Title
Advanced Spectrum Sensing for Multiple Transmitter Identification

Author
Urriza, Paulo Isagani Malijan

Publication Date
2015

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Advanced Spectrum Sensing for Multiple Transmitter Identification

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical Engineering

by

Paulo Isagani Malijan Urriza

2015

Abstract of the Dissertation

Advanced Spectrum Sensing for Multiple Transmitter Identification

by

Paulo Isagani Malijan Urriza

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2015

Professor Danijela Cabric, Chair

The exponential growth in demand for mobile data has led to significant research efforts aimed at more efficient methods of utilizing the scarce RF spectrum resource. One potential solution to this scarcity problem is Cognitive Radio (CR) which involves dynamic spectrum access in which a set of unlicensed users occupy spectrum holes without causing significant degradation of performance to the incumbent users. A key enabling technology for CR networks is accurate spectrum sensing which aims to learn the radio environment in order to adapt the CR transmission. Traditional spectrum sensing techniques have mainly focused on determining only the presence or absence of a licensed user. Recent work in the past few years have shown however that more detailed knowledge pertaining to radio-scene analysis can be used to improve the performance of CR networks. The more the secondary user knows about the active licensed users, the better it can adapt its transmission strategies. In this work, we put forward the concept of advance spectrum sensing which takes a multi-dimensional approach to radio-scene analysis that estimates various parameters of the active transmitters through sensing, localization and tracking, modulation classification, PHY parameter estimation, and MAC-layer classification.

In this work we investigate the elements of such an advance spectrum sensing

system. Firstly, we will look at the problem of conventional spectrum sensing, or detecting the presence or absence of transmitting Primary Users. In particular, we study how detection performance could be improved through the use of cyclostationary feature detection and how it could be made robust to fading, noise uncertainty, and co-channel interferers through the optimal use of multiple sensors. Second, we attack the problem of modulation classification which we argue is a critical piece of information for future cognitive radio systems. We present a new type of pattern classification algorithm based on the concept of sampled distribution distance which offers a low computational complexity alternative to maximum-likelihood based classification. Through our extensive analysis, we have derived the optimal form of this type of classifier and applied it to the modulation classification problem. Finally, we propose a system of MAC-layer classification based on 4th-order cumulants which distinguishes between TDMA, OFDMA, CDMA and contention-based schemes. In addition, it is also able to jointly perform modulation classification with channel access method. The analysis of the statistics of the 4th-order cumulant used in our work also offers large potential for applications in different areas including modulation classification, channel estimation, and estimation of number of users.

The dissertation of Paulo Isagani Malijan Urriza is approved.

Songwu Lu

Babak Daneshrad

Dejan Markovic

Danijela Cabric, Committee Chair

University of California, Los Angeles

2015

For Papa, Mama, and Ate Jo ...

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, all the glory be to God, with whom nothing is impossible.

I would like to acknowledge all the people who supported me through all the years of getting my PhD. My sincerest gratitude to my adviser Prof. Danijela Cabric for her trust, guidance, and confidence in my ability to explore original ideas in a challenging research environment. I would like to thank Prof. Markovic, Prof. Daneshrad, and Prof. Lu for serving in my PhD defense panel.

My most heartfelt thanks go to all the CORES lab members/visitors/supporters that have been with me at one point or another in this arduous journey of getting my doctoral degree. Thanks to Wesam, Varun, Jihoon, and Jared who have made the small office space such a big source of fun and adventure. Thanks to the CORES lab guys, who at the time of writing this are still in the midst of the struggle which is graduate school; Chun-hao, Shailesh, Jason, and Han. Special thanks to my great collaborators, Eric, Przemek, Jun and Mihir. These four people have been instrumental to all the work that I have reported here in this dissertation. Thanks also to all the great friends I have made in UCLA Electrical Engineering, especially Kirti, Yifan, Danesh, and Shaunak.

I'm also greatful for all the people that have helped me outside of UCLA. Thanks to my FDCC family for their spiritual support. To all the great friends I have made throughout my years as a student both outside and inside of the campus setting. Thank you.

I would like to thank my family to whom this disseratation is dedicated. You have been with me through thick and thin and have always believed in me. Thanks for the support from the time I was applying to grad school up to my very last day in UCLA.

Finally, thank you so much to Ms. Sofia Pagdilao who has been my greatest ally throughout times of trials and my dearest friend when I needed one.

Vita

2007	B.S. (Computer Engineering), University of the Philippines Diliman, Quezon City, Philippines
2007-2009	Teaching Associate, University of the Philippines. Taught discussion classes on Logic Design, Proabability and Signals and Systems
2009	M.S. (Electrical Engineering), University of the Philippines Diliman, Quezon City, Philippines
2009-2014	Graduate Student Researcher, UCLA CORES Lab
2010-2012	Teaching Assistant, UCLA. Taught discussion classes on Digital Logic Design and Signals and Systems

Publications

P. Urriza, E. Rebeiz, D. Cabric, *"Optimal Discriminant Functions Based On Sampled Distribution Distance for Modulation Classification"*, IEEE Communications Letters, August 2013.

E. Rebeiz, F.-L. Yuan, P. Urriza, D. Markovic, D. Cabric, *"Energy-Efficient Processor for Blind Signal Classification in Cognitive Radio Networks"*, IEEE Transaction on Circuits and Systems I, June 2013.

P. Urriza, E. Rebeiz, D. Cabric, *"Multiple Antenna Cyclostationary Spectrum*

Sensing Based on the Cyclic Correlation Significance Test", IEEE Journal on Selected Areas in Communications: Cognitive Radio Series, May 2013.

E. Rebeiz, P. Urriza, D. Cabric, *"Optimizing Wideband Cyclostationary Spectrum Sensing under Receiver Impairments"*, IEEE Transactions on Signal Processing, vol. 61, no. 15, pp. 3931-3943. August 2013.

J. Wang, P. Urriza, Y. Han, and Cabric, *"Weighted Centroid Algorithm for Estimating Primary User Location: Theoretical Analysis and Distributed Implementation"*, IEEE Transaction on Wireless Communications, October 2011.

P. Urriza, E. Rebeiz, P. Pawe?czak, and Cabric, *"Computationally Efficient Modulation Level Classification Based on Probability Distribution Distance Functions"*, IEEE Communications Letters, May 2011.

E. Rebeiz, P. Urriza, D. Cabric, *"Experimental Analysis of Cyclostationary Detectors Under Cyclic Frequency Offsets"*, accepted for publication in Asilomar Conference on Signals, Systems and Computers 2012

P. Urriza, E. Rebeiz, D. Cabric, *"Eigenvalue-based Cyclostationary Spectrum Sensing Using Multiple Antennas"*, IEEE Global Communications Conference (GlobeCom). Anaheim, CA. 2012.

P. Urriza, E. Rebeiz, D. Cabric, *"Hardware Implementation of Kuiper-based Modulation Level Classification"*, in Proc. Asilomar Conference on Signals, Systems, and Computers. Pacific Grove, CA, USA. 6-9 Nov. 2011.

J. Wang, P. Urriza, Y. Han, and D. Cabric, *"Performance Analysis of Weighted*

Centroid Algorithm for Primary User Localization in Cognitive Radio Networks", in Proc. Asilomar Conference on Signals, Systems, and Computers. Pacific Grove, CA, USA. 7-10 Nov. 2010.

P. Urriza, J. Marciano, *"Combining Parallel Sequence Spread Spectrum (PSSS) with OFDM - Concept and Simulation Results"*, in Proc. Wireless Communications and Networking Conference (WCNC). Sydney, Australia. 18-21 Apr. 2010.

G. Fadera, L. Ignacio, M. Nastor, P. Urriza, J. Marciano, *"FPGA implementation of space-time encoders"*, in Proc. Intelligent and Advanced Systems (ICIAS). Kuala Lumpur, Malaysia. 25-28 Nov. 2007.

CHAPTER 1

Introduction

There is an ever increasing demand for the Radio Frequency (RF) spectrum resource as a result of higher data rate applications brought about by transitioning from voice-only to multimedia communications. This trend is further amplified by the explosive growth of mobile devices as a result of both their decreasing cost as well as their further integration into our society. However, the RF spectrum resource is naturally limited due to both physics as well as the current state of radio technology. This presents an unsustainable situation for future wireless systems. This problem has led to significant research effort which has been aimed at more efficient methods of utilizing this scarce resource.

One potential solution to the scarcity problem is Cognitive Radio (CR). A popular paradigm for CR involves non-interfering dynamic spectrum access, where a set of unlicensed users occupy spectrum holes left by licensed users without causing significant degradation of performance to the incumbent users. By knowing as much information as possible about nearby licensed users, the probability of finding these opportunities in an accurate and timely manner and subsequently using it with minimal interference to the primary user, can be significantly improved. Traditional spectrum sensing techniques have mainly focused on determining only the presence or absence of a licensed user. In this dissertation, we put forward the concept of advance spectrum sensing which takes a multi-dimensional approach to radio-scene analysis that involves creating an up-to-date "map" of active licensed users, including additional information about their geographical location,

RF transmission parameters, and higher-layer parameters. Our main goal has been to develop accurate yet energy efficient algorithms and architectures to perform RF transmit parameter estimation, modulation classification and MAC-layer classification for use in CR networks.

1.1 Cognitive Radio

In order to address the RF spectrum scarcity problem, innovative techniques must be developed with the aim of providing new and more efficient ways of utilizing the available spectrum. In several spectrum measurement campaigns [Yan05, SCZ10, WRP09, MTM06], it has been shown that the fixed allocation of spectrum currenty employed by the vast majority of the world's regulatory agencies (such as that of the FCC, shown graphically in Fig. 1.1) is very inefficient and lead to most channels being underutilized. One such result, from measurements done in Chicago, IL, is presented in Fig. 1.2. We observe that a significant portion of the spectrum allocated to licensed services show little to no usage over time, with all observed channels being used $< 25\%$ of the time on the average.

Cognitive Radio or CR is a concept that has shown a lot of promise and attracted a lot of interest in recent years. The main idea behind CR networks (CRN) is the exploitation of existing spectrum holes. These spectrum holes are opportunities for a secondary user (SU), either temporally, spectrally, or more recently spatially reuse resources intended for a licensed or primary user (PU). Through CR, the spectrum utilization can be improved significantly and the false scarcity caused by fixed allocation can potentially be eliminated.

While there is no agreement yet on what a cognitive radio specifically entails, the concept has evolved recently to include various meanings in several contexts [Nee06, Sec. 1.1.1]. In this dissertation, we will use the definition adopted by Federal Communications Commission (FCC): "Cognitive radio: A radio or system

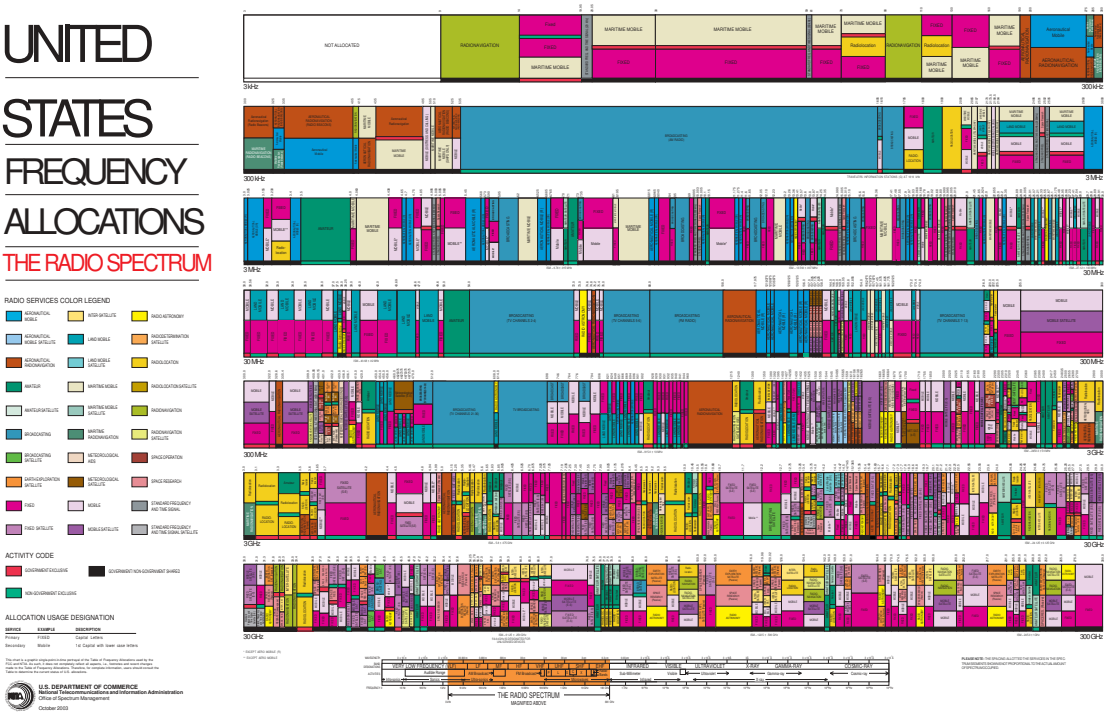


Figure 1.1: The NTIA's frequency allocation chart showing the fixed allocation of the radio spectrum.

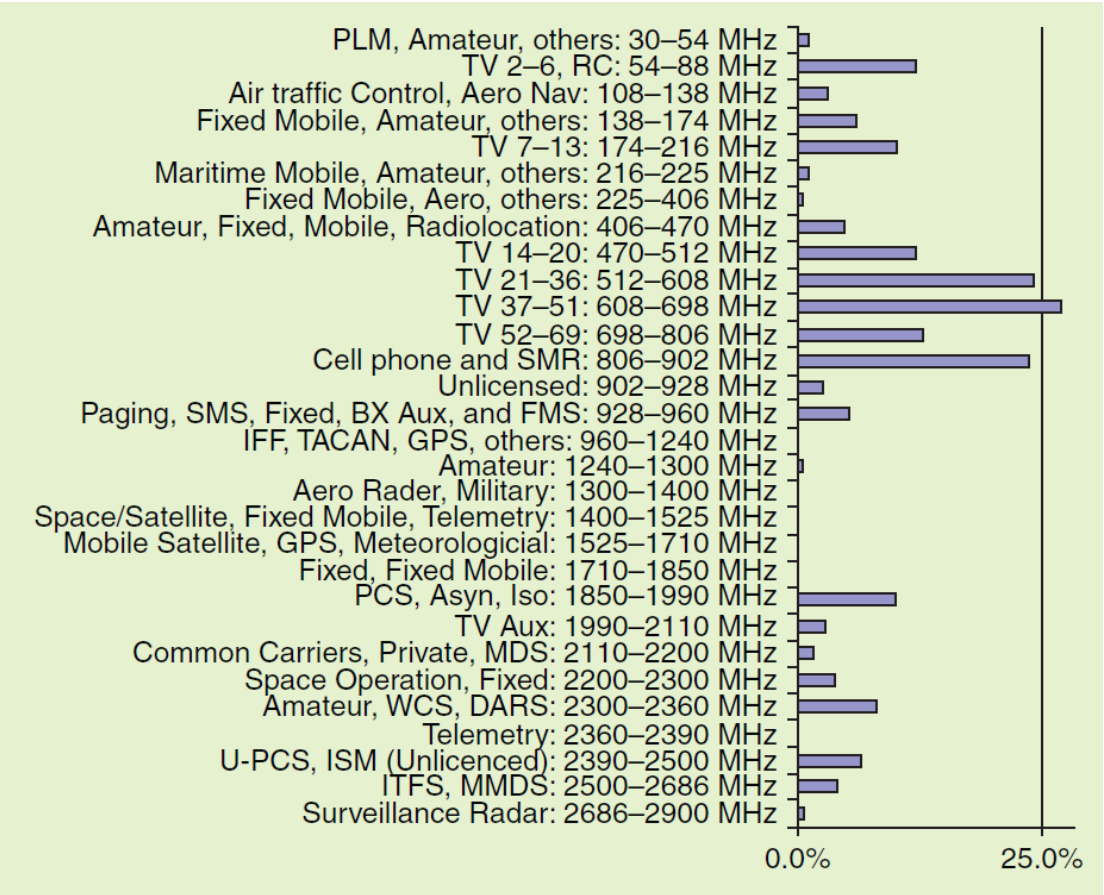


Figure 1.2: Average spectrum utilization taken over multiple locations [MTM06]

that senses its operational electromagnetic environment and can dynamically and autonomously adjust its radio operating parameters to modify system operation, such as maximize throughput, mitigate interference, facilitate interoperability, access secondary markets." [Fed05]. Therefore by this definition, CRs must have the ability to exploit locally unused spectrum with the aim of improving spectral efficiency.

Two major tasks of a CRs become apparent when we apply this definition. A CR must be able to *explore* and *exploit* a spectrum hole or opportunity [20111a]. Spectrum exploration is being able to identify any available and free spectrum that is unused by the licensed user. This involves spectrum sensing techniques in its various forms (see [YA09] and references therein). Exploiting the spectrum hole on the other hand involves, among others, adapting the transmit waveform and other parameters in order to access the available resource while minimizing interference to the PU. The basic outline of how a cognitive radio system is employed to dynamically access spectrum is illustrated in Fig. 1.3. In general we define the transmitters that are licensed to access the specified spectrum as Primary Users (PU). On the other hand, our focus will be on another set of radios that attempt to access the same spectrum opportunistically. We will refer to these radios as Secondary Users (SU) throughout this manuscript. These SUs explore the spectrum environment using a spectrum sensing system and subsequently modify their access strategy to enable transmission with minimal interference to the PU network.

In this dissertation we want to focus on developing practical algorithms for the exploration phase as well as to highlight potential applications of these algorithms in improving the exploitation phase when applicable. In the following subsection we will present the idea of multi-dimensional spectrum sensing which is aimed at further increasing the probability of finding opportunities for resource reuse in contrast to traditional spectrum sensing.

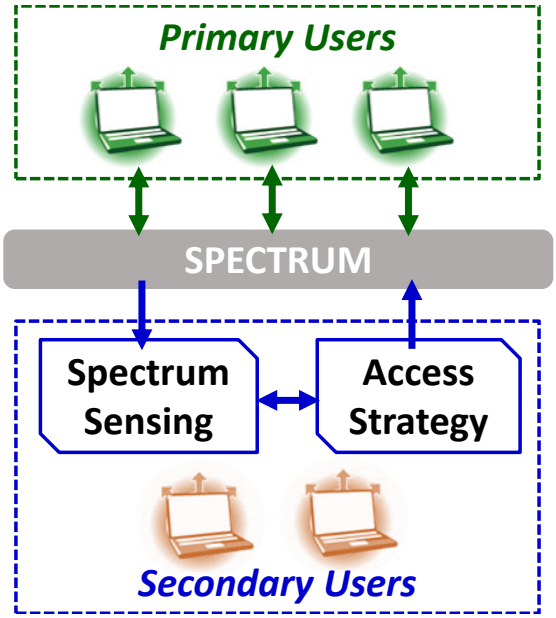


Figure 1.3: General block diagram of Cognitive Radio

1.2 Motivations and Challenges for Advanced Spectrum Sensing

The basic function of any cognitive radio can be simplified into a cycle of exploration and exploitation. This cycle was first popularly presented in the work of Mitola and Maguire [MM99]. A simplified version of this cycle from Haykin [Hay05] is illustrated in Fig. 1.4. The main focus of this dissertation is what Haykin refers to as Radio-Scene analysis. The ultimate goal of which is to accurately quantify the radio environment in order to find as many opportunities where the SU can transmit.

The area of radio-scene analysis has grown over the years to become a very expansive and complicated subject. Many techniques aimed at observing different aspects of the RF environment have been proposed. The most basic of these, and clearly the most important for CR, is detection of spectrum occupancy. Research on this area has mainly focused on the development of techniques that detect the *presence* or *absence* of PU signals with high reliability, specified by the detection

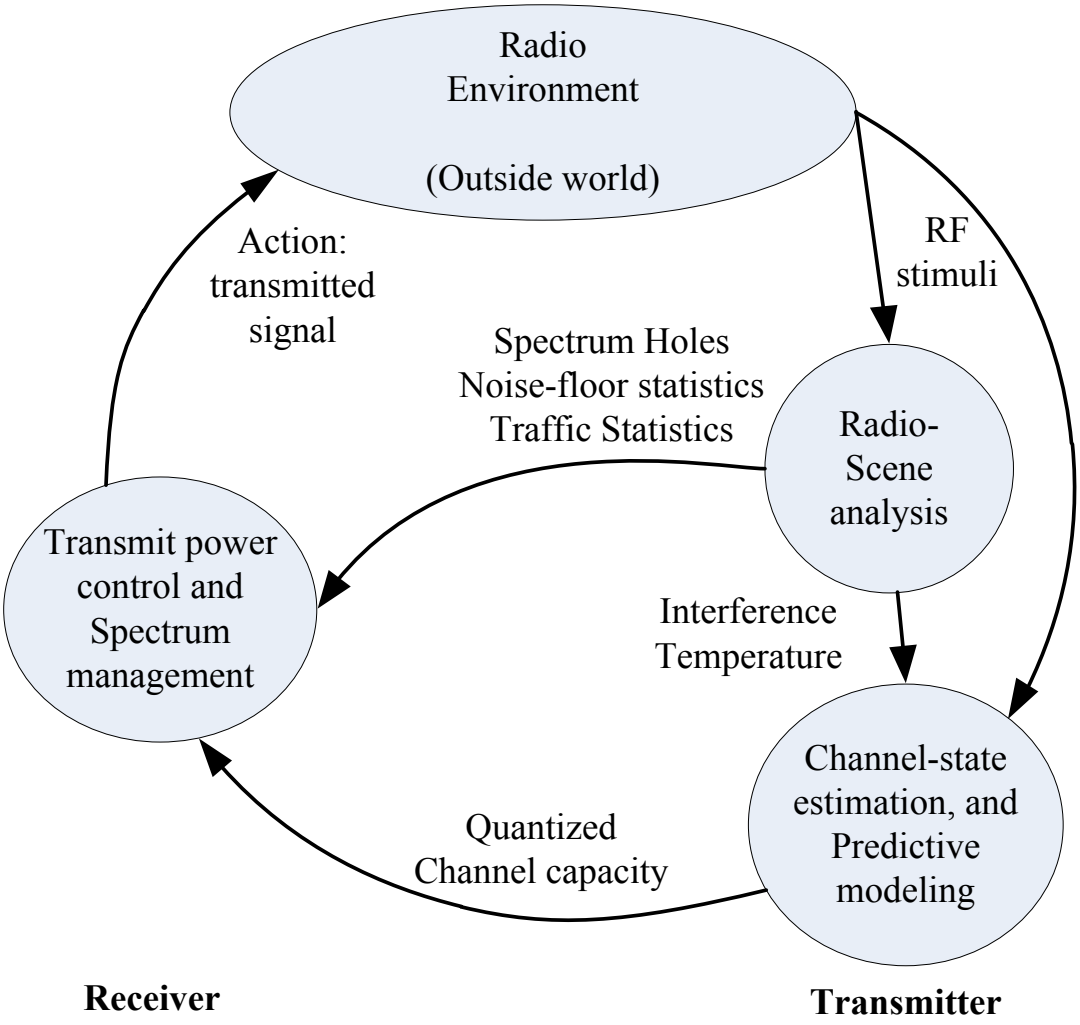


Figure 1.4: Cognitive Cycle [Hay05]

probability and false alarm probability, using the minimum possible number of observations, to minimize the detection delay.

In this dissertation, we put forward the idea of an *Advanced Spectrum Sensing System* which extends the radio-scene analysis into a multi-dimensional problem. Conventionally, spectrum sensing is done as a binary hypothesis test. This means that either of two hypotheses are true. Either the null hypothesis: the primary user is absent or the alternative hypothesis, the primary user is present. There is however, a drawback to this approach because it limits the opportunities available for dynamic spectrum access.

If for example, we had access to the following additional information about the primary user we can adopt more sophisticated methods.

1. **Location:** If we knew where the primary users are located, we can use some form of spatially-aware power control, reducing transmit power to avoid interference.

2. **Modulation Format:** If we knew the modulation type being used by the transmitter we can choose an access method that minimizes the interference to these users or underlay access by performing interference cancellation techniques

3. **RF Transmit Parameters:** If we knew the RF parameters such as the transmit power, and where the signal is located in frequency then we can identify the standards being used and choose our access strategy accordingly.

4. **MAC Scheme:** If we know which multiple access scheme the primary network is using, be it TDMA, CSMA, or OFDMA, then we can tailor our access strategy accordingly.

Knowing this multi-dimensional information about the primary network has been the main goal of this dissertation.

Advance Spectrum Sensing for cognitive radio is particularly challenging because of the following challenges:

1. **Non-cooperative Primary Network:** In CR applications the primary network is typically non-cooperative. This means that there is no exchange of information between primary and secondary networks. This requires the SU network to utilize blind estimation and classification techniques that only require passive information that could be measured without cooperation from the PU network.

2. **Time Varying Environment:** As with all wireless applications our system also need to deal with the rapidly varying radio environment. This motivates us to devolop techniques that work with low computational complexity in real time. We need to have fast converging algorithms because there is an inherent tradeoff between the amount of time we spend observing the spectrum and our throughput. Our algorithms also need to be robust to very low SNR and interference.

3. **Large Variety of Systems:** There is a large variety of systems that follow very different standards and protocols. We need approaches that generalize for a large number of possible standards. In the hypothetical case of a hostile environment, the methods can possibly follow non-standardized protocols.

An example of such a challenging scenario where these issues are present is illustrated in Fig. 1.5. This application is for opportunistic spectrum access in the TV band. The primary users in this band include TV broadcasts and low power devices such as wireless microphones that use analog FM. On the other hand, secondary users such as 802.22 [20111b] compliant systems may now utilize this spectrum. In the future, other ad hoc networks that want to utilize this band will need *advanced spectrum sensing* in order to guarantee non-interference to the primary network as well as coexistence with other secondary networks.

