

# **AMBIGUITY-BASED CLASSIFICATION OF PHASE MODULATED WAVEFORMS**

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Science in Engineering

by

**ANTHONY H. BUCHENROTH**  
B.S. Electrical Engineering, Wright State University, 2012

2015  
Wright State University

WRIGHT STATE UNIVERSITY  
SCHOOL OF GRADUATE STUDIES

December 7, 2015

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Anthony H. Buchenroth ENTITLED Ambiguity-Based Classification of Phase Modulated Waveforms BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science in Engineering.

---

Brian D. Rigling, Ph.D.  
Thesis Director

---

Brian D. Rigling, Ph.D.  
Department Chair

Committee on  
Final Examination

---

Brian D. Rigling, Ph.D.

---

Fred D. Garber, Ph.D.

---

Vasu D. Chakravarthy, Ph.D.

---

Robert E.W. Fyffe, Ph.D.  
Dean, School of Graduate Studies

# ABSTRACT

**Buchenroth, Anthony H. M.S. Egr., Department of Electrical Engineering, 2015. Ambiguity-Based Classification of Phase Modulated Waveforms.**

Accurate classification or recognition of phase modulated radar waveforms, typically accomplished via the combination of pulse parameter estimates and matched filtering, poses a simple problem in ideal conditions. Recognition of these waveforms aids in various spectrum management, surveillance, and electronic warfare (EW) applications. In less than ideal conditions, carrier frequency, time offset, pulse amplitude, initial phase, and bandwidth are unknown to the EW receiver rendering the application of a matched filter futile. This effort investigates the use of features extracted from the ambiguity function of an intercepted pulse. Specifically, this effort will expand upon the methodology of previous work done which uses the autocorrelation as a basis for extracting features. To test the efficacy of this work, extensive Monte Carlo testing employed. Simulation results prove that the methodology implemented herein achieves an overall correct classification rate of about 90% at a signal-to-noise ratio (SNR) of -2 dB on data similar to the training data.

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Previous Work . . . . .	2
1.1.1	Pulse Descriptor Word (PDW)-Based Classification . . . . .	3
1.1.2	Feature-Based Waveform Classification . . . . .	5
1.2	Contribution . . . . .	7
1.3	Outline . . . . .	8
<b>2</b>	<b>Ambiguity Function</b>	<b>9</b>
2.1	Properties . . . . .	11
2.2	Traditional Interpretation . . . . .	12
<b>3</b>	<b>System Design</b>	<b>15</b>
3.1	Signal Interception . . . . .	15
3.2	Feature Extraction . . . . .	16
3.3	Waveform Classification . . . . .	23
<b>4</b>	<b>Experimentation and Results</b>	<b>28</b>
4.1	Experimentation Methodology . . . . .	28
4.1.1	Offline Training . . . . .	29
4.1.2	Online Testing . . . . .	30
4.2	Classification Results . . . . .	31
4.3	Classification Comparison: Ambiguity-Based vs Autocorrelation-Based	34
4.4	Classifier Performance . . . . .	38

4.4.1	Two-Class Classifiers . . . . .	38
4.4.2	Multi-Class Classifiers . . . . .	39
<b>5</b>	<b>Conclusion</b>	<b>50</b>
	<b>References</b>	<b>51</b>

# LIST OF FIGURES

1.1	PDW Clustering Illustration . . . . .	3
1.2	Typical ELINT System . . . . .	4
2.1	Ambiguity Surface . . . . .	13
2.2	Ambiguity Contour . . . . .	14
2.3	Zero-Doppler Cut (top) and Zero-Delay Cut (bottom) . . . . .	14
3.1	Sample of the ambiguity function of a Barker 7 coded pulse given by (3.6). Grid Size: 201 Delay Samples by 101 Doppler Samples . . . . .	19
3.2	Log-scaled ambiguity function of a Barker 7 coded pulse given by (3.7)	20
3.3	Fourier coefficients of the log-scaled ambiguity function of a Barker 7 coded pulse given by (3.8) . . . . .	21
3.4	Selected Fourier coefficients corresponding to 20 delay lags for each of the 11 Doppler lags given by (3.9) . . . . .	22
3.5	Fourier coefficients represented as a column vector concatenation of (3.9) given by (3.10) . . . . .	23
4.1	Confusion matrix at 10 dB SNR with only 1 AF realization. Note: X-axis represents actual class membership, Y-axis represents decided class membership . . . . .	31
4.2	Confusion matrix at -6 dB SNR with 1 AF sequence. Note: X-axis represents actual class membership, Y-axis represents decided class membership . . . . .	32

4.3	Confusion matrix at -6 dB SNR with 5 ensemble average AF sequences. Note: X-axis represents actual class membership, Y-axis represents decided class membership . . . . .	32
4.4	Confusion matrix at -6 dB SNR with 10 ensemble average AF sequences. Note: X-axis represents actual class membership, Y-axis represents decided class membership . . . . .	33
4.5	Confusion matrix at -6 dB SNR with 20 ensemble average AF sequences. Note: X-axis represents actual class membership, Y-axis represents decided class membership . . . . .	33
4.6	A comparison of total classification accuracy as a function of SNR with 1 feature set realization. Observe the effective SNR gain of 2 dB at a fixed classification rate of 90%. . . . .	35
4.7	A comparison of total classification accuracy as a function of SNR with 5 feature set realizations averaged. Observe the effective SNR gain of about 1 dB at a fixed classification rate of 90%. . . . .	36
4.8	A comparison of total classification accuracy as a function of SNR with 10 feature set realizations averaged. Observe the effective SNR gain of about 1 dB at a fixed classification rate of 90%. . . . .	36
4.9	A comparison of total classification accuracy as a function of SNR with 20 feature set realizations averaged. Observe the effective SNR gain of about 1 dB at a fixed classification rate of 90%. . . . .	37
4.10	Relative correct classification rate increase using ambiguity-based features over autocorrelation-based features as a function of SNR. Observe that relative classification accuracy is increased by over 5% within a 4 dB dynamic range. . . . .	37
4.11	All Possible Test Outcomes for the Three-Class Problem Using a Multi-Class Classifier Approach . . . . .	41

4.12 Classifier Precision . . . . .	42
4.13 Classifier Sensitivity . . . . .	43
4.14 Classifier Specificity . . . . .	44
4.15 Classifier F-score . . . . .	45
4.16 Classification System Average Performance . . . . .	46
4.17 ROC Analysis of Discrete Classifiers with 1 AF Sequence . . . . .	47
4.18 ROC Analysis of Discrete Classifiers with 5 Ensemble Average AF Sequences . . . . .	47
4.19 ROC Analysis of Discrete Classifiers with 10 Ensemble Average AF Sequences . . . . .	48
4.20 ROC Analysis of Discrete Classifiers with 20 Ensemble Average AF Sequences . . . . .	48

# ACKNOWLEDGEMENT

I would like to thank my adviser Brian Rigling for guiding me throughout this research. I would also like to thank Vasu Chakravarthy and Sintayehu Dehnie for their continued support and advice throughout this process. A special thanks to all my family and friends who supported and encouraged me. Finally, I would like to thank my thesis committee: Brian Rigling, Fred Garber (WSU), and Vasu Chakravarthy (AFRL).

# CHAPTER 1

## INTRODUCTION

Over time, radar waveforms have become more elaborate. Two of the main causes behind the increasing complexity of modern radar waveforms are the increasing number of radio frequency (RF) spectrum users and the capability of conventional intercept receivers. As the number of RF spectrum users has increased, the amount of available contiguous bandwidth has significantly decreased. This decrease in available spectrum has forced radars to become more efficient at using their available frequencies. To this end, the conventional intercept receivers that once worked extremely well in less dense RF, have become somewhat antiquated against complex radar waveforms operating in dense RF environments. Additionally, the employment of low probability of intercept (LPI) waveforms which operate at a low signal-to-noise ratio (SNR), have further degraded the performance of these electronic support (ES) systems.

Operating in these challenging conditions must be expected of all ES systems, making the need for additional degrees of freedom in all facets of the intercept receiver crucial. The intercept receiver plays an important role in the modern battlefield. Information gained from these receivers aids in mission planning, surveillance, and electronic warfare. The ES system of these receivers must be able to quickly detect and identify adversary waveforms which operate using a wide array of modulations. Many methods have been introduced to aid in the identification of radar waveforms. These methods commonly employ the use of statistical features to describe the waveform. The use of a feature set that best describes the waveform is often coupled with a

trade-off of computational complexity. With computational complexity in mind, an ES system should use a feature set that describes the waveform well and can be computed quickly.

## 1.1 Previous Work

Maintaining the ability to accurately interpret the RF environment has been a focus of research for decades. Within this broad research area, a niche community has focused on methods with which radar emitters, modes, and waveforms can be identified. It is important to note that the methods for classifying emitters and modes differ vastly from classifying waveforms. Additionally, in much of the available open literature, the phrases “radar signal classification” and “radar waveform classification” are often used interchangeably. In an effort to provide clarity, radar signal classification is often times intended to identify emitters and modes. Radar waveform classification and radar modulation classification are phrases used when classifying the modulation on a radar pulse. In this work, it is important to note that classifying modulated radar waveforms such as the length- $L$  Barker codes relies on information contained within a pulse, or intra-pulse information. Conversely, classifying particular emitters such as air traffic control radar or an air defense radar; and particular modes such as acquisition or tracking; rely on intra- and inter-pulse information. Inter-pulse information is comprised of signal parameters including frequency, time-of-arrival (TOA), pulse width, PRI, pulse amplitude, and angle-of-arrival (AOA) [1]. Most conventional ES systems rely on this inter-pulse information, known as a pulse descriptor word (PDW), to sort and classify signals and emitters.

### 1.1.1 Pulse Descriptor Word (PDW)-Based Classification

In less dense RF environments, the conventional emitter and signal identification method uses PDWs to capture information about intercepted signals. This method of classification works by tuning an intercept receiver to a frequency band of interest and sampling all intercepted pulses within that frequency range. The onboard electronic intelligence (ELINT) system obtains measures frequency, pulse amplitude, time-of-arrival (TOA), angle-of-arrival (AOA), and pulse width. This vector of measurements is sent to a clustering block. The clustering, or sorting block, similar to the example in Figure 1.1, keys on certain PDW parameters to group like pulses [1], [2], and [3]. The result of successful pulse sorting allows the pulse repetition interval (PRI) to be computed from TOA measurements. The information from sorted PDWs is compared against a look-up table, often called a mission data file (MDF), made up of *a priori* knowledge about a given RF environment.

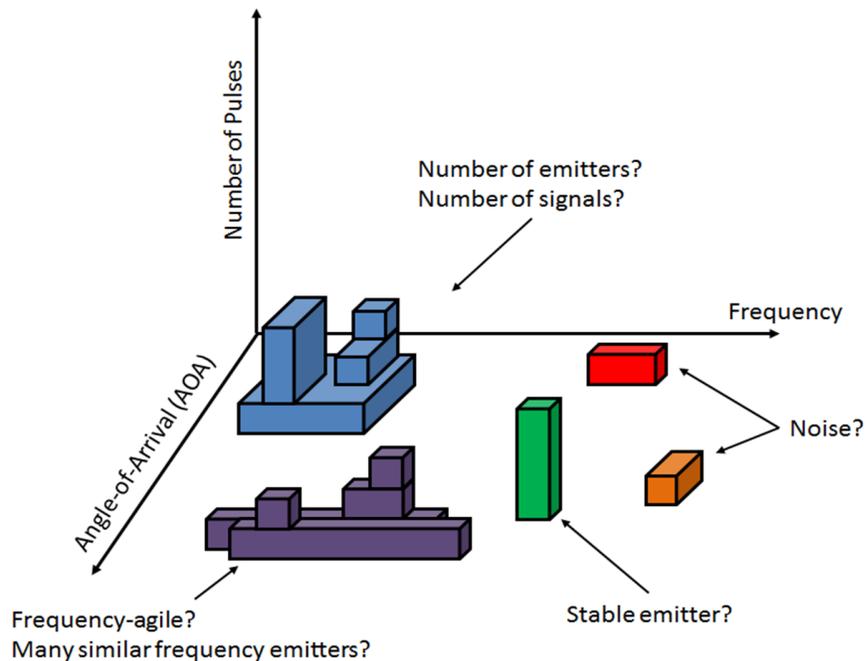


Figure 1.1: PDW Clustering Illustration

The MDF is comprised of a list of all known threat emitters that could potentially

exist in a given scenario. Each emitter in the MDF contains information about all known signals or modes an emitter can achieve. Pulse train information from one or multiple signals, believed to originate from the same source, allow for emitter identification. A block diagram of a typical ELINT system using PDW-based classification method is shown in Figure 1.2.

