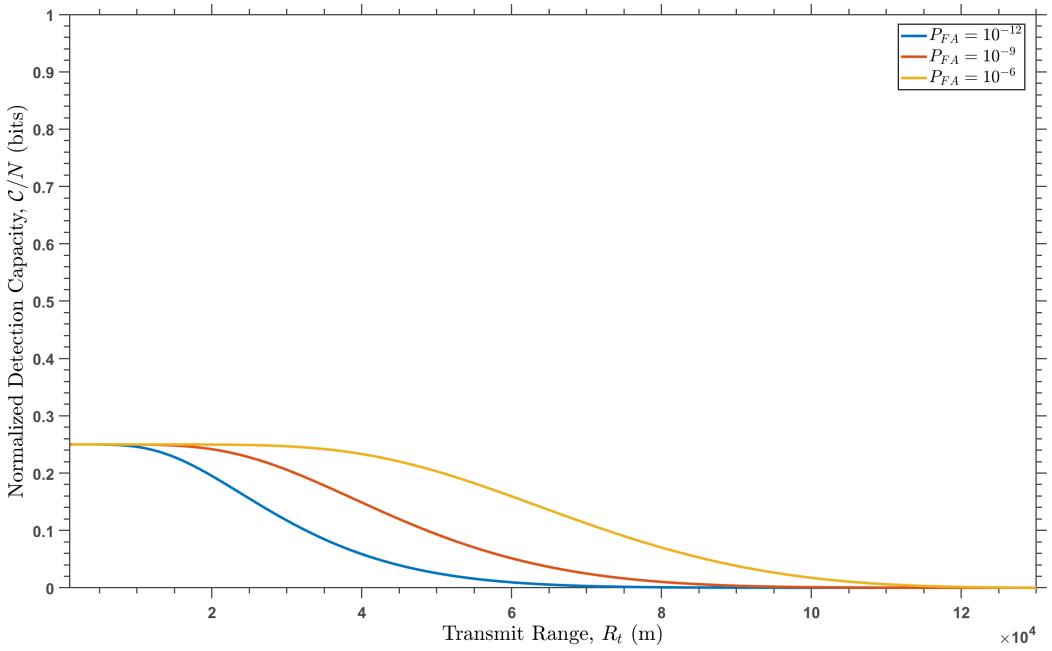

Figure 4.23: Detection capacity v. transmit range for constant, known RCS with known phase in known noise for 1 of N integration.

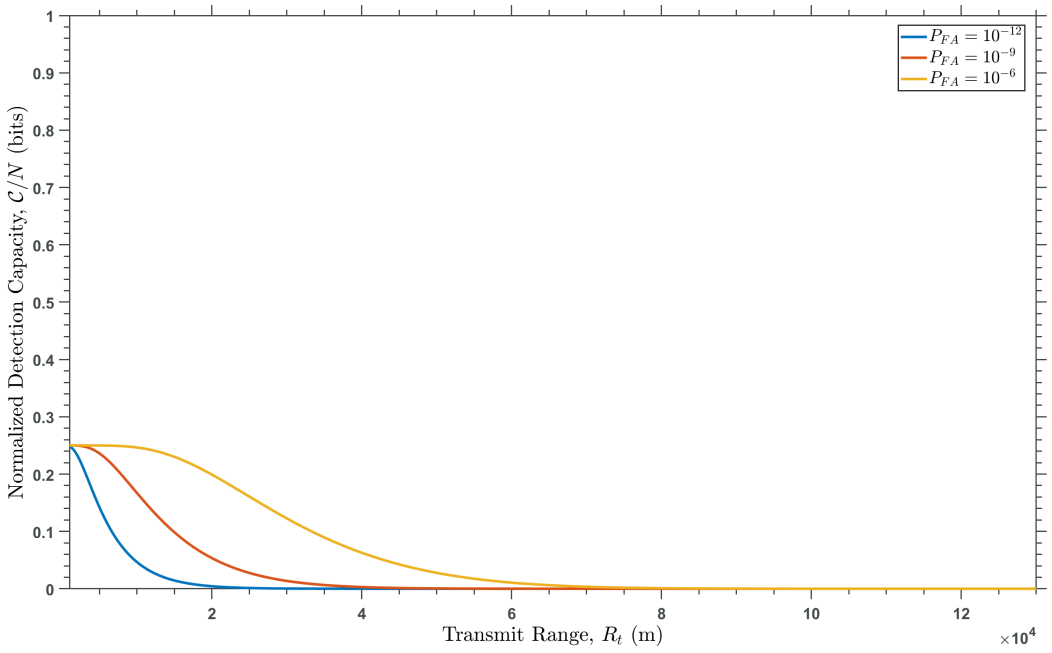

Figure 4.24: Detection capacity v. transmit range for constant, known RCS with known phase in known noise for 2 of N integration.

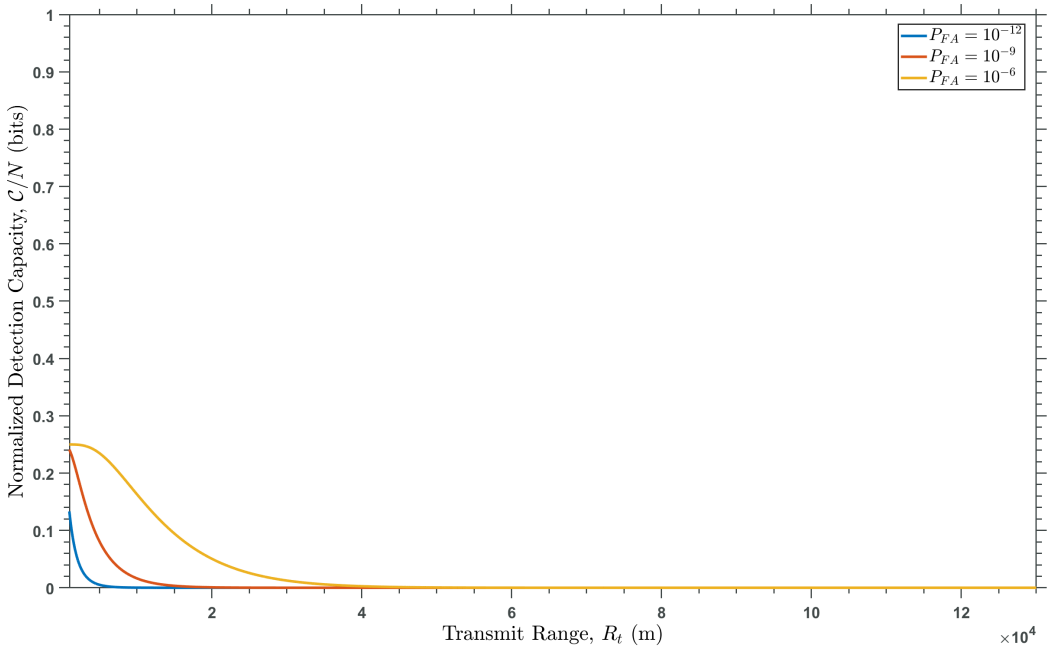

Figure 4.25: Detection capacity v. transmit range for constant, known RCS with known phase in known noise for 3 of N integration.

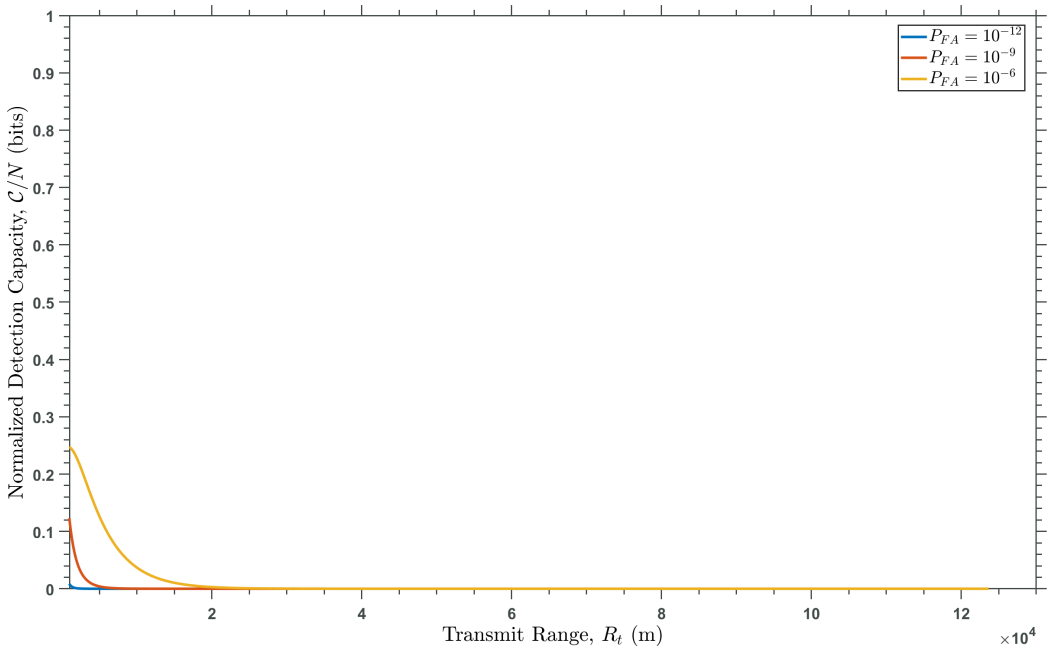

Figure 4.26: Detection capacity v. transmit range for constant, known RCS with known phase in known noise for 4 of N integration.