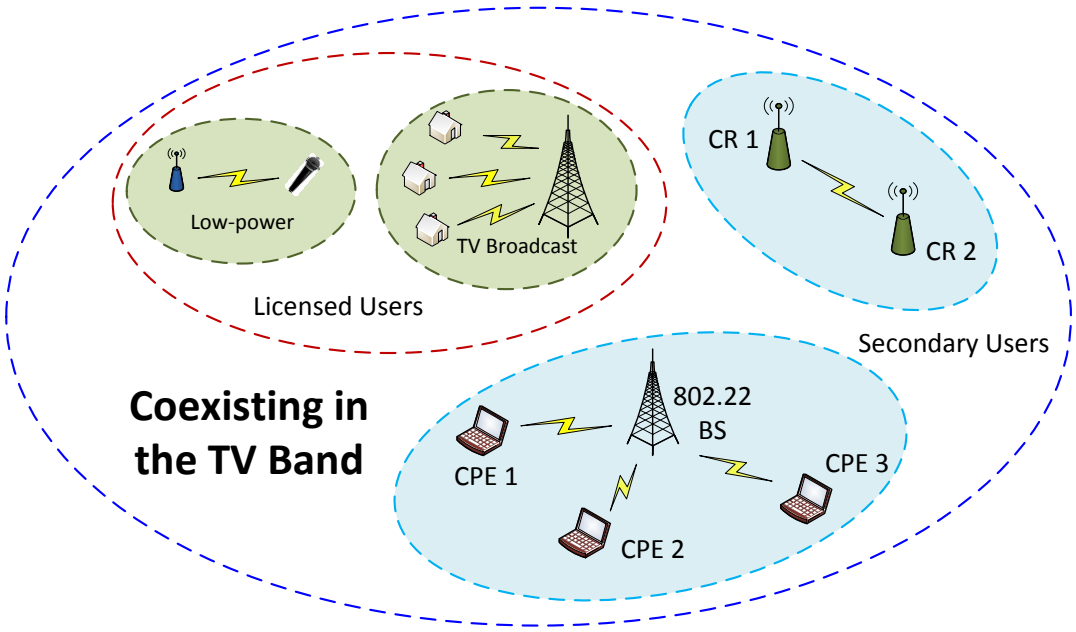


Figure 1.5: Coexistence Scenario in TV Whitespace

Although initially applied to single sensor scenario, it has been shown that cooperative versions of spectrum sensing algorithms provide significantly improved detection performance. Detailed discussions of such algorithms including the accompanying implementation challenges are given in [YA09, GS08, CMB04]. However, according to [GS08], significant gains can be made by shifting from a *reactive* spectrum sensing paradigm to a *proactive* one. In the first case, CRs start to sense the spectrum only when it has some data to transmit and determine occupancy at that given time only; in the second case, CRs maintain an up-to-date "map" or database of PU activity, so as to improve the spectrum sensing responsiveness.

This idea of a PU activity database can further be extended by monitoring not only band occupancy information but also characterizing PUs along multiple dimensions including transmission parameters , modulation, MAC layer scheme, and location, to name a few. This idea of multi-dimensional spectrum awareness is described in [YA09] as a means of finding other venues for resource reuse or

increasing the probability of finding an existing opportunity in the most efficient and timely manner. Examples of these dimensions are multiple codes, angle-of-arrival (AoA), and traffic patterns. In [MS09], the importance of a spatial Received Signal Strength (RSS) profile, and consequently location information was presented. This advanced spatio-temporal spectrum sensing, that is both proactive and multi-dimensional has been seen as the next evolution of this technology, aimed at truly *cognitive* networks.

One aspect of the radio environment that can potentially be utilized by a cognitive radio network (CRN) is the transmission parameters employed by both primary users (PU) as well as competing secondary users (SU). As will be shown in succeeding sections, algorithms that address this issue have received very little attention in the literature. This is especially true when it comes to investigating practical implementation issues in scenarios where there are co-channel or overlapping primary users.

Information about PU location will also enable several key capabilities in CR networks including improved spatio-temporal sensing, intelligent location-aware routing, as well as aiding spectrum policy enforcement. A good example of this is the concept of spatial reuse illustrated in Fig. 1.6. By knowing the location of nearby PUs, a secondary network can theoretically perform power control mechanisms as described in [CGG10] in order to reuse spectrum even when a PU is actively transmitting. This scenario is not possible with the traditional On-Off characterization based on simple energy detection methods.

In this dissertation we investigate practical algorithms for estimating PU transmission parameters which include spectrum activity, modulation class, MAC layer scheme. This forms the basis for a complete advance spectrum sensing system shown in Fig. 1.7. In this block diagram the shaded blocks indicate the focus of this dissertation. We focus on analyzing the context of the primary user activity rather than simply reacting to their absence or presence. The approaches involved

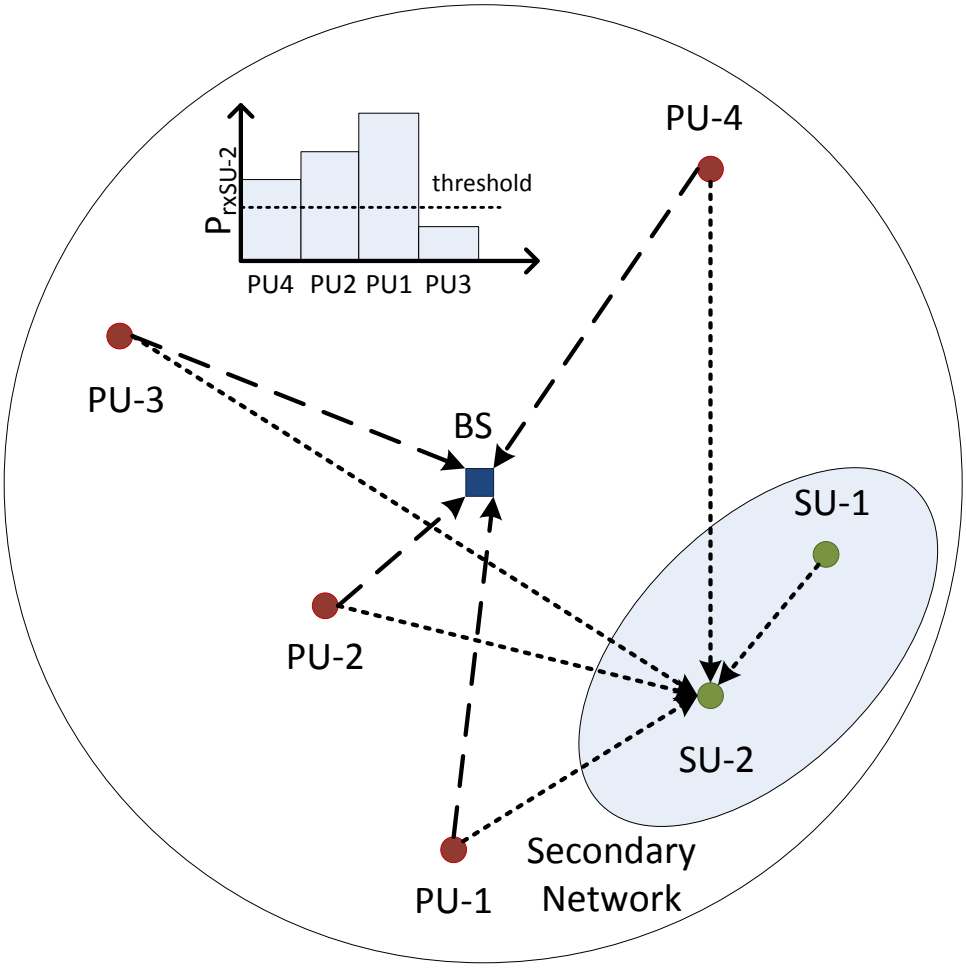


Figure 1.6: Importance of location information for spatial reuse [CGG10].

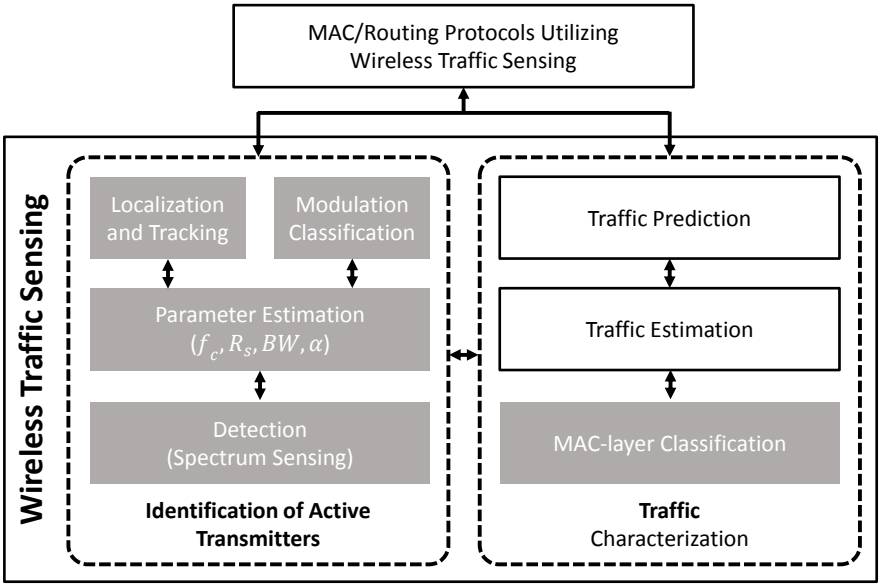


Figure 1.7: Top-level advance spectrum sensing system (shaded blocks are the focus of this dissertation).

in this system ranges from spectrum sensing to estimating PHY layer parameters such as center frequency BW etc. and their modulation type. We also want to try and localize the primary transmitters. On the other side of the equation, we want also want to observe higher level characteristics such as channel access method (what MAC scheme is being used) and there is also work on traffic prediction and estimation. Through a mixture of algorithm development, performance evaluation, and actual hardware experiments we practically address the problem of characterizing the radio environment with as much detail as possible.

1.3 Contributions of this Dissertation

Our contributions comprise of various aspects of the advanced spectrum sensing system which was described in the previous section. These contribution summarized into three main aspects.

Multi-antenna Cyclostationarity Detection

We proposed a cyclostationary spectrum sensing method specifically targeted for a multiple antenna receivers which is highly robust to noise uncertainty because the detection threshold is not dependent on SNR. . We have shown that this method has lower computational complexity than existing methods of comparable performance because it only computes the cyclic covariance matrix once and performs the correlation in time which eliminates the need for taking a high frequency resolution FFT with large number of samples. More importantly we have derived asymptotic theoretical expressions for the probability of detection and false alarm. Using these expressions aided with simulations we evaluated the performance of this method under various scenarios including Rayleigh fading channel, correlated noise environments, and in the presence of a strong interferer. The method was shown to outperform existing techniques in all considered scenarios.

Sampled Distribution Distance Based Classification

We developed a computationally efficient method for modulation level classification based on distribution distance functions. Specifically, we proposed to use a metric based on Kolmogorov-Smirnov and Kuiper distances which exploits the distance properties between CDFs corresponding to different modulation levels. The proposed method results in faster modulation level classification than the commonly used cumulant-based method, by reducing the number of samples needed. It also results in lower computational complexity than the KS-GoF method, by eliminating the need for a sorting operation and using only a limited set of test points, instead of the entire CDF.

We have further verified the practicality of these approach by implementing two modulation level classifiers in a practical real-time hardware platform with radio front-ends and a DSP engine (BEE2), and evaluated their classification

performance under realistic impairments such as quantization errors and timing synchronization errors. Both classifiers were shown to follow similar trends with regard to these impairments, and their hardware computational complexity were proven to be similar.

Finally we have developed the concept of distribution distance based classification. We also derived the optimal discriminant functions for classifying modulation schemes using the sampled distribution distance. This method was shown to provide substantial gains compared to other existing approaches. The performance of this method is also shown to be close to the maximum likelihood classifier but at significantly lower computational complexity. The same classifier can be generalized to any classification problem where the CDF of each class is available.

MAC Classification Based on the Fourth-Order Cumulant

Finally, we have contributed to the problem of channel access method classification. We focus on the development of algorithms for channel access method classification. In particular, we present a novel scheme for acquiring this information comprising of three main stages 1) Band Segmentation, 2) Channel Access Method and Modulation Type Classification, and finally 3) Collision detection. We propose and evaluate a novel method of performing band-segmentation on a wideband spectrum that is able to successfully distinguish between spectrally overlapped signals without the need for multiple antennas as prior methods do. This is achieved by exploiting the temporal independence in primary user activity through the Non-Negative Matrix Factorization (NNMF) method. We extend existing fourth-order cumulant-based methods of modulation type classification to distinguish between multiplexing methods including TDMA, OFDMA, and CDMA. We propose a novel method based on the sample variance of the cumulant estimator in order to identify contention-based systems. Through analysis

and simulations we show that our scheme is capable of distinguishing between TDMA, OFDMA, CDMA, and contention-based channel access methods with a high probability. Further, our proposed method is capable of identifying the modulation type of non-contention-based primary users with high accuracy.

1.4 Organization of this Dissertation

The rest of the dissertation is organized as follows. In Chapter 2 we investigate different techniques that exploit multiple antennas for cyclostationary spectrum sensing. We also discuss a novel method which we have proposed which is aimed at improving the use of multiple antennas in this spectrum sensing problem. The modulation classification approach based on sampled distribution distance is discussed in Chapter 3. In Chapter 4 we present a MAC-layer scheme classification algorithm based on the normalized fourth order cumulant. Finally, we conclude this dissertation in Chapter 5.

CHAPTER 2

Multi-antenna Cyclostationary Spectrum Sensing

Advanced spectrum sensing aims to acquire more information about the primary user beyond their spectrum activity (e.g. whether the transmitter is idle or occupied). However, the traditional spectrum sensing problem still plays a very important role in such a system. Transmitter activity is a key prerequisite knowledge before we can estimate detailed knowledge about the Primary Users. In this chapter we look at various spectrum sensing algorithms existing in the literature. In particular, we highlight the importance of spectrum sensing based on the cyclostationary features embedded in the primary user signal. Specifically, we will address an important problem that has not been tackled extensively in the literature; how to efficiently utilize multiple receive antennas in the detection of cyclostationary features. In this chapter we investigate different techniques that exploit multiple antennas for cyclostationary spectrum sensing. We also discuss a novel method which we have proposed which is aimed at improving the use of multiple antennas in this spectrum sensing problem. Finally, we compare the performance of these various different techniques and discuss the advantages and disadvantages of each.

2.1 Cyclostationary Spectrum Sensing

Spectrum sensing is a key step in effectively realizing cognitive radio networks (CRN). In the CR access paradigm, secondary users (SU) in a CRN are allowed to access spectrum reserved for use by licensed or primary users (PU) given that 1) those resources are either currently unoccupied or 2) interference to the primary network is kept under an acceptable level [Hay05]. The main goal of spectrum sensing is to accurately and efficiently detect the presence or absence of a PU in a given band, usually under the constraint of a low signal-to-noise ratio (SNR).

Several spectrum sensing methods have been proposed in the literature [YA09]. In general, these methods can be categorized as being based on either energy detection, cyclic correlation (cyclostationarity), or matched filtering. Energy detection requires the least prior knowledge about the signal, while matched filtering requires the most. Cyclic correlation-based techniques [LKK10, CGD07] lie in between, requiring either prior knowledge or accurate estimation of the cyclic frequencies present in the PU transmission signal. Although energy detection offers the lowest computational complexity and is the optimal blind detector in the presence of i.i.d. noise, its performance relies on accurate knowledge of noise power due to the SNR wall phenomenon [TS08]. The detection performance of energy detection also degrades in a temporally correlated noise environment.

In some scenarios, such as very low SNR regime or when signal selectivity is important, cyclic correlation-based methods offer several advantages over other spectrum sensing approaches. Unlike energy detection, they do not suffer from the SNR wall issue. These methods are also resilient to temporally correlated noise and enable signal-selective spectrum sensing where the presence of signals-of-interest (SOI) can be detected based on their unique cyclic features due to their modulation type, symbol rate, and carrier frequency [GBC87].

One issue encountered with all spectrum sensing methods is the effect of fading

in the channel between the PU and SU. There is a decrease in the probability of detection whenever the channel is in a deep fade. This can be alleviated by exploiting spatial diversity either through the use of cooperative spectrum sensing [QCP08] or, if available, the use of multiple antennas. As a result, spectrum sensing algorithms exploiting multiple antennas have received considerable interest [TNG10, Tug12].

Algorithms that leverage the cyclostationarity property have been applied in the past for multiple antenna receivers. In [SA08], the sum of the spectral correlation for each antenna was proposed. Such methods are considered *post-combining* techniques since knowledge of the channel state information (CSI) is not exploited. On the other hand, *pre-combining* techniques which utilize an estimate of the CSI to varying degrees have been shown to have better performance in a random channel. A method based on equal gain combining (EGC) was investigated in [CXH08] which uses phase offset estimates to align the raw samples from each antenna. The aligned signals are then summed before finding the spectral correlation. Finally, a blind maximal ratio combining (MRC) scheme was evaluated in [JLZ10] which utilized the singular value decomposition (SVD) to find an estimate of the CSI and applied MRC on the raw samples.

Several issues need to be addressed with these existing schemes. First, none of these works provide an analytical method for setting the decision threshold for fixed probability of false alarm (P_{FA}) since that distribution of the test statistic is not known. Such theoretical expressions are useful because they could be used for further analytical study of these methods from a system level perspective. For example, work such as [GPC11] utilizes the expressions of P_D and P_{FA} as a function of sensing time and SNR to analyze the throughput of CR networks. In addition, these expressions allow the detection threshold to be determined optimally without any prior training of the detector. Second, the threshlold used in prior multi-antenna cyclostationary detectors are dependent on the SNR which

means that a very accurate noise estimation is required to get the desired P_{FA}. A detector is thus very likely to be operating at an incorrect point along the ROC curve because of inaccurate SNR estimate. Third, the spectral correlation function used in all these prior methods require high computational complexity, making them hard to implement in practice. Finally, we also want to improve the detection performance.

These issues have motivated us to propose a spectrum sensing algorithm based on the cyclic correlation significance test (CCST) designed for use in a multiple antenna system which we refer to as Eigenvalue-Based Cyclostationary Spectrum Sensing or EV-CSS. The CCST was used in [SG90] to perform cyclostationary source enumeration using an information-theoretic criterion. However, the use of CCST in the context of multiple antenna cyclostationary spectrum sensing has not been investigated in prior work. The performance of this method in fading channels has also not been evaluated nor compared to other spectrum sensing schemes that exploit cyclostationarity. In this work, we derive the analytical asymptotic performance of this detection method in both AWGN and flat-fading channels. These expressions are then verified through simulations. The results also enable us to investigate sensing performance in a spatially correlated noise environment.

Our main contributions to multi-antenna cyclostationary spectrum sensing are as follows:

1. Proposed a cyclostationary spectrum sensing method specifically targeted for a multiple antenna receivers which is highly robust to noise uncertainty because the detection threshold is not dependent on SNR.

2. Showed that this method has lower computational complexity than existing methods of comparable performance because it only computes the cyclic covariance matrix once and performs the correlation in time which eliminates

the need for taking a high frequency resolution FFT with large number of samples.

3. Derived asymptotic theoretical expressions for the probability of detection and false alarm.

4. Evaluated the performance of this method under various scenarios including Rayleigh fading channel, correlated noise environments, and in the presence of a strong interferer. The method is shown to outperform existing techniques in all considered scenarios.

The rest of this chapter is organized as follows. The system model is introduced in Section 2.2 including a brief discussion of cyclostationarity. The proposed method is detailed in Section 2.3 including analysis of its detection performance under both AWGN and flat fading. Numerical results for various scenarios are presented in Section 4.4.3. Finally, we offer some conclusions in Section 2.5.

Notation: $|\mathbf{A}|$ and $tr(\mathbf{A})$ denote the determinant and trace of square matrix $\mathbf{A}$ respectively. $\mathbf{B}_{ij}$ denotes the (i,j)th element of the matrix $\mathbf{B}$ and $\mathbf{I}$ is the identity matrix. The superscripts $*$ and H denote the complex conjugate and the Hermitian (conjugate transpose) operations, respectively. Given two random vectors $\mathbf{x}$ and $\mathbf{y}$, we define $\text{cov}(\mathbf{x},\mathbf{y}) \triangleq E\{\mathbf{x}\mathbf{y}^H\} - E\{\mathbf{x}\}E\{\mathbf{y}^H\}$. Given column vector $\mathbf{x}$, $\text{diag}\{\mathbf{x}\}$ denotes a square matrix with elements of $\mathbf{x}$ along its main diagonal and zeros everywhere else. We will use the notation $\mathcal{N}_c(\mathbf{m},\mathbf{\Sigma})$ to denote a proper (circularly symmetric) complex multivariate Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$. Finally we use the notation $y = \mathcal{O}(g(x))$ to indicate that there exists some finite real number $b > 0$ such that $\lim_{x\to\infty} |y/g(x)| \leq b$.

2.2 Background and System Model

2.2.1 Background on Cyclostationarity

A signal is considered to be cyclostationary if its statistical properties are periodic. Equivalently, if the cyclic autocorrelation function, defined as:

$$R_x^{\alpha}(\tau) = \lim_{\Delta t \to \infty} \frac{1}{\Delta t} \int_{-\frac{\Delta t}{2}}^{\frac{\Delta t}{2}} x\Big(t + \frac{\tau}{2}\Big) x^*\Big(t - \frac{\tau}{2}\Big) e^{-j2\pi\alpha t} dt, \tag{2.1}$$

is non-zero with some τ for at least one $\alpha \neq 0$, the signal is said to exhibit second-order cyclostationary property with α referred to as the cyclic frequency.

For example, in BPSK signals, cyclostationary features exist at $\alpha = \frac{k}{T_b}$ and at $\alpha = \pm 2f_c + \frac{k}{T_b}$, where T_b is the symbol period, f_c is the carrier frequency, and $k \in \mathbb{Z}$. Detailed analysis of the cyclostationary features for various digital modulations can be found in [GBC87].

2.2.2 Signal Model and Assumptions

We adopt a similar signal model as that used in [JLZ10]. The spectrum sensing problem is to decide between two hypotheses: $\mathcal{H}_0$, where the signal is absent; and $\mathcal{H}_1$, where it is present. The received signal, sampled at a rate of $1/T_s$, forms M streams coming from each antenna with N samples each. This received signal is defined as $\mathbf{x}(n) \triangleq [x_1(n), x_2(n), \ldots, x_M(n)]^T$, $n \in \mathbb{Z}$, $1 \leq n \leq N$, where $x_i(n)$ is the signal coming from the ith antenna. The received signal is the superposition of P signal sources (including both the SOI and any interferer) and receiver noise. The two hypotheses can therefore be expressed in vector form as

$$\mathbf{x}(n) = \begin{cases} \boldsymbol{\eta}(n), & \mathcal{H}_0, \\ \sum_{j=1}^{P} \mathbf{h}_j s_j(n) + \boldsymbol{\eta}(n), & \mathcal{H}_1, \end{cases} \tag{2.2}$$

where $\boldsymbol{\eta}(n)$ is the receiver noise denoted by $\boldsymbol{\eta}(n) \triangleq [\eta_1(n), \eta_2(n), \ldots, \eta_M(n)]^T$, where every η_i is a purely stationary Gaussian random process ($R_{\eta}^{\alpha}(\tau) = 0$ for

any $\alpha \neq 0$) with variance of σ_η^2. For simplicity, we restrict that only one PU transmission, $s_1(n)$, is considered a SOI and that it is cyclostationary with a unique cyclic frequency $\alpha = \alpha_0$. The channel experienced by each of the P sources is given by $\mathbf{h}_j \triangleq [h_{j1}, h_{j2}, \ldots, h_{jM}]^T$, where h_{jk} is the channel between the jth source and the kth antenna. We assume that the channel, although unknown to the receiver, stays constant over the spectrum sensing interval. Subsequently, we define the average signal-to-noise ratio to be

$$SNR \triangleq E\left\{\mathbf{h}^H\mathbf{h}\right\}/E\left\{\boldsymbol{\eta}^H\boldsymbol{\eta}\right\}. \tag{2.3}$$

2.2.3 Spatially Correlated Noise Environments

In the case of spatially correlated noise, which can happen when there is substantial ambient noise in the band, following [SS97], we model $\boldsymbol{\eta}(n)$ to have a covariance matrix given by $\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \text{cov}\{\boldsymbol{\eta}, \boldsymbol{\eta}\}$ where

$$\{\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}}\}_{ij} = E\{\boldsymbol{\eta}_i^H\boldsymbol{\eta}_j\} = \begin{cases} \sigma_\eta^2, & i = j \\ \sigma_\eta^2\rho_s^{|i-j|}, & i \neq j. \end{cases} \tag{2.4}$$

Thus with $\rho_s = 0$, the covariance matrix simplifies to $\sigma_\eta^2\mathbf{I}$ giving spatially white noise, while $\rho_s = 1$ gives fully correlated noise over all antennas. Varying degrees of partial correlation can be achieved by setting $0 < \rho_s < 1$.

2.3 Proposed Detection Statistic

In this section, we describe the proposed method. We focus on a single cycle frequency detection, but this approach could be generalized to multi-cycle detection.

2.3.1 Canonical Correlation Analysis

The key idea of the proposed detection test statistic is based on the theory of canonical correlation analysis and the concept of common factors. A common

factor is a signal component that is common to two data sets and can be estimated from either of these data sets by a linear transformation. Since common factors are defined to be independent of each other, the number of common factors is equal to the rank of the cross-correlation matrix of the two data sets. As discussed in [Law59], and subsequently utilized in [SG90], the number of common factors between two $M \times 1$ time-series vectors $\mathbf{x}(n)$ and $\mathbf{y}(n)$ is the rank of the matrix

$$\mathbf{R} = \mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{R}_{\mathbf{xy}}\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{yx}}, \tag{2.5}$$

where we define $\mathbf{R}_{\mathbf{xy}} \triangleq \text{cov}(\mathbf{x}, \mathbf{y})$ given the two random vectors $\mathbf{x}$ and $\mathbf{y}$.

Canonical analysis (see [And03] for details) aims to find the relationships between two groups of variables in a data set. Given two random vectors $\mathbf{x}$ and $\mathbf{y}$, of length m and n respectively, canonical analysis aims to find at most $\min\{m, n\}$ pairs of $(\mathbf{u}_i, \mathbf{v}_i)$ such that the correlation between the linear combinations, $U_i \triangleq \mathbf{u}_i^H\mathbf{x}$ and $V_i \triangleq \mathbf{v}_i^H\mathbf{y}$, is maximized. An additional restriction is that U_i and V_i must be uncorrelated with U_j and V_j for $i \neq j$. These linear combinations are referred to as canonical variates. The canonical variates are sorted in decreasing order of correlation such that the first canonical variates, U_1 and V_1, have the highest correlation. The correlation coefficient, $\rho_i \triangleq \text{cov}(U_i, V_i)/(\sigma_{U_i}\sigma_{V_i})$, between U_i and V_i is referred to as the ith canonical correlation. The procedure of finding $\mathbf{u}_i$ and $\mathbf{v}_i$ can be efficiently performed using a singular value decomposition (SVD) and the square of the canonical correlations can be found by finding the eigenvalues of (2.5).