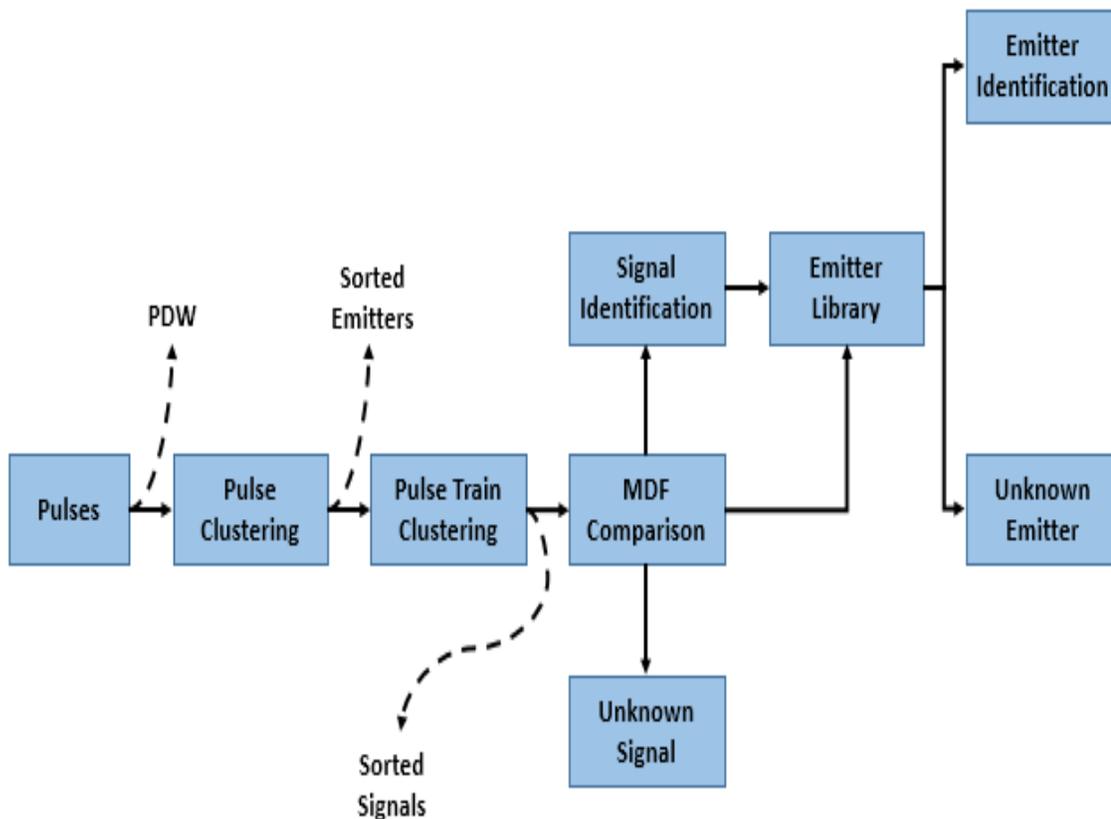


Figure 1.2: Typical ELINT System

The ability to correctly identify signals and emitters using a system similar to that shown in Figure 1.2 hinges upon the quality of information contained within the MDF and the congestion of the RF system in which the system is deployed. In ideal scenarios, emitters and signals are well separated in time, frequency, and space, allowing PDW-based classification to work well. In congested scenarios, many emitters and signals exist, and the classification accuracy is highly dependent on the processing power of the onboard ELINT system. The congested scenario, further degraded

by noise and interference, may render the typical ELINT system inoperable. The system's inability to operate in harsh conditions detracts from mission planning and electronic warfare efforts. To regain EW capabilities, many researchers have considered the use of intra-pulse features. These intra-pulse features are used to represent the intercepted radar signal in a concise way to minimize processing requirements.

### 1.1.2 Feature-Based Waveform Classification

Modern signal processing provides many methods with which one can analyze or interpret signals. Many of these methods are implemented as some sort of transform or statistical manipulation. The goal of choosing a manipulation that best represents a waveform becomes foremost. The feature set chosen to represent the waveform must be strict enough that it cannot be confused with the feature set of a different waveform but lenient enough to account for variations caused by changing environmental conditions or signal parameters.

#### Time-Frequency Features

Features extracted from various time-frequency distributions are an attractive option for many researchers. Under the umbrella of time-frequency analysis, [4] incorporates features extracted from the Wigner distribution, the Choi-Williams distribution (CWD), and Quadrature Mirror Filter Bank (QMFB) implemented on a field programmable gate array (FPGA). While multiple frequency and phase modulations are tested, the authors do not present the results of the classification; only the accuracy of extracting PDW-like parameters from time-frequency features.

Wavelet-based features have been used for waveform classification with generally good results [5], [6], [7], [8]. While [6] lays a framework for using wavelet decomposition to analyze radar waveforms, test waveforms are limited and classification is not performed. Later research conducted in [5], [7], and [8] use wavelet-based features

with fuzzy clustering [5],[8] and fuzzy SVM [7] classification algorithms. Among these, [5] and [8] are very similar and use energy entropy from wavelet-based transforms, specifically, wavelet packet decomposition combined with principal component analysis (PCA) [5] and direct computation of the wavelet transform. The wavelet-based features perform very well for SNR above 0 dB but are limited to *M*-ary PSK, FSK, CW, and various instantiations of LFM.

### Statistical Features

Within the category of statistical features, many possibilities exist for formulating feature sets. These feature sets can be comprised of a single statistical feature vector [9] or a combination of multiple feature vectors [10]. The methodology used in [10] is attractive because it has the ability to use any combination of second-order or higher-order features as well as instantaneously measured features to achieve good performance. This system works by computing all possible features and pruning the overall feature set by using an information theoretic selection algorithm. The classification algorithm incorporates two independently operating parallel multilayer perceptron networks (MLP). This system was tested with eight waveform classes, using frequency modulations and phase modulations, achieving very good results. The author, however, does mention that the performance of the system is dependent upon the quality of the carrier frequency estimate. Further, the majority of the results are based on perfect knowledge of the intercepted signal's carrier frequency. In test cases where the carrier frequency is estimated, accomplished by computing the mean of the power spectral density (PSD), the performance suffers when attempting to classify polyphase waveforms whose power spectrum is non-symmetric.

Alternatively, [9] uses one statistical feature, the autocorrelation function (ACF), of the intercepted pulse as a basis for classification. While this work does not incorporate frequency modulated radar waveforms, it is successful in its ability to distinguish

between 23 phase modulated radar waveforms. Additionally, the performance of the autocorrelation-based classifier is near perfect at SNR above 4 dB. Furthermore, the system is able to increase its performance at low SNR by computing the ensemble average of the autocorrelation over multiple pulses. The result of this effectively reduces noise variance, achieving near perfection at SNR above -6 dB when computing the ensemble average of 20 autocorrelation sequences. This methodology seems attractive due to its success with such a broad range of waveforms. Additionally, the system's only dependency is on the autocorrelation function, making the ability to average multiple autocorrelation sequences together to increase performance noteworthy.

With regard to the work done in [9], a natural extension would be to consider the ambiguity function as a possible feature set. Naturally, if the ACF is a unique feature set invariant to nuisance parameters, so too is the ambiguity function of an intercepted waveform. Further, because the ambiguity function contains all of the information of the ACF, the proposed performance of such a system should at least achieve results similar to [9]. The next chapter will discuss the implementation of the proposed system whose methodology will mirror that of [9] with the exception of the ambiguity function substituted for the autocorrelation function.

## 1.2 Contribution

It is imperative that an ES system be able to accurately identify adversary waveforms. To date, there is no consensus method with which this task is accomplished. Because of the wide array of radar waveforms in existence, the selection of a worthy feature set paired with a robust classification algorithm must be considered. In this thesis, we begin by investigating promising feature sets used in open literature. These algorithms are assessed by a number of factors. An ideal feature set should have a limited number of dependencies, have the ability to operate on a wide array of waveforms (i.e., uniqueness), maintain invariance to nuisance parameters (e.g., car-

rier frequency, time offset, amplitude, initial phase, bandwidth, etc.), and should be robust to low SNR [10],[9].

From these candidate feature sets, the autocorrelation function is chosen, which we conclude best represents the waveform while satisfying our assessment criteria. In this work, we incorporate a feature set modified from the autocorrelation function used in [9]. We follow a similar methodology, operate on the same waveforms, and perform the same evaluation methods. This work, which parallels the work in [9], can be compared fairly, in a head-to-head manner. Furthermore, we provide standard metrics from the machine learning community to evaluate our classifier [11],[12],[13].

### 1.3 Outline

The remainder of this thesis will be organized as follows. Chapter 2 will provide a mathematical basis of our feature set, the ambiguity function, and discuss its traditional interpretation and application. In chapter 3, we will develop our implementation methodology, and the testing strategy. Chapter 4 will provide an analysis of our results, significant findings, and comparison to the work of [9]. Finally, Chapter 5 will provide concluding remarks followed by the account of open problems for future research.

In this thesis, we will use the following notation. Bold capital letters (e.g.,  $\mathbf{X}$ ) represent matrices. Overlined lowercase letters (e.g.,  $\bar{x}$ ) indicate vectors, and lower and upper case letters (e.g.,  $x$  and  $X$ ) represent scalars. Also,  $(\cdot)^T$  represents the transpose operator, and  $(\cdot)^H$  represents the complex transpose operator.

## CHAPTER 2

# AMBIGUITY FUNCTION

Given that radar signals are designed in such a fashion to meet both hardware constraints and a tactical goal within an acceptable margin of error, some method must be used to evaluate the performance of such signals. The ambiguity function is the traditional method by which both waveforms and signals are evaluated. Using the ambiguity function, radar waveform designers can gain insight into a particular waveform in the form of resolution, side-lobe level, ambiguity spacing, and range-Doppler coupling. These measures help to determine if a given waveform is suitable for a given radar application [14]. The subject of this work, however, pertains the use of the ambiguity function with respect to radar waveform classification. The remainder of this chapter will focus on the mathematical development of the ambiguity function and its properties with respect to its traditional use in signal analysis.

The ambiguity function is defined as the time response of a filter matched to a given waveform when the waveform is received with time delay,  $\tau$ , and Doppler shift,  $\nu$ , relative to the nominal values expected by the filter [15]. Consider a radar system to employ a prototype pulse given by

$$s_b(t) = u(t)e^{j\theta(t)} \quad (2.1)$$

where  $u(t)$  is the signal's complex envelope and  $\theta(t)$  is the phase modulation. To

transmit,  $s_b(t)$  is mixed to some RF carrier given by

$$s(t) = u(t)e^{j(\Omega t + \theta(t))} \quad (2.2)$$

where  $\Omega$  is the carrier frequency. Upon reception of the reflected signal,  $s(t)$  is mixed back to baseband. If we consider a matched filter for the waveform  $s_b(t)$  when the input is a Doppler-shifted version,  $s_r(t)$ , the filter output will be given by

$$h(\tau, \nu) = \int_{-\infty}^{\infty} s_b(t)s_r^*(t - \tau)e^{j2\pi\nu t} dt \equiv \Phi(\tau, \nu). \quad (2.3)$$

According to Parseval's theorem, the complex ambiguity or time-frequency autocorrelation function (TFACF),  $\Phi(\tau, \nu)$ , can equivalently be expressed in the frequency domain by applying a Fourier transform to the signal

$$\Phi(\tau, \nu) = \int_{-\infty}^{\infty} S_b(f)S_r^*(f - \nu)e^{j2\pi f\tau} df. \quad (2.4)$$

At this point, it is appropriate to note that some authors define the ambiguity function in slightly different ways. In this work, we define the ambiguity function to be consistent with [16] as the magnitude of the complex ambiguity function given by

$$\chi(\tau, \nu) = \left| \Phi(\tau, \nu) \right| = \left| \int_{-\infty}^{\infty} s_b(t)s_r^*(t - \tau)e^{j2\pi\nu t} dt \right| \quad (2.5)$$

For clarity, we denote the difference between the ambiguity function and its complex variant, the time-frequency autocorrelation (TFACF) as

$$\chi(\tau, \nu) = \left| \Phi(\tau, \nu) \right|. \quad (2.6)$$

## 2.1 Properties

There are five main properties of the ambiguity function,  $\chi(\tau, \nu)$ , that not only aid our conceptual understanding, but also provide insight to its use as a source of features, described in the next chapter.

1. The maximum value of  $\chi(\tau, \nu)$  is  $\chi(0, 0)$  and is given by

$$\chi(\tau, \nu) \leq \chi(0, 0) = \left| \int_{-\infty}^{\infty} s_b(t) s_r^*(t) dt \right| = A \quad (2.7)$$

2. The volume under  $\chi(\tau, \nu)$  is constant

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \chi(\tau, \nu)^2 d\tau d\nu = A \quad (2.8)$$

3.  $\chi(\tau, \nu)$  is symmetric with respect to the origin

$$\chi(\tau, \nu) = \chi(-\tau, -\nu) \quad (2.9)$$

4. The zero-Doppler cut ( $\tau$ -axis) is the autocorrelation function

$$\chi(\tau, 0) = \left| \int_{-\infty}^{\infty} s_b(t) s_r^*(t - \tau) dt \right| = \left| R(\tau) \right| \quad (2.10)$$

5. The zero-Delay cut ( $\nu$ -axis) is the Fourier Transform of the magnitude squared of the complex envelope  $s_r(t)$

$$\chi(0, \nu) = \left| \int_{-\infty}^{\infty} s_r^2(t) e^{j2\pi\nu t} dt \right| \quad (2.11)$$

## 2.2 Traditional Interpretation

The premise of analyzing the waveform in terms of its ambiguity function evokes the notion of an optimal or ideal ambiguity function. Depending on the intent of the waveform, the term ideal can vary. For instance, an ambiguity function can be ideal in terms of resolution or tolerance, but not both. For this reason, we can qualitatively analyze waveforms in terms of ideal resolution or ideal tolerance, making note of ambiguity side-lobe level. For the first extrema, an ambiguity function ideal in resolution would exhibit a peak at the origin of infinitesimally narrow width and zero elsewhere. Known as the “thumbtack” ambiguity function, resolution in both range and Doppler would be highly accurate. The lack of secondary peaks would also imply that no range or Doppler ambiguities exist. A radar waveform that exhibits these qualities would be beneficial for a system requiring fine measurements such as a tracking or imaging radar. Conversely, some waveforms may be intended to be more tolerant of Doppler mismatch. Such a waveform may exhibit a peak at the origin of considerable width in Doppler in order to be responsive to this broad range of Doppler frequencies. The Doppler tolerance for this type of waveform can be ideal for a system attempting to detect targets whose velocity is not known. Since a target of unknown velocity is likely to be detected with these Doppler tolerant waveforms, a system employing this waveform would be ideal for surveillance applications.

With this template for comparison, we can qualitatively analyze the ambiguity function of the simple unmodulated pulse. With regard to the ideal thumbtack ambiguity function, it is apparent that this waveform has limited resolution capabilities in both delay and Doppler. Figures 2.1 - 2.3 show the ambiguity of an unmodulated pulse normalized in both delay by sampling interval,  $t_b$ , and Doppler by waveform length,  $Mt_b$ . From the ambiguity contour, it is apparent that this waveform would not be ideal for applications requiring fine resolution. But, for a given Doppler mis-

match, the overall shape of the ambiguity function remains consistent with a peak at the correct time delay, this waveform is able to detect a broader range of target velocities. With this qualitative analysis, the unmodulated pulse may be suitable for some target detection applications due to its rank as a relatively Doppler tolerant waveform.