4.4.2 Constant, Known RCS in Known Noise with Unknown Phase

To find the detection capacity for a constant, known RCS with unknown phase in known noise detector using integration, we can plug the modified binary false alarm probability, Equation (4.11), into the expression for P_D in the constant, known RCS with unknown phase in known noise, Equation (4.14), to find the probability of detection with integration. Figures 4.27, 4.28, 4.29, and 4.30 show the effect of integration on detection capacity for M of N processing.

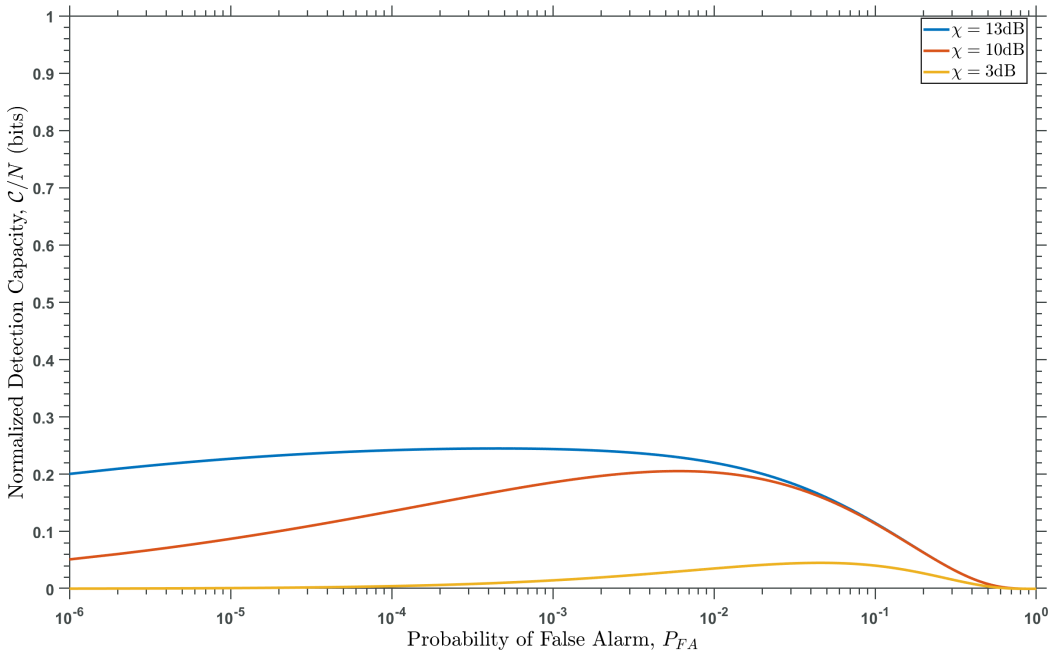

Figure 4.27: Detection capacity for constant, known RCS with unknown phase in known noise for 1 of N integration.

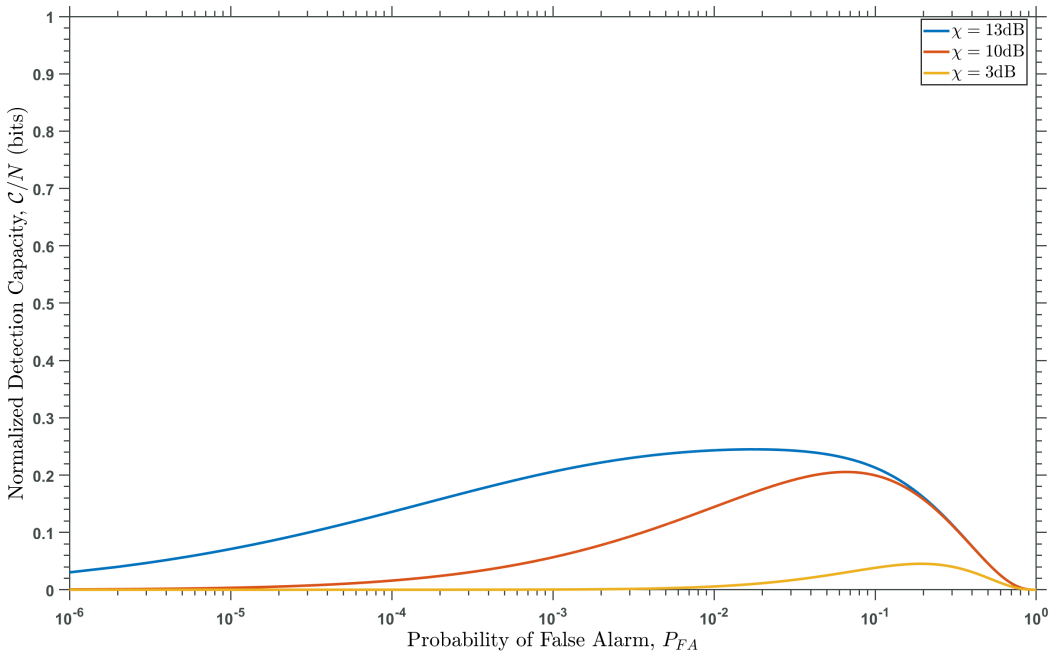

Figure 4.28: Detection capacity for constant, known RCS with unknown phase in known noise for 2 of N integration.

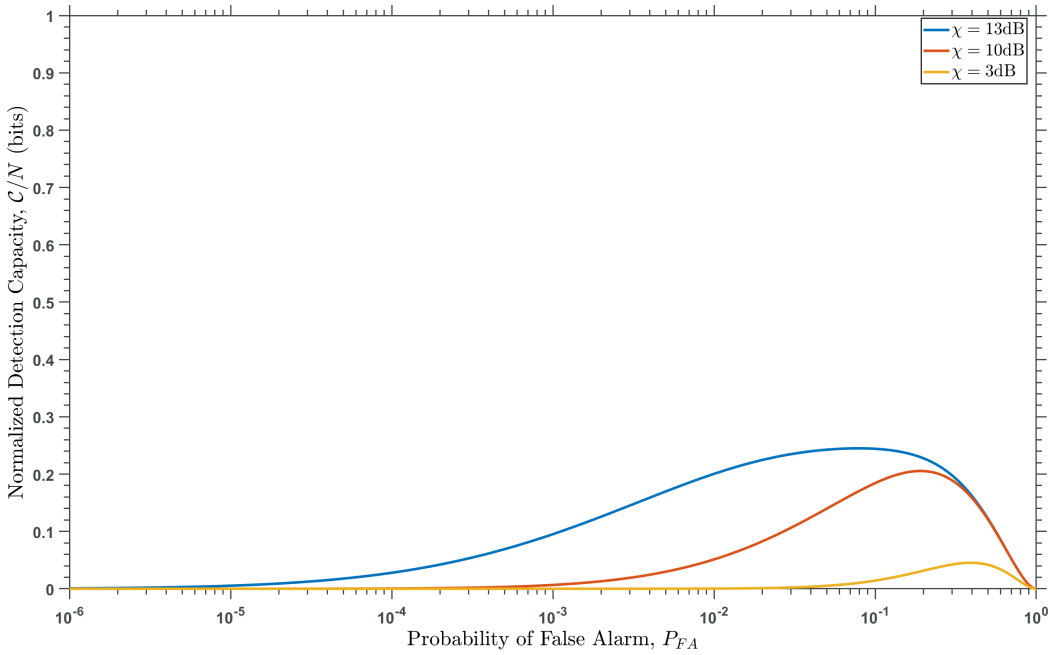

Figure 4.29: Detection capacity for constant, known RCS with unknown phase in known noise for 3 of N integration.

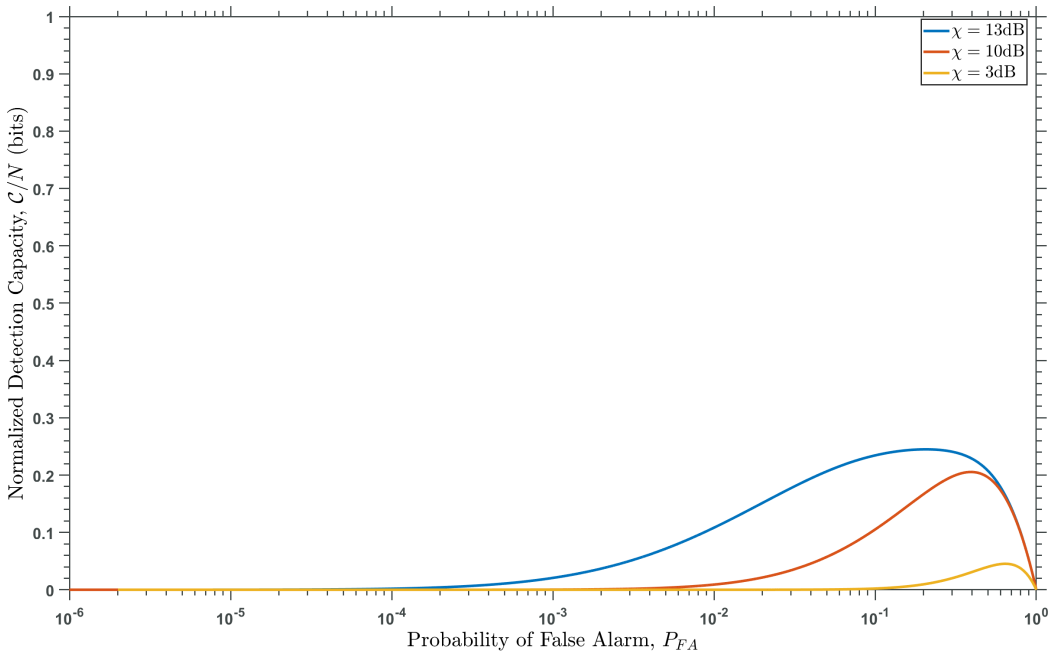

Figure 4.30: Detection capacity for constant, known RCS with unknown phase in known noise for 4 of N integration.