In the context of cyclostationary spectrum spectrum sensing, canonical analysis provides us with a very powerful tool to optimally combine the signals from M antennas and find the canonical correlations, ρ_i, resulting from up to M mutually uncorrelated cyclostationary signals. This can be accomplished using the Cyclic Correlation Significance Test (CCST) [SG90] by performing canonical analysis on $\mathbf{x}(n)$ and $\mathbf{x}(n-\tau)e^{-j2\pi\alpha nT_s}$ for a given lag τ and cyclic frequency α. By finding the

canonical correlations between these two sets of data, we are in effect measuring the maximum amount of cyclic correlation for all possible linear combinations of the signals coming from the M antennas. A threshold can then be applied on the combined ρ_i's to determine the presence or absence of the PU. Additionally, some cyclic frequencies, such as those located on $\alpha = \pm 2f_c$ for BPSK, only appear in the conjugate cyclic correlation. These can also be detected by instead performing the canonical analysis with $\mathbf{x}^*(n-\tau)e^{-j2\pi\alpha nT_s}$.

Prior to performing the detection, we pick the lag τ that provides the best detection performance based on the modulation format used by the PU. This could be done off-line by performing the maximization, $\tau_0 = \arg\max_\tau |R_s^{\alpha_0}(\tau)|$.

2.3.2 Detection Test Statistic

The steps for computation of the proposed test statistic are summarized as follows:

1. Estimate the covariance matrix of size $M \times M$ at lag τ_0

$$\hat{\mathbf{R}}_{\mathbf{xx}}(\tau_0) = \frac{1}{N'} \sum_{n=0}^{N'} \mathbf{x}\,(n)\,\mathbf{x}^H\,(n-\tau_0)\,, \tag{2.6}$$

   where $N' = N - 1 - \tau_0$.

2. Estimate the cyclic correlation matrix using a cyclic cross-correlogram at cyclic frequency α_0 and lag τ_0, defined as

$$\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0}(\tau_0) = \frac{1}{N'} \sum_{n=0}^{N'} \mathbf{x}\,(n)\,\mathbf{x}^H(n-\tau_0)\,e^{-j2\pi\alpha_0 nT_s}. \tag{2.7}$$

   We will refer to the τ_0-lag covariance matrices for both conventional and cyclic autocorrelation function simply as $\hat{\mathbf{R}}_{\mathbf{xx}}$ and $\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0}$ from this point for the sake of brevity, since other τ are not utilized by the test statistic. The CCST is then calculated by finding the matrix

$$\hat{\mathbf{R}} = \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0} \hat{\mathbf{R}}_{\mathbf{xx}}^{-1} \left(\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0}\right)^H. \tag{2.8}$$

3. Find the eigenvalues, $\boldsymbol{\mu} = [\mu_1^2, \mu_2^2, \ldots, \mu_M^2]^T$, of $\hat{\mathbf{R}}$.

4. Combine eigenvalues as

$$\lambda \triangleq \prod_{i=1}^{M} \left(1 - \mu_i^2\right), \tag{2.9}$$

and finally calculate the test statistic:

$$\mathcal{T}_{\mathbf{xx}}^{\alpha} \triangleq -m \ln \lambda. \tag{2.10}$$

The factor $m \triangleq N - M - 1$ is used to scale the test statistic so that its distribution is independent of the number of samples used [Bar38, Sec. 8].

5. Decision: $\mathcal{T}_{\mathbf{xx}}^{\alpha} \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} \gamma$, where $\gamma > 0$ is a threshold chosen to achieve constant false alarm rate (CFAR) which will be discussed in the following section.

Note that all $\mathbf{x}^H(n)$ can be replaced with $\mathbf{x}^T(n)$ if the conjugate cyclic correlation matrix is needed. We refer to the version of the test statistic that uses $\mathbf{x}^H(n)$ as the non-conjugate cyclic correlation significance test (NC-CCST) while the other is the conjugate cyclic correlation significance test (C-CCST). For the test statistic of each, we will use the notations $\mathcal{T}_{\mathbf{xx}}^{\alpha}$ and $\mathcal{T}_{\mathbf{xx}^*}^{\alpha}$ respectively.

2.3.3 Distribution of Test Statistic Under $\mathcal{H}_0$ and Constant False-Alarm Rate

Two key parameters are used to evaluate the performance of spectrum sensing algorithms. The detection probability, P_D, is the probability of being at $\mathcal{H}_1$ and accurately detecting the PU ($P_D \triangleq \Pr(\mathcal{T}_{\mathbf{xx}}^{\alpha} > \gamma \mid \mathcal{H}_1)$). On the other hand, the false alarm probability, P_{FA}, is the probability of being at $\mathcal{H}_0$ and mistakenly detecting a PU ($P_{FA} \triangleq \Pr(\mathcal{T}_{\mathbf{xx}}^{\alpha} > \gamma \mid \mathcal{H}_0)$).

It has been shown in [Bar38, Sec. 8] that the limiting distribution ($N \to \infty$) of the test statistic (2.10) for real and normally distributed random vectors approaches a χ^2 distribution with degree-of-freedom M^2. Following a similar proof,

it can also be shown that for zero mean, complex Gaussian random variables, the distribution is also χ^2 with degree-of-freedom M^2 when using the NC-CCST and $M(M+1)$ for the C-CCST.

Based on the distribution of $\mathcal{T}_{\mathbf{xx}}^{\alpha}$ under $\mathcal{H}_0$, the detection threshold γ can be set to achieve a desired P_{FA} by satisfying

$$\int_{\gamma}^{\infty} f_{\chi_k^2}(x)\, dx = P_{FA}, \tag{2.11}$$

where

$$k = \begin{cases} M^2 & \text{NC-CCST} \\ M(M+1) & \text{C-CCST} \end{cases}, \tag{2.12}$$

and $f_{\chi_k^2}(\cdot)$ is the probability density function (pdf) of a χ^2 random variable with degree-of-freedom k.

These asymptotic distributions are verified to closely match simulation in Fig. 2.1 for $N = 1000$. Due to the scaling factor in (2.10), the distribution is independent of N. The empirical pdfs for two different σ_η^2 values are also shown to demonstrate that the test statistic's distribution under $\mathcal{H}_0$ is independent of noise power.

As introduced in Section 2.2.3, spatially correlated noise happens whenever $\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ is non-diagonal. This could also be interpreted as having a transformed noise vector

$$\boldsymbol{\eta}' = \mathbf{A}\boldsymbol{\eta} \quad s.t. \quad \sigma_\eta \mathbf{A}\mathbf{A}^H = \mathbf{R}_{\boldsymbol{\eta}'\boldsymbol{\eta}'}, \tag{2.13}$$

where $\mathbf{A}$ is an $M \times M$ matrix that determines the spatial correlation among antennas. In order to see the effect of correlated noise on the distribution of $\mathcal{T}_{\mathbf{xx}}^{\alpha}$ under $\mathcal{H}_0$, it is helpful to use an alternate interpretation of canonical correlation as given in [SF69, Eqn. 2.1] such that

$$\lambda = \frac{1}{|\hat{\mathbf{R}}_{\mathbf{xx}}|^2} \begin{vmatrix} \hat{\mathbf{R}}_{\mathbf{xx}} & \hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0} \\ \hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0 H} & \hat{\mathbf{R}}_{\mathbf{xx}}^{H} \end{vmatrix}, \tag{2.14}$$

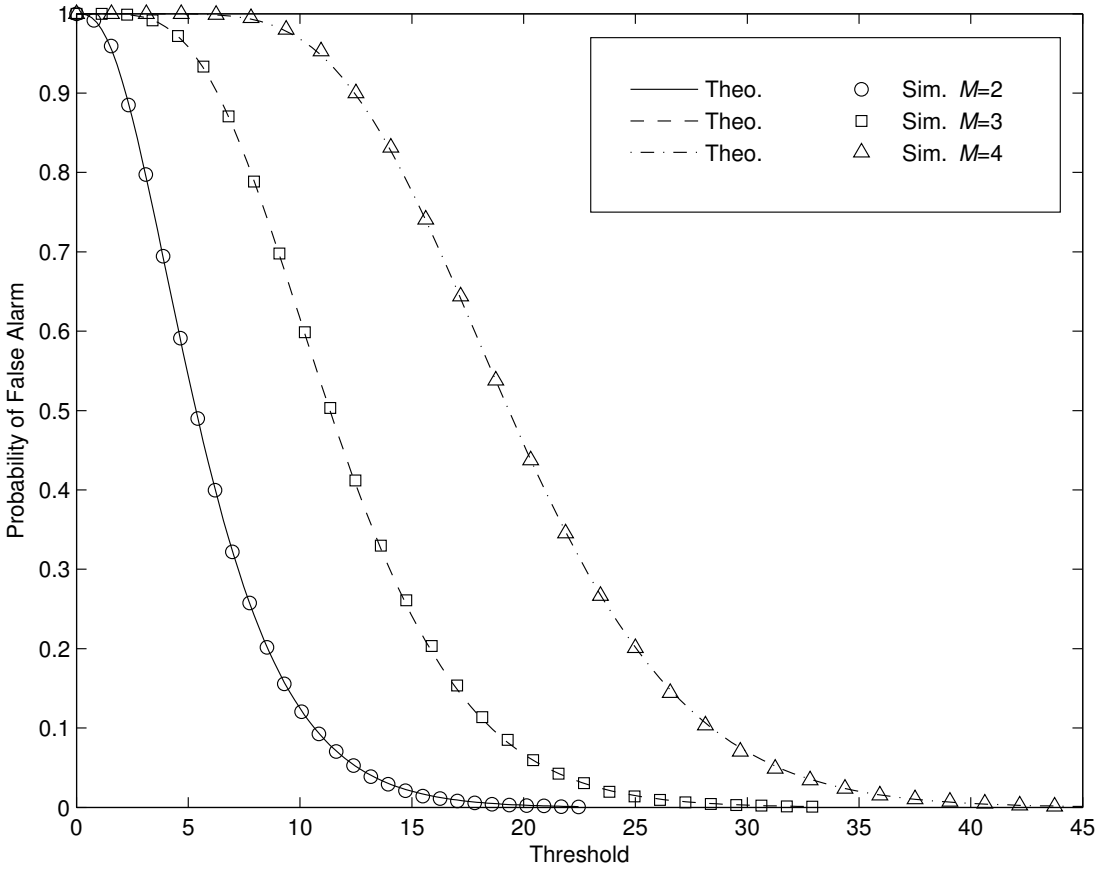


Figure 2.1: Verification of the asymptotic distribution of the proposed test statistic (C-CCST) under $\mathcal{H}_0$ with different number of antennas ($M = \{2, 3, 4\}$). These plots show the accuracy of the analytical expression under $N = 1000$ number of samples per antenna (SNR=-10 dB).

using the covariance matrix estimates given in (2.6) and (2.7) and λ as defined in (2.9). Using (2.13), in conjunction with the multiplication property of determinants, and determinants of block diagonal matrices, we find the value of λ for correlated noise (λ_c) to be

$$\lambda_c = \frac{\left| \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{R}}_{\mathbf{xx}} & \hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0} \\ \hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0 H} & \hat{\mathbf{R}}_{\mathbf{xx}}^{H} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix}^H \right|}{\left| \mathbf{A} \hat{\mathbf{R}}_{\mathbf{xx}} \mathbf{A}^H \right|^2} \tag{2.15}$$

$$= \begin{vmatrix} \hat{\mathbf{R}}_{\mathbf{xx}} & \hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0} \\ \hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0 H} & \hat{\mathbf{R}}_{\mathbf{xx}}^{H} \end{vmatrix} |\hat{\mathbf{R}}_{\mathbf{xx}}|^{-2} = \lambda. \tag{2.16}$$

Therefore the test statistic under $\mathcal{H}_0$ is invariant to any linear transformation on the noise measurements. As such the same expression for P_{FA} as well as the threshold for maintaining constant P_{FA} is applicable even for correlated noise environments.

2.3.4 Distribution of Test Statistic Under $\mathcal{H}_1$ and Probability of Detection

In this section we derive the distribution of the proposed test statistic, $\mathcal{T}_{\mathbf{xx}}^{\alpha}$, under $\mathcal{H}_1$. We begin by summarizing the prior work in statistics leading to the derivation of the complete non-null distribution of the canonical correlations. We show how this result is parameterized by the canonical correlation, ρ, between the two signals being tested. In the context of cyclostationary spectrum sensing, this corresponds to the value of the signal's cyclic autocorrelation, $R_{ss}^{\alpha} \triangleq E\left[s(n)s^*(n)e^{-j2\pi\alpha_0 n T_s}\right]$. We then derive the canonical correlation, ρ, resulting from R_{ss}^{α} with a given σ_η^2, N, $\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ and $\mathbf{h}$. Using this canonical correlation, the complete distribution of the test statistic under $\mathcal{H}_1$ is derived. Once the distribution is found, it allows us to find the theoretical probability of detection given as:

$$P_D = \Pr(\mathcal{T}_{\mathbf{xx}}^{\alpha} > \gamma \mid \mathcal{H}_1) = 1 - F_{\mathcal{T}_{\mathbf{xx}}^{\alpha}|\mathcal{H}_1}(\gamma), \tag{2.17}$$

where $F_{\mathcal{T}^{\alpha}_{\mathbf{xx}}|\mathcal{H}_1}(\cdot)$ is the cdf of the test statistic under $\mathcal{H}_1$ which we find in the rest of this subsection.

Several works in the past have contributed to deriving the non-null distribution of the test of significance of the canonical correlations originally proposed by Bartlett [Bar47]. In [Law59], the approximate means and variances of the highest eigenvalue were found in the asymptotic case with only one non-zero eigenvalue ($\mu_1 \neq 0, \mu_i = 0, i > 1$) and normality assumption. The distribution of the actual likelihood-ratio criteria, of the same form as (2.10), was found in [SF69]. However, both of these results break down in the case of local alternatives, which correspond to the cases when the distribution under $\mathcal{H}_1$ is very close to the null hypothesis. In relation to the CCST, this corresponds to having a very-low SNR, which is clearly the case of interest in the spectrum sensing problem.

Finally, the non-null distribution of the likelihood ratio criteria for covariance matrix under local alternatives were found in [Sug73]. This criteria is used to test the independence between two multivariate random variables by combining the canonical correlation into a single test statistic. We use the same criteria to test for independence between the signal of interest and a frequency shifted version of itself. After the publication of this distribution, several works have focused on eliminating the normality assumption [MW80, Oga07]. However, in the case of spectrum sensing for CR, it is more likely to deal with detection under very-low SNR. If the noise is assumed AWGN and the $\sigma^2_{\eta} \geq \sigma^2_s$, the received signal, $\mathbf{x}(n)$, is approximately normal and the results for canonical analysis assuming normality can be used.

Following [Sug73, Thm. 4.1], the cdf of the the test statistic under local alternatives is found to be asymptotically distributed as a non-central chi-square. We have,

$$F_{\mathcal{T}^{\alpha}_{\mathbf{xx}^*}|\mathcal{H}_1}(x) = F_{\chi'^2}(x, M(M+1), \delta^2) + \mathcal{O}(m^{-1}), \tag{2.18}$$

where $F_{\chi'^2}(x, d, \delta^2)$ is the cdf of the non-central chi-square random variable with

degree-of-freedom of d and non-centrality parameter δ^2. The actual value of the non-centrality parameter can be found as $\delta^2 = tr(\Theta^2)$, where $\Theta = \sqrt{m}\,\mathrm{diag}\{\rho_1, \rho_2, \ldots, \rho_M\}$. Where, ρ_i^2 for $1 < i < M$ are the eigenvalues of (2.8) using ensemble averages. Thus, they are solutions to the determininant equation

$$\left|\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0}\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha_0 H} - \rho^2\mathbf{I}\right| = 0, \tag{2.19}$$

which has M roots.

The true distribution deviates from the non-central chi-square with lower number of samples as indicated by the additional $\mathcal{O}(m^{-1})$ term in (2.18). The complete distribution up to the order $\mathcal{O}(m^{-2})$ can be found in [Sug73, Thm. 4.1] which for the sake of brevity is no longer presented here. This more accurate distribution is an expansion based on non-central chi-square random variables of higher degrees-of-freedom with non-centrality parameters of $tr(\Theta^4)$ and $tr(\Theta^6)$. As such, a very accurate expression for $F_{\mathcal{T}_{\mathbf{xx}}^{\alpha}|\mathcal{H}_1}(x)$ can be calculated with knowledge of Θ.

Let's assume, for simplicity, that only one signal of interest $s(n)$ has cyclic frequency α. Therefore, $\rho_i = 0$ for $i > 1$. This assumption applies in almost all cases since two communication signals will, with high probability, have different cyclic frequencies. However, it is straightforward to extend these results to the case of multiple signals with exactly the same α, since this simply corresponds to additional non-zero ρ_i and the derivations presented here are still applicable. Thus, these scenarios can be treated theoretically as being single signal scenarios as long as the number of signals is less than or equal to the number of antennas M.

The channel, $\mathbf{h}$, is also assumed to be constant over one sensing period. Therefore, the expression for the distribution derived in this subsection is for a particular channel instance. Later we extend these results for flat-fading channels by integrating the distributions over the statistics of the fading channel.

With these assumptions, we proceed with finding the values of ρ_i as a func-

tion of the channel $\mathbf{h}$, the signal of interest $s(n)$ and the noise covariance $\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}}$. The received signal can be expressed as $\mathbf{x}(n) = s(n)\mathbf{h} + \boldsymbol{\eta}(n)$. Using our initial assumption that the signal has unit power and is uncorrelated with the noise, the asymptotic zero-lag covariance matrix then becomes

$$\begin{aligned} \mathbf{R}_{\mathbf{xx}} &= \operatorname{cov}(\mathbf{x}(n), \mathbf{x}(n)) && (2.20)\\ &= \mathbf{h}\mathbf{h}^H E\left[s(n)s^*(n)\right] + \mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}} && (2.21)\\ &= \mathbf{h}\mathbf{h}^H + \mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}}. && (2.22) \end{aligned}$$

On the other hand, recalling that noise has no cyclic features, the asymptotic cyclic cross covariance matrix can be found as

$$\begin{aligned} \mathbf{R}_{\mathbf{xx}}^{\alpha_0} &= \operatorname{cov}\left\{\mathbf{x}(n), \mathbf{x}(n)e^{-j2\pi\alpha_0 nT_s}\right\} && (2.23)\\ &= \mathbf{h}\mathbf{h}^H E\left[s(n)s^*(n)e^{-j2\pi\alpha_0 nT_s}\right] && (2.24)\\ &= \mathbf{h}\mathbf{h}^H R_{ss}^{\alpha}. && (2.25) \end{aligned}$$

Subsequently the conjugate version, $R_{ss^*}^{\alpha}$, could also be used by replacing $s^*(n)$ with $s(n)$.

We again use the alternate interpretation in (2.14) to find that the canonical correlation under $\mathcal{H}_1$ as

$$\rho = \left\|\mathbf{A}^{-1}\mathbf{h}\right\|^2 |R_{ss}^{\alpha}| / (\|\mathbf{A}^{-1}\mathbf{h}\|^2 + \sigma_\eta^2), \tag{2.26}$$

where the noise covariance matrix is $\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \sigma_\eta^2 \mathbf{A}\mathbf{A}^H$. The details of the derivation of (2.26) are provided in Appendix A. In the case of spatially uncorrelated noise, we have $\mathbf{A} = \mathbf{I}_M$ so that the true correlation becomes

$$\rho = \|\mathbf{h}\|^2 |R_{ss}^{\alpha}| / (\|\mathbf{h}\|^2 + \sigma_\eta^2), \tag{2.27}$$

which is only dependent on the 2-norm of the channel coefficient, the noise variance σ_η^2, and the cyclic correlation for the chosen α. A derivation of R_{ss}^{α} is given in [RC11] for various modulation schemes including BPSK, MSK, and QAM.

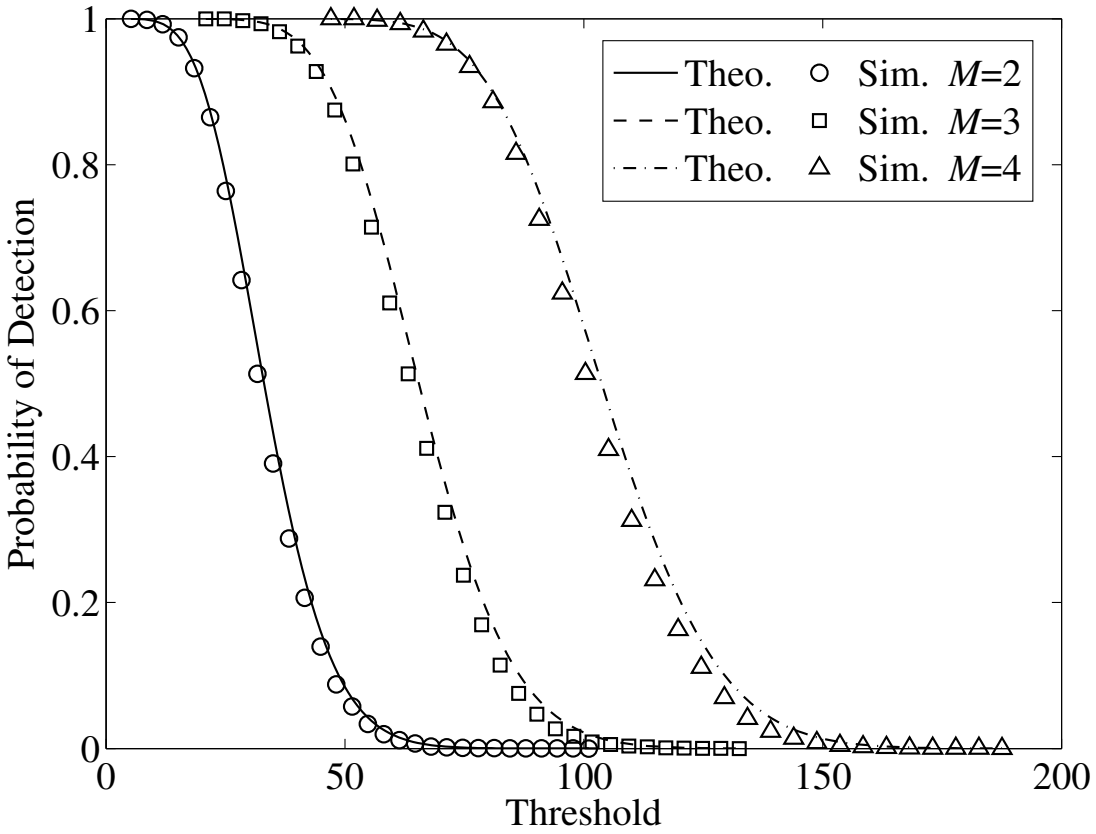


Figure 2.2: Verification of the asymptotic distribution of the proposed test statistic (C-CCST) under $\mathcal{H}_1$ with different number of antennas ($M = \{2, 3, 4\}$). These plots show the accuracy of the analytical expression under $N = 1000$ number of samples per antenna (SNR=-10 dB).

A similar derivation can be done for C-CCST by replacing R_{ss}^{α} with $R_{ss^*}^{\alpha}$. Thus for a single SOI scenario we have

$$\Theta = \sqrt{m}\,[\rho, 0, \ldots, 0]\,\mathbf{I}_M. \tag{2.28}$$

Which when substituted to (2.18) gives us the complete distribution. In Fig. 2.2, we show that the theoretical results are in very close agreement to the simulation.

2.3.5 Rayleigh Fading

Now that we have the complete distribution for both the null and non-null hypothesis with AWGN parameterized by a particular instance of the channel $\mathbf{h}$ and the noise covariance $\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}}$, we can proceed to evaluate the performance of the proposed test statistic under a flat-fading environment.

In particular we use a flat fading channel model where the channel vector, $\mathbf{h}$, remains constant during the whole frame of samples used in detection. This is described using a channel vector for the ith frame as $\mathbf{h}_i = [r_1 e^{j\theta_1}, r_2 e^{j\theta_2}, \ldots, r_M e^{j\theta_M}]^T$, where r_n is a Rayleigh distributed random variable of unit variance and θ_n is a uniformly distributed random variable in $[0, 2\pi]$.

We first assume spatially uncorrelated noise ($\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \sigma_\eta^2 \mathbf{I}$). Based on (2.27), the true correlation coefficient, which eventually determines the non-centrality parameter in (2.18) is a function of $\|\mathbf{h}\|^2 = \mathbf{h}^H \mathbf{h}$. The distribution of $\mathbf{h}$ can be seen as a complex normal random vector of size $M \times 1$ which results in $\|\mathbf{h}\|^2 \sim \chi^2_{2M}$. Combining (2.18) and (2.27), we find the new cdf of our test statistic under Rayleigh flat-fading and spatially uncorrelated noise as:

$$F^*_{\mathcal{T}^\alpha_{\mathbf{xx}}|\mathcal{H}_1}(x) = \int_0^\infty F_{\chi'^2_{M^2+M}}\left(x, \frac{\beta \left|R^\alpha_{ss}(\tau)\right|}{\beta + \sigma_\eta^2}\right) f_{\chi^2_{2M}}(\beta) d\beta. \tag{2.29}$$

In the case of spatially correlated noise, the random variable $\beta = \|\mathbf{A}^{-1}\mathbf{h}\|^2 = \mathbf{h}^H \left(\mathbf{A}\mathbf{A}^H\right)^{-1} \mathbf{h}$ becomes a generalized χ^2 r.v. instead.

Although a closed form expression is very difficult to derive for such an expression due to the presence of the non-central chi-square, there is still some insight to be gained by numerically integrating (2.29) to arrive at P_D and P_{FA} expressions under flat fading.

2.3.6 Comparison With Existing Approaches

The algorithms for multiple antenna spectrum sensing based on cyclostationarity that are currently in the literature can generally be classified into two categories. The simplest method to do this is to find the sum of the spectral correlation test statistic estimated individually from each antenna [SA08]. We refer to this approach as SUM-MSDF (where MSDF means Modified Spectral Density Function). The MSDF is defined as the spectral correlation function (SCF) normalized by signal energy as discussed in [JLZ10].