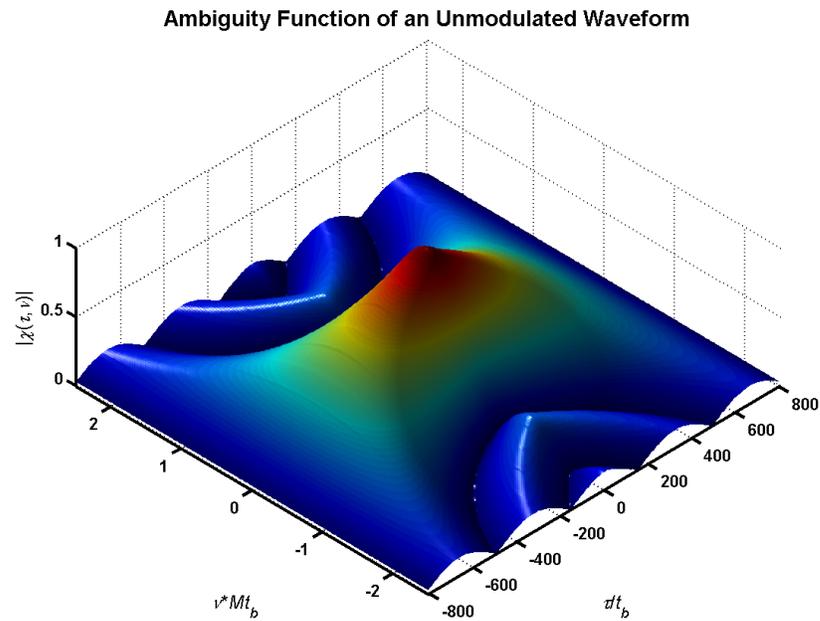


Figure 2.1: Ambiguity Surface

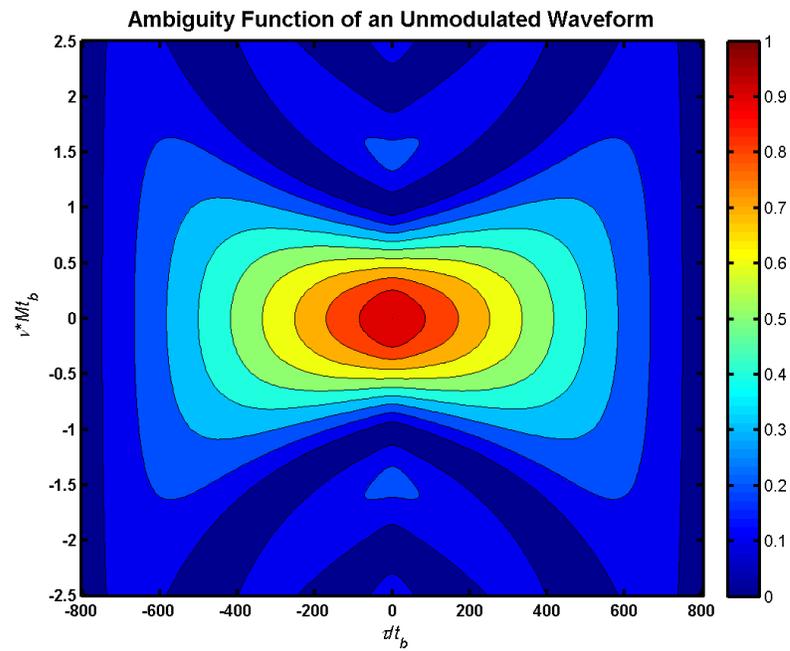


Figure 2.2: Ambiguity Contour

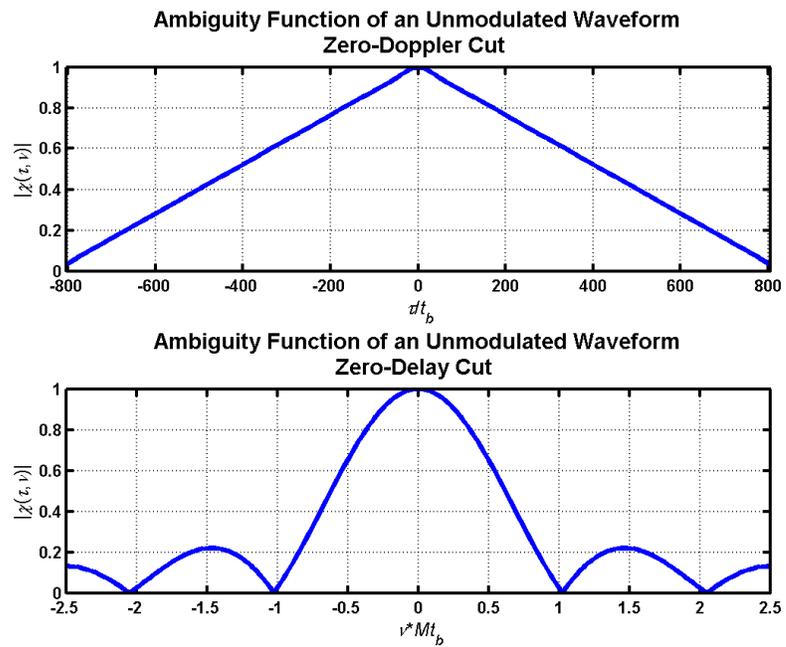


Figure 2.3: Zero-Doppler Cut (top) and Zero-Delay Cut (bottom)

## CHAPTER 3

# SYSTEM DESIGN

While the ambiguity function, developed in the previous chapter, models a matched filter based on a template of the transmitted waveform matched to a waveform received via target reflection as expressed in (2.3), the ambiguity function used for waveform classification is quite different. Additionally, the way the ambiguity function is perceived by the system differs from its traditional interpretation. While this chapter discusses aspects of signal interception and deinterleaving processes, no such algorithms are implemented. The focus of this thesis is on the feature extraction and classification of an intercepted radar waveform. Note, that for the following development, many of the equations are transcribed from [9]. This is done to maintain consistency in the mathematical development and testing in order to fairly compare the two feature sets.

### 3.1 Signal Interception

As defined in [9], a threat radar is assumed to employ a continuous-time prototype pulse

$$x_c(t) = \begin{cases} \exp\{j\phi_c(t)\}, & t \in [-\tau/2, \tau/2] \\ 0, & \text{otherwise} \end{cases} \quad (3.1)$$

where  $\phi_c(t)$  is a phase modulation belonging to one of  $C$  classes. The threat radar transmits a chain of  $P$  pulses with pulse repetition interval (PRI),  $T$ . The intercept receiver observes the signal

$$x(t) = \sum_{p=0}^{P-1} A_p x_c(t/\alpha - pT) e^{j\omega_0 t} + v(t) \quad (3.2)$$

corrupted by white Gaussian noise  $v(t)$  with variance  $\sigma_v^2$  and potentially degraded by model mismatch in the receiver. Further, the observed pulses are altered by jitter in complex amplitude  $A_p$ , time scaled by  $\alpha$  to achieve the designed pulse width, and modulated to carrier frequency,  $\omega_0$ ; all of which are unknown to the intercept receiver. From the received signal, envelope detection is performed to identify leading and falling edges of each pulse in order to create a bank of stacked pulses

$$y_p(t) = A_p x_c(t/\alpha - t_p) e^{j\omega_0 t} + v_p(t), \quad p = 0 \dots P - 1, \quad (3.3)$$

where  $v_p(t)$  represents noise samples from  $v(t)$ . Additionally, each pulse from the threat radar is contained within a window slightly larger than the pulse width to allow for errors in edge detection. The offset within the window,  $t_p$ , is assumed to be unknown. This ensemble of radar pulse observations are then input into the phase modulated waveform classifier in digitized form,  $y_p[n] = y_p(nT_s)$ ,  $n = 0 \dots N - 1$ , where  $T_s$  is the sampling interval assumed to satisfy the Nyquist requirement.

## 3.2 Feature Extraction

In order to identify the phase modulation of the waveform, [9] uses a variation of the waveform's autocorrelation function that is invariant to unknown signal parameters, including complex amplitude ( $A_p$ ), frequency offset ( $\omega_0$ ), time offset ( $t_p$ ), and time scaling ( $\alpha$ ). At this point, we choose to substitute the autocorrelation-based

feature for the ambiguity-based feature. This seems to be a natural extension to the work done in [9] given that the autocorrelation sequence is contained within the ambiguity function according to Property 2.10. While each phase modulated waveform yields a unique ambiguity function, the ambiguity-based feature set is also lenient enough to account for variation in the waveform (e.g., pulse width). The waveform classification system begins with the feature extraction process by computing the discretized complex ambiguity function, or time-frequency autocorrelation function (TFACF) of each intercepted pulse. The model for the discrete TFACF is given by

$$\begin{aligned}\Phi(\tau, \nu) &= \int_{-\infty}^{\infty} s(t)s^*(t - \tau)e^{j2\pi\nu t} dt \\ &\approx \sum_{n=0}^{N-1} T_s s(nT_s)s^*(nT_s - \tau)e^{j2\pi\nu nT_s} \equiv \hat{\Phi}(\tau, \nu)\end{aligned}\quad (3.4)$$

where we incorporate the total number of samples,  $N$ , and the sampling interval,  $T_s$ . Using modified code from [16],  $y_p$  is substituted for  $s$  and vectors for discrete  $\tau$  and  $\nu$  are computed. The computed TFACF of each pulse will be of the form

$$\begin{aligned}\hat{\Phi}_p[m, k] &= \sum_{n=0}^{N-1} T_s y_p[nT_s]y_p^*[nT_s - m\Delta t]e^{j2\pi k\Delta f nT_s} \\ &= |A_p|^2 \hat{\Phi}_c(m\Delta t, k\Delta f)e^{j\omega_0 nT_s} + w_p[m\Delta t, k\Delta f], \\ m &= -M + 1 \dots M - 1, \quad \Delta t = \frac{N}{\alpha M}, \\ k &= -K + 1 \dots K - 1, \quad \Delta f = F/K/N,\end{aligned}\quad (3.5)$$

where  $T_s$  is the sampling interval of the input pulse,  $N$  is the total number of samples,  $m\Delta t$  is the discrete sample delay,  $k\Delta f$  is the discrete Doppler shift, and noise terms are grouped into  $w_p[m\Delta t, k\Delta f]$ . The ambiguity's invariance to time offset is used to neglect  $t_p$  and the time scaling term,  $\alpha$ , is dropped in the delay resolution normalization,  $\Delta t = N/\alpha M$ . It is important to note that while the length of the signal can change, the size of the TFACF is constant due to hard-coded parameters for number

of delay and Doppler shifts,  $M$  and  $K$ , respectively. Also, the maximum normalized Doppler shift,  $F$ , is hard-coded. The effect of hard-coding these parameters results in variable delay and Doppler resolution dependent on the length of the waveform,  $N$ .

Because the ambiguity function is symmetric, this algorithm computes the first two quadrants (i.e., positive and negative delay for positive Doppler shifts) efficiently using a sparse matrix representation of the waveform correlation. Again, note that for the computation of the complex ambiguity, parameters for maximal delay and Doppler shift as well as the number of grid points for each is hard-coded. Although the choice for these parameters is largely arbitrary, the values were chosen to capture the entire behavior of the waveform with qualitatively good resolution while also adhering to the recommended parameters given by [16]. For clarity, images of the feature extraction process are included throughout this chapter using a Barker-7 waveform as a model. To begin, Figure 3.1 displays two quadrants of the ambiguity function of the Barker 7 waveform at the chosen grid size of 101 Doppler samples by 201 delay samples. Accordingly, these values allowed for variable levels of resolution depending on sampling frequency and pulse width of the waveform.

The prototype signal TFACF for the  $c$ th waveform is expressed as  $\hat{\Phi}_c(\tau, \nu)$ . A noise suppression technique is then employed to compute the ensemble average TFACF given by

$$\begin{aligned} \hat{\Phi}(m, k) &= \frac{1}{P} \sum_{p=0}^{P-1} \hat{\Phi}_p[m, k] & (3.6) \\ &\equiv |A|^2 \hat{\Phi}_c[m\Delta t, k\Delta f] e^{j\omega_0 m \Delta t} + w[m, k], \\ & \quad m = -M + 1 \dots M - 1, \\ & \quad k = -K + 1 \dots K - 1, \end{aligned}$$

where the variance of  $w[m, k]$  has been reduced by a factor of  $P$  relative to the variance of  $w_p[m, k]$  and the average waveform intensity is  $|A|^2 = P^{-1} \sum_p |A_p|^2$ . Computing

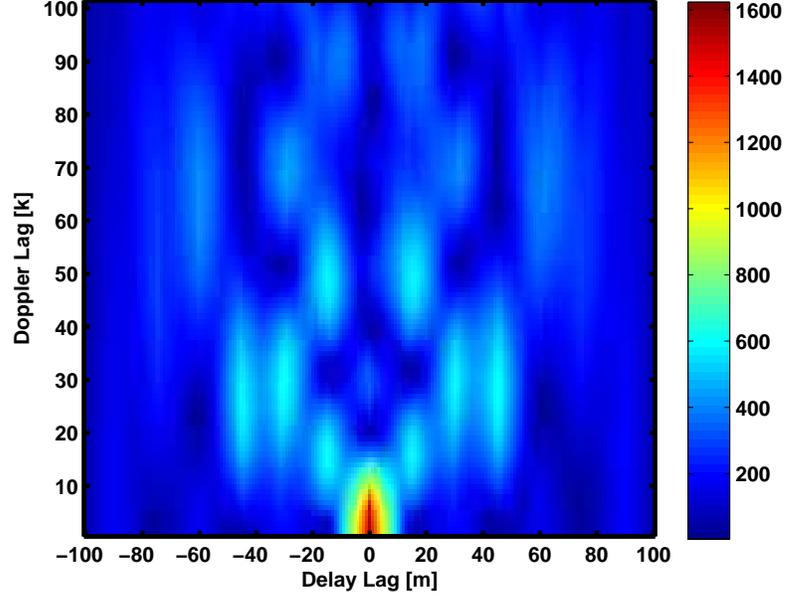


Figure 3.1: Sample of the ambiguity function of a Barker 7 coded pulse given by (3.6). Grid Size: 201 Delay Samples by 101 Doppler Samples

the log-magnitude of the ensemble average TFACF produces a unity peak normalization with enhanced sidelobe structure, shown in Figure 3.2, given by

$$l_{\hat{\Phi}}[m, k] = \log \left| \hat{\Phi}_c[m\Delta t, k\Delta f] \right| + \log \frac{|A|^2}{\hat{\Phi}_c[0, 0]} + z[m, k] \quad (3.7)$$

$$m = -M + 1 \dots M - 1,$$

$$k = -K + 1 \dots K - 1,$$

where  $z[m, k]$  is introduced as an additive term accounting for corruption due to noise and model mismatch. As stated in [9] and assumed here, the statistics of the noise are difficult to track but are treated as zero-mean.