The case of constant, known RCS with unknown phase in known noise with integration follows the same trend as the case of constant, known noise in known noise with integration, provided in Section 4.4.1. The notable difference between the two detectors is a slightly lower detection capacity for case of unknown phase.

The detection capacity of a single receiver with integration deciding only hard decisions is plotted against range in Figures 4.31, 4.32, 4.33, and 4.34 for P_{FA} values of 10^{-12}, 10^{-9}, and 10^{-6}.

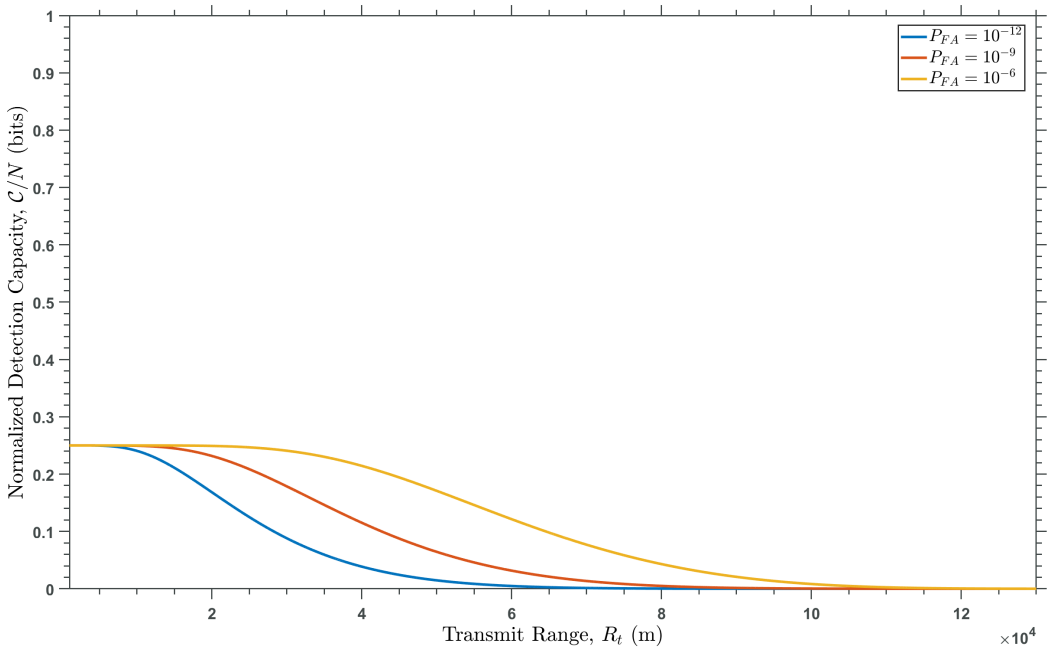

Figure 4.31: Detection capacity v. transmit range for constant, known RCS with unknown phase in known noise for 1 of N integration.

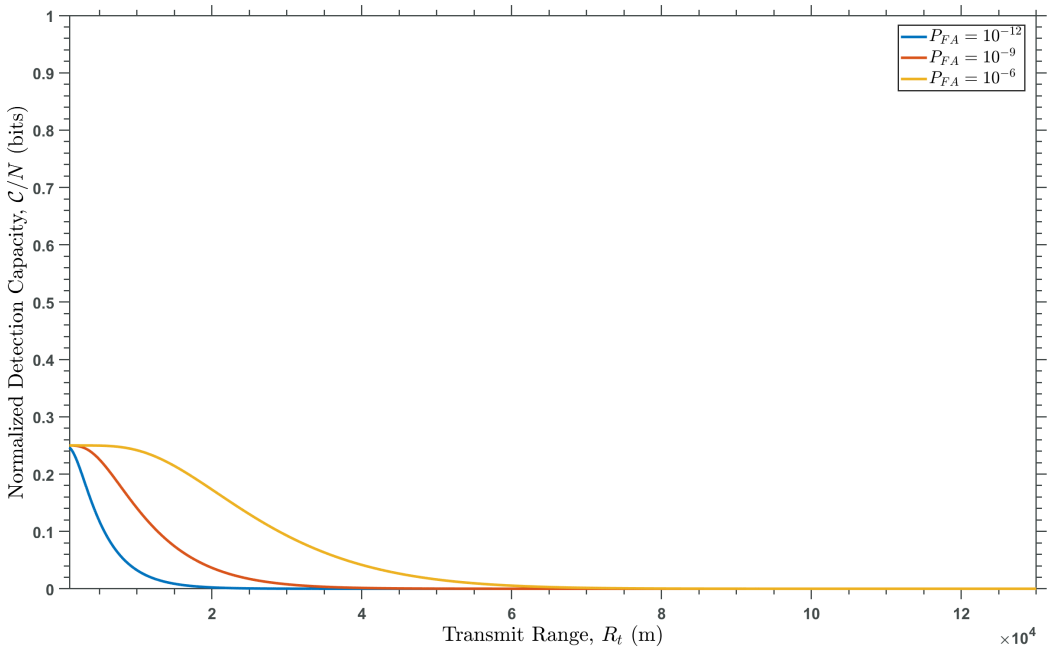

Figure 4.32: Detection capacity v. transmit range for constant, known RCS with unknown phase in known noise for 2 of N integration.

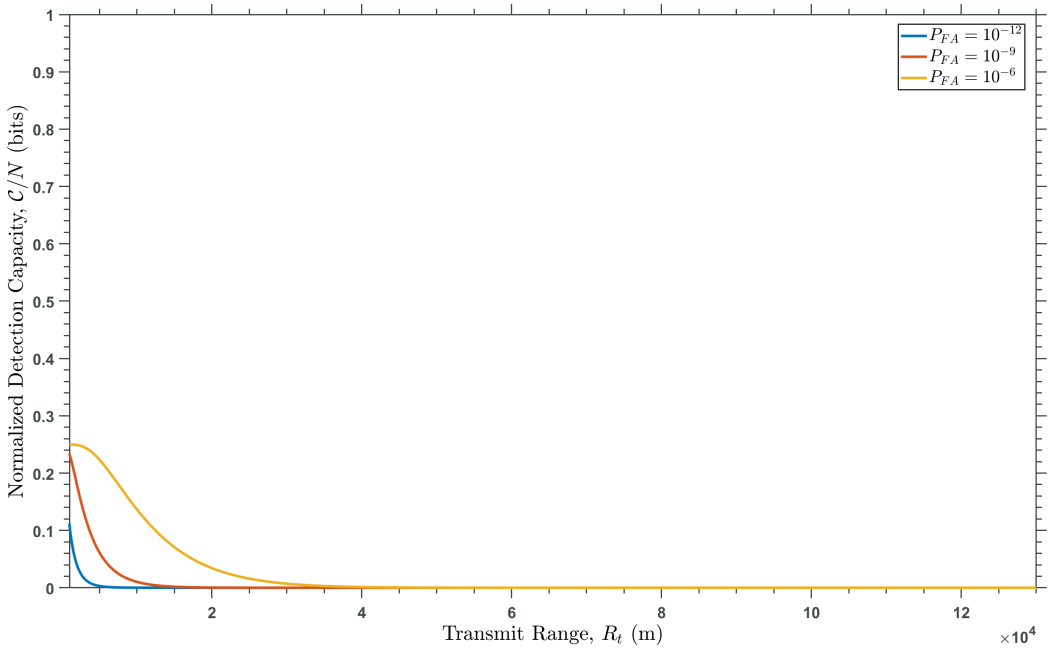

Figure 4.33: Detection capacity v. transmit range for constant, known RCS with unknown phase in known noise for 3 of N integration.