Another existing approach is to sum the raw samples from each antenna and then perform a single spectral correlation test. However, we encounter a problem when the channel is not simply AWGN but instead has random fading. In this case, each antenna will have some unknown phase offset and attenuation. Thus, simply adding the raw samples non-coherently would decrease the probability of detection. This problem is addressed in [CXH08] by first eliminating the phase rotation of signal samples coming from each antenna. An estimate of the relative phase difference between each antenna is calculated by finding both the cyclic correlation of one antenna chosen as reference (auto-spectral correlation function or auto-SCF) and the cross-cyclic correlation of every other antenna and the reference antenna. The phase difference can then be extracted from these two. We refer to this method in our comparisons as Equal Gain Combining (EGC).

Finally, Maximal Ratio Combining (MRC) is used in [JLZ10]. Blind channel estimation is achieved by taking the vector corresponding to the highest singular value of (2.7) as an estimate of the channel, $\hat{\mathbf{h}}$. The raw samples from each antenna are combined using

$$y(n) = \hat{\mathbf{h}}^H \mathbf{x}(n)/\|\hat{\mathbf{h}}\|. \tag{2.30}$$

The cyclic correlation test is then performed on the combined samples $y(n)$. This method is called MSDF with blind maximal ratio combining or BMRC-MSDF.

It was shown to outperform the other techniques but at the cost of additional complexity due to the channel estimation and combining. One issue with this approach is the fact that the cyclic correlation is calculated twice. The first is used to blindly estimate the channel and the second to perform the detection on the combined samples. In contrast, the method we proposed only needs to perform the first part of BMRC-MSDF, finding the eigenvalues, and then uses the eigenvalues themselves to infer the presence or absence of the PU.

2.3.7 Advantages of the Proposed Method

As with other cyclostationarity-based spectrum sensing methods, one major advantage of the proposed method is its robustness to the noise uncertainty problem. Since the noise is assumed to be stationary and does not exhibit cyclostationarity at any $\alpha \neq 0$, its cyclic correlation approaches zero as $N \to \infty$. Thus, the effect of any error in the noise power estimate on the detection probability can be eliminated by taking more samples. However, in the interest of conserving power and arriving at a timely decision, both of which are high priority in the case of CR applications, we aim to minimize N needed to achieve a target P_D. This presents another, more subtle, issue related to noise uncertainty.

In the non-asymptotic scenario, the methods based on the SCF (BMRC-MSDF, EGC and SUM-MSDF) under $\mathcal{H}_0$ have been shown to depend on both N and the noise power σ_η^2 [RC11]. Therefore, the proper detection threshold is still a function of the noise variance. By incorrectly specifying this threshold, the detector could be at the wrong point in the receiver operating characteristic (ROC) curve. Equivalently, the target CFAR cannot be achieved. However, as previously discussed and demonstrated in Fig. 2.1, the proposed test statistic is independent of both σ_η^2 and N. Consequently, the threshold γ only needs to be chosen once for a given number of antennas M to guarantee CFAR. This property has been shown for other eigenvalue-based approaches [ZL09]. It derives from the

fact that noise power estimation is built-in to the test statistic.

2.3.8 A Note on Complexity

We provide an approximate complexity comparison of the proposed method with the best performing existing method (BMRC-MSDF) by taking number of complex multiplications required for each under the same number of samples N. Since the cyclic covariance operation and the SVD are common to both methods, they are not included in the analysis.

Assuming the MSDF is calculated using an N_S-point Fast Fourier Transform (FFT) it requires in the order of $N \log_2(N_S)$ multiplications. In addition, $(M+1)N$ multiplications are needed to perform the MRC and normalization. Finally, the correlation in frequency uses $NN_S/2$ multiplications. Thus, the BMRC-MSDF approach performs in the order of $N(\log_2(N_S) + N_S/2 + M + 1)$ multiplications without taking into account the SVD and the cyclic covariance.

In comparison, the proposed EV-CSS method finds the conventional covariance in addition to an eigenvalue decomposition (EVD) and the same cyclic covariance as BMRC-MSDF, or in the order of NM^2 multiplications. The operation $\hat{\mathbf{R}}_{\mathbf{xx}}^{-1}\hat{\mathbf{R}}_{\mathbf{xx}}^{\alpha}$ in (2.8) is essentially the solution to a generalized linear system which can be seen as an LU decomposition requiring approximately $2M^3/3$ multiplications. Therefore, the EV-CSS approach requires in the order of $NM^2 + 2M^3/3$ multiplications in addition to the common operations with BMRC-MSDF. Since M is typically much less than both N and N_S, there is overall a significant decrease in complexity with the proposed method. For example, if we take $N = 4000$, $N_S = 128$, and $M = 2$, (same parameters used in [JLZ10]), the BMRC-MSDF requires $\sim$296K multiplications while EV-CSS needs only $\sim$16K multiplications, without counting the common operations.

2.4 Numerical Results and Discussion

In this section, simulation results are presented in order to compare the performance of the proposed test statistic with the various existing techniques discussed in Section 2.3.6. In addition, theoretical plots are included to further verify the analytical expressions of the proposed method's performance.

For these simulations we assume that only one PU has a feature at the chosen cyclic frequency, α_0. This PU is assumed to be transmitting a BPSK signal at a carrier frequency $f_c = 80$ KHz with symbol period of 25 μs. Each antenna of the SU is sampled at a rate $f_s = 320$ kHz. For all methods, the same cyclic frequency located at $\alpha_0 = 2f_c$ is used. This cyclic feature is only present in the conjugate cyclic autocorrelation which means the C-CCST statistic is used. This feature is chosen because it is the highest magnitude feature among all cyclic frequencies. The maximum cyclic autocorrelation at this cyclic frequency is observed at $\tau_0 = 0$ which is the lag we will be using for all EV-CSS simulations.

2.4.1 Threshold Selection for EV-CSS

One key advantage of the proposed EV-CSS scheme over other spectrum sensing schemes is the simplicity of threshold selection to achieve CFAR. As discussed in Section 2.3.3 the detection threshold, γ, is only dependent on M and not on N or σ_η^2. As a result of this, in a practical implementation of EV-CSS, the threshold can be pre-calculated using only the χ^2 distribution parameterized by the number of antennas. As we will show in the following subsections, the theoretically determined threshold achieves the desired P_{FA} and matches very well with simulations. Thus, only a single threshold needs to be stored. For the following simulations we set the CFAR to $P_{FA} = 0.1$ unless otherwise stated.

2.4.2 ROC and Detection Probability Versus SNR

We first consider $M = 2$ antennas in the SU. The channel between the PU and each antenna of the SU, is modeled as a quasi-static Rayleigh fading channel with channel $\mathbf{h}$ as described in Section 2.3.5. The fading is assumed to be frequency-flat and remains constant during the whole frame of $N = 1000$ samples per antenna used for detection. The noise in the antennas is assumed to be distributed as a zero-mean circularly symmetric complex-Gaussian, $\boldsymbol{\eta} \sim \mathcal{N}_c(\mathbf{0}, \sigma_\eta^2 \mathbf{I})$. The PU signal energy is assumed to be unit energy and σ_η^2 is chosen to achieve an average SNR across antennas as defined in (2.3).

Using these assumptions the ROC curves under SNR $= -10$ dB and $N = 1000$ samples for the proposed method and the other cyclic-based approaches are shown in Fig. 2.3 for comparison. Also included in this figure is the ROC curve for an energy detection (ED) based technique designed for multiple antennas described in [ZL09] referred to as Minimum-Maximum Eigenvalue (MME) detection. The MME method does not utilize any cyclostationarity in the PU signal. Both theoretical results and Monte Carlo simulations (with 50,000 trials) are shown for EV-CSS, while only the Monte Carlo simulations are shown for the other methods since these are already analyzed in the respective works that proposed them. We have verified that these results agree with simulations presented in these prior work under similar assumptions.

As seen in Fig. 2.3, there is very strong agreement in the theoretical and simulation results for the proposed method. The performance of the proposed method is clearly degraded when compared to a simple AWGN channel. However, it performs better than all the other techniques. Interestingly, the method also also outperforms BMRC-MSDF which, as discussed in Section 2.3.8, has significantly higher computational complexity. Although this result initially appears to be counter-intuitive, further experiments where only an AWGN channel is consid-

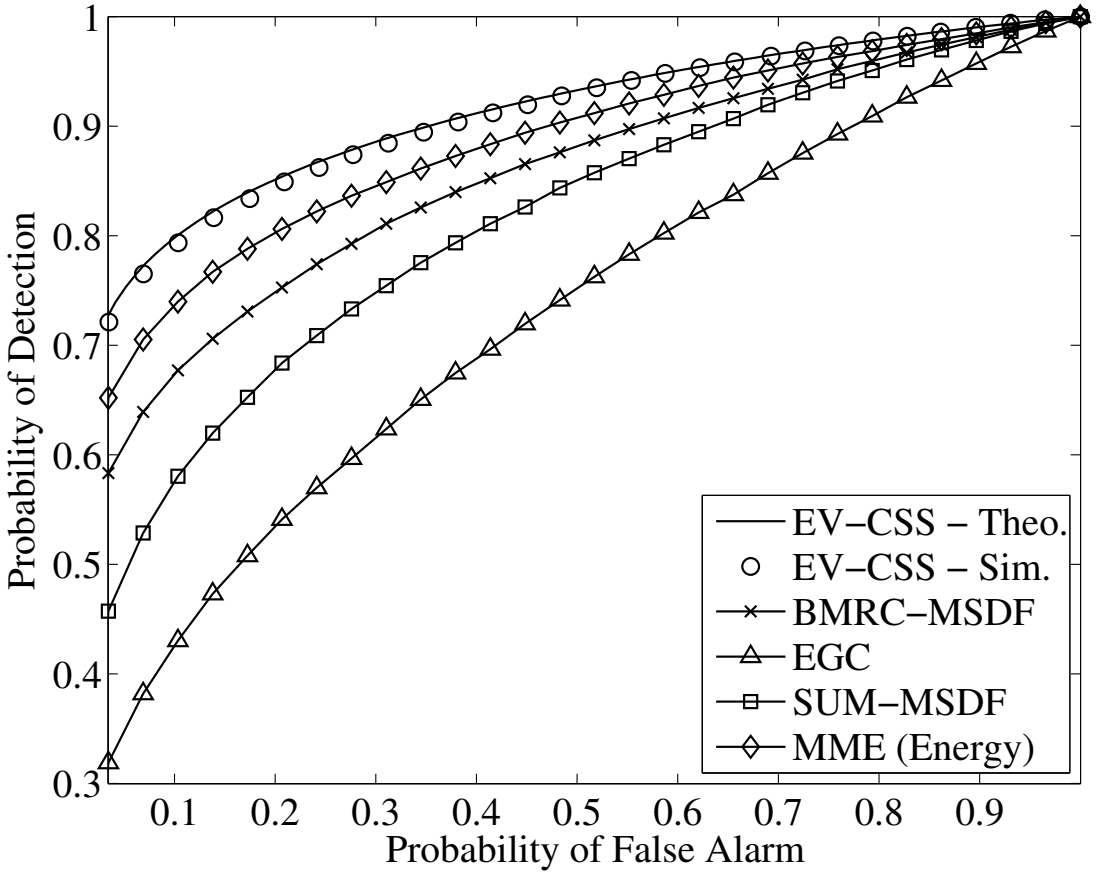


Figure 2.3: Receiver Operating Characteristic (ROC) of different cyclostationary-based spectrum sensing algorithms under Rayleigh flat-fading (SNR=-10 dB, $N = 1000$).

ered or if perfect channel state information (CSI) is assumed, show comparable performance between BMRC-MSDF and EV-CSS. Therefore, we can conclude that at very low SNR the blind channel estimates based on the SVD have large errors and the full benefit of MRC is not achieved. In contrast, the EV-CSS is able to fully take advantage of the information from all antennas because the test statistic works directly with the covariance matrices instead of utilizing an estimated CSI to pre-combine the signals.

Another reason for the performance gain is that EV-CSS works directly with the time-domain cyclic autocorrelation of the PU signal which is particularly effective if some prior knowledge about the cyclic frequency α_0 is provided. An FFT based scheme requires the use of some form of frequency smoothing which degrades the cyclic feature. However, BMRC-MSDF can be made more robust to inaccurate knowledge of the cyclic features through these smoothing techniques as described in [JLZ10]. This has no bearing in the scenarios presented in this work since perfect knowledge of α_0 is assumed and the parameters are chosen such that the cyclic frequency is perfectly aligned with an FFT bin.

The EGC approach performs worst among all the techniques in a flat-fading environment because at very low SNR the estimation of the phase of $\mathbf{h}$ has very substantial error resulting in the test statistic being degraded in most cases. In contrast to this SUM-MSDF is able to separately calculate the spectral correlation of the signal on each antenna which offers significant gain after combining.

An interesting result to note here is that EV-CSS performs slightly better than the MME technique. If there was perfect channel state information and the noise variance was also known precisely, the energy detection method is known to be the optimal detector. However, to deal with fading and unknown noise variance, the MME technique does not use the measured energy as the test statistic but instead uses the ratio of the maximum and minimum eigenvalue. This results in a slight degradation in performance as compared to a fixed threshold on the energy.

The effect of varying SNR on probability of detection is also shown in Fig. 2.4. In both analysis and simulation plots for EV-CSS, γ is set directly using (2.11) in order to maintain a CFAR of $P_{FA} = 0.1$. We again observe the strong agreement between theory and simulation for EV-CSS. In addition to this, the threshold selection is shown to be very effective in achieving the desired theoretical P_{FA}. As for the other techniques, the threshold needs to be determined as a function of SNR. This was achieved empirically in our simulations by Monte Carlo simulations of each method under $\mathcal{H}_0$. Although the distributions used for $\mathcal{H}_1$ are designed for local alternatives (very low SNR), the analysis still matches simulation very accurately. This is true even at SNR ≥ 0 dB since the distribution converges to a non-zero mean Gaussian at very high non-centrality parameter which accurately describes the test statistic at both high N and high SNR.

2.4.3 Varying Sample Size and Varying Number of Antennas

The probability of detection over varying number of samples $N \in [1000, 5000]$ under an SNR of -10 dB and random flat-fading is shown in Fig. 2.5. In these simulations the $P_{FA} = 0.1$ and the threshold is theoretically determined for EV-CSS. For the other techniques, the CFAR threshold is determined empirically through Monte Carlo simulations. As with varying SNR, only one threshold calculation is done for EV-CSS due to its independence to N. Significant improvements in detection probability of detection can be seen with increasing sample size. However, these gains tend to slow down with higher number of samples.

The effect of number of antennas, M, on detection accuracy is studied in Fig. 2.6. In this figure only the best two approaches (EV-CSS and BMRC-MSDF) are shown to facilitate the comparison. Note that for EV-CSS, to keep the the P_{FA} constant at 0.1 the threshold must be set to a new value based on (2.11). On the other hand, for BMRC-MSDF, the threshold is set for different SNR and M. The σ_η^2 across all antennas is assumed to be the same and an SNR is set

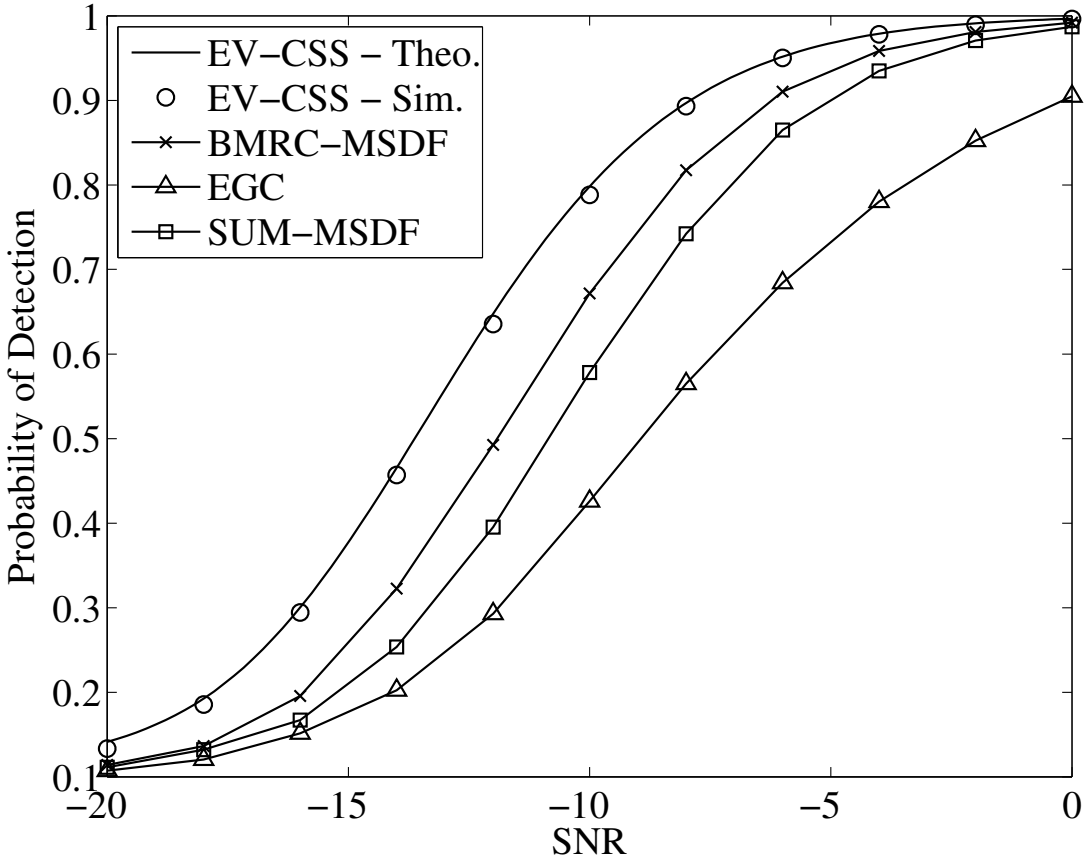


Figure 2.4: Comparison of multiple antenna cyclostationary spectrum sensing techniques with varying SNR under Rayleigh flat-fading ($M = 2$, $N = 1000$, $P_{FA} = 0.1$, uncorrelated noise).

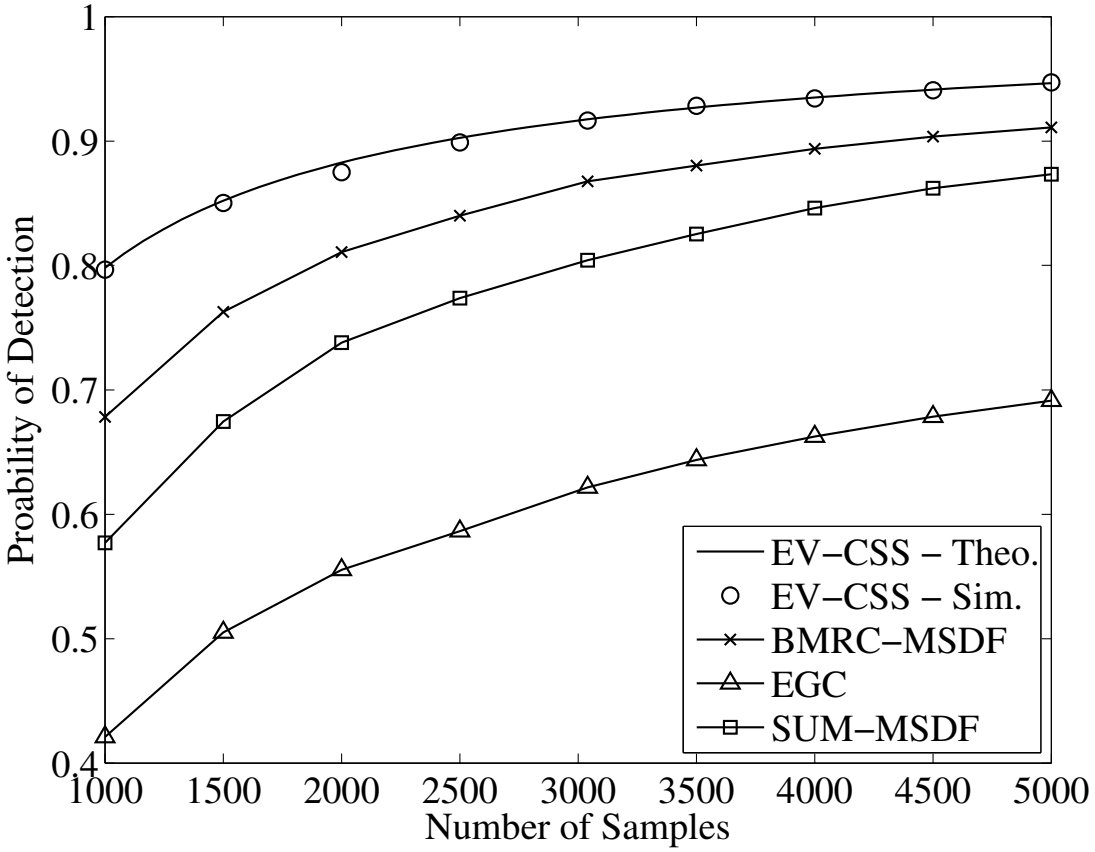


Figure 2.5: Comparison of multiple antenna cyclostationary spectrum sensing techniques with varying sample size N under Rayleigh flat-fading channel ($M = 2$, $N = 1000$, SNR=-10 dB, $P_{FA} = 0.1$, uncorrelated noise).

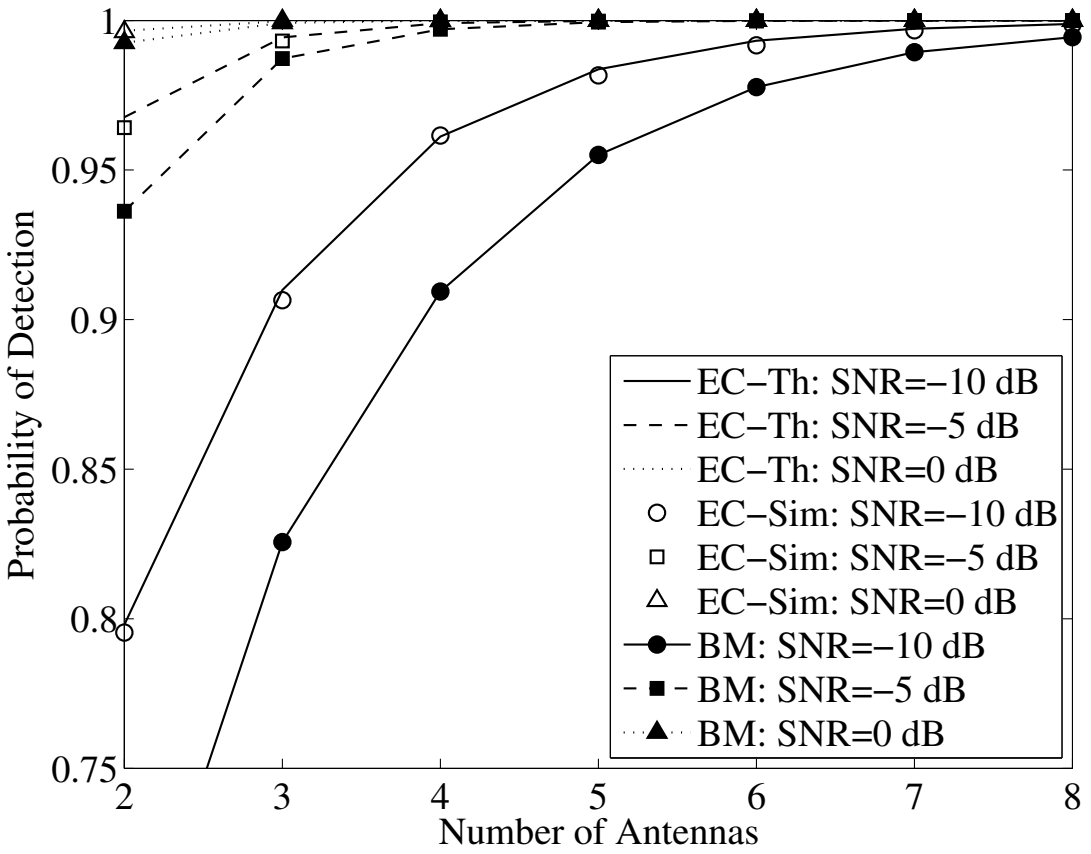


Figure 2.6: The effect of the number of antennas on the detection probability of EV-CSS (EC) and BMRC-MSDF (BM) at different values of SNR. The same number of samples per antenna, $N = 1000$, is used ($P_{FA} = 0.1$).

using (2.3). Since both algorithms utilize some form of blind channel estimate we expect both to perform successively better as M is increased due to the improved spatial diversity provided by multiple antennas due to independent fading. Similar to previous results, the EV-CSS has better performance than BMRC-MSDF for different values of M at low SNRs. The agreement between theory and simulation for EV-CSS remains very strong even with higher M.

2.4.4 Effect of Interfering Signal

In this section we investigate the performance of these spectrum sensing methods in the presence of multiple PUs. For this scenario we regard only one PU as the SOI and other PUs are seen as interfers. We test the robustness of these algorithms in the presence of a strong co-channel interferer by introducing another BPSK signal with the same symbol rate and with 30% spectral overlap. The effect of the interferer on the detection performance is shown in Fig. 2.7 as a function of the signal-to-interferer ratio (SIR), defined as the ratio of the interferer power to SOI power, is varied from -20 dB to 0 dB. The noise is kept constant at $\sigma_\eta^2 = 1$. The proposed algorithm shows the best signal selectivity among all the techniques. By performing the correlation entirely in time domain, the proposed method is able to suppress the interferer much better than the MSDF.

2.4.5 Spatially Correlated Noise

The effect of varying spatial correlation, ρ_s, is shown in Fig. 2.8. In this simulation we again have $M = 2$ antennas. In the BMRC-MSDF simulations we assume perfect knowledge of the channel with $\mathbf{h} = [1, 1]^T$ (no fading) and therefore the EVD-based blind channel estimation is no longer performed. This also results in BMRC-MSDF and SUM-MSDF having very comparable performance and thus, only one of them is shown. All methods, are degraded by increasing levels of spatial correlation. However, we see that EV-CSS consistently performs better at all levels of correlation. Also the performance of BMRC-MSDF degrades much faster and has comparable performance to SUM-MSDF at very high levels of correlation. SUM-MSDF is seen to be the most robust to spatially correlated noise since it calculates the spectral correlation for each antenna individually.