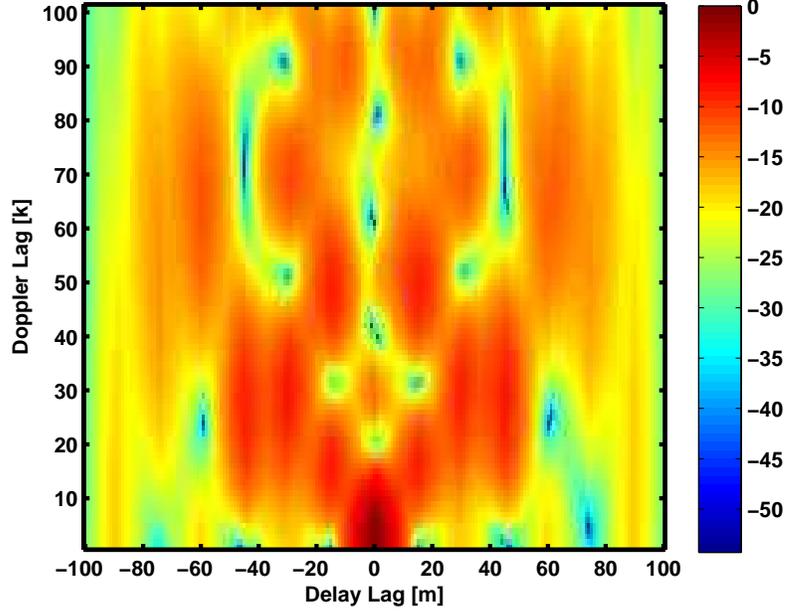


Figure 3.2: Log-scaled ambiguity function of a Barker 7 coded pulse given by (3.7)

By computing the two-dimensional discrete Fourier transform of (3.7) we obtain a compact set of Fourier coefficients given by

$$\begin{aligned}
 L_{\hat{\Phi}}[u, v] &= \frac{1}{MK} \sum_{m=0}^{2M-1} \sum_{k=0}^{2K-1} \left[ \log \left| \hat{\Phi}_c[m\Delta t, k\Delta f] \right| \right. \\
 &\quad \left. + \log \frac{|A|^2}{\hat{\Phi}_c[0, 0]} + z[m, k] \right] e^{-j2\pi \left( \frac{u}{2M-1} m + \frac{v}{2K-1} k \right)}, \quad (3.8) \\
 u &\in \left[ \frac{-1}{2\Delta t}, \frac{1}{2\Delta t} \right], \quad v \in \left[ \frac{-1}{2\Delta f}, \frac{1}{2\Delta f} \right].
 \end{aligned}$$

At this point, we now have an expression that is invariant to unknown time shifts and frequency offsets. Also, the unknown complex-valued amplitude has been relegated into a DC offset. By considering the complex ambiguity function as a log-scaled, two-dimensional discrete Fourier transform, as shown in Figure 3.3, we now have a feature set that is not only content-rich, but is invariant to the nuisance parameters introduced at the beginning of this chapter.

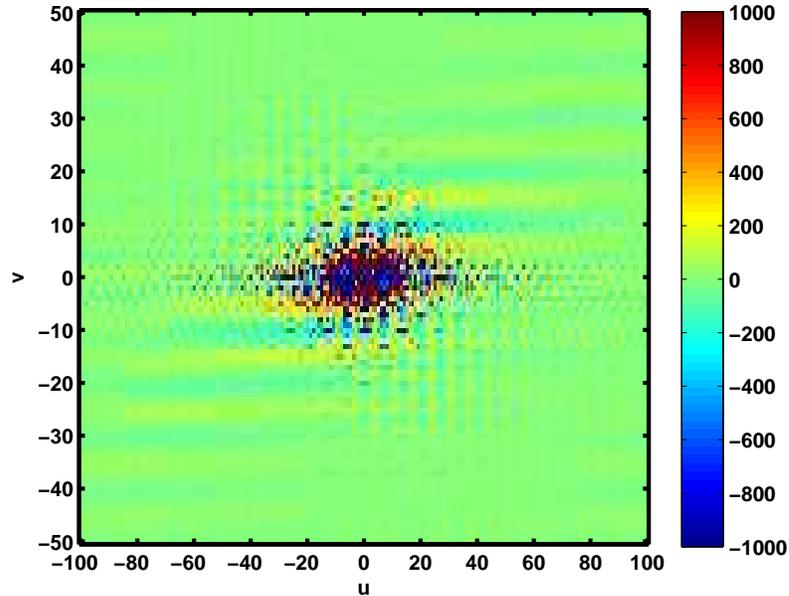


Figure 3.3: Fourier coefficients of the log-scaled ambiguity function of a Barker 7 coded pulse given by (3.8)

Now, because we have a feature matrix that exists in a two-dimensional frequency space,  $u$  and  $v$ , we want to capture a subset of the information that will fully describe a waveform while keeping the size of the subset manageable to maintain the robustness of the algorithm, shown in Figure 3.4. It is noted in [9] that low-frequency information is generally sufficient for classification purposes. With this in mind, we can sample a small number of terms in both dimensions to capture the principal sidelobe structure of the log-TFACF while omitting the DC coefficient that remains ambiguous. Specifically, we choose the block of samples that consist of all of the 20 ACF coefficients from [9] along with the 20 coefficients from each of the 5 Doppler lag rows above and below, shown in Figure 3.4. Choosing samples at  $p = 1 \dots P$  with

$P \ll 2M - 1$  and  $q = -Q \dots Q$  with  $|Q| \ll 2K - 1$ , we can rewrite (3.8) as

$$L_{\hat{\Phi}}[p, q] = \frac{1}{MK} \sum_{m=0}^{2M-1} \sum_{k=0}^{2K-1} \left[ \log \left| \hat{\Phi}_c[m\Delta t, k\Delta f] \right| + \log \frac{|A|^2}{\hat{\Phi}_c[0, 0]} + z[m, k] \right] e^{-j2\pi \left( \frac{p}{2M-1} m + \frac{q}{2K-1} k \right)}, \quad (3.9)$$

$$p = 1 \dots P, \quad q = -Q \dots Q.$$

Finally, we normalize, then manipulate  $L_{\hat{\Phi}}[p, q]$  by concatenating its column vectors to create a  $PQ$ -dimensional feature vector, shown in Figure 3.5, given by

$$x[k] = \text{vec}(L_{\hat{\Phi}}[p, q]) = \begin{bmatrix} L_{\hat{\Phi}}[1, 1] \\ \vdots \\ L_{\hat{\Phi}}[p, q] \end{bmatrix} \quad (3.10)$$

to be input into a linear classifier.

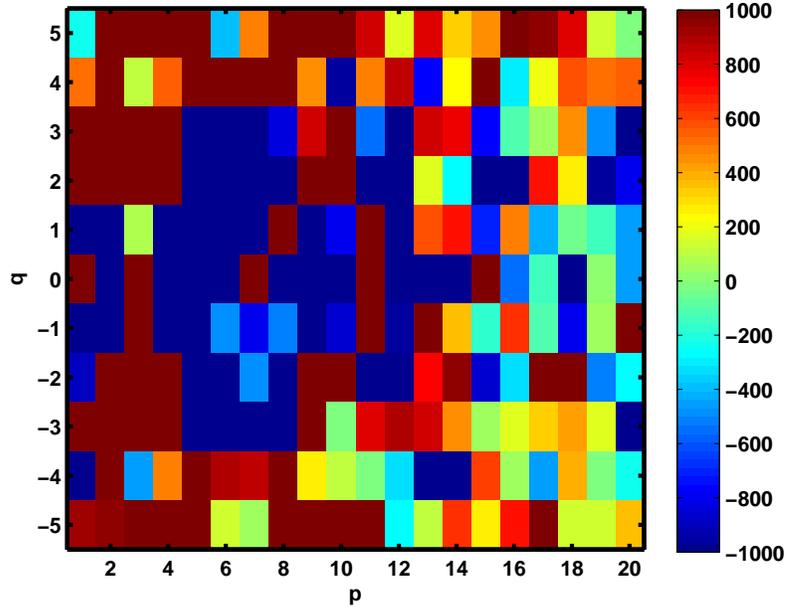


Figure 3.4: Selected Fourier coefficients corresponding to 20 delay lags for each of the 11 Doppler lags given by (3.9)

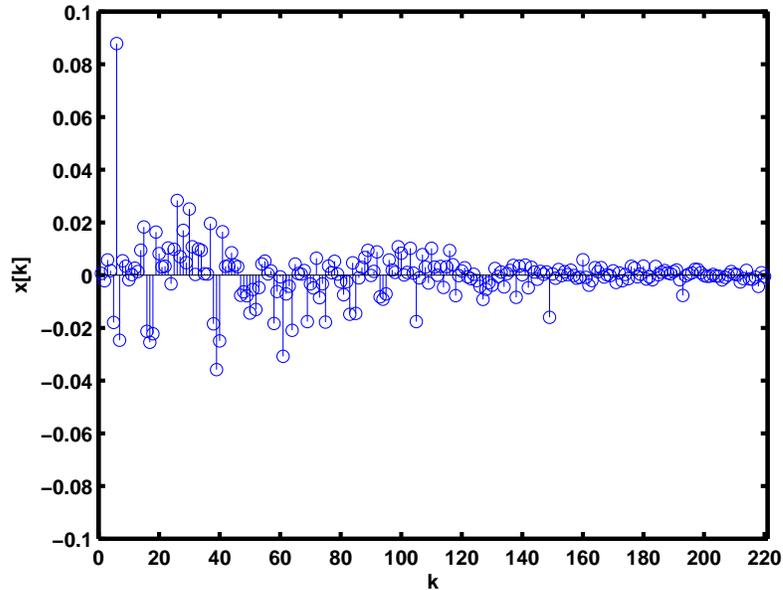


Figure 3.5: Fourier coefficients represented as a column vector concatenation of (3.9) given by (3.10)

### 3.3 Waveform Classification

We now have a  $PQ$ -dimensional feature vector belonging to class  $c$ , to which one must apply some classification method to discriminate between classes. For fair comparison, we align our classification method with that of [9], the Fisher Linear Discriminant (FLD) [17]. Since we are interested in a  $c$ -class problem, we must decide between two classification methodologies, one-versus-one (OVO) or one-versus-all (OVA). In the OVO or two-class methodology, we can apply the FLD directly. Applying the OVO to the  $c$ -class problem would in turn yield  $c(c-1)/2$  binary classifiers. Thus, for a classification problem with many classes, the number of classifiers grows quickly. Instead, we use the OVA methodology where we employ  $c-1$  classifiers to the multiclass problem. This also allows for simple analysis of each class as shown in the next chapter.

For the multiclass problem, we use a generalized variant of the FLD in the OVA paradigm. This classifier seeks to project feature vectors onto a weighting vector

that will provide the maximal separation between the one within-class subset and the out-of-class subset samples. Suppose we have a set of  $n$   $PQ$ -dimensional samples  $\bar{x}_1, \dots, \bar{x}_n$ , belonging to the within-class subset,  $D_1$ , and the out-of-class subset,  $D_0$ . We can then express the projection of  $PQ$ -dimensional samples onto the weighting vector as

$$y_i = \bar{w}_i^T \bar{x}, \quad (3.11)$$

where  $\bar{w}_i$  is the weighting vector of the  $i$ th class,  $i = 1, \dots, c - 1$ . Because  $\|\bar{w}_i\|$  simply scales the data, the magnitude of  $\bar{w}_i$  is of no consequence. The direction of  $\bar{w}_i$  is significant however, and must be chosen in a way to best separate the within-class and out-of-class samples. In order to find the best weighting vector, the FLD begins by computing the  $PQ$ -dimensional sample means of the within-class and out-of-class data, respectively, given by

$$\bar{\mu}_1 = \frac{1}{n_1} \sum_{\bar{x} \in D_1} \bar{x}, \quad (3.12)$$

$$\bar{\mu}_0 = \frac{1}{n_0} \sum_{\bar{x} \in D_0} \bar{x}, \quad (3.13)$$

where  $\bar{x} \in D_1$  represents the samples from one class, denoted  $D_1$ , and  $\bar{x} \in D_0$  represents the samples from the remaining  $c - 1$  classes, denoted  $D_0$ . To obtain the best separation, the maximum difference between the sample means is expressed in terms of standard deviations for each class. We then define the total within-class scatter matrix as

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_0, \quad (3.14)$$

where  $\mathbf{S}_1$  and  $\mathbf{S}_0$  are the scatter matrices of each subset,  $D_1$  and  $D_0$ , respectively,

given by

$$\mathbf{S}_i = \sum_{\bar{x} \in D_i} (\bar{x} - \bar{\mu}_i)(\bar{x} - \bar{\mu}_i)^T \quad (3.15)$$

Similarly, the variance of the projected samples are expressed as

$$s_i^2 = \sum_{\bar{x} \in D_i} (\bar{w}_i^T \bar{x} - \bar{w}_i^T \bar{\mu}_i)^2 \equiv \sum_{\bar{x} \in D_i} \bar{w}_i^T (\bar{x} - \bar{\mu}_i)(\bar{x} - \bar{\mu}_i)^T \bar{w}_i \equiv \bar{w}_i^T \mathbf{S}_i \bar{w}_i. \quad (3.16)$$

Therefore, the sum of the projected scatters becomes

$$s_1^2 + s_0^2 = \bar{w}_i^T (\mathbf{S}_1 + \mathbf{S}_0) \bar{w}_i \equiv \bar{w}_i^T \mathbf{S}_W \bar{w}_i. \quad (3.17)$$