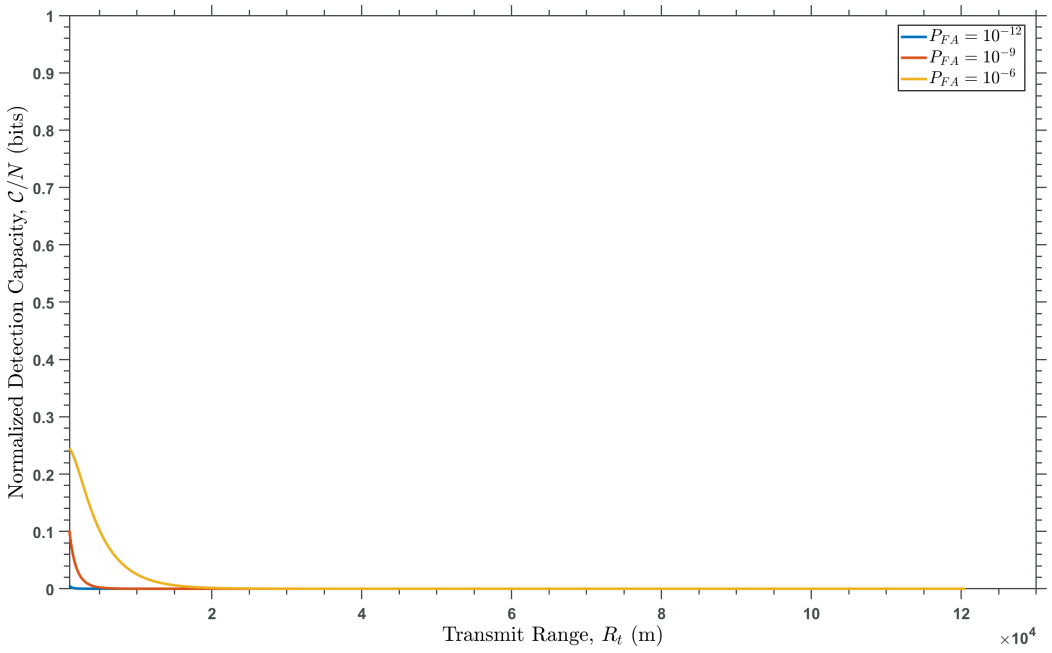

Figure 4.34: Detection capacity v. transmit range for constant, known RCS with unknown phase in known noise for 4 of N integration.

Again, the detection capacity is slightly lower than the previous detector but follows the same trends. This is due to the SNR being lower for this detector.

4.5 Detection Capacity of CFAR

From our elementary understanding of CFAR from Section 2.3.4, we can find the detection capacity for different CFAR assumptions such as differing CUT amount and threshold levels. To explore various detector and clutter assumptions of CFAR, we can refer to Table 4.3.

Table 4.3: Target category map [14].

Clutter	Constant Target	Fluctuating Target
No Clutter	I	II
Zero Mean Clutter	III	IV

The target parameters are provided in Table 4.4. Parameters N_t, X_t, and Y will be used

for the fixed threshold, no target fluctuation case. Adding target fluctuation will add target fluctuation parameter K. Changing from a fixed threshold to an adaptive threshold introduces N_r and replaces Y with α.

Table 4.4: Target parameters [14].

Target Parameter	Description
N_t	Number of target cells integrated
N_r	Total number of target-free cells averaged
X_t	Total signal level from all N_t target cells
Y	Fixed threshold, compared to X_t
α	Threshold parameter, replaces normalized threshold parameter Y, by αy, where y is a random variable representing a target free cell's response
K	Target fluctuation parameter

The primary target fluctuation model is based on the standard chi-square/Gamma density function with parameters X_t and K [14]

$$p(x|X_t, K) = \frac{x^{K-1}}{\Gamma(K)} \left(\frac{K}{X_t}\right)^K \exp(-Kx/X_t), \quad x \geq 0\,. \tag{4.16}$$

4.5.1 Category I: Constant Target, No Clutter

Before analyzing CFAR, we should first analyze the case in which we have a constant target with no clutter to compare between clutter/non-clutter target scenarios. The probability of detection and probability of false alarm for the case of category I, the constant target and

no clutter case, are [14]

$$P_{D,i}(N_t, X_t, Y) = \sum_{m=0}^{N_t-1} e^{-Y}\frac{Y^m}{m!} + \sum_{m=N_t}^{\infty} e^{-Y}\frac{Y^m}{m!} \times \left(1 - \sum_{i=0}^{m-N_t} e^{-X_t}\frac{(X_t)^i}{i!}\right), \tag{4.17}$$

$$P_{FA,i}(N_t, Y) = \sum_{m=0}^{N_t-1} e^{-Y}\frac{Y^m}{m!}. \tag{4.18}$$

The first term of Equation (4.17), $\sum_{m=0}^{N_t-1} e^{-Y}\frac{Y^m}{m!}$, is the noise term which will only change if clutter is introduced. The term, $\left(1 - \sum_{i=0}^{m-N_t} e^{-X_t}\frac{(X_t)^i}{i!}\right)$, is the target term, which will only change if target fluctuation is introduced. The probability of false alarm, Equation (4.18), is calculated by setting the total signal level, X_t, equal to zero.

Plugging in Equations (4.17) and (4.18) into Equation (4.7) allows us to find the detection capacity of a constant target, no clutter detector. Figure 4.35 shows the detection capacity for varying target cells given threshold $Y = 13$. We can see that higher total signal levels correspond to higher capacity peaks but lower to 0 at around $N_t = 25$. Detection capacity flattens out at $N_t = 25$ because the probability of false alarm gets too high due to the ratio of threshold and number of target cells. Another interesting trend that we can see is where each total signal level has peak capacity. For example, for a total signal level $X_t = 10$, the capacity peaks at around 10 target cells averaged. This is due to X_t again being the total received SNR from the samples containing the target, meaning lower SNR values require more target cells to be averaged to achieve higher probabilities of detection and to meet the threshold.

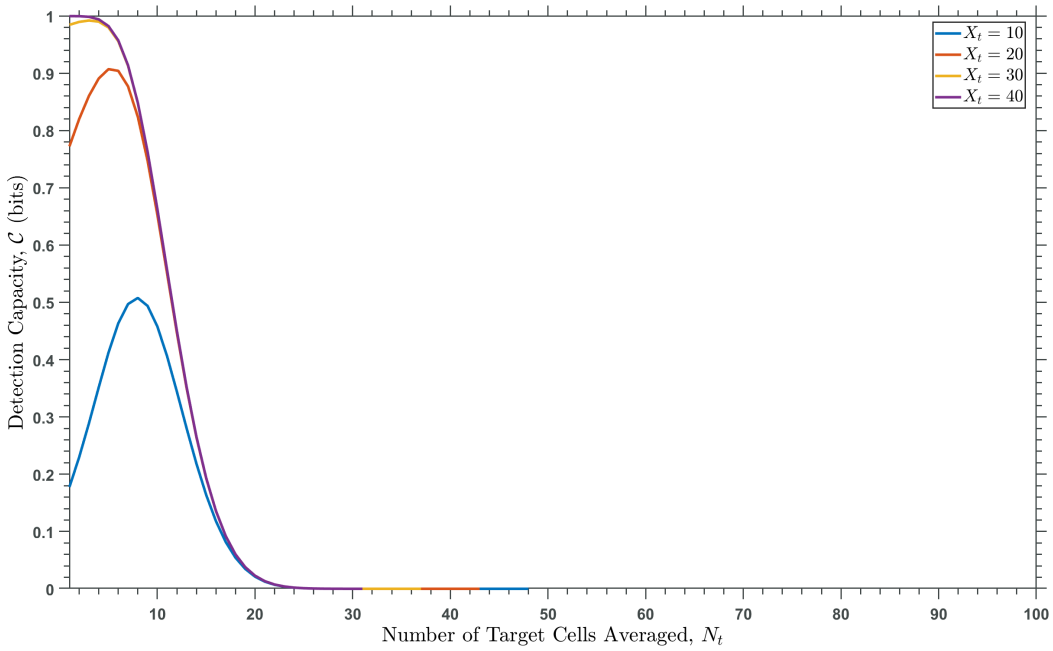

Figure 4.35: Detection capacity v. number of target cells averaged for $Y = 13$.

Figure 4.36 shows the detection capacity for varying thresholds for a total target cell of $N_t = 1$. Higher total signal levels stay at peak detection capacity for a higher threshold range due to the fact that it takes longer for the second term in Equation 4.17 to go towards 0.