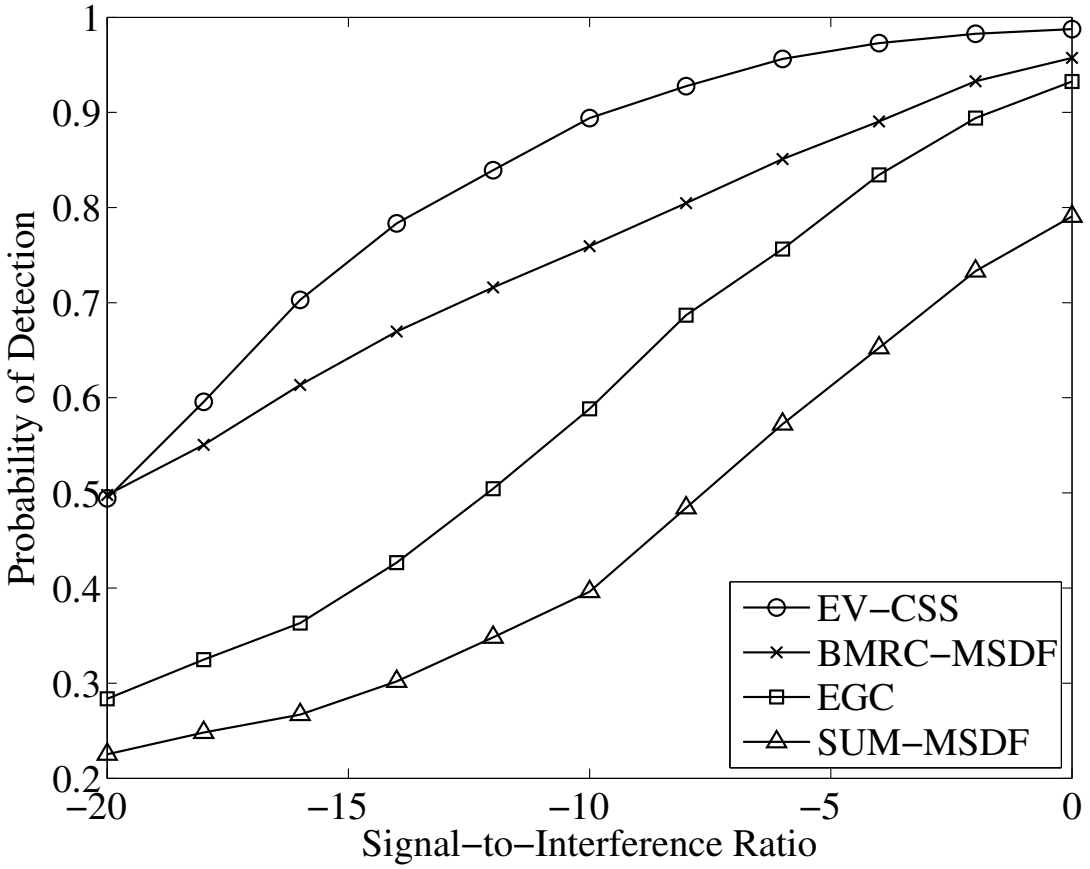


Figure 2.7: The effect a co-channel BPSK interfer on detection probability with 30% spectral overlap. The number of samples used for all antennas is $N = 1000$ and the noise level is kept constant at $\sigma_\eta = 1$ (P_{FA}).

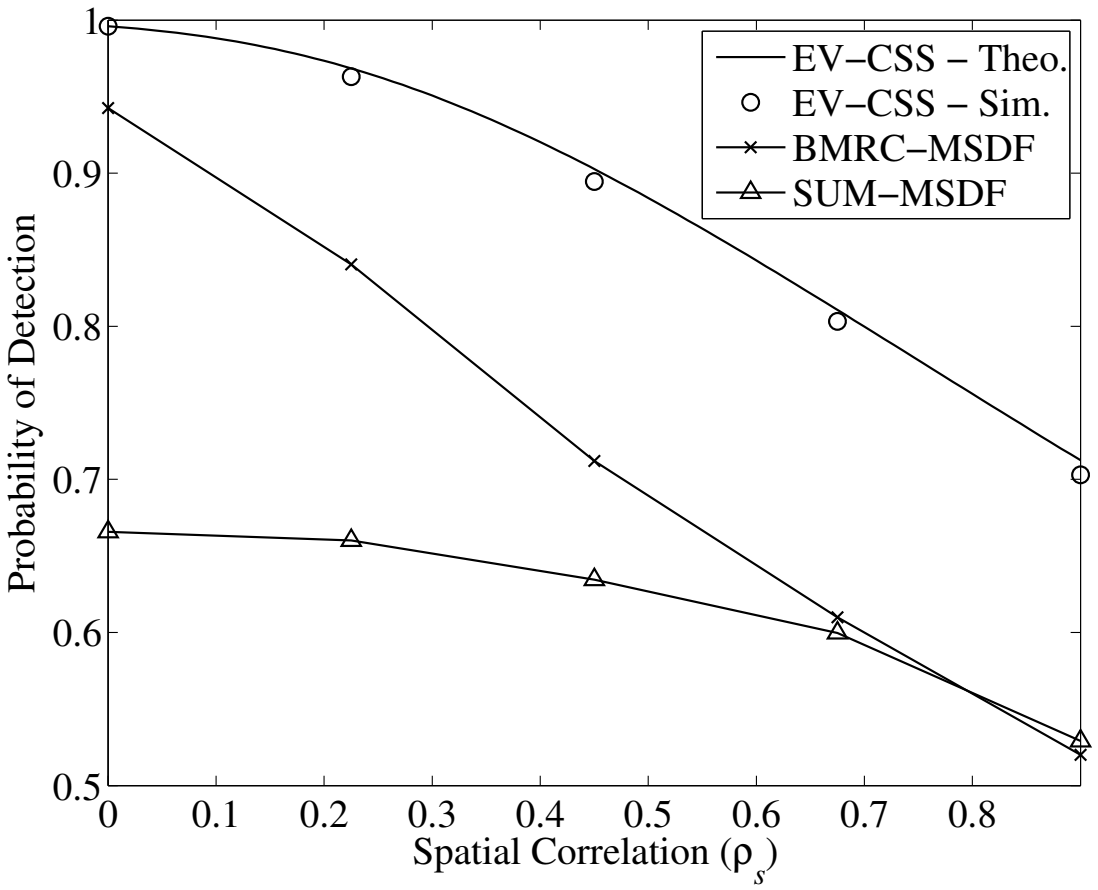


Figure 2.8: The effect a increasing spatial correlation on detection probability. A simple AWGN channel is used in these simulations to highlight the effect of spatial correlation ($N = 1000$, SNR=-10 dB, $P_{FA} = 0.1$).

2.4.6 Robustness to Noise Uncertainty

Finally, we test the proposed test statistic's robustness under noise uncertainty and compare it with other cyclostationary-based spectrum sensing techniques. Noise uncertatinty arises from the noise power level varying due to changes in the amount of thermal noise, amplifier gain, calibration error, and fluctuating interference [TS08]. The impact of the noise uncertainty is evaluated using the Bayesian statistics approach [TS06], where a prior distribution on the noise power is assumed. In our simulations this is achieved by having an SNR with uniform distribution.

$$f_{SNR}(x) = \begin{cases} \frac{1}{2\Delta}, & \overline{SNR} - \Delta \leq x \leq \overline{SNR} + \Delta \\ 0, & otherwise \end{cases}, \tag{2.31}$$

where $\overline{SNR}$ is the average SNR which we set to be -10 dB. We then calculate the P_{FA} and P_D by averaging over $f_{SNR}(x)$. The results for $\Delta \in [0, 3]$ dB are presented in Fig. 2.9 where the average $P_{FA} = 0.1$. As can be seen on the figure, all methods have similar robustness to noise uncertainty. The EV-CSS method offers performance gains over the two other methods even under very high levels of noise uncertainty. The slight degradation in EV-CSS is caused by averaging P_D over a uniformly distributed random variable SNR where the relationship between P_D and SNR is non-linear. As such, the values below $\overline{SNR}$ have a higher effect on the average P_D resulting in a net degradation.

2.5 Summary

In this chapter, a multi-antenna cyclostationary-based spectrum sensing method based on the cyclic correlation significance test was proposed and evaluated both analytically and through simulations. Conventional spectrum sensing forms the foundation of the advanced spectrum sensing system and it is crucial to solve

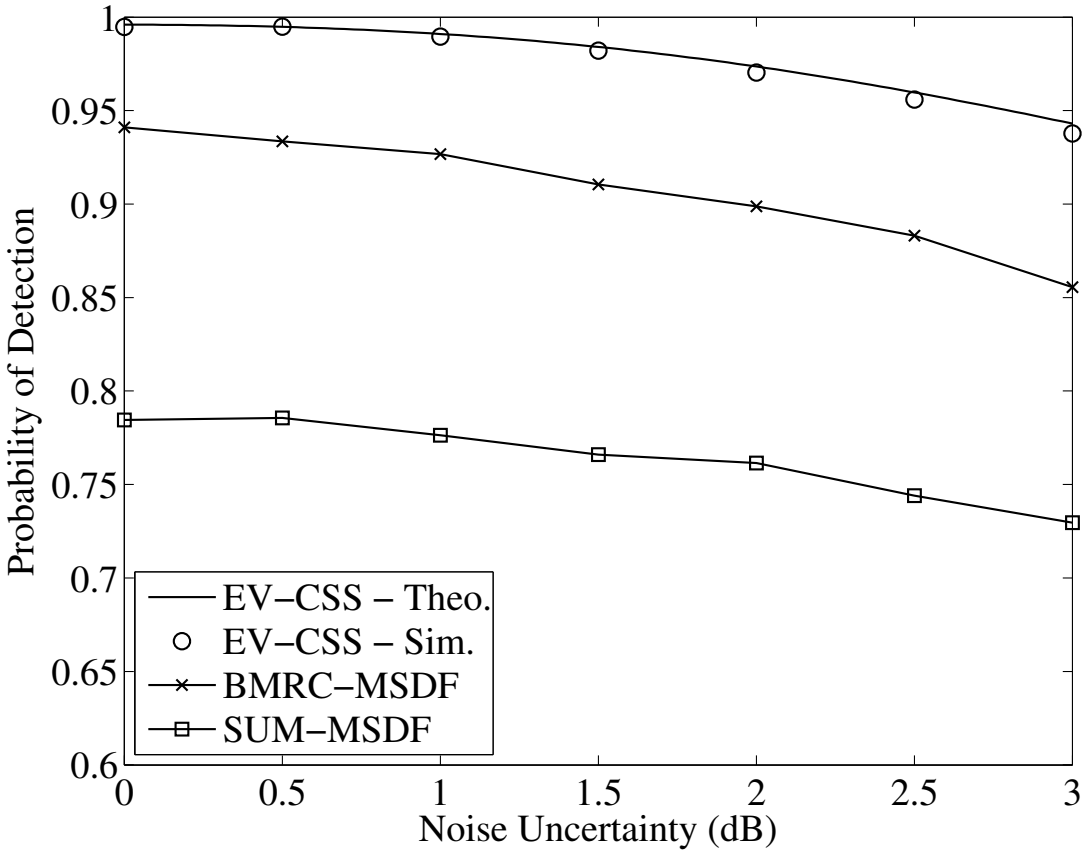


Figure 2.9: The effect of noise uncertainty, Δ, on the performance of various spectrum sensing schemes based on cyclostationarity. Noise uncertainty is assume to be uniformly distributed over an interval of $\left[\overline{SNR} - \Delta, \overline{SNR} + \Delta\right]$ dB. $\overline{SNR}$=-10 dB and $N = 1000$ samples (spatially uncorrelated noise)

issues relating to fading and interference. The proposed method was shown to outperform existing multiple antenna signal-selective spectrum sensing methods in the literature. In line with our key objectives for the advanced spectrum sensing system, the proposed approach reduces the computational complexity of current multi-antenna cyclostationary approaches. The computational complexity of the test statistic was also compared with that of the best performing existing algorithm that uses MRC by blindly estimating the CSI and was shown to require substantially less multiplications. The detection threshold for CFAR was also determined, both theoretically and via simulations, to be independent of the noise variance or the number of samples. This means that a single threshold is required for a given number of antenna, eliminating the need for separate noise estimation. Our proposed method has also been shown to be highly robust to the effects of noise uncertainty. Techniques such as EV-CSS allow for both accurate activity detection as well as feature detection which is an important part of solving the modulation classification problem which will be tackled in the next chapter.

CHAPTER 3

Sampled Distribution Distance-Based Classification

In this chapter we will focus on the modulation classification problem. Blind modulation classification allows us to further improve our knowledge of the behavior of primary users. In this chapter, we review various approaches for performing modulation classification in the context of advanced spectrum sensing. In particular, we present a novel approach to classifying the modulation order (number of constellation points). We refer to this as modulation level classification (MLC) method. This is a particularly difficult class of modulation classification problem which has received minimal treatment in prior work. The approach we will discuss in this chapter is based on concept of probability distribution distance functions. Recent work have shown the effectiveness of this type of classification technique. We further improve the practicality of these techniques by proposing key design changes that achieve low computational complexity and outperforms the state of the art methods based on cumulants and goodness-of-fit tests. These techniques are also verified in practice using a hardware implementation using an FPGA platform.

Later in the chapter we tackle the question of finding the optimal classifier based on distribution distances only. This method, based on the Bayesian decision criteria, asymptotically provides the minimum classification error possible given a set of testpoints. Testpoint locations are also optimized to improve classification performance. The method provides significant gains over existing approaches that

also use the distribution distance of the signal features.

3.1 Modulation Classification

Modulation level classification (MLC) is a process which detects the transmitter's digital modulation level from a received signal, using a priori knowledge of the modulation class and signal characteristics needed for downconversion and sampling. Among many modulation classification methods [DAB07], a cumulant (Cm) based classification [SS00b] is one of the most widespread for its ability to identify both the modulation class and level. However, differentiating among cumulants of the same modulation class, but with different levels, i.e. 16QAM vs. 64QAM, requires a large number of samples. A recently proposed method [WW10] based on a goodness-of-fit (GoF) test using Kolmogorov-Smirnov (KS) statistic has been suggested as an alternative to the Cm-based level classification which require lower number of samples to achieve accurate MLC.

We propose a novel MLC method based on distribution distance functions, namely Kuiper (K) [CDG04] [Ste74, Sec. 3.1] and KS distances, which is a significant simplification of methods based on GoF. We show that using a classifier based only on K-distance achieves better classification than the KS-based GoF classifier. At the same time, our method requires only $2ML$ additions in contrast to $2M(\log 2M + 2K)$ additions for the KS-based GoF test, where K is the number of distinct modulation levels, M is the sample size and $L \ll M$ is the number of test points used by our method.

3.2 Modulation Level Classification

3.2.1 System Model

Following [WW10], we assume a sequence of M discrete, complex, i.i.d. and sampled baseband symbols, $\mathbf{s}^{(k)} \triangleq [s_1^{(k)} \cdots s_M^{(k)}]$, drawn from a modulation order $\mathcal{M}_k \in \{\mathcal{M}_1, \ldots, \mathcal{M}_K\}$, transmitted over AWGN channel, perturbed by uniformly distributed phase jitter and attenuated by an unknown factor $A > 0$. Therefore, the received signal is given as $\mathbf{r} \triangleq [r_1 \cdots r_M]$, where $r_n = Ae^{j\Phi_n} s_n + g_n$, $\{g_n\}_{n=1}^M \sim \mathcal{CN}(0, \sigma^2)$ and $\{\Phi_n\}_{n=1}^M \sim \mathcal{U}(-\phi, +\phi)$. The task of the modulation classifier is to find $\mathcal{M}_k$, from which $\mathbf{s}^{(k)}$ was drawn, given $\mathbf{r}$. Without loss of generality, we consider unit power constellations and define SNR as $\gamma \triangleq A^2/\sigma^2$.

3.2.2 Classification based on Distribution Distance Function

The proposed method improves upon the MLC technique based on GoF testing using the KS statistic [WW10]. Since the KS statistic, which computes the minimum distance between theoretical and empirical cumulative distribution function (ECDF), requires *all* CDF points, we postulate that similarly accurate classification can be obtained by evaluating this distance using a smaller set of points in the CDF.

Let $\mathbf{z} \triangleq [z_1 \cdots z_N] = f(\mathbf{r})$ where $f(\cdot)$ is the chosen feature map and N is the number extracted features. Possible feature maps include $|\mathbf{r}|$ (magnitude, $N = M$) or the concatenation of $\Re\{\mathbf{r}\}$ and $\Im\{\mathbf{r}\}$ (quadrature, $N = 2M$). The theoretical CDF of z given $\mathcal{M}_k$ and γ, $F_0^k(z)$, is assumed to be known a priori (methods of obtaining these distributions, both empirically and theoretically, are presented in [WW10, Sec. III-A]). The K CDFs, one for each modulation level, define a set of test points

$$t_{ij}^{(\epsilon)} = \arg\max_z D_{ij}^{(\epsilon)}(z), \tag{3.1}$$

with the distribution distances given by

$$D_{ij}^{(\epsilon)}(z) = (-1)^{\epsilon} \left(F_0^i(z) - F_0^j(z) \right), \tag{3.2}$$

for $1 \leq i, j \leq K$, $i \neq j$, and $\epsilon \in \{0, 1\}$, corresponding to the maximum positive and negative deviations, respectively. Note the symmetry in the test points such that $t_{ji}^{(0)} = t_{ij}^{(1)}$. Thus, there are $L \triangleq 2\binom{K}{2}$ test points for a K order classification.

The ECDF, given as

$$F_N(t) = \frac{1}{N} \sum_{n=1}^{N} \mathbb{I}(z_n \leq t), \tag{3.3}$$

is evaluated at the test points to form $\mathbf{F}_N \triangleq \{F_N(t_{ij}^{(\epsilon)})\}$, $1 \leq i, j \leq K, i \neq j$. Here, $\mathbb{I}(\cdot)$ equals to one if the input is true, and zero otherwise. By evaluating $F_N(t)$ only at the test points in (3.1), we get

$$\hat{D}_{ij}^{(\epsilon)} = (-1)^{\epsilon} \left(F_N \left(t_{ij}^{(\epsilon)} \right) - F_0^j \left(t_{ij}^{(\epsilon)} \right) \right) \tag{3.4}$$

which are then used to find an estimate of the maximum positive and negative deviations

$$\hat{D}_j^{(\epsilon)} = \max_{1 \leq i \leq K, i \neq j} \hat{D}_{ij}^{(\epsilon)}, \quad 1 \leq j \leq K, \tag{3.5}$$

of the ECDF to the true CDFs. The operation of finding the ECDF at the given testpoints (3.4) can be implemented using a simple thresholding and counting operation and does not require samples to be sorted as in [WW10]. The metrics in (3.5) are used to find the final distribution distance metrics

$$\hat{D}_j = \max \left(\left| \hat{D}_j^{(0)} \right|, \left| \hat{D}_j^{(1)} \right| \right), \quad \hat{V}_j = \left| \hat{D}_j^{(0)} + \hat{D}_j^{(1)} \right|, \tag{3.6}$$

which are the reduced complexity versions of the KS distance (rcKS) and the K distance (rcK), respectively[1]. Finally, we use the metrics in (3.6) as substitutes to the true distance-based classifiers with the following rule: choose $\mathcal{M}_{\hat{k}}$ such that

$$\hat{k}_D = \arg \min_{1 \leq j \leq K} \hat{D}_j, \ \hat{k}_V = \arg \min_{1 \leq j \leq K} \hat{V}_j. \tag{3.7}$$

[1]Note that other non-parametric distances used in hypothesis testing exist (see introduction in e.g. [CDG04]), although for brevity they are not addressed here. We note, however, that our approach is easily applied to any assumed distance metric.

In the remainder of the chapter, we define $h_{\hat{D}}(\mathbf{F}_N) = \hat{k}_D$ and $h_{\hat{V}}(\mathbf{F}_N) = \hat{k}_V$, where $\hat{k}_D, \hat{k}_\mathrm{V} \in \{1, \ldots, K\}$.

3.2.3 Analysis of Classification Accuracy

Let $\mathbf{t} \triangleq [t_1 \cdots t_L]$ denote the set of test points, $\{t_{ij}^{(\epsilon)}\}$, sorted in ascending order. For notational consistency, we also define the following points, $t_0 \triangleq -\infty$ and $t_{L+1} \triangleq +\infty$. Given that these points are distinct, they partition $\mathbf{z}$ into $L+1$ regions. An individual sample, z_n, can be in region l, such that $t_{l-1} < z_n \leq t_l$, with a given probability, determined by $F_0^k(z)$.

Assuming z_n are independent of each other, we can conclude that given $\mathbf{z}$, the number of samples that fall into each of the $L+1$ regions, $\mathbf{n} \triangleq [n_1 \cdots n_{L+1}]$, is jointly distributed according to a multinomial PMF given as

$$f(\mathbf{n}|N, \mathbf{p}) = \begin{cases} \frac{N! p_1^{n_1} \cdots p_{L+1}^{n_{L+1}}}{n_1! \cdots n_{L+1}!}, & \text{if } \sum_{i=1}^{L+1} n_i = N, \\ 0, & \text{otherwise,} \end{cases} \tag{3.8}$$

where $\mathbf{p} \triangleq [p_1 \cdots p_{L+1}]$, and p_l is the probability of an individual sample being in region l. Given that $\mathbf{z}$ is drawn from $\mathcal{M}_k$, $p_l = F_0^k(t_l) - F_0^k(t_{l-1})$, for $0 < l \leq L+1$.

Now, with particular $\mathbf{n}$, the ECDF at all the test points is

$$\mathbf{F}_N(\mathbf{n}) \triangleq [F_N(t_1) \cdots F_N(t_L)], \quad F_N(t_l) = \frac{1}{N} \sum_{i=1}^{l} n_i. \tag{3.9}$$

Therefore, we can analytically find the probability of classification to each of the K classes as

$$\Pr(\hat{k} = \kappa | \mathcal{M}_k) = \sum_{\mathbf{n} \in \mathbb{N}^{L+1}} \mathbb{I}(h_{\hat{V}}(\mathbf{F}_N(\mathbf{n})) = \kappa) f(\mathbf{n}|N, \mathbf{p}), \tag{3.10}$$

for the rcK classifier. A similar expression can be applied to rcKS, replacing $h_{\hat{V}}(\cdot)$ with $h_{\hat{D}}(\cdot)$ in (3.10).

Table 3.1: Number of Operations and Memory Usage for Various Modulation Level Classifier Algorithms

Method	Multiply	Add	Memory
Cm	$6M$	$6M$	K
rcKS/rcK	0	$2ML$	$WL(K+1)$
KS/K	0	$2M(\log 2M + 2K)$	$KW\bar{N}$
rcKS/rcK (mag)	$2M$	$M(L+1)$	$WL(K+1)$
KS/K (mag)	$2M$	$M(\log M + 2K + 1)$	$KW\bar{N}$

3.2.4 Complexity Analysis

Given that the theoretical CDFs change with SNR, we store distinct CDFs for W SNR values for each modulation level (impact of the selection of W on the accuracy is discussed further in Section 3.2.5.2.) Further, we store KW theoretical CDFs of length $\bar{N}$ each. For the non-reduced complexity classifiers that require sorting samples, we use a sorting algorithm whose complexity is $N \log N$. From Table 3.1, we see that for $K \leq 3$ rcK/rcKS tests use less addition operations than K/KS-based methods [WW10] and Cm-based classification [SS00b]. For $K > 3$, the rcK method is more computationally efficient when implemented in ASIC/FPGA, and is comparable to Cm in complexity when implemented on a CPU. In addition, the processing time would be shorter for an ASIC/FPGA implementation, which is an important requirement for cognitive radio applications. Furthermore, their memory requirements are also smaller since $\bar{N}$ has to be large for a smooth CDF.

It is worth mentioning that the authors in [WW10] used the theoretical CDF, but used $\bar{N}$ as the number of samples to generate the CDF in their complexity figures. The same observation favoring the proposed rcK/rcKS methods holds for the magnitude-based (mag) classifiers [WW10, Sec III-A].

3.2.5 Results

As an example, we assume that the classification task is to distinguish between M-QAM, where M $\in \{4, 16, 64\}$. For comparison we also present classification result based on maximum likelihood estimation (ML).

3.2.5.1 Detection Performance versus SNR

In the first set of experiments we evaluate the performance of the proposed classification method for different values of SNR. The results are presented in Fig. 3.6. We assume fixed sample size of $M = 50$, in contrast to [WW10, Fig. 1] to evaluate classification accuracy for a smaller sample size. We confirm that even for small sample size, as shown in [WW10, Fig. 1], Cm has unsatisfying classification accuracy at high SNR. In (10,17) dB region rcK clearly outperforms all detection techniques, while as SNR exceeds ≈17 dB all classification methods (except Cm) converge to one. In low SNR region, (0,10) dB, KS, rcKS, rcK perform equally well, with Cm having comparable performance. The same observation holds for larger sample sizes, not shown here due to space constraints. Note that the analytical performance metric developed in Section 3.2.3 for rcK and rcKS matches perfectly with the simulations. For the remaining results, we set $\gamma = 12$ dB, unless otherwise stated.