Next, in order to find the between-class scatter or difference between projected mean vectors,  $\mathbf{S}_B$ , we define the separation as

$$\begin{aligned} (\mu_1 - \mu_0)^2 &= (\bar{w}_i^T \bar{\mu}_1 - \bar{w}_i^T \bar{\mu}_0)^2 \\ &= \bar{w}_i^T (\bar{\mu}_1 - \bar{\mu}_0)(\bar{\mu}_1 - \bar{\mu}_0)^T \bar{w}_i \\ &= \bar{w}_i^T \mathbf{S}_B \bar{w}_i \end{aligned} \quad (3.18)$$

Given that we now have expressions for both within-class scatter,  $\mathbf{S}_W$ , and between-class scatter,  $\mathbf{S}_B$ , we seek the criterion that maximizes the ratio of between-class scatter to within-class scatter, given by

$$J(\bar{w}_i) = \frac{\bar{w}_i^T \mathbf{S}_B \bar{w}_i}{\bar{w}_i^T \mathbf{S}_W \bar{w}_i}. \quad (3.19)$$

The maximization of the criterion function,  $J(\cdot)$ , is a traditional optimization problem

solved by evaluating the derivative of the criterion function at zero, given by

$$\begin{aligned}
\nabla J(\bar{w}) &= \nabla \frac{\bar{w}^T \mathbf{S}_B \bar{w}}{\bar{w}^T \mathbf{S}_W \bar{w}} \equiv 0 \\
&= (\bar{w}^T \mathbf{S}_W \bar{w}) \frac{\delta \bar{w}^T \mathbf{S}_B \bar{w}}{\delta \bar{w}} - (\bar{w}^T \mathbf{S}_B \bar{w}) \frac{\delta \bar{w}^T \mathbf{S}_W \bar{w}}{\delta \bar{w}} = 0 \\
&= (\bar{w}^T \mathbf{S}_W \bar{w}) 2 \mathbf{S}_W \bar{w} - (\bar{w}^T \mathbf{S}_B \bar{w}) 2 \mathbf{S}_B \bar{w} = 0.
\end{aligned} \tag{3.20}$$

If we then divide (3.20) by  $2\bar{w}^T \mathbf{S}_W \bar{w}$ , we see that  $\mathbf{S}_W$  and  $\mathbf{S}_B$  are in the same direction and now have

$$\frac{\bar{w}^T \mathbf{S}_W \bar{w}}{\bar{w}^T \mathbf{S}_W \bar{w}} \bar{w} \mathbf{S}_B \bar{w} - \frac{\bar{w}^T \mathbf{S}_B \bar{w}}{\bar{w}^T \mathbf{S}_W \bar{w}} \bar{w} \mathbf{S}_W \bar{w} = 0, \tag{3.21}$$

which reduces to

$$\mathbf{S}_B \bar{w} - J \mathbf{S}_W \bar{w} = 0. \tag{3.22}$$

Observe that (3.22) is the generalized eigenvalue problem

$$\mathbf{S}_W^{-1} \mathbf{S}_B \bar{w} = J(\bar{w}) \bar{w}, \tag{3.23}$$

and the optimal  $\bar{w}$  is found via

$$\operatorname{argmax}_{\bar{w}} J(\bar{w}) = \mathbf{S}_W^{-1} \mathbf{S}_B = \mathbf{S}_W^{-1} (\mu_1 - \mu_0). \tag{3.24}$$

Now that we have found the optimal weighting vector,  $\bar{w}$ , the determination threshold,  $T$ , can be defined as the midpoint between the projected mean vectors given by

$$T = \bar{w}^T \frac{\bar{\mu}_1 - \bar{\mu}_0}{2}, \tag{3.25}$$

and the linear decision function is thus given by

$$d(\bar{x}) = \begin{cases} 1, & \bar{w}^T \bar{x} > T \\ 0, & \bar{w}^T \bar{x} < T \end{cases} \quad (3.26)$$

## CHAPTER 4

# EXPERIMENTATION AND RESULTS

In Section 4.1, the methodology in which we train and test the classification system is discussed. In Section 4.2, we present the results of our testing by means of confusion matrices. Using these confusion matrices, in Section 4.3, we compare our ambiguity-based classification system with the autocorrelation-based classification system in [9]. Lastly, in Section 4.4, we assess our classification system using the standard metrics used in the machine learning community [11],[12],[13].

### 4.1 Experimentation Methodology

To fairly compare the two feature sets, classifier training and testing is conducted to mirror the methodology of [9]. Today’s EW receivers operate at a frequency band of interest, down convert the instantaneous bandwidth to an intermediate frequency (IF), and digitize. The digitized samples are input into an energy detection subsystem to determine the presence of a signal. A PDW encoder then estimates signal pulse parameters. In our simulation, we assume a single signal detection occurred within an AWGN channel, the signal has been time gated, initially identified as a phase modulated waveform, and mixed to baseband. Additionally, our only other assumptions are that the input waveform is one of the 23 different modulation classes shown in Table 4.1 and that the feature vectors corresponding to the 23 different modulation classes are linearly separable.

Table 4.1: Phase modulation types addressed in [9]

$c$	Modulation Type	Code Length	Training $\tau$ ( $\mu\text{sec}$ )	Testing $\tau$ ( $\mu\text{sec}$ )
1	Barker	7	1.75	7.0
2	Barker	11	2.75	11.0
3	Barker	13	3.25	13.0
4	Combined Barker	16	2.0	8.0
5	Combined Barker	49	6.13	2.11
6	Combined Barker	169	22.1	84.6
7	Maximum Length Pseudo Random	16	1.5	4.5
8	Maximum Length Pseudo Random	64	3.5	10.5
9	Maximum Length Pseudo Random	256	6.3	18.9
10	Minimum Peak Sidelobe	10	1.4	4.2
11	Minimum Peak Sidelobe	25	2.5	10.0
12	Minimum Peak Sidelobe	48	4.8	19.2
13	T1	NA	4.0	16.0
14	T2	NA	3.0	12.0
15	T3	NA	8.0	2.0
16	Polyphase Barker	7	1.75	7.0
17	Polyphase Barker	20	2.0	8.0
18	Polyphase Barker	40	4.0	16.0
19	P1	NA	10.0	20.0
20	P2	NA	6.4	25.6
21	P3	NA	6.4	25.6
22	P4	NA	10.0	29.0
23	Minimum Shift Key	64	18.9	8.0

### 4.1.1 Offline Training

The FLD classifier is trained with the 23 phase modulations displayed in Table 4.1. As noted in [9], the modulation classes are chosen to represent the majority of phase modulated radar waveforms an EW receiver is likely to encounter. To train the classifier, we perform a 1000 iteration Monte Carlo simulation at 10 dB SNR. In each iteration, pseudo-random noise and initial phase realizations are generated for each waveform and training pulse width pair. Additionally, several different sampling rates are used to safeguard against bias being introduced into the training coefficients.

It is important to note the Monte Carlo training methodology. Because the FLD is a deterministic system whose performance is tied to computing the best weighting vector,  $\bar{w}$ , and corresponding thresholds,  $T$ , the manner in which we train our classifiers is correlated to our expected performance. Variance in the input feature

vectors, caused mainly by the noise component within the signal, creates some unknown probability distribution in the  $PQ$ -dimensional feature space. Since the FLD is nonparametric in nature, obtaining a closed form expression for the distribution of each class's features is unnecessary. Instead, by training our classifier with many instantiations of each feature vector, we need only to account for the probability distributions in terms of sample mean and sample covariance, as shown in (3.24) and (3.25). To obtain the projection that best separates classes of waveforms, leading to optimal FLD performance, training in a Monte Carlo fashion, where the most accurate sample means and sample covariances are computed using many instantiations of each feature vector, will generally increase expected performance.

#### 4.1.2 Online Testing

To test our classification algorithm, we perform simulations using the 23 phase modulations found in Table 4.1 with different pulse widths ( $\tau$ ) than we used in our offline training phase. The effect of changing the pulse width of the input waveform under test from the training set, which alters the time bandwidth product of the pulse, allows us to gauge the performance of the algorithm independent of any specific pulse parameters. Good performance under this independence criterion suggests algorithm flexibility and feature set leniency. To this end, a random initial phase is imparted on the waveform to account for fluctuations in pulse transmissions and a random time shift is applied to simulate imperfect TOA and time-of-departure (TOD) estimates. Additionally, because this algorithm is assumed to operate as a back-end process to some antenna hardware and front-end signal processing, a fixed test sampling rate of 105 MHz is used. To gauge overall performance, a 1000 iteration Monte Carlo simulation is employed with different noise realizations.

## 4.2 Classification Results

Figure 4.1 displays the confusion matrix, also known as a contingency table, for the 23 class problem at 10 dB SNR with only 1 AF realization. For all of the following confusion matrices, the X-axis represents actual class membership while the Y-axis represents the class membership decided by the classifier. As shown, with one realization of the ambiguity-based features, our classifier performs perfectly at 10 dB SNR. The perfect classification rate in Figure 4.1 is noteworthy because the varying nuisance parameters do not affect classification rate when the SNR of the training and testing features are equivalent.

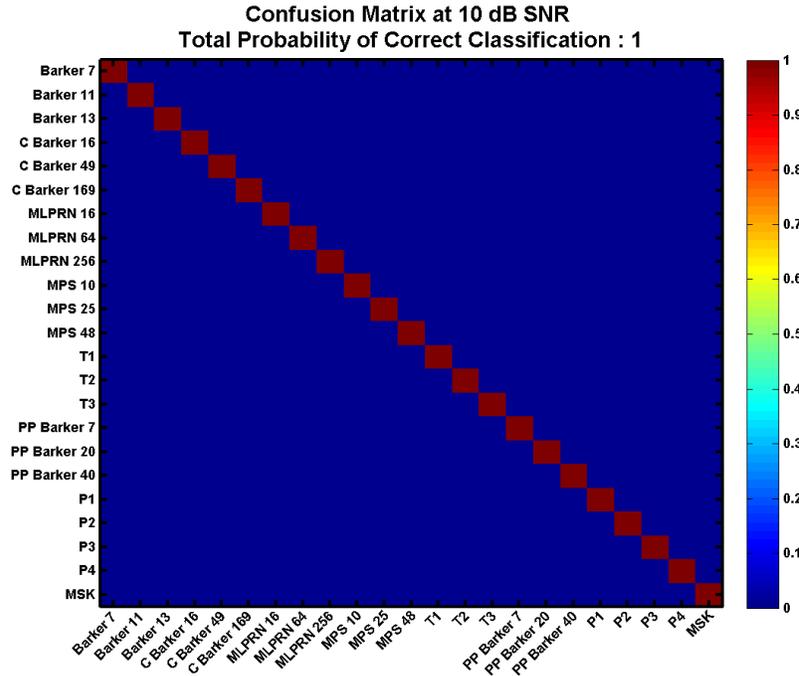


Figure 4.1: Confusion matrix at 10 dB SNR with only 1 AF realization. Note: X-axis represents actual class membership, Y-axis represents decided class membership

Again, aligning our display of results to that of [9], below in Figures 4.2-4.5 shows the confusion matrix with 1, 5, 10, and 20 AF features averaged together at -6 dB SNR. As shown, averaging just 5 received AF sequences boosts classifier accuracy

from 11% to 81%. What's more, averaging all 20 AF sequences achieves 96% correct classification.

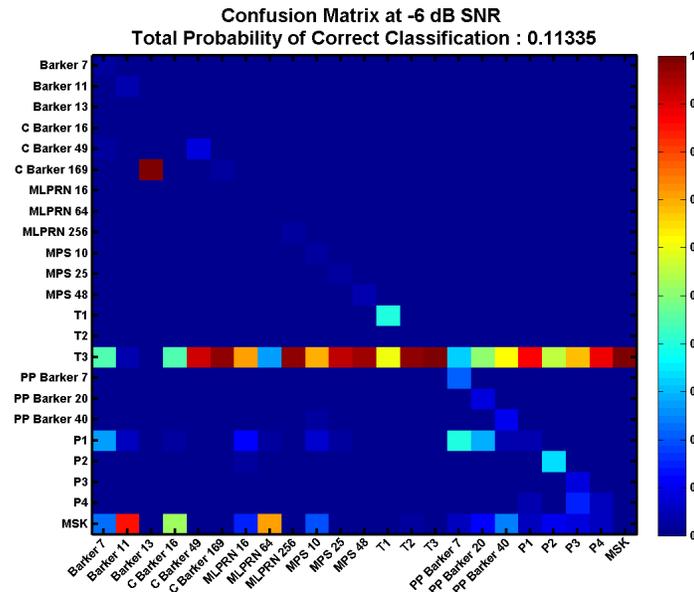


Figure 4.2: Confusion matrix at -6 dB SNR with 1 AF sequence. Note: X-axis represents actual class membership, Y-axis represents decided class membership

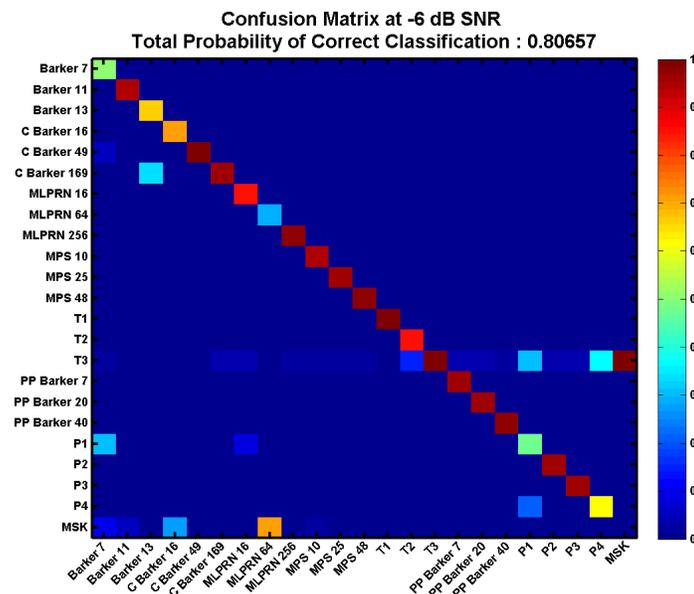


Figure 4.3: Confusion matrix at -6 dB SNR with 5 ensemble average AF sequences. Note: X-axis represents actual class membership, Y-axis represents decided class membership