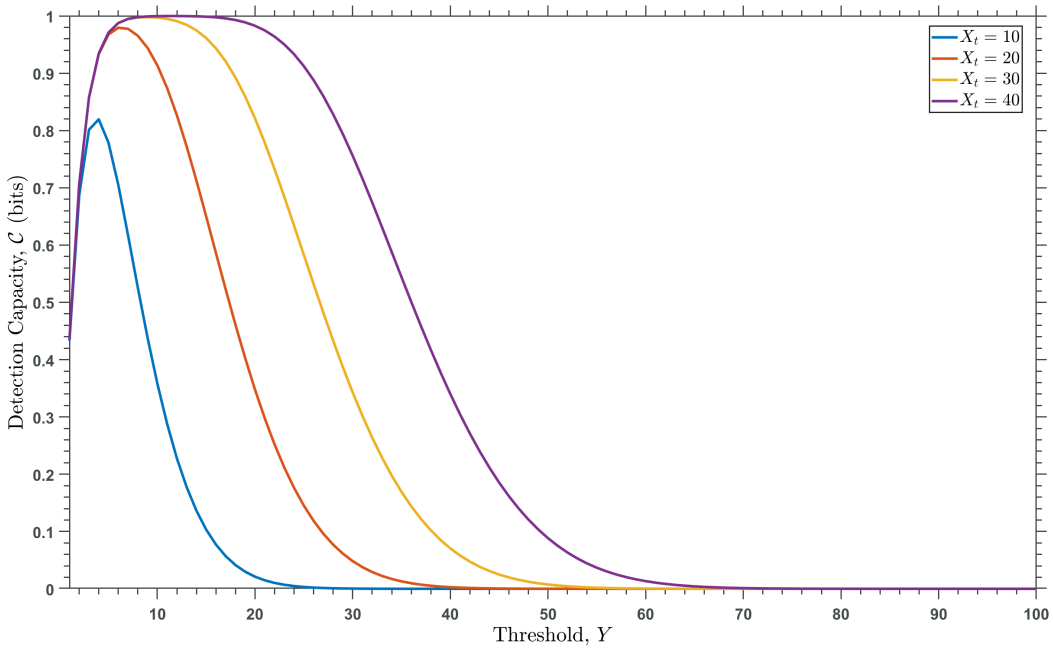

Figure 4.36: Detection capacity v. threshold for $N_t = 1$ and $\alpha = 0.148$.

4.5.2 Category II: Fluctuating Target, No Clutter

The probability of detection for the case of category II, fluctuating target and no clutter, is [14]

$$P_{D,ii}(N_t, X_t, Y, K) = \sum_{m=0}^{N_t-1} e^{-Y}\frac{Y^m}{m!} + \sum_{m=N_t}^{\infty} e^{-Y}\frac{Y^m}{m!} \times \left[1 - \sum_{i=0}^{m-N_t} \frac{\Gamma(k+i)}{i!\Gamma(K)} \left(\frac{1}{1+X_t/K}\right)^K \left(\frac{X_t/K}{1+X_t/K}\right)^i\right] . \tag{4.19}$$

Due to the probability of false alarm simply being the probability of detection with X_t set to zero, the probability of false alarm of category II simplifies to category I, provided by Equation (4.18). We can see that our target term has changed from adding fluctuation to $\left[1 - \sum_{i=0}^{m-N_t} \frac{\Gamma(k+i)}{i!\Gamma(K)} \left(\frac{1}{1+X_t/K}\right)^K \left(\frac{X_t/K}{1+X_t/K}\right)^i\right]$ but the first term has stayed the same due to the lack of clutter. Substituting Equations (4.19) and (4.18) into Equation (4.7) will give us a detection capacity expression reliant on N_t, X_t, Y, and fluctuation K. The four special cases

of the model, $K = 1$, N_t, 2, and $2N_t$ are the Swerling fluctuation model cases of I, II, III, and IV respectively. A summary of the different Swerling models is provided in Table 4.5.

Table 4.5: Summary of Swerling models [4].

Swerling Model	Description
I	Echo pulses received on any one scan are of constant amplitude throughout the entire scan but are independent (uncorrelated) from scan-to-scan.
II	Fluctuations are independent from pulse to pulse rather than scan to scan.
III	Constant RCS within a scan and independent from scan to scan like case I but aims to represent targets that can be modeled together as one large scatterer together with a number of small scatterers.
IV	Pulse to pulse fluctuation but aims to represent the same types of targets as case III.

Figure 4.37 shows the effect of fluctuating targets on detection capacity. Detection capacity v. total signal power is plotted for $N_t = 3$ and $Y = 13$ for various fluctuation values. From the plots, we can see that the Swerling models that are independent from pulse-to-pulse (Swerling models II and IV) have the highest detection capacity relative to their scan to scan counterparts (Swerling models I and III). This high channel capacity is due to the two Swerling models having higher fluctuation/K values than the other two models. As $K \to \infty$, the fluctuating target simplifies to a constant target due to the number of target cells increasing. Furthermore, we see that targets that can be modeled together as one large scatterer will some small scatterers (Swerling models III and IV) have higher capacities than scatterers that are modeled as many small and independent scatters (Swerling models I and II). This is due to less fluctuation and information required for the larger scatter, so there's a smaller chance of error, giving a smaller information penalty.

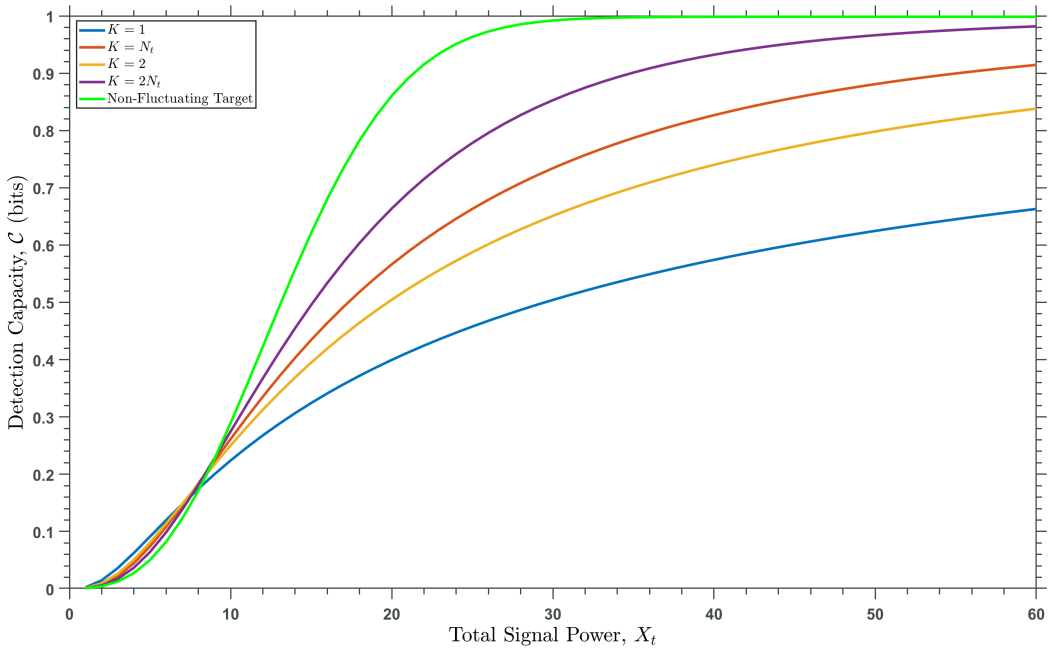

Figure 4.37: Detection capacity v. total signal power for $N_t = 3$ and $Y = 13$.

4.5.3 Category III: Constant Target, Zero Mean Clutter

The probability of detection and probability of false alarm for the case of category III, the constant radar cross section and CFAR case with zero mean clutter, are [14]

$$\begin{aligned} P_{D,iii}(N_t, N_r, X_t, \alpha) = &\sum_{m=0}^{N_t-1} \frac{(N_r+m-1)!}{(N_r-1)!m!} \left(\frac{1}{1+\alpha}\right)^{N_r} \left(\frac{\alpha}{1+\alpha}\right)^{m} \\ &+ \sum_{m=N_t}^{\infty} \frac{(N_r+m-1)!}{(N_r-1)!m!} \left(\frac{1}{1+\alpha}\right)^{N_r} \left(\frac{\alpha}{1+\alpha}\right)^{m} \\ &\times \left(1 - \sum_{i=0}^{m-N_t} e^{-X_t} \frac{(X_t)^i}{i!}\right), \end{aligned} \tag{4.20}$$

$$P_{FA,iii}(N_t, N_r, \alpha) = \sum_{m=0}^{N_t-1} \frac{(N_r+m-1)!}{(N_r-1)!m!} \left(\frac{1}{1+\alpha}\right)^{N_r} \left(\frac{\alpha}{1+\alpha}\right)^{m}. \tag{4.21}$$

We can see that the target term, $\left(1 - \sum_{i=0}^{m-N_t} e^{-X_t} \frac{(X_t)^i}{i!}\right)$ is the same as the target term in category I, as there is no fluctuation. The noise term has changed from category I/II

due to the addition of clutter on the model. Furthermore, we can see that our fixed threshold parameter Y has been changed to an adaptive threshold α, as described in Table 4.4. Substituting Equations (4.20) and (4.21) into Equation (4.7) gives us a detection capacity expression dependent on the number of target cells integrated N_t, total signal level from all target cells X_t, number of clutter cells averaged N_r, and the threshold parameter α.

Figure 4.38 shows the relationship between the number of clutter cells averaged and detection capacity for threshold $\alpha = 0.148$ and target cells averaged $N_t = 1$. We can see high total signal level X_t and number of clutter cells averaged correspond with the highest detection capacity. This is due to the accuracy increasing, as P_{FA} gets smaller as the number of clutter cells averaged increases for a set threshold with a high total signal level. Moreover, we can see that for lower total signal levels from the target cells such as $X_t = 10$ and $X_t = 20$, there seems to be diminishing returns as the cells could be filled with clutter only.