3.2.5.2 Detection Performance versus Sample Size

In the second set of experiments, we evaluate the performance of the proposed classification method as a function of sample size M. The result is presented in Fig. 3.2. As observed in Fig. 3.6, also here Cm has the worst classification accuracy, e.g. 5% below upper bound at $M = 1000$. The rcK method performs best at small sample sizes, $50 \leq M \leq 300$. With $M > 300$, the accuracy of rcK and KS is equal. Classification based on rcKS method consistently falls slightly

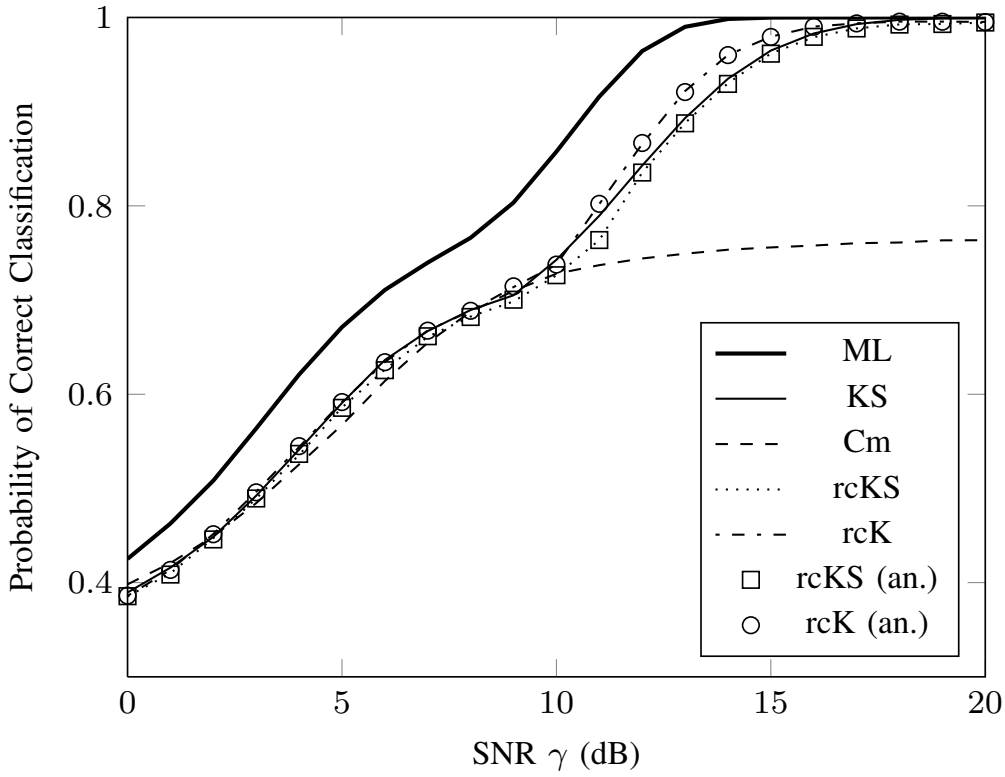


Figure 3.1: Effect of varying SNR on the probability of classification of various MLC techniques with M=50; (an.) – analytical result using (3.10).

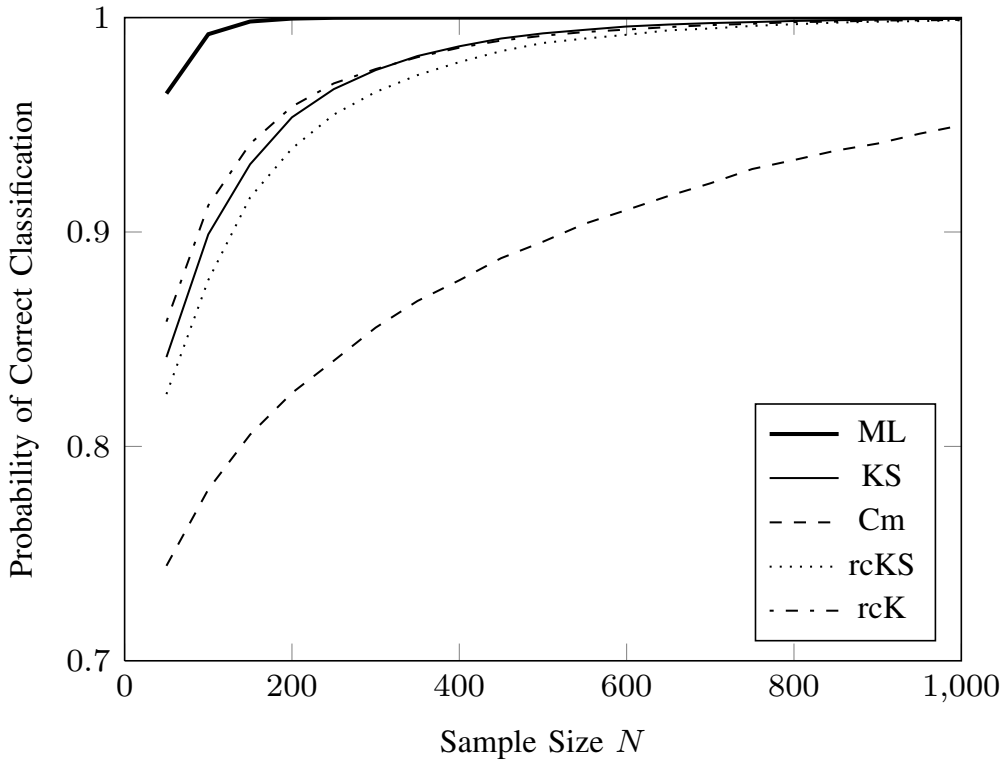
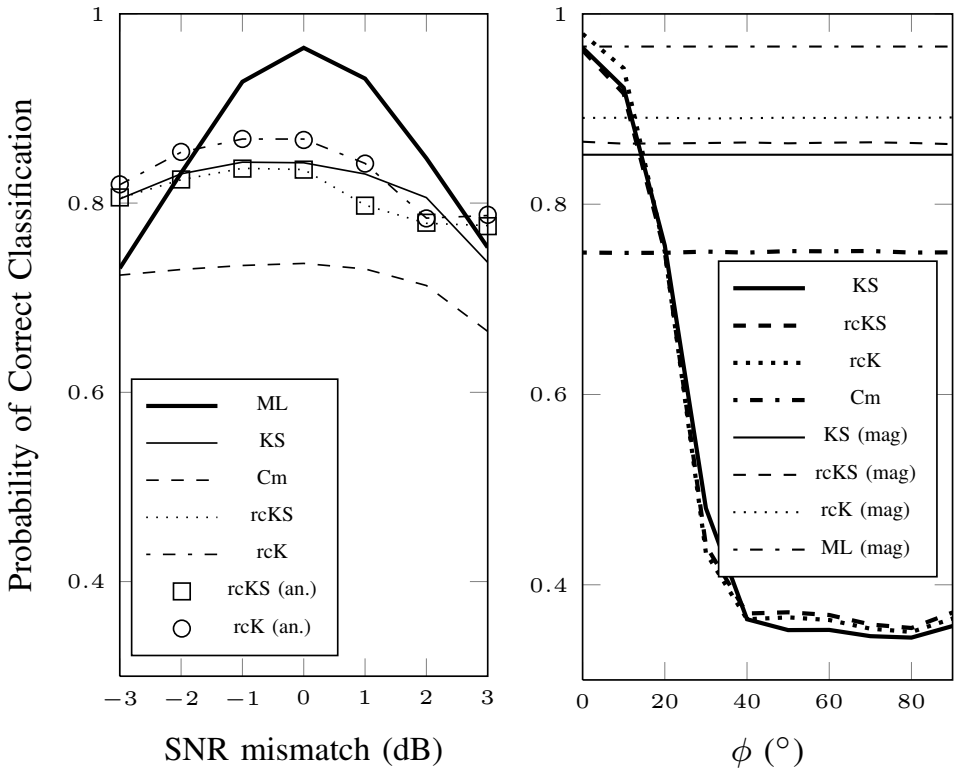


Figure 3.2: Effect of varying sample size on the probability of classification of various MLC techniques with $\gamma = 12\,\mathrm{dB}$.

below rcK and KS methods. In general, rcKS, rcK and KS converge to one at the same rate.

3.2.5.3 Detection Performance vs SNR Mismatch and Phase Jitter

In the third set of experiments we evaluate the performance of the proposed classification method as a function of SNR mismatch and phase jitter. The result is presented in Fig. 3.3. In case of SNR mismatch, Fig. 3.3(a), our results show the same trends as in [WW10, Fig. 4]; that is, all classification methods are relatively immune to SNR mismatch, i.e. the difference between actual and maximum SNR mismatch is less than 10% in the considered range of SNR values. This justifies the selection of the limited set of SNR values W for complexity evaluation used



(a) (b)

Figure 3.3: (a) Effect of SNR mismatch, nominal/true SNR=12dB; (b) effect of phase jitter, nominal SNR=15dB.; (an.) – analytical result using (3.10), (mag) – magnitude.

in Section 3.2.4. As expected, ML shows very high sensitivity to SNR mismatch. Note again the perfect match of analytical result presented in Section 3.2.3 with the simulations.

In the case of phase jitter caused by imperfect downconversion, we present results in Fig. 3.3(b) for $\gamma = 15\,\text{dB}$ as in [SS00b], in contrast to $\gamma = 12\,\text{dB}$ used earlier, for comparison purposes. We observe that our method using the magnitude feature, rcK/rcKS (mag), as well as the Cm method, are invariant to phase jitter. rcK and rcKS perform almost equally well, while Cm is worse than the other three methods by ≈10%. As expected, the ML performs better than all other methods. Quadrature-based classifiers, as expected, are highly sensitive

to phase jitter. Note that in the small phase jitter, $\phi < 10°$, quadrature-based classifiers perform better than others, since the sample size is twice as large as in the former case.

3.3 Optimal Sampled Distribution Distance Classification

We have thus presented a computationally efficient method for modulation level classification based on distribution distance functions. Specifically, we proposed to use a metric based on Kolmogorov-Smirnov and Kuiper distances which exploits the distance properties between CDFs corresponding to different modulation levels. The proposed method results in faster MLC than the cumulant-based method, by reducing the number of samples needed. It also results in lower computational complexity than the KS-GoF method, by eliminating the need for a sorting operation and using only a limited set of test points, instead of the entire CDF. However, the choice of the KS-GoF test as the basis of our reduced complexity approach might strike the reader as arbitrary. The idea is motivated mainly by the earlier work in [WW10] which focused on KS-GoF. The question remains as to which GoF test statistic gives the optimal performance. Alternatively, this problem is equivalent to finding the optimal classifier based only on the sampled ECDF. The ECDF forms the basis of all the GoF tests and is therefore the true underlying test statistic. So, given a set of measurements of the ECDF at particular test points, we want to find the optimal discriminant function that based on the classification accuracy.

Since we already have shown the effectivity of the rcK classifier, it stands to reason that we could use this as the basis of the optimal classifier. In fact, the idea of improving the classification accuracy of the rcK classifier by using more testpoints was proposed in [WC12]. The method is referred to as Variational Distance (VD) classifier where testpoints are selected to be the pdf-crossings of the

two classes being recognized. To generalize, we refer to methods such as rcK and VD, that utilize the ECDF at a small number of testpoints, as sampled distribution distance classifiers. In the rest of this chapter, we derive the optimal discriminant functions for classification with the sampled distribution distance given a set of testpoint locations. We also provide a systematic way of finding testpoints that provide near optimal performance by maximizing the Bhattacharyya distance between features. Finally, we present results that compare the performance of this approach with existing techniques.

3.3.1 Proposed Optimal Classifier

3.3.1.1 Classification Based on Sampled Distribution Distance

The common idea behind the techniques such as KS, rcK, and rcKS, which we have discussed so far, is that they are all based on the ECDF defined as

$$F_N(t) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{I}(z_n \leq t), \tag{3.11}$$

as the discriminating feature for classification. Here, $\mathbb{I}(\cdot)$ is the indicator function whose value is 1 if the function argument is true, and 0 otherwise. Once the ECDF is found and the distribution distance is calculated, the candidate constellation with minimum distance is chosen. However, we have shown that improved classification accuracy can be achieved at lower computational complexity by finding the value of the ECDF at a small number of testpoints.

We now describe these methods formally by defining a set of L testpoints: $\mathbf{t} = [t_1 \cdots t_L]$, with $t_{i+1} \geq t_i$ and $t_1 \ldots t_L \in \mathbb{R}$. We also define the following virtual test points, $t_0 \triangleq -\infty$ and $t_{L+1} \triangleq +\infty$ in addition to $\mathbf{t}$. Evaluating the ECDF from (3.11) at $\mathbf{t}$ gives us $\mathbf{x} = [x_1 \cdots x_L]$, $x_i \triangleq F_N(t_i)$. We refer to any classifier that utilizes the feature vector $\mathbf{x}$ as a *sampled distribution distance-based classifier*. An example of such a classifier is the variational distance (VD) classifier proposed

in [WC12], which chooses $\mathbf{t}$ to be at the local maxima or minima of the difference between two theoretical CDFs of the candidate classes. It could be shown that this is equivalent to finding the points where the two theoretical pdfs cross each other.

Our goal is to optimize the classification accuracy

$$P_C = \sum_{i=1}^{K} \Pr(\hat{k} = i \,|\mathcal{M}_i) \Pr(\mathcal{M}_i) \,. \tag{3.12}$$

Intuitively, there are two ways to improve P_C. First, since different testpoints have varying distribution distance, different weights should be assigned to each. Second, the number and location of testpoints should also be optimized to find the proper balance between complexity and classification accuracy. Both of these improvements are addressed in the following subsection.

3.3.1.2 Proposed Classifier

We first assume that $\mathbf{t}$ has been selected *a priori* and our goal is to find the optimal classifier for the resulting feature vector $\mathbf{x}$. We want to find a discriminant function $g_k(\mathbf{x})$, $k \in [1, K]$, for every candidate constellation $\mathcal{M}_k$. The classification rule is given as:

$$\text{Choose: } \mathcal{M}_i \text{ s.t. } g_i(\mathbf{x}) > g_j(\mathbf{x}) \,\forall\, j \neq i. \tag{3.13}$$

According to decision theory, if the average classification error is used as the performance metric, the optimal classifier is the *Bayes decision procedure* [DHS01] stated as:

$$\text{Choose: } \mathcal{M}_i \text{ s.t. } \Pr(\mathcal{M}_i \,|\mathbf{x}) > \Pr(\mathcal{M}_j \,|\mathbf{x}) \,\forall\, j \neq i. \tag{3.14}$$

Using the prior probabilities $\Pr(\mathcal{M}_i)$, the posterior probabilities $\Pr(\mathcal{M}_i \,|\mathbf{x})$ could be found from $\Pr(\mathbf{x} \,|\mathcal{M}_i)$ using Bayes formula. Thus, finding the pdf of the feature vector conditioned on the modulation scheme, $\Pr(\mathbf{x} \,|\mathcal{M}_i)$, effectively gives us the optimal classifier.

The testpoints partition $\mathbf{z}$, which is the feature vector space, into $L+1$ disjoint regions. An individual sample, z_n, falls into region l if $t_{l-1} < z_n \leq t_l$. This happens with a given probability determined by $F_0^k(z)$. We define the number of samples that fall into each of the regions as $\mathbf{n} \triangleq [n_1 \cdots n_{L+1}]$, where n_i corresponds to region i. This is analogous to a random experiment with N independent trials, each of which leads to a success for exactly one region, with each region having a given fixed success probability, p_l. The number of successes for each region is described by a multinomial probability mass function (pmf) given as

$$f(\mathbf{n}|N,\mathbf{p}) = \begin{cases} \frac{N! p_1^{n_1} \cdots p_{L+1}^{n_{L+1}}}{n_1! \cdots n_{L+1}!}, & \text{if } \sum_{i=1}^{L+1} n_i = N, \\ 0, & \text{otherwise,} \end{cases} \tag{3.15}$$

where $\mathbf{p} \triangleq [p_1 \cdots p_{L+1}]$. Given that $\mathbf{z}$ is drawn from $\mathcal{M}_k$, $p_l = F_0^k(t_l) - F_0^k(t_{l-1})$, for $0 < l \leq L+1$.

Given a particular $\mathbf{x}$, the number of samples in each region could be found as $n_i = N(x_i - x_{i-1})$ where $x_0 \triangleq 0$ and $x_{L+1} \triangleq 1$. This gives a mapping from any given $\mathbf{x}$ to $\mathbf{n}$ and therefore to the pmf $f(\mathbf{n}|N,\mathbf{p})$ as defined in (3.15). Combining this mapping with (3.15), we have the complete class-conditional pdf, $\Pr(\mathbf{x}|\mathcal{M}_k)$ with $\mathbf{p}$ determined by $F_0^k(z)$. Thus we have the optimal classifier given $\mathbf{t}$ and $F_0^k(z)$. We will refer to $\mathbf{x}$ and $\mathbf{n}$ conditioned on class $\mathcal{M}_k$ as $\mathbf{x}^{(k)}$ and $\mathbf{n}^{(k)}$.

Although the multinomial pmf in (3.15) gives the optimal classifier, its calculation is very computationally intensive. To address this issue we note that asymptotically the multinomial pmf, $f(\mathbf{n}|N,\mathbf{p})$ in (3.15), approaches a multivariate Gaussian distribution, $\mathbf{n}^{(k)} \sim \mathcal{N}(\boldsymbol{\mu}_k^{(n)}, \boldsymbol{\Sigma}_k^{(n)})$ as $N \to \infty$ [Are77], with

$$\boldsymbol{\mu}_k^{(n)} = N\mathbf{p} \tag{3.16}$$

$$\{\boldsymbol{\Sigma}_k^{(n)}\}_{ij} = \begin{cases} Np_i(1-p_i), & \text{if } i = j, \\ -Np_ip_j, & \text{if } i \neq j. \end{cases} \tag{3.17}$$

Since $\mathbf{x}$ is the cumulative sum of $\mathbf{n}$ divided by N (i.e. $x_i = \frac{1}{N}\sum_{j=1}^{i} n_j$), which is

a linear operation, it follows that $\mathbf{x}^{(k)} \sim \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ where,

$$\{\boldsymbol{\mu}_k\}_i = \sum_{j=1}^{i} p_j = F_0^k(t_i), \tag{3.18}$$

$$\{\boldsymbol{\Sigma}_k\}_{ij} = \frac{1}{N^2} \sum_{l=1}^{i} \sum_{m=1}^{j} \{\boldsymbol{\Sigma}_k^{(n)}\}_{lm}. \tag{3.19}$$

Having shown that the feature vector $\mathbf{x}$ is asymptotically Gaussian distributed, we can proceed to apply the *Bayes decision procedure* in (3.14). However, the full multivariate pdfs are no longer required because the optimal discriminant functions for Gaussian feature vectors are known to be [DHS01]:

$$g_k(\mathbf{x}) = \mathbf{x}^T \mathbf{W}_k \mathbf{x} + \mathbf{w}_k^T \mathbf{x} + w_{k0}, \tag{3.20}$$

where

$$\mathbf{W}_k = -\frac{1}{2}\boldsymbol{\Sigma}_k^{-1},\ \mathbf{w}_k = \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k, \tag{3.21}$$

and

$$w_{k0} = -\frac{1}{2}\boldsymbol{\mu}_k^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k - \frac{1}{2}\ln|\boldsymbol{\Sigma}_k| + \ln \Pr(\mathcal{M}_k). \tag{3.22}$$

In the following sections we will simply refer to this classifier as the Bayesian approach.

3.3.1.3 Testpoint Selection

In this subsection we present a method for choosing testpoint locations, $\mathbf{t}$, that provide good classification performance. The method of using the pdf-crossings makes intuitive sense, since it tries to find the testpoints that provide the maximum difference in the theoretical CDF while providing some heuristic rule that the testpoints will be sufficiently far apart. Testpoints that are close to each other are not as effective because the ECDF tends to be highly correlated.

Another issue with using the pdf-crossing is that it does not factor in knowledge of the correlation between testpoints. As we have shown in Section 3.3.1.2,

the distribution of $\mathbf{x}$ follows an approximate multivariate Gaussian with statistics given in (3.18) and (3.19). Therefore, the class-conditional means $\boldsymbol{\mu}_k$ and covariance matrices $\boldsymbol{\Sigma}_k$ are sufficient to completely describe the distribution of the feature vectors conditioned on $\mathcal{M}_k$. Thus, these statistics are also sufficient to find the optimal testpoint locations, $\mathbf{t}^*$.

To find the classification accuracy a K-dimensional integration is required and the limits are determined by the decision boundaries defined by (3.20), which in general are non-trivial. We replace exact P_C with a sub-optimum distance metric that is easier to evaluate and does not require integration. In particular, we use the Bhattacharyya distance first studied for signal selection in [Kai67]. Clearly, this approximation does not guarantee the optimal testpoint locations based on P_C. However, the K-dimensional integration is infeasible due to its intractability. The metric is shown here for reference:

$$D_B = \frac{1}{8}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right) + \frac{1}{2}\ln\left(\frac{\left|\left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right)/2\right|}{\sqrt{\left|\boldsymbol{\Sigma}_1\right|\left|\boldsymbol{\Sigma}_2\right|}}\right). \tag{3.23}$$

Note that the Bhattacharyya distance is calculated between 2 classes. As a result, the search for testpoints can only be performed for the $K = 2$ case. However, this could be done sequentially through all the possible pairs of $\mathcal{M}_k$. As D_B is a function of $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ which are functions of our testpoint selection, $\mathbf{t}$, then we can express it as $D_B(\mathbf{t})$. We thus find the good candidate testpoint by

$$\mathbf{t}^* = \arg\max_{\mathbf{t}} D_B\left(\mathbf{t}\right), \tag{3.24}$$

under the constraint $t_{i+1} \geq t_i$.

The testpoint selection method for our proposed classifier and the VD classifier is only applicable to the 2-class problem. As such, both approaches need to be adapted to handle the multi-class problem. One approach is to merge the testpoints for each possible pair of classes. However, this method results in a large

increase in complexity. Instead we follow the sequential classification procedure proposed in [WC12, Table. I] which involves first distinguishing between 4-QAM and 16-QAM, and then distinguishing between 16-QAM vs. 64-QAM if the first test results in 16-QAM. The 4-QAM vs. 16-QAM test is performed first because it has the highest classification accuracy. The testpoint selection is performed for these two tests independently.

As this is an L-dimensional optimization problem, a closed-form solution is beyond the scope of this work. Instead, we turn to numerical optimization methods (gradient descent methods) to find local maxima. The initial point of these procedures could be chosen to coincide with the pdf-crossings or equally spaced over some interval.

3.3.1.4 Note on Implementation

Similar to rcK discussed in Sec. 3.2 and VD [WC12], the Bayesian approach only needs to store the testpoint locations for a fixed set of SNRs since the theoretical CDF is dependent on SNR. In contrast, the ML classifier requires the pdf of each class under each SNR to be either calculated online or stored with high resolution. Thus the primary cost in implementing the ML classifier is the large look-up-table (LUT). Given a $\mathbf{t}$ of size L, VD and rcK require both $\mathbf{t}$ and $\boldsymbol{\mu}_k$ for each class $\mathcal{M}_k$. In contrast, the Bayesian approach requires the same vector $\mathbf{t}$, an $L \times L$ matrix $\mathbf{W}_k$, a vector $\mathbf{w}_k$ of size L, and a scalar w_{k0} for each class $\mathcal{M}_k$. For performance comparisons in Sec. 3.3.2, the same number of testpoints are used for both VD and Bayesian method. Based on the criteria used for testpoint selection for VD, which is to find the local maxima and minima of the difference between the theoretical CDFs, there are 4 testpoints for the 4-QAM vs. 16-QAM test (for all SNRs), and 4 testpoints for the 16-QAM vs. 64-QAM test for SNR ≤ 7 dB and 8 testpoints for SNR > 7 dB. Therefore there are no more than 12 testpoints per SNR, meaning that additional storage requirements are negligible.

In terms of additional computational complexity, the proposed approach requires the calculation of a quadratic form expression (3.20) which results in $L(L+1)$ additional multiplications. Again, due to the fact that only a relatively small number of testpoints is used, the additional complexity is minimal and is not a function of N but only of the number of testpoints L.

3.3.2 Results and Discussion

3.3.2.1 Testpoint Selection

For the results section we focus on the quadrature feature which is a concatenation of the I and Q component of each symbol. In Fig. 3.4, we show the results of the testpoint selection procedure with $M = 200$, under 0 dB SNR, for varying number of testpoints with the two class being 4-QAM and 16-QAM. The solid line plot corresponds to the difference of the two theoretical CDFs. We note that in the VD classifier the local maxima and minima of this plot are used as the testpoints. However, we find that the numerical optimization finds "good" testpoints to be close, but not exactly at the local maxima and minima. This is due to the additional information provided by the covariance matrices.

In contrast to VD classifier that has a fixed number of testpoints (4 for this particular problem) corresponding to the number of local maxima and minima, the optimization procedure allows more flexibility in choosing the number of testpoints. In Fig. 3.4, we show the result of the optimization procedure for a range of 1 to 8 testpoints. It confirms our intuition that "good" testpoints tend to be 1) spaced apart to avoid high correlation, 2) concentrated around locations that have high CDF difference, and 3) are not necessarily the same for different values of L. Intuitively, testpoints being located at the local minima and maxima yield suboptimal results once the covariance matrix is accounted for. These results confirm the need to jointly optimize the testpoint locations.

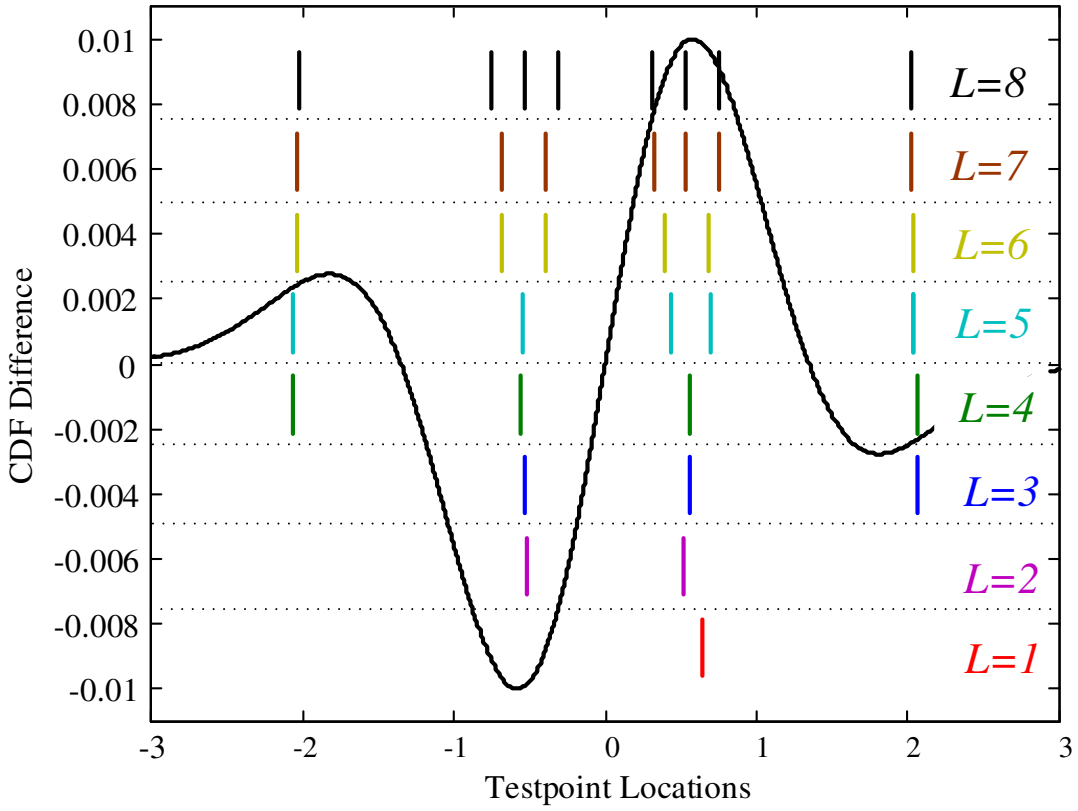


Figure 3.4: Optimized testpoint locations for varying number of testpoints, L. The solid line shows the CDF difference between the two classes (4-QAM and 16-QAM, under SNR $= 0$ dB, $M = 200$)

3.3.2.2 Comparison With Existing Techniques

As mentioned in the previous section, the proposed approach has the flexibility of varying the number of testpoints. This effectively gives more flexibility to trade-off classification accuracy with computational complexity. This idea is illustrated in Fig. 3.5. For $M = 1000$ and SNR $= 0$ dB, we show the classification accuracy of the proposed method as the number of testpoints is increased from 1 to 16, for all possible pairs of $\mathcal{M}_k$. The dotted lines correspond to the accuracy of the ML classifier which serves as an upperbound to classification accuracy, while the dashed lines correspond to that of the VD classifier. Note that both are plotted as horizontal lines because ML does not utilize testpoints, while VD has a fixed number of testpoints.