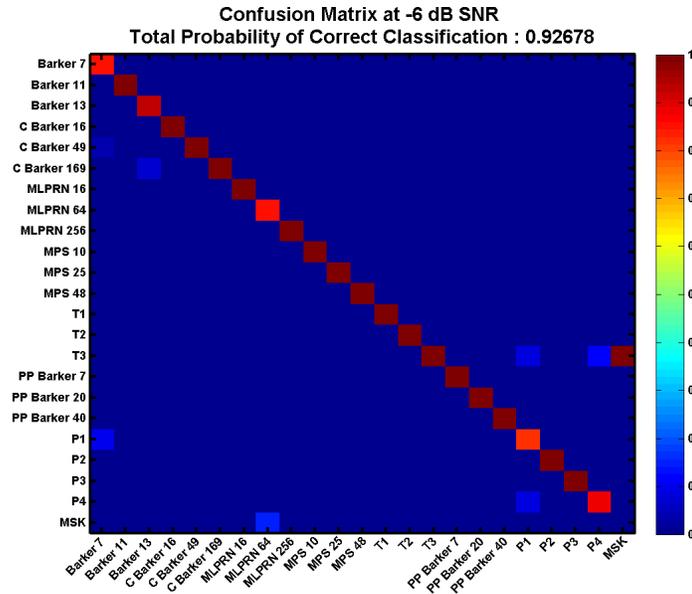


Figure 4.4: Confusion matrix at -6 dB SNR with 10 ensemble average AF sequences.  
Note: X-axis represents actual class membership, Y-axis represents decided class membership

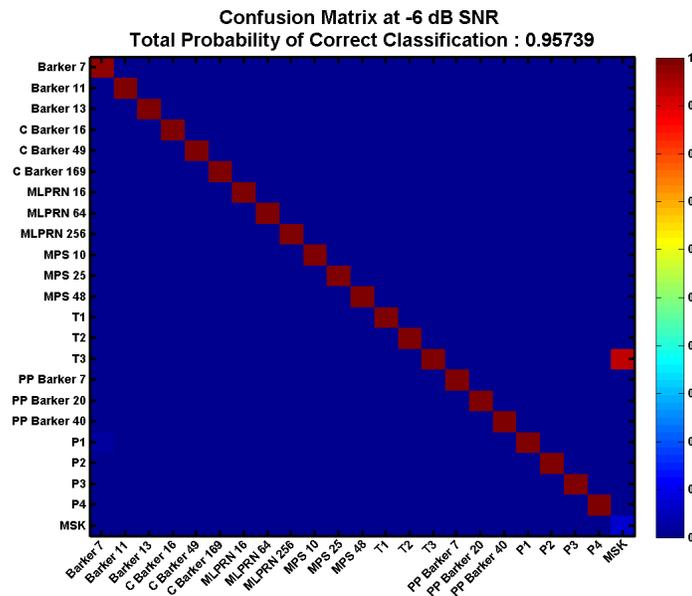


Figure 4.5: Confusion matrix at -6 dB SNR with 20 ensemble average AF sequences.  
Note: X-axis represents actual class membership, Y-axis represents decided class membership

It is interesting to note, however, that while the correct classification rate increases from Figure 4.2 to Figure 4.5, we only observe modest improvement in the classifier’s ability to correctly identify MSK 64 waveforms. Table 4.2 shows the increase in total probability of correct classification at a range of different SNRs when averaging 1, 5, 10, and 20 AF features together. As the number of averaging operations increases, performance reaches a 100% correct classification rate.

Table 4.2: Probability of Correct Classification for 23 Different Modulations Types Across Multiple SNRs and Increasing Number of AF Averaging

SNR (dB)	1 Chain $P_t$	5 Chain $P_t$	10 Chain $P_t$	20 Chain $P_t$
-10	.0432	.0525	.1023	.3205
-8	.0467	.2518	.5744	.8068
-6	.1134	.8066	.9268	.9574
-4	.5667	.9488	.9622	.9782
-2	.8929	.9673	.9882	.9987
0	.9484	.9921	.9999	1
2	.9700	.9998	1	1
4	.9925	1	1	1
6	.9993	1	1	1
8	.9999	1	1	1
10	1	1	1	1

### 4.3 Classification Comparison: Ambiguity-Based vs Autocorrelation-Based

Until this point, our system design and experimentation methodology has mirrored that of [9]. Now, we can fairly compare the performance of the two different feature sets, holding all else constant. By comparing the correct classification rates in Table 4.2 with its counterpart in [9], the ambiguity-based features provide an apparent increase in correct classification rate. If we consider a “good” classifier as one that achieves a correct classification rate of 90%, the ambiguity-based classifier has an effective SNR gain of 2 dB when using one feature set realization. Interestingly,

when averaging multiple feature set realizations, the ambiguity-based classifier only has about a 1 dB effective SNR gain over the autocorrelation-based classifier. This provokes further comparison of total performance with respect to SNR for 1, 5, 10, and 20 AF ensemble averages, as shown in Figures 4.6-4.9. In theory, a perfect classifier would resemble a step function in shape where the correct classification rate is always maximum. In reality, such systems are not achievable, so instead we seek a system in which the area under the curve is greatest. In each of the figures, the AF-based classifier out-performs the ACF-based classifier in terms of greatest area under the curve.

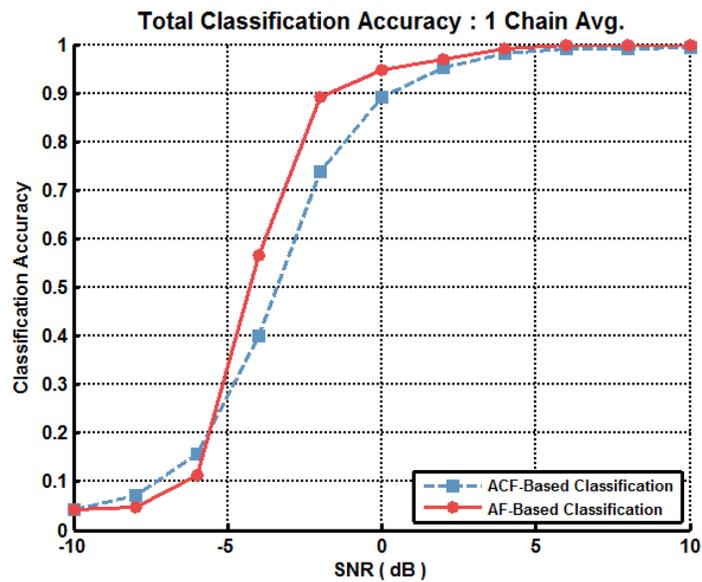


Figure 4.6: A comparison of total classification accuracy as a function of SNR with 1 feature set realization. Observe the effective SNR gain of 2 dB at a fixed classification rate of 90%.

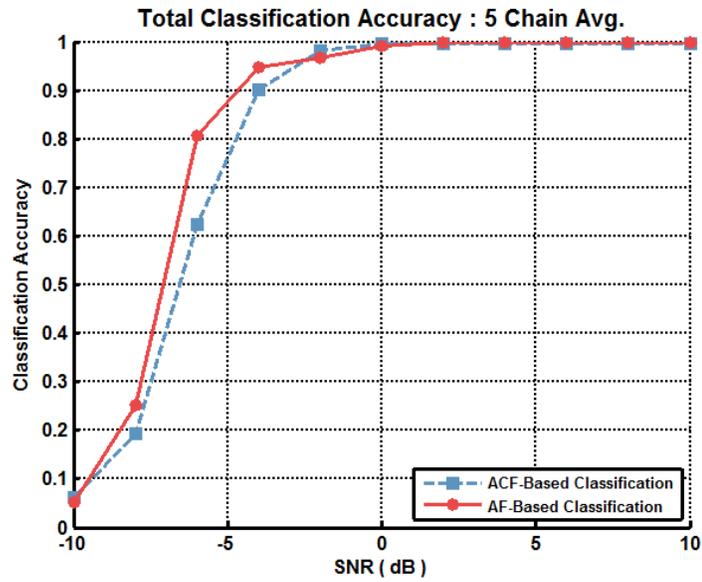


Figure 4.7: A comparison of total classification accuracy as a function of SNR with 5 feature set realizations averaged. Observe the effective SNR gain of about 1 dB at a fixed classification rate of 90%.

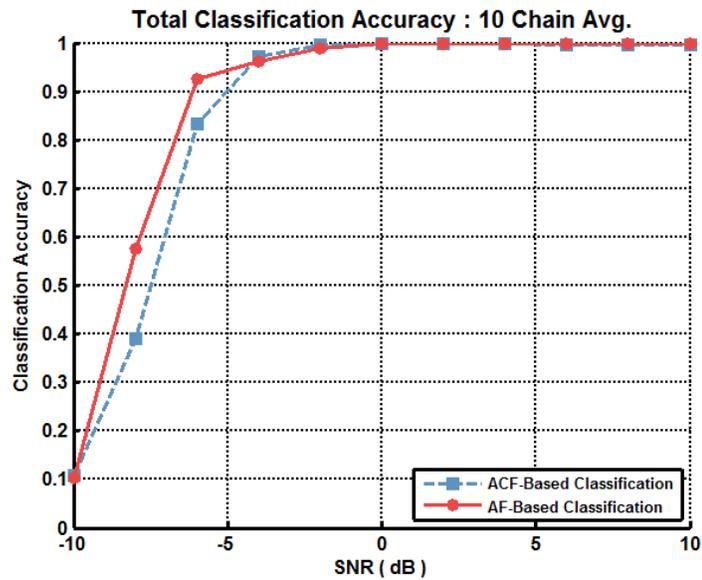


Figure 4.8: A comparison of total classification accuracy as a function of SNR with 10 feature set realizations averaged. Observe the effective SNR gain of about 1 dB at a fixed classification rate of 90%.

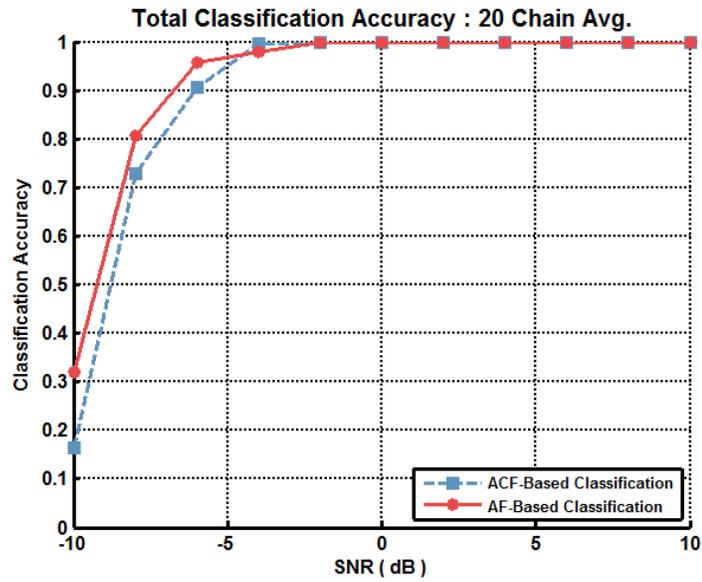


Figure 4.9: A comparison of total classification accuracy as a function of SNR with 20 feature set realizations averaged. Observe the effective SNR gain of about 1 dB at a fixed classification rate of 90%.

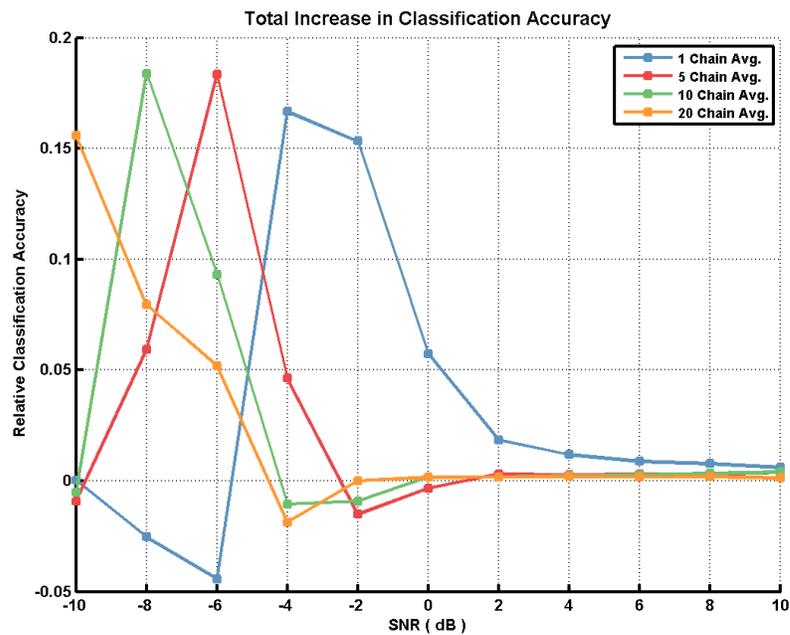


Figure 4.10: Relative correct classification rate increase using ambiguity-based features over autocorrelation-based features as a function of SNR. Observe that relative classification accuracy is increased by over 5% within a 4 dB dynamic range.

A more telling curve to consider is the overall performance difference with respect to SNR. In Figure 4.10, each curve represents a different amount of ensemble averaging across the SNR range. It is evident that the AF features, which contain all of the information in the ACF features, perform better in terms of correct classification rate by about 15% at low SNRs.

## 4.4 Classifier Performance

### 4.4.1 Two-Class Classifiers

Apart from determining classifier performance through correct classification rate, or accuracy, given by a confusion matrix, there exists a standard metric system within the machine learning community to assess classifier performance [11],[12],[13]. While most of these metrics have been defined for two-class, or binary, classifiers, [11] has extended these same metrics to multiclass classifiers. For a standard binary classifier, a confusion matrix will exhibit the form as shown in Table 4.3.