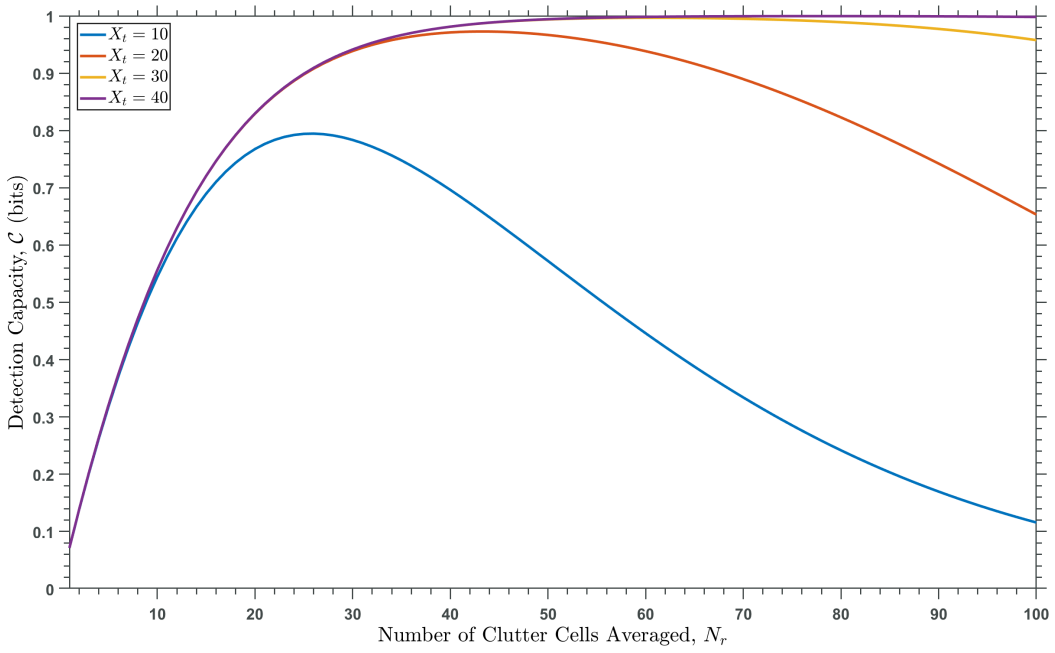

Figure 4.38: Detection capacity v. number of clutter cells averaged for $N_t = 1$ and $\alpha = 0.148$.

Figure 4.39 shows the relationship between the number of target cells averaged and detection capacity for threshold $\alpha = 0.148$ and clutter cells averaged $N_r = 100$. Again, we

see that high total signal level corresponds to higher detection capacity but it seems that having higher target cells averaged reduces detection capacity significantly. This is due to the increase of P_{FA} that having a N_t to N_r ratio causes for a set threshold α. When target cells increase relative to clutter cells, it follows that there would be a higher probability of error, or false alarm, as there are more cells to possibly flag as a target when it is only clutter.

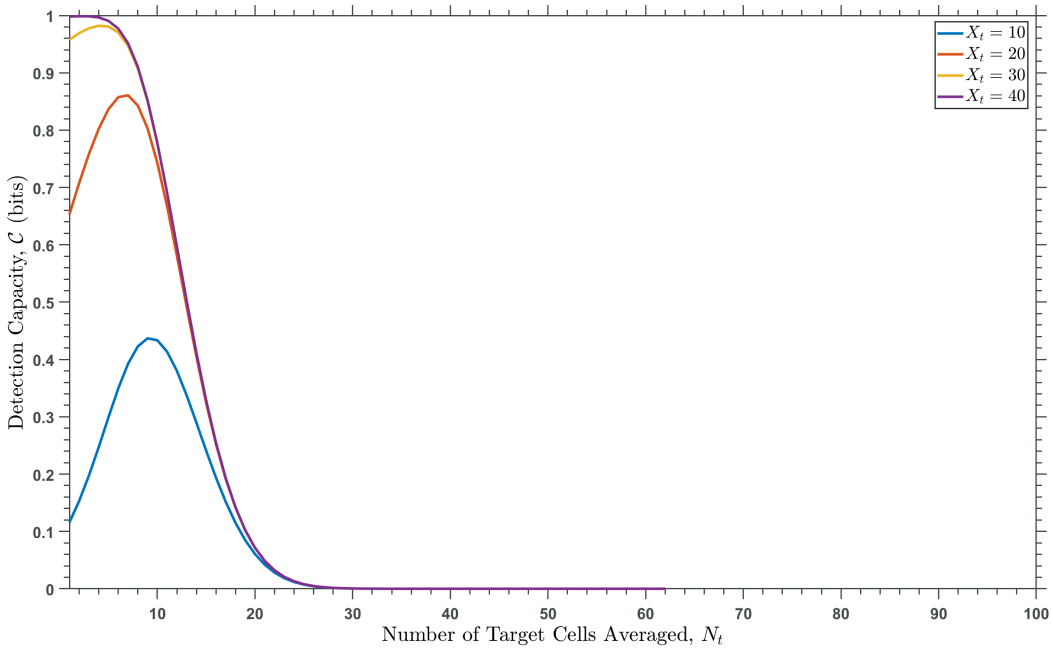

Figure 4.39: Detection capacity v. number of target cells averaged for $N_r = 100$ and $\alpha = 0.148$.

4.5.4 Category IV: Fluctuating Target, Zero Mean Clutter

For the chi-square fluctuating target case of CFAR with zero mean clutter, we can express the probability of detection as [14]

$$\begin{aligned} P_{D,iv}(N_t, X_t, N_r, \alpha, K) = &\sum_{m=0}^{N_t-1} \frac{(N_r+m-1)!}{(N_r-1)!m!} \left(\frac{1}{1+\alpha}\right)^{N_r} \left(\frac{\alpha}{1+\alpha}\right)^{m} \\ &+ \sum_{m=N_t}^{\infty} \frac{(N_r+m-1)!}{(N_r-1)!m!} \left(\frac{1}{1+\alpha}\right)^{N_r} \left(\frac{\alpha}{1+\alpha}\right)^{m} \\ &\times \left[1 - \sum_{i=0}^{m-N_t} \frac{\Gamma(K+i)}{i!\Gamma(K)} \left(\frac{1}{1+X_t/K}\right)^{K} \left(\frac{X_t/K}{1+X_t/K}\right)^{i}\right]. \end{aligned} \tag{4.22}$$

Due to the probability of false alarm being derived by setting X_t to zero in Equation (4.22), it follows that the probability of false alarm will be the same as the case of a non fluctuating target, given by Equation (4.21). Similarly, the detection capacity can be calculated by substituting Equations (4.22) and (4.21) into Equation (4.7). This gives us a similar expression as before, but introduces the target fluctuation parameter, K.

To compare between fluctuating and constant targets for CFAR, Figures 4.40, 4.41, 4.42, and 4.43 show the detection capacity for various total clutter cells for Swerling models I, II, III, and IV with the capacity of a non-fluctuating target shown for $N_t = 3$ and $\alpha = 0.148$. The fluctuating and constant target models for CFAR follow the same trends as categories I and II. Swerling models II and IV again have the highest channel capacity relative to the non-fluctuating target due to the high K value. Again, there is an information penalty for adding noise and fluctuation, which agrees with our intuition as more uncertainty increases the information penalty as we saw in the previous sections with the addition of random phase.

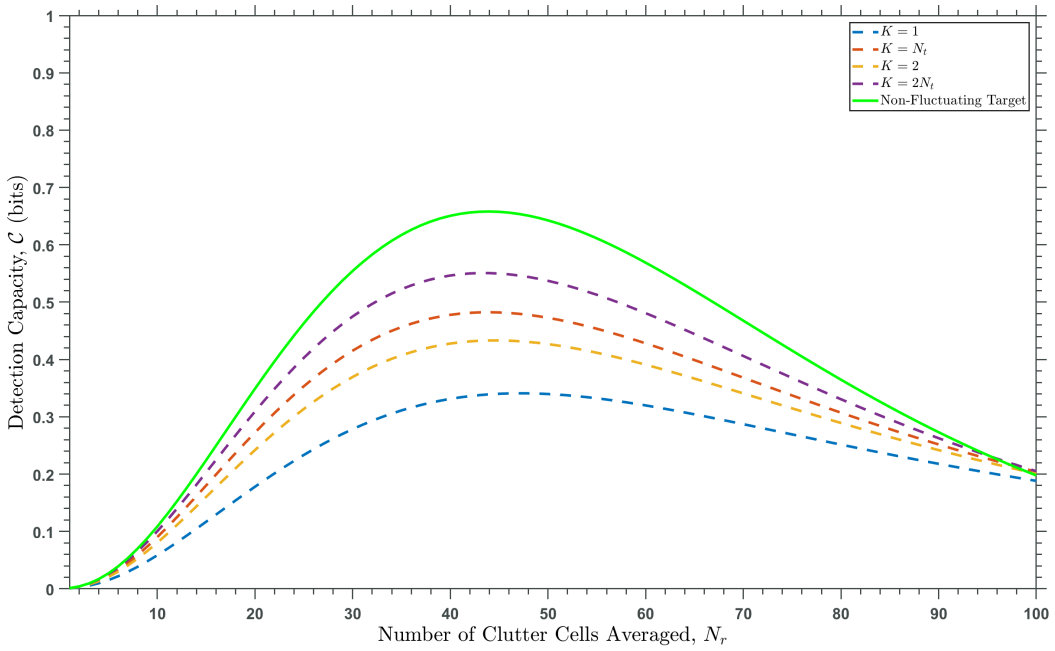

Figure 4.40: Detection capacity v. number of clutter cells averaged for $N_t = 3$, $\alpha = 0.148$, and $X_t = 10$.