We see that the proposed method is able to match the accuracy of the VD classifier with only 3 testpoints. Further, the method's accuracy could be improved by adding more testpoints but at the cost of higher complexity. However, there exists a gap in performance between the Bayesian and the optimal ML classifier even with high number of testpoints. Increasing the number of testpoints even further increases the computational burden due to the need to perform an $L \times L$ matrix multiplication.

Finally, in Fig. 3.6, we compare the performance of the proposed method with the existing techniques under varying SNR with $M = 200$ symbols used for classification. To have a fair comparison, the same number of testpoints are used for both VD and Bayesian. For the entire range of SNR, the proposed Bayesian approach is shown to provide substantial gains over the VD classifier. We emphasize again that asymptotically, the proposed approach is the optimal classifier when using the sampled distribution distance as the discriminating feature. Also shown in the plot are the classification accuracy of the ML classifier which acts as the upperbound, and the conventional Kuiper classifier.

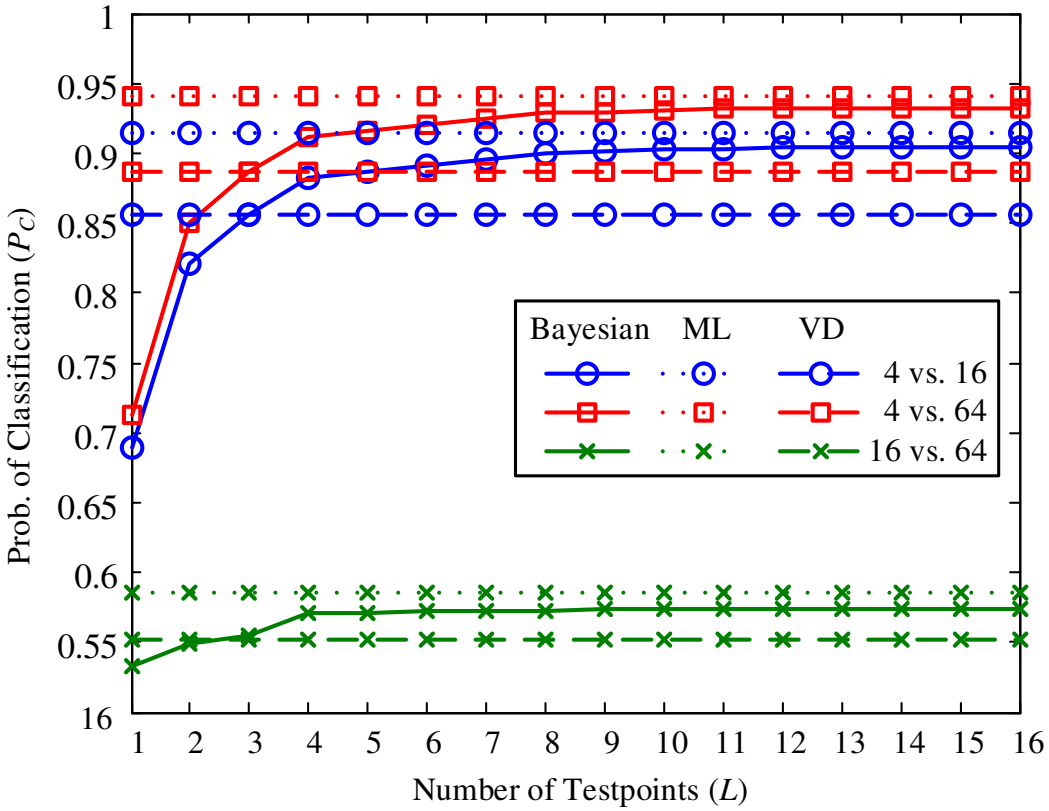


Figure 3.5: Effect of increasing number of testpoints on P_C for all possible pairs of constellations of interest. The classification accuracy of both ML and VD classifiers are also shown for comparison. (SNR = 0 dB, $M = 1000$)

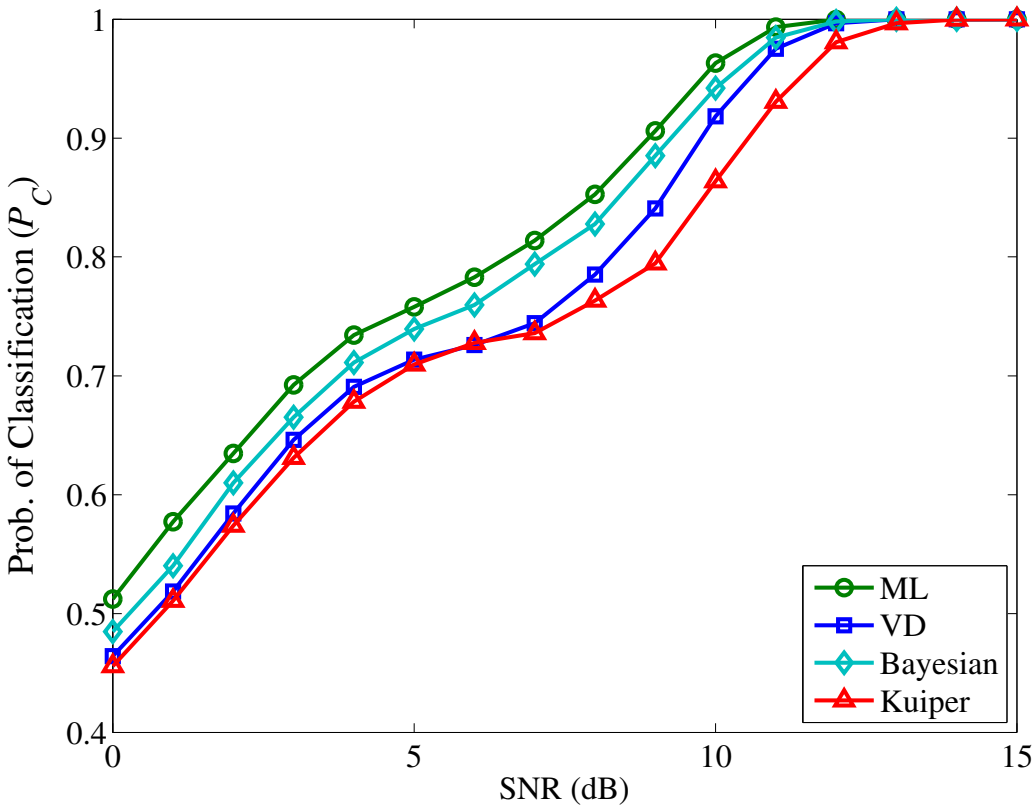


Figure 3.6: Comparison of the proposed Bayesian method with other existing approaches under varying SNR with $M = 200$ symbols used for classification. The same number of testpoints are used for both VD and Bayesian.

3.4 Hardware Implementation

In order to validate the idea of MLC in a practical platform, we implemented the reduced complexity Kuiper test on the BEE2 hardware prototyping platform and evaluated its performance. In particular, the classification accuracy of the proposed classifier in distinguishing among 4, 16, and 64-QAM is analyzed under varying SNR, varying number of symbols, and different timing offsets. The performance of the proposed classifier was also compared to that of the widely-used Cumulant-based classifier. Next, we propose possible architectures and their classification accuracy and computational complexity are compared. The proposed method achieves a probability of correct classification of 93% using 512 symbols at an SNR of 14dB as compared to an accuracy of 73% for the Cumulants method. The two classifiers were shown to have comparable hardware utilizations.

3.4.1 Motivation for Hardware Implementation

Although the algorithms presented in this chapter have been evaluated extensively through simulations, they all work under the assumption that the modulation classifier has access to the transmitted symbol corrupted by only additive white Gaussian noise (AWGN). This assumption implies perfect timing synchronization recovery and perfect channel state information which is unrealistic in a practical signal interception scenario. In order to understand the practicality of synchronous modulation classification, the effects of quantization errors, frequency offsets, and timing offsets must be quantified. In the remainder of this chapter, the effects of practical impairments on the classification performance are evaluated by implementing both the reduced complexity Kuiper and the Cm-based classifiers on hardware prototyping platform. The experiments were conducted on the Berkeley Emulation Engine 2 (BEE2) [CWB05] platform using two radio front-ends [Mis05] operating in the ISM band.

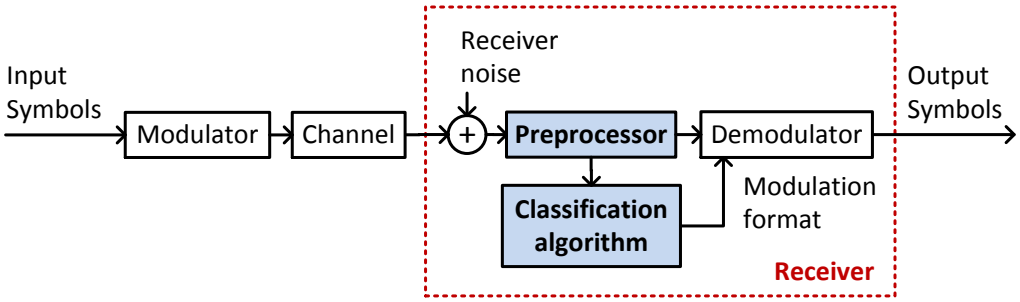


Figure 3.7: General system model for adaptive demodulators with semi-blind modulation classifier.

3.4.2 System Model and Overview of Modulation Classification

3.4.2.1 General Modulation Classification Flow

Most modulation classifiers rely on obtaining information symbols, which could be acquired by passing the incoming signal through a pre-processor which estimates parameters that are not known a priori. In this work, we limit the scope by considering imperfect frequency and phase estimates, SNR mismatches, and inaccurate timing synchronization. The general block diagram for an automatic modulation classification (AMC) system is shown in Fig. 3.7. This figure illustrates the basic concepts of all existing modulation classifiers, as well as blind demodulators. It consists of a preprocessing block that conditions the signal based on the particular classification algorithm to be used, and a classification block which we will be discussing in the next section.

3.4.2.2 Kuiper-based Modulation Classification

The algorithm to be implemented is based on the Goodness of Fit (GoF) features as discussed in this chapter. The idea behind GoF-based classification is to map the empirical cumulative distribution function (ECDF) of the received symbols

to the closest known CDF. Different criteria, such as magnitude and/or phase of received symbols, can be chosen as the statistical variable for classification. This criteria will be referred to as the classification feature. Each feature has a distinct CDF; once the feature has been selected, the ECDF for that feature is computed, and the *distance* between the ECDF and all theoretical CDFs for modulations of interests is calculated. Different *distance* computation methods will result in different performance classifications. For instance, the Kolmogorov-Smirnov method is based on the maximum deviation between the ECDF and every CDF [WW10]. Finally, the modulation level whose CDF is closest to the ECDF according to the selection criterion is chosen as the classified modulation level.

All existing GoF algorithms require two hardware intensive operations: 1) sorting the received symbols to form the ECDF, and 2) storing the theoretical CDF at all points. On the other hand, our proposed approach, the reduced Kuiper test, requires computing the ECDF at a limited number of points, which we call test-points. In addition, the computational complexity of our algorithm requires no multiplications, and simulation results show that it outperforms all existing algorithms. For further details about the algorithm and its theoretical classification performance, the reader is referred to Sec. 3.3.1.

It is important to note that the performance of all GoF-based tests is directly related to the feature that will be used for classification. For instance, one could obtain N complex information symbols s_k, and concatenate the real and imaginary values of each symbol yielding a vector of length $2N$. This is referred to as quadrature classification. In contrast, one could also compute the magnitude of each of the received symbols, and obtain a vector of length N. Classification using magnitude vectors is referred to as magnitude classification.

Although quadrature classifiers outperform magnitude classifiers for the same vector length, quadrature classifiers are very sensitive to frequency offsets. In

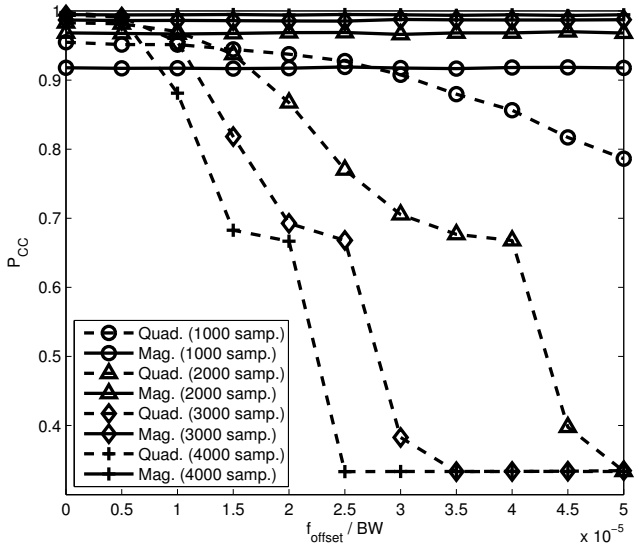


Figure 3.8: Tolerance of the reduced complexity Kuiper classifier to frequency offset at SNR = 10dB.

the context of signal interception, such offsets are hard to fully mitigate, and the performance of quadrature classifiers under slight offsets degrades at a fast rate as shown in Fig. 3.8, where the classification performance is plotted as a function of normalized frequency offset with respect to the signal bandwidth. In contrast, magnitude-based classifier are robust to both frequency and phase offset. Therefore, implementing magnitude classifiers circumvents the need to finely estimate the frequency and phase offsets, and the performance drop can always be mitigated by increasing the number of information symbols used for classification. We wish to note that the Cm-based classifier is robust to frequency offsets, and therefore is suitable to use under frequency and phase offsets as well.

Finally, we would like to note that the information symbols have to be normalized to unit energy before computing the ECDF in order to have a proper comparison against the stored known CDFs, assumed to be obtained from unit-energy information symbols. After downconversion from IF to baseband, the signal is sampled, passed through a matched filter assumed to be known in order

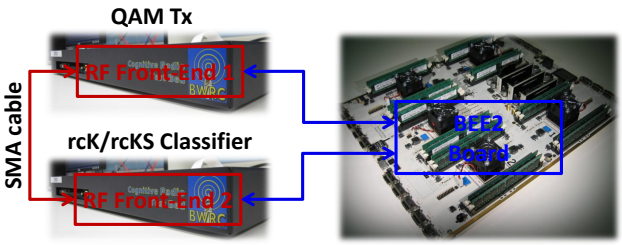


Figure 3.9: Experimental Setup for Modulation Level Classification

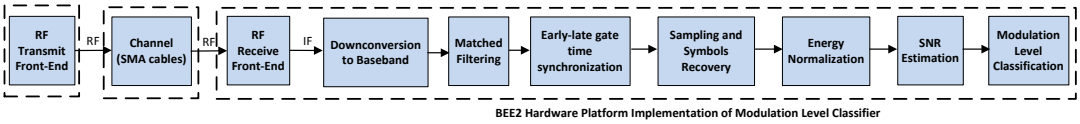


Figure 3.10: Architectural building blocks

to maximize the SNR. Note that the pulse shape was assumed to be known, but could be estimated blindly as well. Next, the symbols are normalized and passed to the classifier.

3.4.3 Experimental Setup and Modulation Classification Architecture

3.4.3.1 Experimental Setup

In our experiments, we use the Berkeley Emulation Engine 2 (BEE2) [CWB05] as the prototyping platform which we connect to two RF front-ends that transmit in the 2.39-2.49 GHz range. We designed our transceivers to transmit and receive QAM waveforms modulated using a square root rate cosine pulse shape filter with rolloff factor $\beta = 0.2$ at a rate of 2 Msymbols/sec. In all experiments, the transmit center frequency was set to 2.485 GHz to minimize the interference with other signals in the ISM band, and the receiver front-end downconverted the incoming signal to an intermediate frequency of 4 MHz. The setup is shown Fig. 3.9 where the two front-ends are connected via SMA cables to emulate a flat fading channel

with Additive White Gaussian Noise (AWGN).

Our Tx design, which is implemented on the RF front-end [Mis05], allows us to configure the transmitter to choose among 4, 16 and 64 QAM at different SNR levels. Given that we will be implementing the magnitude-based Kuiper classifier, fine carrier frequency and phase offsets is no longer required. In Fig. 3.11, we present the main architectural building blocks of our proposed classifier. Assuming perfect knowledge of the transmit pulse shape at the receive side, we matched-filter the incoming signal in order to maximize the receive SNR. The main architectural blocks are blind timing synchronization, SNR estimation and energy normalization, followed by the modulation classifier. The setup allows us to study the effect of timing offsets at the receiver end, and energy normalization errors on the classification performance, imperfections that often arise in the context of signal interception.

3.4.3.2 Early-Late Gate Timing Synchronization Recovery

Timing synchronization, which entails selecting the optimal sampling point after matched-filtering, is essential in the sampling process. In the context of signal interception, training sequences cannot be acquired in order to start the sampling process at the right sampling instant, and therefore timing synchronization has to be performed blindly. One of the possibly candidates for timing synchronization recovery is the early-late gate (ELG) sampler. The sampling rate and the symbol rate were chosen so that we obtain 16 samples per information symbol, and therefore there are 16 possible sampling instances. We refer to each of the sampling instances by an address $u \in [0, ..., 15]$. The ELG takes advantage of the correlation of the processed signal with the pulse shaping filter, the result of which produces a peak. The idea behind ELG is that it samples the output of the matched-filter at two timing instances, compares their values, and feeds back the values to the address list u. The sampling instance is shifted adaptively until the

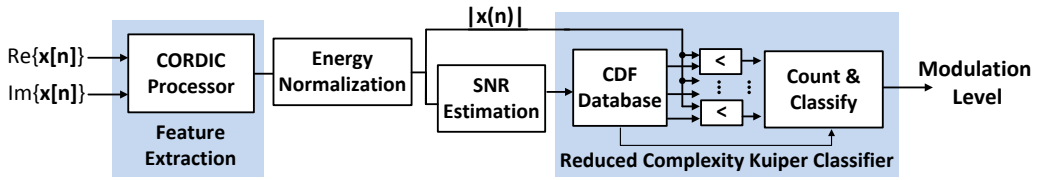


Figure 3.11: Architecture of Proposed Modulation Level Classification Algorithm

difference between the values of the processed signal is minimized.

Before obtaining any set of information symbols for classification, the ELG block is enabled for a fixed processing time required for the iteration-based timing synchronization recovery method to converge.

3.4.3.3 Modulation Level Classification

The reduced complexity Kuiper (rcK) classifier is composed of a feature extraction block, followed by an energy normalization block that measures the average symbol energy, and scales the symbols accordingly to achieve unit average energy. Next, the receiver SNR is estimated, and used to select the corresponding test-points and theoretical CDF from a look-up table stored in a ROM. Finally, a bank of comparators are applied to the incoming feature samples for classification.

Fig. 3.11 depicts the high level block diagram of the proposed classifier. The result of each threshold operation is used to increment several counters. The value of these counters after N symbols have been processed is used as the classification statistic. This block of comparators is repeated for each modulation level that we need to classify, and the outputs of which are compared to find the minimum rcK distance and subsequently to classify.

3.4.4 Results and Discussion

In this section, we first show the hardware resources in terms of occupied area for each of the implemented classifiers. Next, we focus our efforts on showing effect of the impairments on the classification performance. The considered impairments are: 1) quantization effects, which can be seen in all results below, 2) timing synchronization impairments, and 3) energy normalization impairments. We wish to note that it has been proven in [SS00b] and [URP11] that both classifiers studied in this section are robust to SNR mismatches, which is proportional to the error between true and estimated SNR, and therefore this aspect of impairments were not studied experimentally.

3.4.4.1 Hardware Resources Comparison

The FPGA hardware resource breakdown estimates of the proposed architectures are tabulated in Table 3.2. The bulk of the FPGA area is primarily allocated to the Preprocessing block which is utilized by both classifiers, where the matched-filtering and timing synchronization are performed. This can be attributed mainly to two reasons, first is that the preprocessing block needs to operate at the highest sampling rate (64 MHz) in order to give the oversampling ratio required for the ELG. Secondly, both the downconversion and filtering blocks require several multipliers that operate at the same high frequency. On the other hand, no dedicated multipliers were used for both classifiers since they operate at a slower rate than the system clock.

As for the two classifiers, they occupy comparable FPGA area. Although the two algorithms are inherently different, the word-lengths at various points in the datapath were selected to match as closely as possible. We note that the cumulant classifier utilizes fewer memory elements (FFs and BRAMs) while the Kuiper classifier has less Look-Up-Tables. This can be attributed to the need to

Classifier	Slices	FFs	LUT	BRAMs	Mult
Proprocessing	7882	14709	13127	5	3
Reduced Comp. Kuiper	2904	6647	4431	24	0
Cumulant	2724	4416	4585	0	0

Table 3.2: Hardware resource breakdown estimates of the proposed classifiers

store the testpoint and ECDF values. Although the FPGA area provides us with a rough estimate of the complexity of each classifier, we also need to take into account their latency. The cumulant classifier has much higher latency due to the need to calculate several *magnitude* and *square* operations.

3.4.4.2 Classification Performance versus Sample Size

We start by defining the probability of correct classification P_c as

$$P_c = \frac{1}{3}\left[\frac{1}{N_s}\sum_{n=1}^{N_s}\mathcal{I}\{o = m|M = m\}\right], \tag{3.25}$$

where $\mathcal{I}\{\}$ is the indicator function, o is the output of the modulation classifier based on N symbols, $M \in [4, 16, 64]$ is the modulation level used by the transmitter, and N_s is the number of realizations used in the experiments, set to $N_s = 10000$ for all experimental results given below. In the first set of experiments, we compare (in Fig. 3.12) the classification performance obtained through the hardware platform to the theoretical classification performance of both classifiers as a function of number of symbols, varied from 64 to 4096 at a fixed SNR of 14 dB. The gap between the theoretical and experimental results is a direct consequence of the hardware impairments including residual timing offset that result from the limited time resolution, and the inherent quantization effects of a hardware implementation. However, given that we are operating on a magnitude-based classifier, the impairments that arise from frequency and phase offsets do not affect the classification performance, and the only remaining impairments are

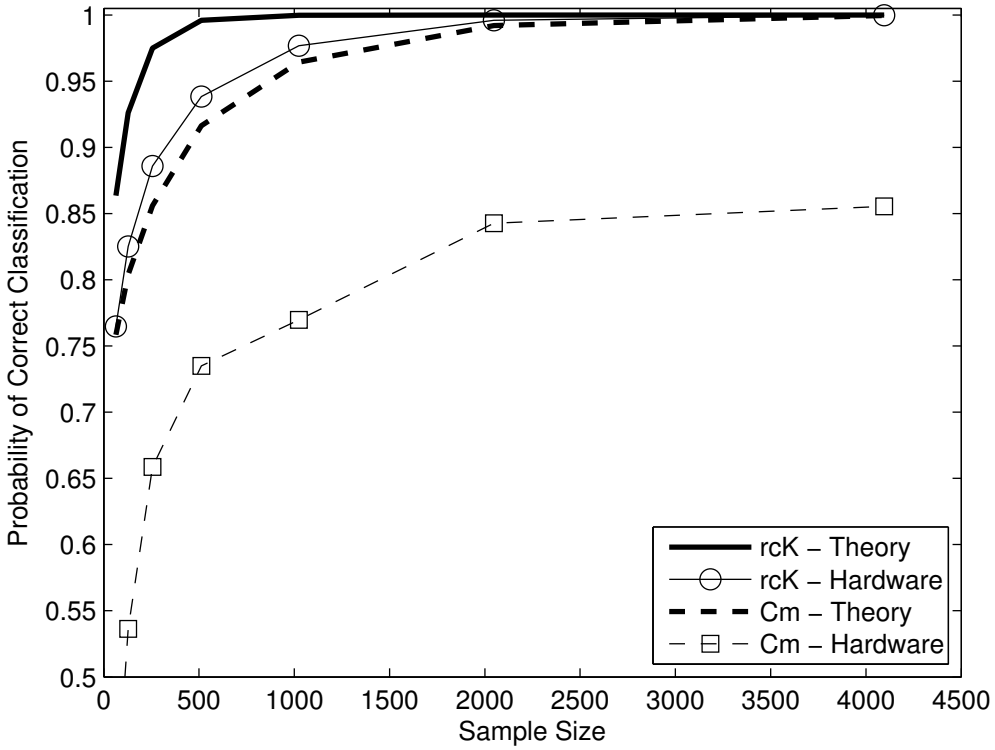


Figure 3.12: Probability of correct classification versus number of symbols at a fixed SNR = 14 dB.

quantization errors, errors in normalizing the incoming symbols to unit energy, and errors in estimating the SNR. Note that all three impairments are common to both classifiers, which makes the performance comparison fair. We note that the theoretical classification performance is an upper bound for the experimental results. In addition, the results in Fig. 3.12 show that the Cm classifier is very sensitive to the normalization factor that is used to normalize the symbol energies. In fact, given that the Cm classifier is based on fourth order cumulants, which is sample mean of the fourth moment of the incoming symbols. Therefore, any normalization error due to the finite number of samples would have a compounded effect on the cumulant, which proves that the Cm is very sensitive to normalization impairments. This effect is further studied in Section 3.4.4.4.