Table 4.3: Confusion Matrix for a Binary Classifier

Data Class	(+) Data	(-) Data
(+) Decision	True Positive ( $t_p$ )	False Positive ( $f_p$ )
(-) Decision	False Negative ( $f_n$ )	True Negative ( $t_n$ )

Because accuracy alone does not provide all of the necessary information to determine how well a classifier performs, metrics including precision, sensitivity (recall), specificity, and F-score are used to capture more of the information contained within the confusion matrix. These metrics exist as proportions of confusion matrix data that aim to answer questions regarding the predictive abilities of a classifier.

For example, it may be important to know the probability that a positive prediction, or decision, is truly positive, defined as precision (positive predictive value). Precision is computed as the ratio between the true number of positive samples and

the number of positive predictions made by the classifier, given by

$$Precision = \frac{t_p}{t_p + f_p}. \quad (4.1)$$

It also may be important to know how well a classifier recognizes positive samples, defined as sensitivity (recall, true positive rate). Sensitivity is computed as the ratio between the true positive predictions and the sum of true positive predictions and false negative predictions, given by

$$Sensitivity = \frac{t_p}{t_p + f_n}. \quad (4.2)$$

Additionally, we may want to know how well a classifier recognizes negative samples, defined as specificity (true negative rate). Specificity is computed as the ratio between true negative predictions and the sum of true negative predictions and false positive predictions, given by

$$Specificity = \frac{t_n}{t_n + f_p}. \quad (4.3)$$

Finally, the F-score ( $F_1$ -score) accounts for the classifier's accuracy in terms of the relation between the actual positive labels and those given by the classifier. The F-score is computed as the harmonic mean of precision and sensitivity, given by

$$F\ score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity}. \quad (4.4)$$

#### 4.4.2 Multi-Class Classifiers

As shown, these standard metrics provide a concise performance analysis tool for the two-class problem. To account for the multi-class problem, [11] and [13] have expanded these equations. This expansion is achieved by redefining the test outcomes

true positive ( $t_p$ ), true negative ( $t_n$ ), false positive ( $f_p$ ), and false negative ( $f_n$ ) with respect to each of the  $C$  classes (i.e.  $t_{p_i}$ ,  $i = 1 \dots C$ ). As an example, Figure 4.11 displays a three class confusion matrix with these redefined parameters.

While the expressions for the standard metrics remain the same, Figure 4.11 displays how test outcomes change for each of the OVA classifiers. As an example, the true negatives for class  $A$  ( $t_{n_A}$ ) are defined as samples that do not belong to class  $A$  and are not classified to class  $A$ . The cells of the confusion matrix that meet that criteria exist in the pink shaded region, and thus, the sum of instances found in these cells produce the number of true negatives for class  $A$ .

What's more, we assess each of our multi-class classifiers individually, in what is referred to as micro-level metrics, and our classification system as a whole, or macro-average metrics [11]. Lastly, because each of our classifiers only output a decided class label, known as a discrete classifier, we also display the operating points in a receiver operating characteristics (ROC) graph [12].

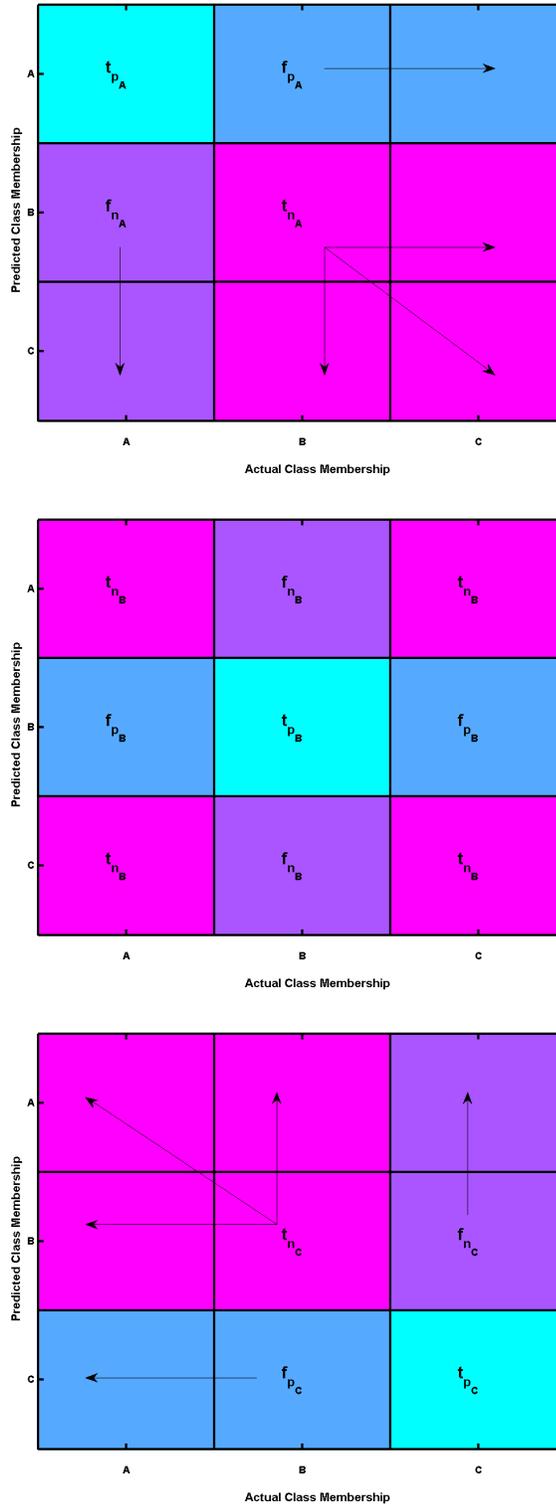


Figure 4.11: All Possible Test Outcomes for the Three-Class Problem Using a Multi-Class Classifier Approach

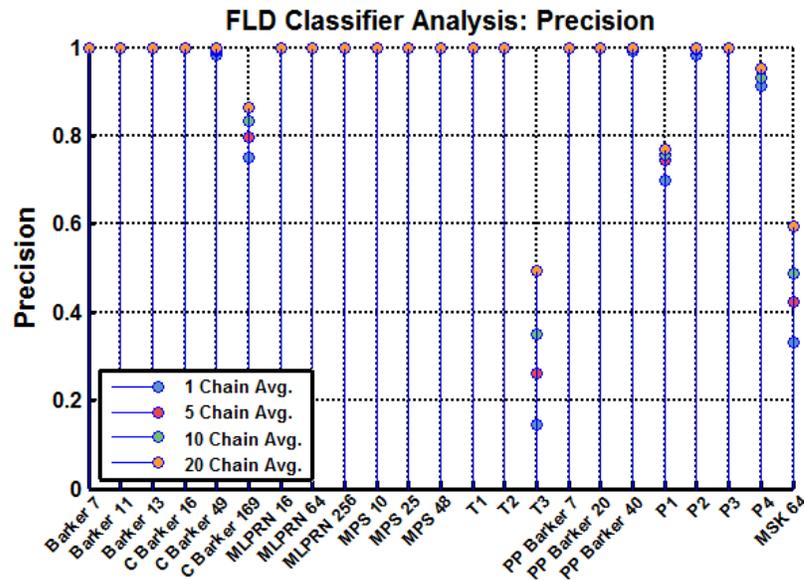


Figure 4.12: Classifier Precision

The precision metric, displayed in Figure 4.12, which denotes the proportion of positive decisions that belong to the positive class, is viewed in the machine learning community as a confidence measure. For all of the micro-level metrics, all of the confusion matrix data from Table 4.2 is used. Micro-level metrics are computed using the aggregate of test data from all SNR levels. As shown, we see that the majority of classifiers have a very high degree of precision, meaning that when a positive decision is made by one of those classifiers, confidence in that decision is strong. Alternatively, several classifiers (e.g. C Barker 169, T3, P1, and MSK 64) have relatively poor precision, thus, when a positive decision is made by one of those classifiers, confidence in that decision is weak. This phenomenon occurs when noise degrades feature vectors to the point where its projection onto a line converges to a certain region, or regions, outside the boundaries of its respective class.

Conversely, classifier sensitivity provides a measure of how well it is able to recognize positive samples. Because the T3 classifier has the worst precision, we expect

many samples to be classified as T3, both correctly and incorrectly. This is proven in Figure 4.13 where the T3 classifier is best at recognizing T3 samples.

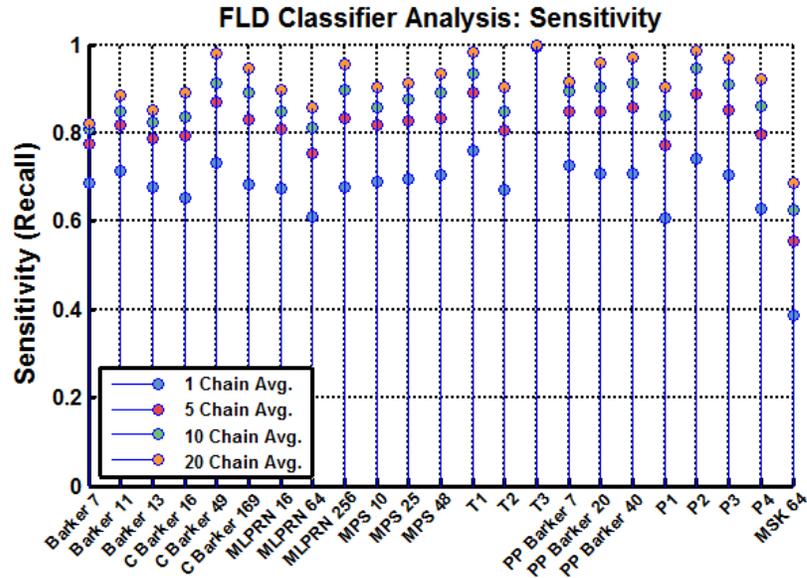


Figure 4.13: Classifier Sensitivity

Additionally, we notice that sensitivity for MSK 64 is comparatively low. Interestingly, low precision and low sensitivity of MSK 64 decisions are products of our classification system. Because we only train  $C - 1$ , or 22, classifiers, an MSK 64 classification is only made if a positive decision for the other 22 classes cannot be made, thus, treating the MSK 64 as an *else* decision. As an inverse metric, specificity assesses our classifiers in terms of its ability to recognize negative samples. As shown in Figure 4.14, most of our classifiers are near perfect at recognizing negative samples.

From Figure 4.14, only the T3 classifier has comparatively low specificity, confirming our conclusion that as noise degrades our test features, the projection of those feature vectors converge to the T3 region of our one-dimensional decision region.

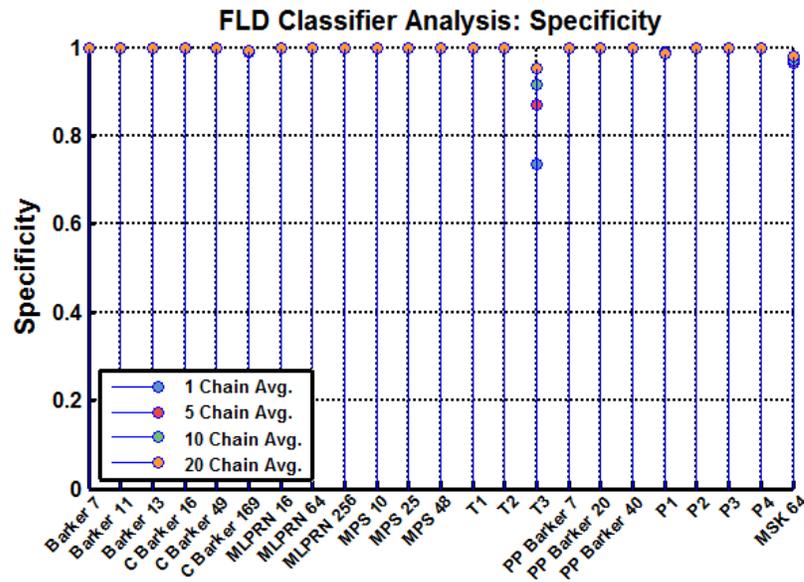


Figure 4.14: Classifier Specificity

Lastly, the F-score, or  $F_1$ -score, is the harmonic mean of precision and sensitivity. In the machine learning community, it is interpreted as a weighted average where both precision and sensitivity are equally important, thus, equally weighted. Note, that depending on the intended application, equal weights may not be appropriate, therefore  $F_2$ -score or  $F_{0.5}$ -score may be used, which will weigh precision lower or higher, respectively. However, in this context, it seems appropriate to evaluate each classifier in terms of requiring that we remain both highly confident in decisions while also being highly sensitive.

Because precision and sensitivity are equally weighted, a performance lapse in any of the two metrics will equally degrade of the resulting F-score. As an example, shown in Figure 4.15, the T3 classifier has such low precision, but high sensitivity, and thus, the resulting F-score, or harmonic mean, suffers. What's more, because most of the classifiers have such a high degree of precision, with varying levels of sensitivity, the F-score provides us the ability to recognize the strongest classifiers.

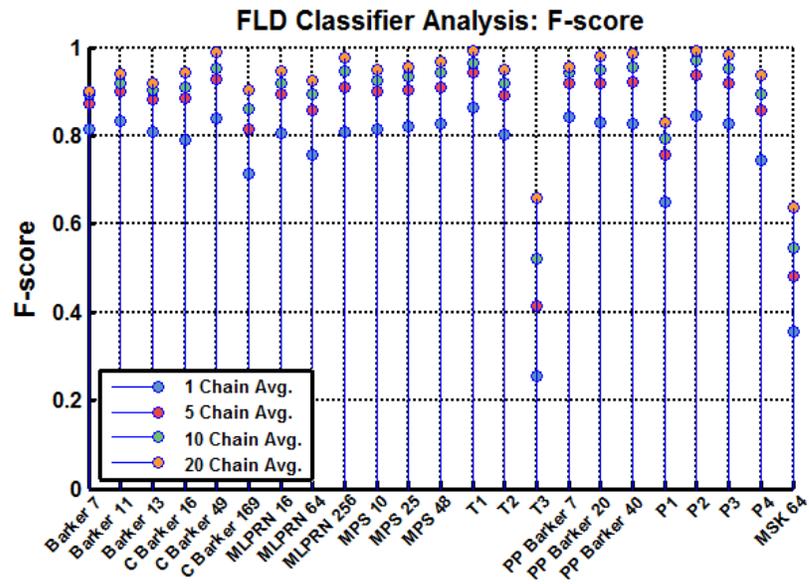


Figure 4.15: Classifier F-score

As shown in Figures 4.12-4.15, most of the individual classifiers perform extremely well in each of the standard metrics. Accordingly, because our classifiers have such strong performance on an individual basis, it follows that the average, or macro-level performance, should also be strong. Below, in Figure 4.16, is the average performance for the classification system in terms of precision, sensitivity, specificity, and F-score.