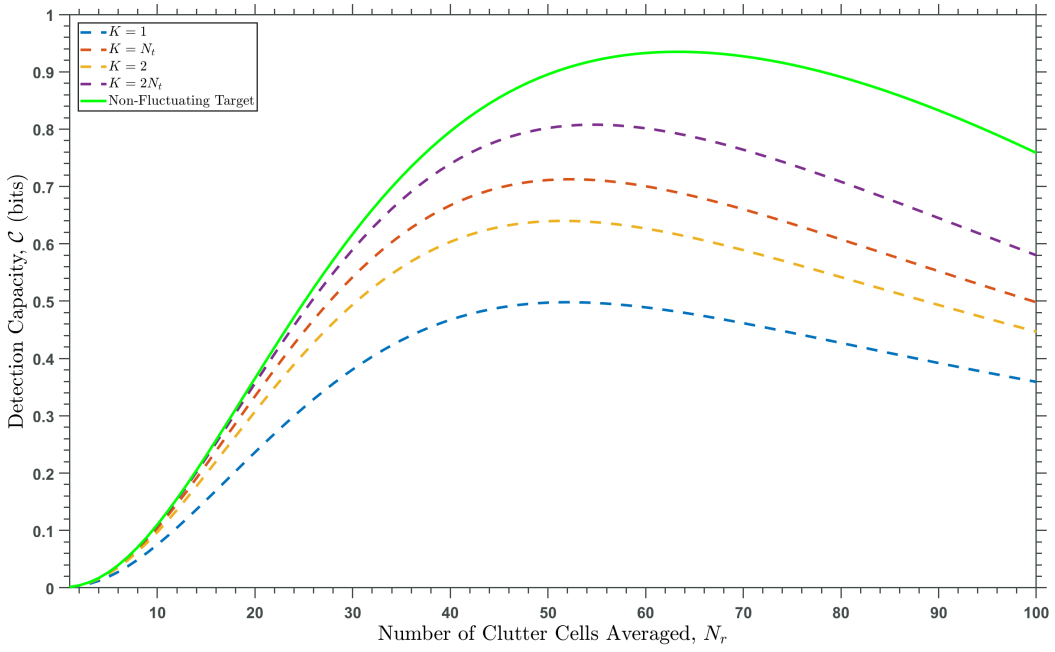

Figure 4.41: Detection capacity v. number of clutter cells averaged for $N_t = 3$, $\alpha = 0.148$, and $X_t = 20$.

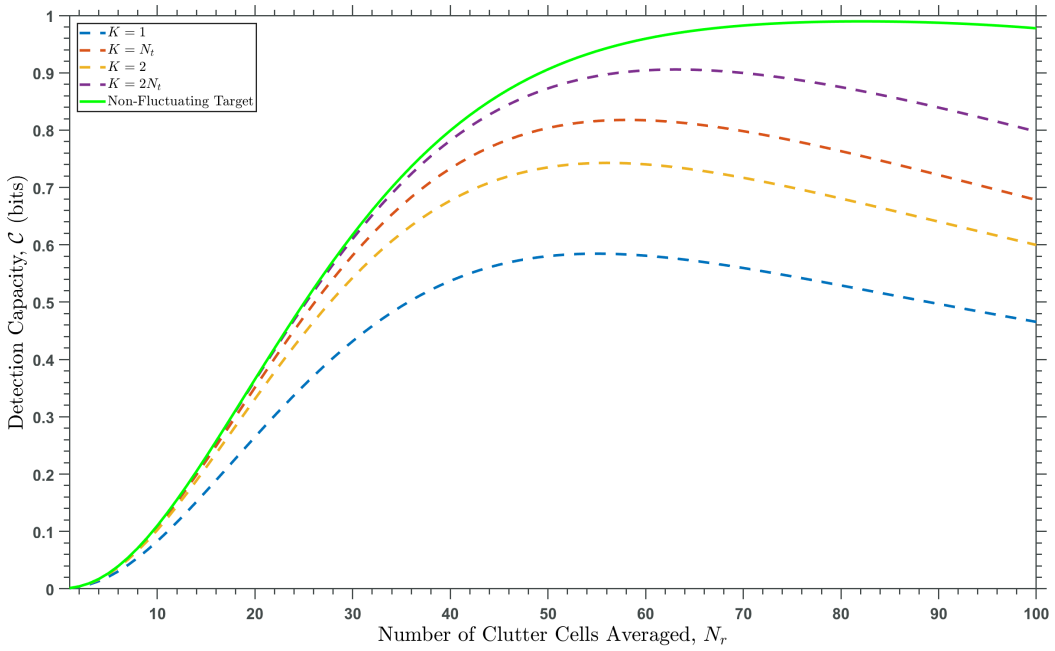

Figure 4.42: Detection capacity v. number of clutter cells averaged for $N_t = 3$, $\alpha = 0.148$, and $X_t = 30$.

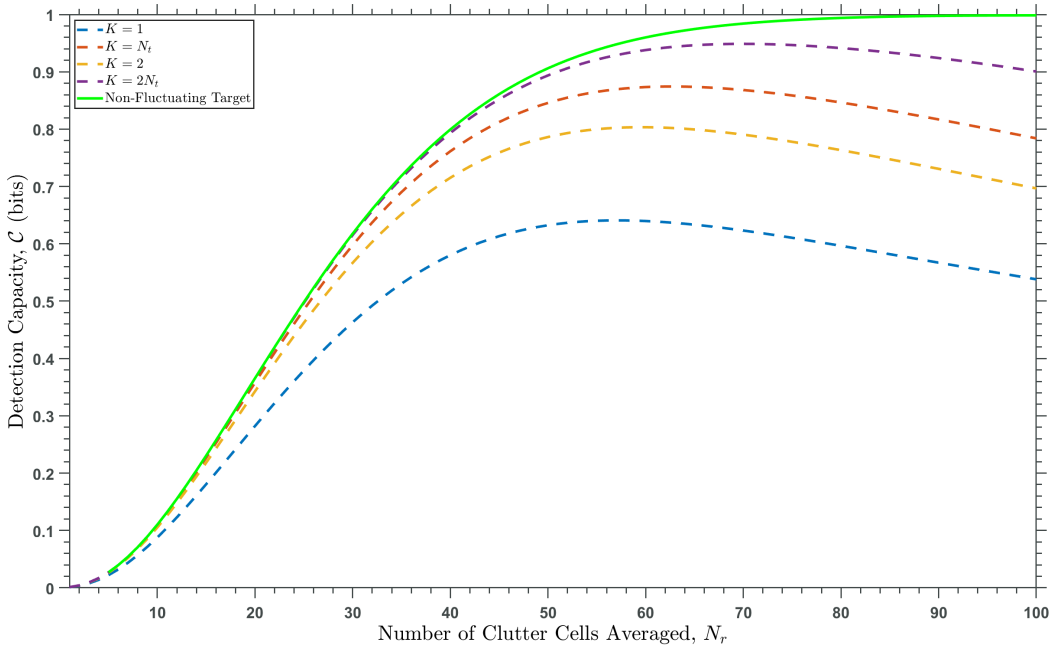

Figure 4.43: Detection capacity v. number of clutter cells averaged for $N_t = 3$, $\alpha = 0.148$, and $X_t = 40$.

Chapter 5

Conclusion

To support the growing use of the spectrum, bandwidth and information usage need to be analyzed for different operations and applications. Previous work has started to analyze pulse-Doppler radar from an information theoretic perspective, with key contributions of providing a fundamental decision bound of 1 decision per second per Hz of transmitted bandwidth and a Bayesian detection capacity expression [1]. This thesis has the novel contributions of improving on previously derived detection capacity expressions to find an expression best suited for radar (i.e. does not rely on *a priori* probabilities). Furthermore, detection capacity has been expanded for different radar parameters such as multiple receivers and M of N integration. Detection capacity has been analyzed for different detectors such as known and unknown phase for different radar assumptions.

Some of the key points from this thesis are the relationships between range, signal-to-noise ratio, P_{FA}, P_D, and detection capacity. Analysis of existing literature on radar signal processing has shown the relationship between SNR and range, primarily that a decrease in SNR leads to higher transmit range. Conversely, decreasing SNR leads to a lower detection capacity due to the minimum detectable signal threshold being reduced. Furthermore, it leads to the reduction of detection capacity at higher transmit ranges, with higher probabilities of false alarm corresponding to a more gradual reduction in detection capacity over transmit range. From the analysis of different detectors, we can see that the introduction of

random/unknown phase reduces the detection capacity, which agrees with our intuition as the increase in uncertainty causes an information penalty.