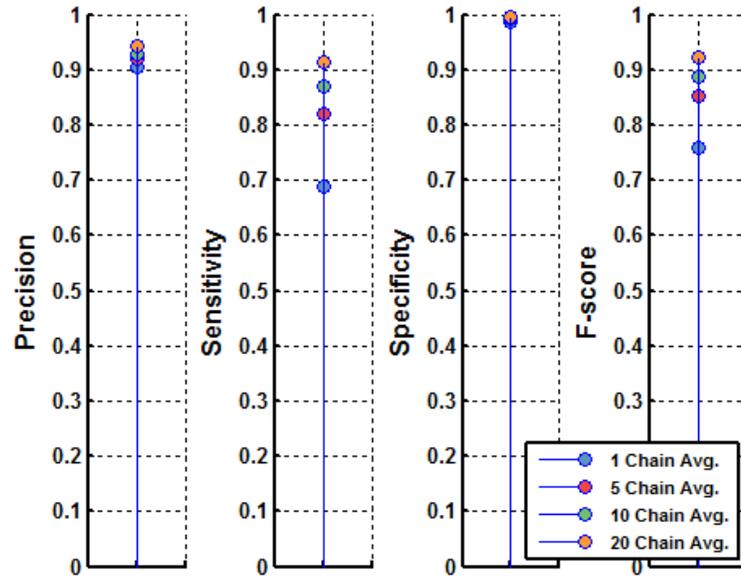


Figure 4.16: Classification System Average Performance

It is important to note that in all of the preceding performance figures, not only is classification accuracy aided by the ensemble averaging of the AF-based features, but also each of the performance metrics is benefited. Most notably, we are able to show how performance is improved. Because precision and specificity are already strong, ensemble averaging does little to improve these. Alternatively, sensitivity, and therefore F-score, is improved drastically. The ability of the feature extraction process to ensemble average multiple ambiguity surfaces allows the system to essentially pull waveforms from the noise, boosting SNR, increasing the dynamic range of effective operation. This can be more traditionally displayed in the discrete ROC graph shown below.

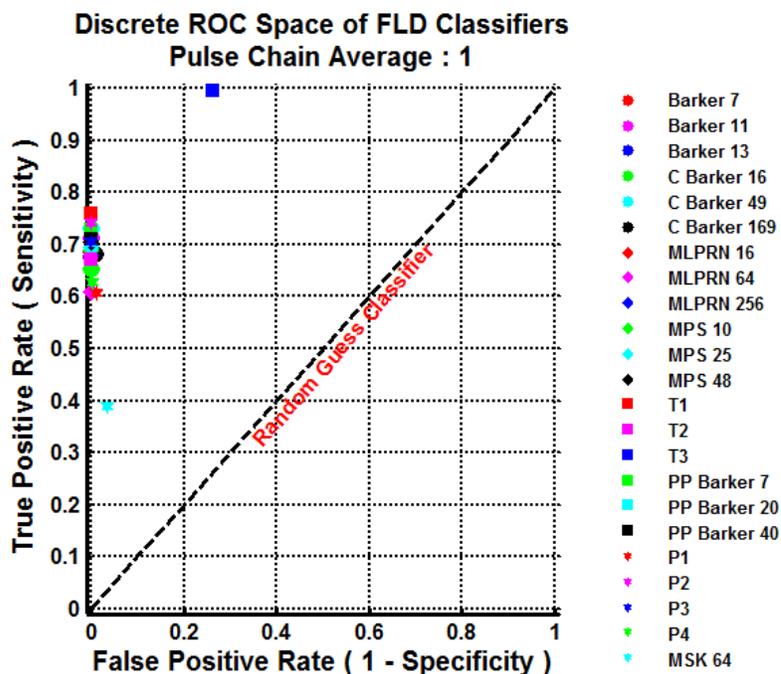


Figure 4.17: ROC Analysis of Discrete Classifiers with 1 AF Sequence

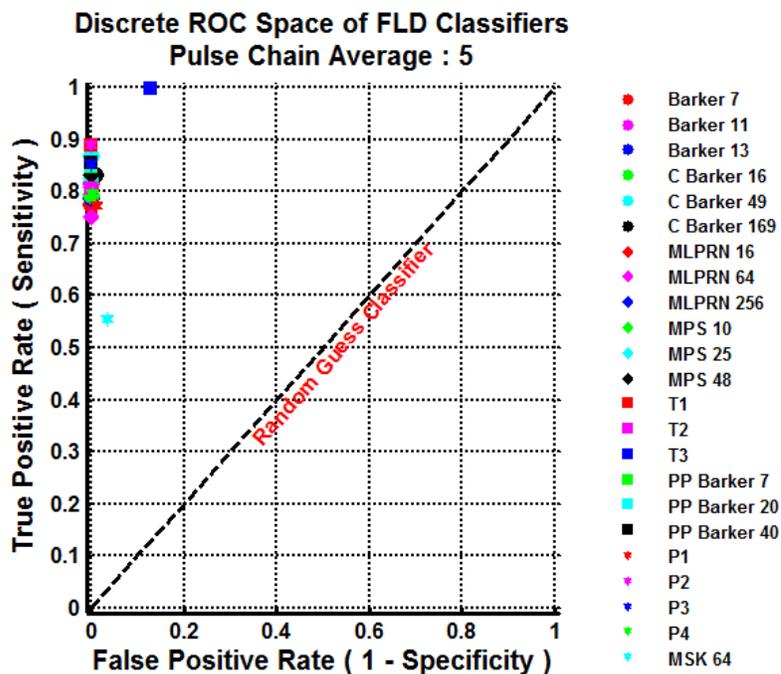


Figure 4.18: ROC Analysis of Discrete Classifiers with 5 Ensemble Average AF Sequences

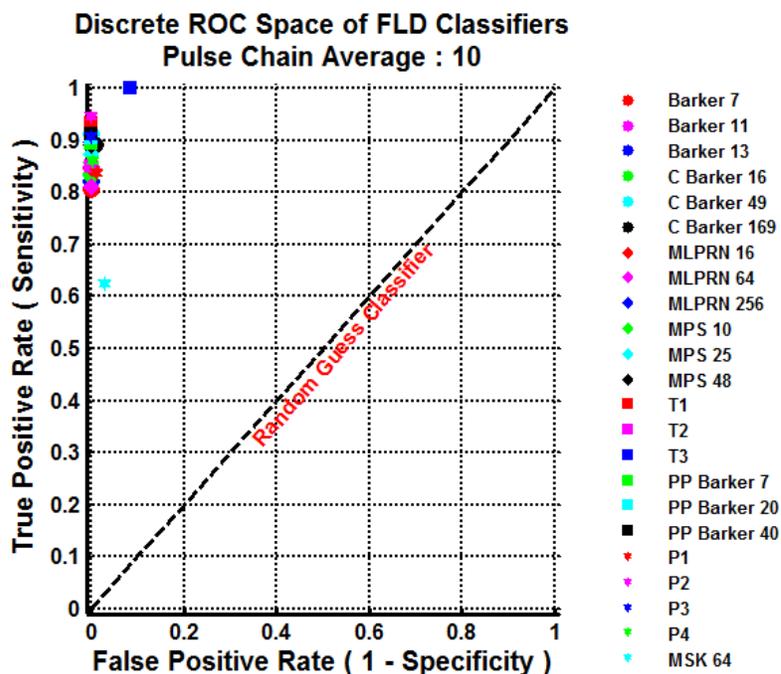


Figure 4.19: ROC Analysis of Discrete Classifiers with 10 Ensemble Average AF Sequences

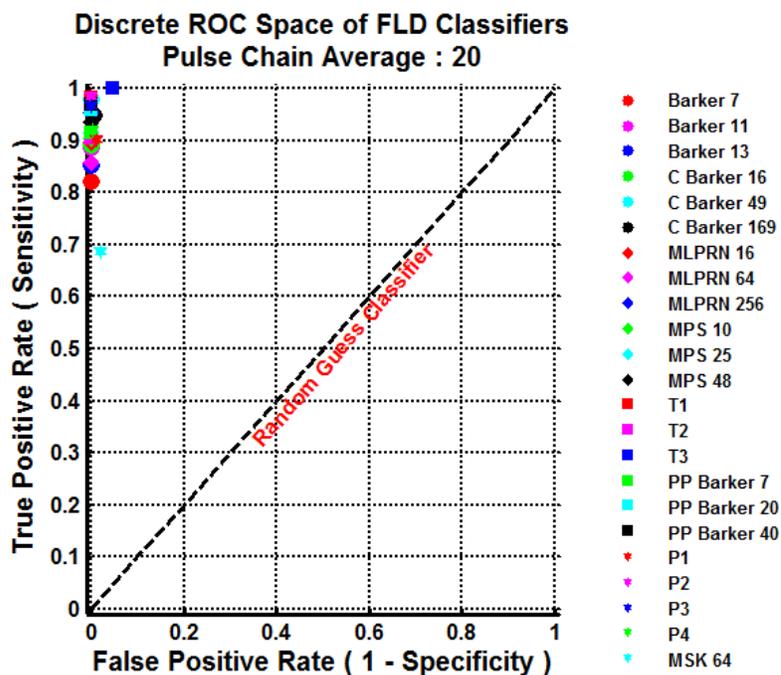


Figure 4.20: ROC Analysis of Discrete Classifiers with 20 Ensemble Average AF Sequences

From the ROC figures, we are again able to see the improvement made by averaging multiple AF features. Because the false positive rate is already near zero, much of the improvement is seen in the true positive rate, or sensitivity. It is evident that we approach what is termed as a perfect classifier (true positive rate  $\rightarrow 1$ , false positive rate  $\rightarrow 0$ ) as the effective SNR of our feature set approaches that of our training data.

## CHAPTER 5

# CONCLUSION

The waveform classification system detailed in this thesis shows promising results even while compared to the autocorrelation-based classifier in [9]. Our initial hypothesis stated that a waveform classification system using features derived from the ambiguity function of a radar pulse should at least achieve similar results to that of a system using the autocorrelation of a radar pulse given that all of the features of the autocorrelation are contained within the corresponding ambiguity-based features. In this work, we test our hypothesis by designing a system which parallels the design of the autocorrelation-based classification system, found in [9]. Our only alteration to the system design is in the feature extraction process where we substitute in ambiguity-based features with the caveat that all of the features of the autocorrelation are contained within the ambiguity-based features.

From our analysis, we have shown that radar waveform classification is improved by using the ambiguity-based features. While the main objective of this work was to provide a fair comparison of two waveform classification systems, we provide further insight into the robustness of the classifier by evaluating it against standard metrics used in the machine learning community.

Future work may consist of evaluating the robustness of this algorithm to various degradations of the pulse profile. These degradations could include unintended modulations induced on the pulse as a result of imperfections in the transmitter and receiver hardware, imperfections in envelope detection that result in partial pulse testing, and even the effect of channel conditions including fading, clutter, and multipath.

# REFERENCES

- [1] N. Whittall, “Signal sorting in esm systems,” *Communications, Radar and Signal Processing, IEE Proceedings F*, vol. 132, no. 4, pp. 226–228, 1985.
- [2] J. Rogers, “Esm processor system for high pulse density radar environments,” *Communications, Radar and Signal Processing, IEE Proceedings F*, vol. 132, no. 7, pp. 621–625, 1985.
- [3] D. Wilkinson and A. Watson, “Use of metric techniques in esm data processing,” *Communications, Radar and Signal Processing, IEE Proceedings F*, vol. 132, no. 4, pp. 229–232, 1985.
- [4] A. Singh and K. Subba Rao, “Detection, identification and classification of intra pulse modulated lpi radar signal using digital receiver,” *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, pp. 433–440, 2012.
- [5] Z. Yu, C. Chen, and W. Jin, “Radar signal automatic classification based on pca,” in *Intelligent Systems, 2009. GCIS’09. WRI Global Congress on*, vol. 3. IEEE, 2009, pp. 216–220.
- [6] Y. Fei, L. Jingqing, and L. Jiuming, “Radar emitter signal fractal feature based on wavelet transform,” in *Radar, 2006. CIE’06. International Conference on*. IEEE, 2006, pp. 1–4.
- [7] M. Ren, J. Cai, Y. Zhu, and M. He, “Radar emitter signal classification based on mutual information and fuzzy support vector machines,” in *Signal Processing, 2008. ICSP 2008. 9th International Conference on*. IEEE, 2008, pp. 1641–1646.
- [8] T. Chen, W. Jin, and Z. Chen, “Feature extraction using wavelet transform for radar emitter signals,” in *Communications and Mobile Computing, 2009. CMC’09. WRI International Conference on*, vol. 1. IEEE, 2009, pp. 414–418.
- [9] B. D. Rigling and C. Roush, “Acf-based classification of phase modulated waveforms,” in *Radar Conference, 2010 IEEE*. IEEE, 2010, pp. 287–291.
- [10] J. Lundén and V. Koivunen, “Automatic radar waveform recognition,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 1, pp. 124–136, 2007.
- [11] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.

- [12] T. Fawcett, “An introduction to roc analysis,” *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [13] C. Beleites, R. Salzer, and V. Sergo, “Validation of soft classification models using partial class memberships: An extended concept of sensitivity & co. applied to grading of astrocytoma tissues,” *Chemometrics and Intelligent Laboratory Systems*, vol. 122, pp. 12–22, 2013.
- [14] B. R. Mahafza and A. Elsherbeni, *MATLAB simulations for radar systems design*. CRC press, 2003.
- [15] N. Levanon and E. Mozeson, *Radar signals*. John Wiley & Sons, 2004.
- [16] E. Mozeson and N. Levanon, “Matlab code for plotting ambiguity functions,” *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1064–1068, 2002.
- [17] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.