The effect of M of N integration on channel capacity also has been analyzed for different detectors. Comparing the detection capacity performance of M of N processing has shown that 1 of N processing improves non-normalized detection capacity compared to non-integrated detection. 1 of N processing has higher capacity peaks at the cost of capacity at higher probabilities of false alarm. 2 of N, 3 of N, and 4 of N processing have higher capacity at higher probabilities of false alarm but suffer from lower capacity at lower probabilities of false alarm. Moreover, range plots show that 1 of N processing has higher non-normalized detection capacity compared to non-integrated detection. 2 of N, 3 of N, and 4 of N processing suffer in capacity at higher transmit ranges, due to the rapid decrease in SNR at certain P_{FA}/P_D.

Basic target fluctuation environments have been covered for standard chi-square/Gamma density functions. Swerling models I and IV have the highest detection capacity peaks relative to non-fluctuating targets. This is due to the fact that as $K \to \infty$ there is less detection capacity degradation as they become closer to the case of a non-fluctuating target. For lower total signal power, the fluctuation/non-fluctuating target models converge to the same capacity as $N_r \to \infty$. The performance of CFAR is also analyzed for different numbers of clutter cells and target cells averaged. We can see that as the total number of clutter cells averaged increases, the detection capacity also increases. This is because P_{FA} decreases as the total number of clutter cells averaged increases. Conversely, as the number of target cells averaged increases, detection capacity decreases. Again, this is due to the relationship that N_t and N_r have on P_{FA}. For higher N_t, P_{FA} increases rapidly, causing the decrease in channel capacity.

This thesis has expanded on various important target and radar assumptions. However, it has also opened up many avenues for future work. Much of the future work could focus on the analysis of soft decision frameworks and further target assumptions, using the

unnormalized M of N plots as an example of how soft decision frameworks can improve detection capacity. Target assumptions such as different fluctuation and environments could be a good next step. Furthermore, connecting this work to distributed detection networks should be the key goal, with the ultimate goal of optimizing a fusion center based on information/capacity/bandwidth cost.

Appendix A

Derivation of Detection Capacity

The mutual information between the true hypothesis, $\mathcal{H}$, and decision, μ, is defined as

$$I(\mathcal{H};\mu) = \sum_{\mathcal{H}} \sum_{\mu} P(\mathcal{H},\mu) \log\left(\frac{P(\mathcal{H}|\mu)}{P(\mathcal{H})}\right). \tag{A.1}$$

The input distribution of priors can be expressed as

$$P(\mathcal{H}) = \begin{cases} P_0 & \mathcal{H}=0 \\ \\ P_1 & \mathcal{H}=1 \end{cases}, \quad P_0 + P_1 = 1. \tag{A.2}$$

The *a posteriori* probability of deciding no target is present is the marginalization

$$\begin{aligned} P(\mu_0) &= P(\mu=0) = P(\mathcal{H}_0)P(\mu=0|\mathcal{H}_0) + P(\mathcal{H}_1)P(\mu=0|\mathcal{H}_1) \\ P(\mu_0) &= P_0(1-P_{FA}) + (1-P_0)(1-P_D). \end{aligned} \tag{A.3}$$

Similarly, the *a posteriori* probability of deciding a target is present is the marginalization

$$\begin{aligned} P(\mu_1) &= P(\mu=1) = P(\mathcal{H}_0)P(\mu=1|\mathcal{H}_0) + P(\mathcal{H}_1)P(\mu=1|\mathcal{H}_1) \\ P(\mu_1) &= P_0 P_{FA} + (1-P_0)P_D. \end{aligned} \tag{A.4}$$

The mutual information can now be expressed as

$$\begin{aligned} I(\mathcal{H};\mu) &= \sum_{H}\sum_{\mu} P(\mathcal{H},\mu)\log\frac{P(\mathcal{H}|\mu)}{P(\mathcal{H})} \\ &= P_0(1-P_{FA})[\log(1-P_{FA})-\log(P_{\mu_0})] \\ &+ (1-P_0)(1-P_D)[\log(1-P_D)-\log(P_{\mu_0})] \\ &+ (1-P_0)P_D[\log(P_D)-\log(P_{\mu_1})] \\ &+ P_0P_{FA}[\log(P_{FA})-\log(P_{\mu_1})]. \end{aligned} \tag{A.5}$$

The channel capacity is defined as the maximum (over the input distribution) of the mutual information between the input and output

$$\mathcal{C} = \max_{p(H)} I(\mathcal{H};\mu) = \max_{p(H)} \left[H(\mu) - H(\mu|\mathcal{H})\right]. \tag{A.6}$$

We can define the conditional entropy as

$$\begin{aligned} H(\mu|\mathcal{H}) &= P(\mathcal{H}_0)h(P(\mu=1|\mathcal{H}_0) + P(\mathcal{H}_1)h(P(\mu=0|\mathcal{H}_1) \\ &= P_0h(P_{FA}) + (1-P_0)h(1-P_D), \end{aligned} \tag{A.7}$$

where $h(\bullet)$ is defined as the binary entropy function, provided by Equation (2.23). We can now rewrite the mutual information between the input and output distribution as

$$\begin{aligned} I(\mathcal{H};\mu) &= H(\mu) - H(\mu|\mathcal{H}) \\ &= h(P_0(1-P_{FA}) + (1-P_0)(1-P_D)) \\ &- P_0h(P_{FA}) - (1-P_0)h(1-P_D) \\ &= h(P_0[1-P_{FA}-(1-P_D)] + (1-P_D)) \\ &- P_0[h(P_{FA}) - h(1-P_D)] - h(1-P_D). \end{aligned} \tag{A.8}$$

We want to find the *a priori* probability, P_0, that maximizes the mutual information

$$\begin{aligned} I(\mathcal{H};\mu) &= h(P_0[1-P_{FA}-(1-P_D)]+(1-P_D)) \\ &\quad - P_0[h(P_{FA})-h(1-P_D)]-h(1-P_D). \end{aligned} \tag{A.9}$$

Following calculus, to find the *a priori* probability that maximizes the mutual information, we want to take the derivative of the function with respect to the variable, set the derivative equal to zero, and solve for the variable. Taking the derivative of the mutual information with respect to the *a priori* probability, $\frac{\partial}{\partial P_0} I(\mathcal{H};\mu)$, gives

$$\begin{aligned} I'(\mathcal{H};\mu) &= (1-P_{FA}-(1-P_D))\log_2\left(\frac{1}{P_0[1-P_{FA}-(1-P_D)]+(1-P_D)}-1\right) \\ &\quad - [h(P_{FA})-h(1-P_D)]. \end{aligned} \tag{A.10}$$

Setting the derivative $I'(\mathcal{H};\mu)=0$ gives

$$\frac{1}{P_0[1-P_{FA}-(1-P_D)]+(1-P_D)}-1 = 2^{\frac{h(P_{FA})-h(1-P_D)}{1-P_{FA}-(1-P_D)}}. \tag{A.11}$$

Solving for the *a priori* probability P_0 gives us

$$P_0 = \frac{1}{1-P_{FA}-(1-P_D)}\left(\frac{1}{2^{\frac{h(P_{FA})-h(1-P_D)}{1-P_{FA}-(1-P_D)}}+1}-(1-P_D)\right). \tag{A.12}$$

Plugging the a priori probability into the derivative $I'(\mathcal{H};\mu)$ gives us a channel capacity of

$$\mathcal{C}_{HD} = \frac{P_{FA}}{P_D-P_{FA}}h(1-P_D)-\frac{P_D}{P_D-P_{FA}}h(P_{FA})+\log\left[1+2^{\frac{h(P_{FA})-h(1-P_D)}{P_D-P_{FA}}}\right]. \tag{A.13}$$

Bibliography

[1] J. G. Metcalf, "An examination of the spectral utility of radar," in *2022 IEEE Radar Conference (RadarConf22)*, 2022, pp. 1–6.

[2] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[3] I. Y. Hoballah and P. K. Varshney, "An information theoretic approach to the distributed detection problem," *IEEE Trans. Inf. Theory*, vol. 35, pp. 988–994, 1989.

[4] M. Skolnik, *Introduction to Radar Systems*, 3rd ed. New York, NY, USA: McGraw-Hill, 2002.

[5] M. A. Richards, *Fundamentals of Radar Signal Processing*, 2nd ed. New York, NY, USA: McGraw-Hill, 2014.

[6] T. K. G.A. Korn, *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*, 1st ed. New York, NY, USA: Dover Publications, 2000.

[7] K. W. A. Charlish and H. Griffiths, "Information theoretic measures for mfr tracking control." 2010 IEEE Radar Conference, 2010, pp. 987–992.

[8] P. K. Varshney, *Distributed Detection and Data Fusion*, 1st ed. Berlin, Heidelberg: Springer-Verlag, 1996.

[9] H. Van Trees, Z. Tian, and K. Bell, *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*, ser. Detection Estimation and Modulation Theory. Wiley, 2013.

[10] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing (4th Edition)*, 4th ed. Prentice Hall, 2006.

[11] J. V. Michalowicz, J. M. Nichols, and F. Bucholtz, "Handbook of differential entropy," 2013.

[12] S. M. Moser, P.-N. Chen, H.-Y. Lin, and Organization, "Error probability analysis of binary asymmetric channels," 2010.

[13] R. Blahut, "Hypothesis testing and information theory," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 405–417, 1974.

[14] D. Shnidman, "Radar detection probabilities and their calculation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 31, no. 3, pp. 928–950, 1995.