3.4.4.3 Classification Performance versus Timing Offset

In this subsection, we evaluate the performance of both classifiers under inaccurate timing synchronization, which could arise due to the lack of synchronization, limited number of samples, or a low SNR. In this experiment, we shifted the sampling instances to span half the symbol period, beyond which the effect of sampling synchronization would be symmetric. Fig. 3.13 shows the classification performance of both classifiers as a function of the normalized timing offset with respect to the symbol period. Both classifiers show a similar trend, and the classification performance drops to 0.33 at about a ratio of $0.25\frac{\text{Timing Offset}}{\text{Symbol Period}}$. This result shows that if the classification were to be done without any timing synchronization, the sampling instance will be uniformly distributed between 0 and T, and the classifier will yield a probability of classification of 0.33 60% of the time.

3.4.4.4 Classification Performance versus Normalization Error

After recovering the N information symbols, the modulation classifier block first estimates the symbol energy by $\hat{\mathcal{E}} = \frac{1}{N}\sum_{n=1}^{N}|x(n)|^2$, and normalizes the complex-valued symbols by $\sqrt{\mathcal{E}}$. It is obvious that increasing the sample size will improve the estimate of $\hat{\mathcal{E}}$, and improve classification accuracy, but will also reduce the normalization mismatch. As $N \to \infty$, the estimated symbol energy $\hat{\mathcal{E}} \to \mathcal{E}$, where $\mathcal{E}$ is the asymptotic symbols average energy. In order to isolate the normalization effect, we keep the number of symbols used for classification as $N = 4096$ symbols, and vary the number of symbols used for the energy estimator. Let $\delta_{\mathcal{E}} = \mathcal{E}/\hat{\mathcal{E}}$ be the normalization mismatch. Fig. 3.14 shows the effect that the normalization mismatch on the experimental classification performance for both classifiers. This shows that the claim made earlier is indeed valid, and proves that Cm classifiers are less robust to normalization mismatch.

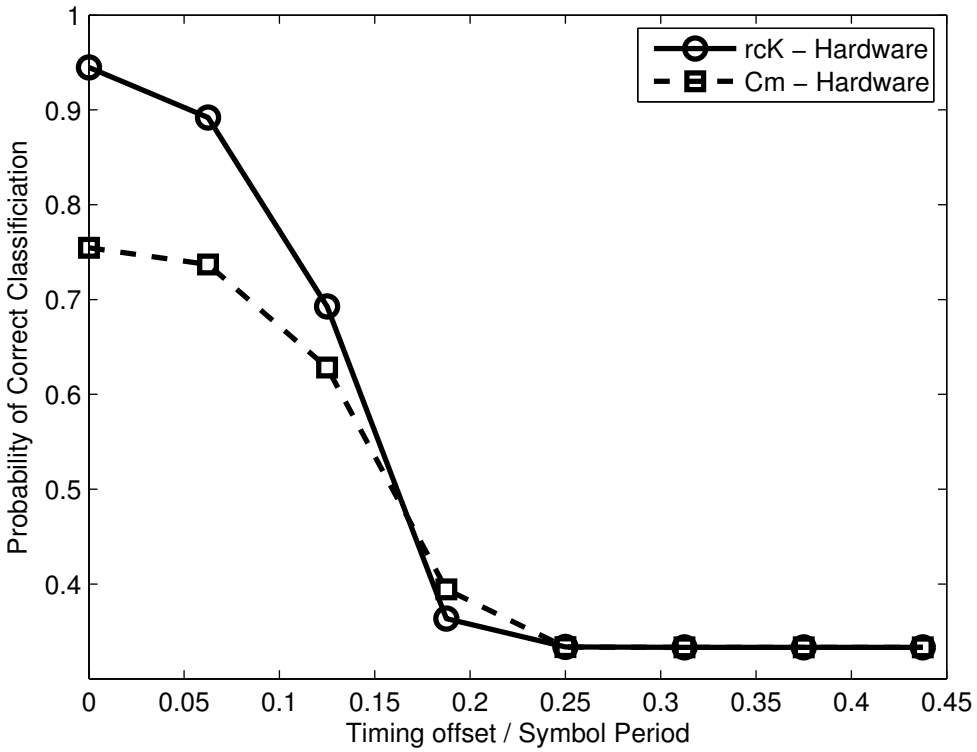


Figure 3.13: Probability of correct classification versus relative timing offset at a fixed SNR = 14 dB, and N = 1024.

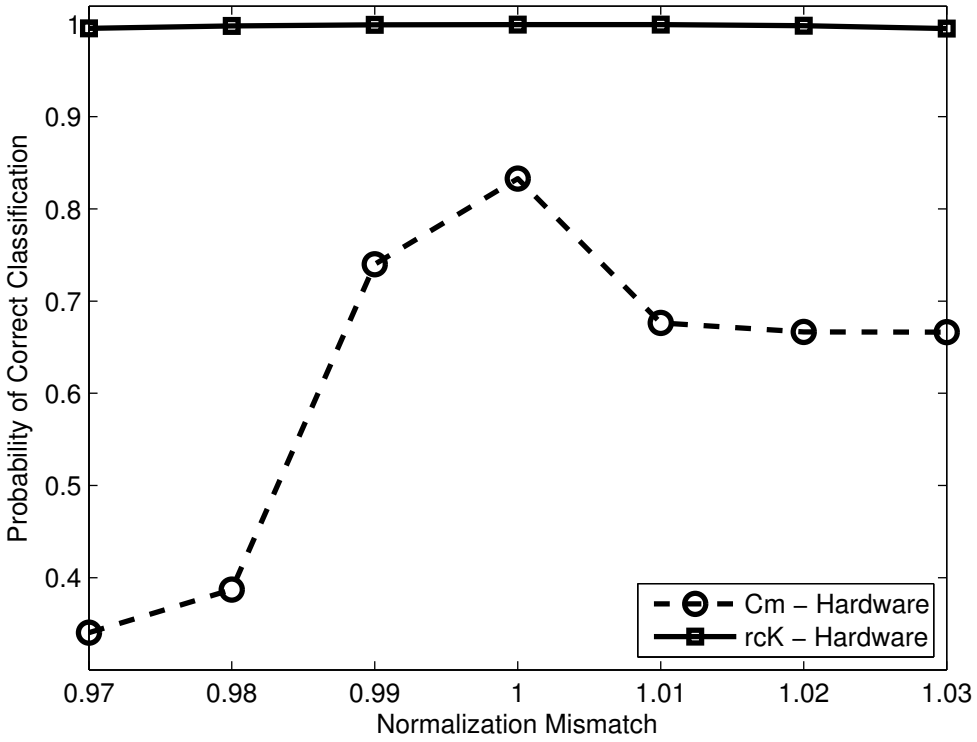


Figure 3.14: Probability of correct classification versus normalization mismatch at a fixed SNR = 14 dB, and N = 4096 symbols.

3.4.5 Conclusion

In this last section, we implemented two modulation level classifiers in a practical real-time hardware platform with radio front-ends and a DSP engine (BEE2), and evaluated their classification performance under realistic impairments such as quantization errors and timing synchronization errors. Both classifiers show similar trends with regard to these impairments, and their hardware computational complexity were proven to be similar. One interesting result presented in this section shows that the cumulant-based technique is much more sensitive to energy normalization errors than the rcK classifier. It was also shown that rcK performs better than Cm for a fixed number of symbols, as predicted by theory and simulations.

3.5 Summary

In this chapter we presented the concept of distribution distance based classification. In particular, this approach was applied to the modulation level classification probem. We also derived the optimal discriminant functions for classifying modulation schemes using the sampled distribution distance. This method was shown to provide substantial gains compared to other existing approaches. The performance of this method is also shown to be close to the ML classifier but at significantly lower computational complexity. Although modulation classification is presented in this work to illustrate the basic concept, the approach is not limited to this application. The same classifier can be generalized to any classification problem where the CDF of each class is available. Finally, we presented a hardware implementation to verify the practicality of the algorithm.

CHAPTER 4

MAC-layer Classification

Various works that dealt with the radio-scene analysis problem have mainly focused on detecting the presence or absence of spectrum opportunities, i.e., spectrum holes, through various spectrum sensing methods. The output of the majority of these algorithms is binary; either the spectrum is busy or idle. However, the concept of exploiting spectrum opportunity beyond the three dimensions of space, time, and freqeuncy, also known as multi-dimensional spectrum awareness, has increasingly gained research interest in recent years [YA09].

4.1 Introduction

The benefits of a multi-dimensional spectrum awareness approach have been shown in several prior works. For example, time-domain waveform information has been shown to improve spectrum sensing performance [Tan05]. Knowledge of second-order statistics of primary user (PU) traffic parameters significantly reduces the probability of mis-detecting the presence of the primary user [KYY12]. Information about higher layer features used by a primary network has also been shown to be beneficial to CR. It was shown by [SFT11] that if the primary network is known to be using a legacy 802.11 MAC protocol then the CR transmission parameters can be chosen in such a way so as to maximize throughput. However, utilizing information related to the specific standard used by a primary network has limited applicability. In general, there is very minimal prior information available about the primary users in a secondary access scenario a problem commonly

referred to as blind signal classification. To address this issue, more general characteristics common to a large set of standards can instead be used. One such feature is the channel access method utilized by the primary network.

4.1.1 Related Work

Channel access method (CAM) is defined as the protocol used by the wireless network in order to arbitrate and share the use of the common physical medium [Min09]. These methods, which are also referred to as multiple access or multiplexing schemes, give the ability to support more than one user over a single radio channel. At the highest level, there are two main categories of channel access methods: conflict-free and contention protocols. In conflict-free protocols, a user is given complete control of a particular channel and other users are prohibited from sharing that channel. Techniques including Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) fall into this category. On the other hand, in contention protocols there is some probability that one or more users will transmit in the same channel at the same time instant and thus cause a collision. Methods such as Carrier Sense Multiple Access (CSMA) fall into this category.

Various works in the CR literature have pointed out that knowledge about the channel access method employed by the primary network can be utilized to improve CR performance. For example, if it is known that the primary network employs a time division multiple access (TDMA) system then [WB12] proposes a wideband time-frequency analysis method to infer the center frequency, bandwidth, and occupancy of active channels with higher accuracy. The same information can also be used by the CR to synchronize both its sensing and access to the time slot thus reducing the overall need to perform sensing. In fact, several proposed CR strategies are based on the assumption that PUs operate under a TDMA or slotted access scheme [PSM05, ZTS07, SZ09, CXH08].

There have also been several proposals of exploiting unused subcarriers whenever the channel access method used is orthogonal frequency division multiple access (OFDMA) [TCP09,BAP09]. For code division multiple access (CDMA) networks, there is limited work on performing underlay spectrum access using spread spectrum. Finally, there are several prior works tackling how to efficiently utilize the white space in contention based methods [MB08, LKT14].

ToAlthough such methods exploiting channel access method knowledge are available in the literature, there has been very little work on the problem of actually identifying the channel access method used in CR context. The existing works fall into two general categories. First, work in [DBM10] and [HPM07] aim to determine the specific standard used by the primary network. These approaches require very detailed knowledge of PHY/MAC characteristics such as packet structure and preamble format. The second category of approaches, such as the method proposed in [HYY14], employ a more general channel access method (CAM) classification which do not try to determine the exact standard in use. In their work a Support Vector Machine (SVM) approach is employed in order to classify between TDMA, carrier sense multiple access with collision avoidance (CSMA/CA), slotted ALOHA, and pure ALOHA networks. However, we show in our work that this approach has very poor performance in the presence of fading. Further, CDMA and OFDMA systems which are very commonly used channel access method, are not addressed.

Another key limitation of these prior works is the assumption that the frequency band occupied by the primary user is known exactly. However, this is not true in cognitive radio network. In particular, wideband or multi-band spectrum sensing has been shown to significantly increase the amount of opportunities available to the CR network [QCS09, ALL12]. Therefore, this assumption makes these works impractical in the CR scenario. In fact, it makes band segmentation necessary.

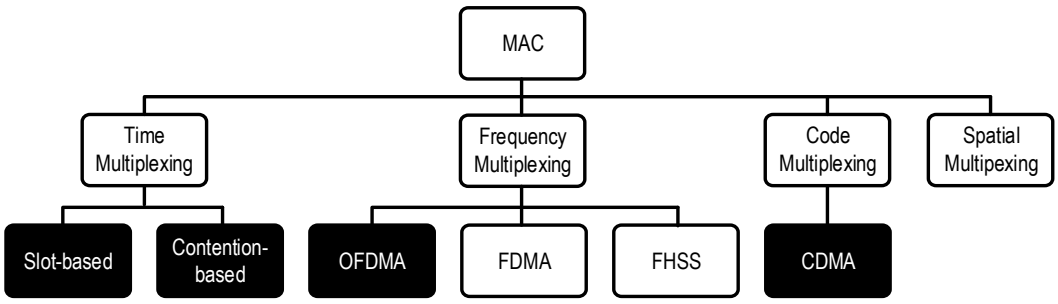


Figure 4.1: Taxonomy of different channel acceess methods. Techniques in black are addressed in this work.

4.1.2 Contribution

In this work, we approach the channel access method classification in 3 stages. First, the frequency bands occupied by individual PUs are identified (Band Segmentation Stage). Second, we identify the channel access method and modulation utilized by the primary network (Channel Access Method and Modulation Type Classification Stage). Finally, we distinguish between collision-free and contention based protocols (Collision Detection Stage). The range of channel access schemes identified in our scheme is shown in Fig. 4.1.

Analogous to the 3 stages of classification, the key contributions of this work is three-fold. First, we propose and evaluate a novel method of band segmentation designed for single antenna receivers based on the Non-negative Matrix Factorization (NNMF) approach. Second, we present and analyze the performance of a fourth-order cumulant based classifier for identifying both channelization schemes and modulation type. Finally, we developed a new collision detection method based on normalized fourth-order cumulants which substantially improves on the performance of the technique in [HYY14].

The rest of this chapter is organized as follows. We first present the system model and the channel access method classification problem in Section 4.2. Next,

we present and evaluate the band segmentation scheme in Section 4.3. Since our approach to channel access scheme classification is largely based on the cumulants, we present some theoretical background on them in Section 4.4.1. The proposed methods for both collision detection, modulation classification and channelization identification is detailed in Section 4.4. Numerical results and comparisons to existing approaches are presented in Section 4.4.3. Finally, the chapter is concluded in Section 2.5.

4.2 System Model

In this section, we describe the problem of channel access method classification and band segmentation in detail. We will detail the system model used and state the channel access method classification problem in the context of this system model.

4.2.1 Network Model

Consider an environment where at least one PU network is present. The goal of the proposed system is to identify the PU networks observed and the channel access method used by these networks. We assume that there is only one spectrum sensing node that is able to observe the transmitted signals from the primary network. However, there is no cooperation between the primary networks and the sensing node. Thus the classification needs to be done blindly using only passive information. We assume that there are a total of N_{total} primary users which are part of a primary network. These nodes are distributed in a uniform random manner over a given area such that $\mathbf{L}_i = (L_x^{(i)}, L_y^{(i)}), i \in \{1, 2, \ldots, N_{\text{total}}\}$ is the location of the ith node with $L_x^{(i)}$ and $L_y^{(i)}$ being its x-coordinate and y-coordinate respectively. Assume further that a spectrum sensing node located at $\mathbf{L}_0 = (L_x^{(0)}, L_y^{(0)})$ is able to observe the transmitted signals from the primary

network. The signal received by the spectrum sensing node from the ith node experiences path-loss which is distance dependent and represented as

$$\alpha_i = \left(\frac{d_0}{\|\mathbf{L}_i - \mathbf{L}_0\|} \right)^{-\gamma}. \tag{4.1}$$

where γ is the path-loss exponent, d_0 is the reference distance which we assume to be 1m, and $\|\cdot\|$ is the Euclidean distance.

4.2.2 PU signal model

Each PU node in the network transmits a signal of the following form:

$$x_i(n) = a_i(n)s_i(n)e^{-j2\pi\omega_{c_i}n}. \tag{4.2}$$

where $a_i(n) \in \{0, 1\}$ represents the activity (idle or busy) of the ith node. The activity of each node is determined by two key parameters: 1) the Channel Access Method in use by the PU network and 2) the traffic load or total traffic offered to the channel by the PU network. We will define these two parameters shortly and describe how they affect $a_i(n)$. The actual baseband signal for the ith node is $s_i(n)$ which depends primarily on the modulation format used as well as the channel access method.We assume that the $s_i(n)$ has a bandwidth of $\Delta\omega_{BW_i}$ and center frequency ω_{c_i}.

4.2.3 Channel Access Methods Being Considered

Consider a single primary network with N_{total} users. The transmited signal $x_i(n)$ for each PU depends on the channel access method employed by the primary network. This is the key parameter we are trying to classify in this work. As we have explained in Section 4.1, we are interested in distinguishing between Contention-based, TDMA, OFDMA, and CDMA. At the highest level, there are two general categories of channel access methods, packet-based and channelization-based protocols. Packet-based methods use a single channel which is shared by multiple

PU nodes through a contention protocol. We will refer to these techniques simply as contention based approach. The second category of channel access methods are, channelization or circuit based approaches. These protocols assign orthogonal channels to each node in which they have guaranteed transmission privileges. The orthogonality can be achieved through one of several dimensions such as time (TDMA), frequency (FDMA), code (CDMA) or space (SDMA). We focus on the first three of these. The signal model for each class of interest in this work is as follows:

4.2.3.1 TDMA

In this approach a particular timeslot is assigned to each user. During this time a given node is guaranteed to not interfere with any other node. In a TDMA system, there will be at most one active user at a time or $\sum_{i=1}^{N_{\text{total}}} a_i(n) \leq 1$, $\forall n \in \{1, \ldots, N\}$.

4.2.3.2 Contention-based

In contrast, contention-based channel access methods such as CSMA do not guarantee zero interference for each transmission. Each node can randomly start a transmission resulting in some non-zero probability that a collision occurs. The key difference between contention-based approaches and TDMA is that in the former more than one user can be active at any given time. The amount of temporal overlap in which $\sum_{i=1}^{N} a_i(n) > 1$, for any given $n \geq 0$, is a function of the network traffic and the particular back-off strategy chosen.

4.2.3.3 OFDMA

OFDMA allows for simultaneous transmissions to a single transmitter by assigning a unique set of subcarriers to each individual node. The received signal model

from one particular PU in a single OFDM symbol is given by

$$s_i(n) = \frac{1}{\sqrt{N_{sc}}} \sum_{k \in S_i} x_k e^{\frac{j2\pi kn}{N_{sc}}}, \quad 0 \leq n \leq N_{sc} - 1, \tag{4.3}$$

where x_k is the modulated data to be transmitted on the kth subcarrier, N_{sc} is the total number of subcarriers and S_i is the set of subcarriers assigned to the ith user.

4.2.3.4 CDMA

Finally, code division multiple access (CDMA) allows all users to access the same bandwidth at the same time but they are distinguished from each other by a unique code. Each assigned code is used to transform the particular users data stream into a spread-spectrum signal, using direct sequence spread spectrum. Another method to achieve spread-spectrum is by using Frequency Hopping (FHSS). However, this latter approach is beyond the scope of our work due since it requires an entirely different band segmentation approach than the one presented in this work. The signal model for a single transmitter using spread spectrum or CDMA system is as follows:

$$s_i(n) = c_i \left(n \mod L_c\right) x_i(n) \tag{4.4}$$

where $x_i(n)$ is the data stream of the ith user and $c_i(n)$ is the code of that particular user and L_c is the code length.

4.2.4 Traffic Model

The second key parameter that affects the behavior of the channel access method is the amount of traffic offered to the channel. In particular, the rate of collisions in a contention based scheme will be a function of the number of messages transmitted by each node. This is referred to as offered load (G) and is measured in Erlangs. For simplicity, we assume the following conditions: 1) all packets have the same length, 2) a packet is generated from each node according to a Poisson process with

rate λ_i, 3) packets that collide are not retransmitted. As a result, the aggregate messages to the channel will also be a Poisson arrival process and is simply the sum of all the rates λ_i, $G = \sum_{i=1}^{N} \lambda_i$.

4.2.5 Sensing Node

Finally, we combine the entire system by expressing the signal received by a randomly placed spectrum sensing node. The sensing node downconverts spectrum of bandwidth $\Delta\omega_{BW}$ to baseband. The downconverted signal is composed of the transmissions from all the PUs. We assume that the channel between the sensing node is the combined effect of random Rayleigh flat fading and path loss. Thus the baseband signal at the spectrum sensing node can be written as follows:

$$r\left(n\right) = \sum_{i=1}^{N_{\text{total}}} h_i\left(n\right) x_i(n) + \nu\left(n\right) \tag{4.5}$$

where $x_i(n)$ represents the primary transmitted signal given in (4.2) at time n and $\nu(n)$ is i.i.d. additive complex white Gaussian noise with zero mean and variance σ_ν^2, i.e., $\nu(n) \sim \mathcal{CN}(0, \sigma_\nu^2)$. The channel coefficient $h_i(n) \sim \mathcal{CN}(0, \sigma_i^2)$ where σ_i^2 is determined by nodes transmit power and the pathloss experienced from the transmitter to the sensing node given in (4.1).

4.3 Band Segmentation

An essential part of any wideband signal processing algorithm is to accurately idenfity the frequency bands occupied by each active transmitter. This process is referred to as Band Segmentation and is defined in [Jon85] as the process of decomposing the observed wideband spectrum into individual receiving information channels. In order to achieve this, we need to estimate the bandwidth and center frequency of each of the N_{total} active transmitters. A hypothetical scenario in which band segmentation is particularly challenging is illus-

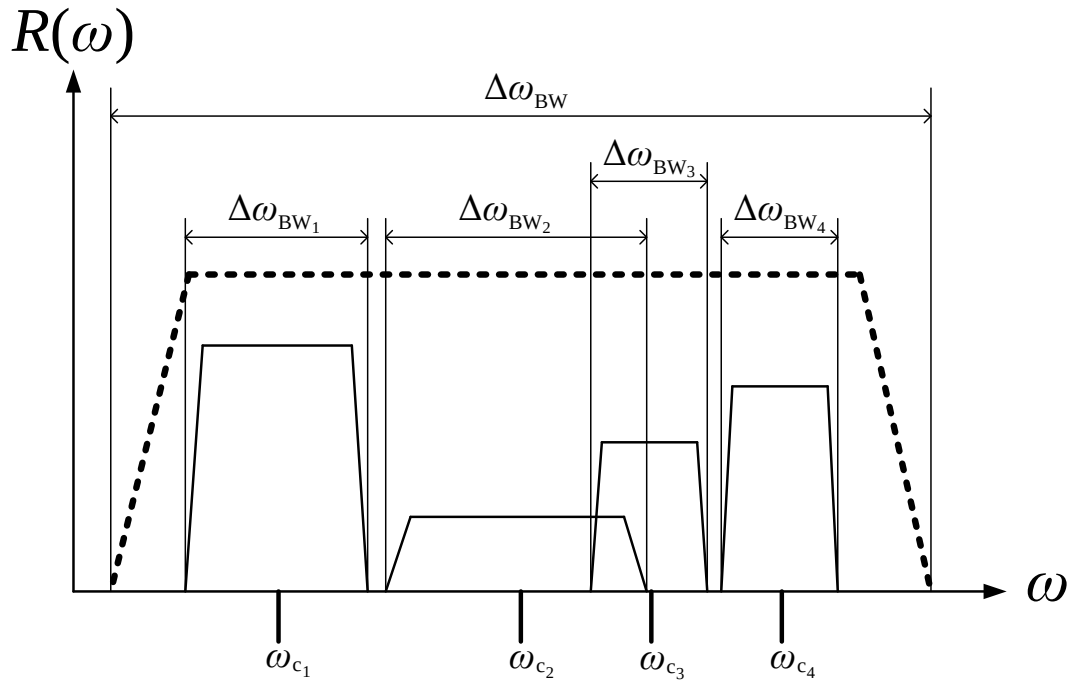


Figure 4.2: Example scenario for band segmentation.The dotted line indicates the entire RF front-end bandwidth while the solid lines represent the individual bands occupied by the four different PU.

trated in Fig. 4.2 in which the 2nd and 3rd signal have partial spectral overlap. The key objective of Band Segmentation is to give accurate estimates $\Delta\hat{\omega}_{BW_i}$, $\hat{\omega}_{c_i}, \forall i \in \{1, 2, \ldots, N_{total}\}$.

The Band Segmentation problem can be addressed through the use of multiple antennas as proposed in [ES99]. In their work they assumed that signals from various transmitters can be distinguished on the basis of their angle-of-arrival (AoA) which is found through the popular MUSIC algorithm [Sch86]. However, such an approach requires multiple antennas and is therefore very costly to implement especially in the wideband case. Similarly, other methods based on eigenvalue decomposition that utilize the raw downconverted wideband signal in order to separate the individual transmitters result in very high computational complexity.

One promising approach which requires only a single sensor and which offers significantly low computational complexity is to use Frequency-Domain Power Detection (FPD) [Yu13, Chapter 6]. In this method a Fast Fourier Transform is used to channelize the entire spectrum and perform narrowband signal detection on each FFT bin [QCS09, YSR11, QCP08, QCS08]. The result of this FPD procedure is then used to give a rough estimate of center frequency and bandwidth. We first describe this approach which forms the foundation of the methods we present later in this section. By using information about the traffic behavior of each PU we are able to extend FPD band segmentation to correctly identify bands that have partial spectral overlap.

4.3.1 Band Segmentation Using Frequency-Domain Power Detection

In FPD the problem of detecting the activity for each subchannel k can be modeled as a binary hypothesis test, where hypothesis $H_0(k)$ stands for noise only, and hypothesis $H_1(k)$ indicates both noise and signal are present. FPD is based on estimating the power spectral density (PSD) and applying a threshold which yields the binary estimate of activity. We will represent the estimated activity of the kth bin as $\hat{A}(k)$. The FPD test statistic is given in [YSR11] as

$$T(k) \triangleq \sum_{m=0}^{N_f-1} |\hat{X}_m(k)|^2 \underset{\hat{A}(k)=0}{\overset{\hat{A}(k)=1}{\gtrless}} \tau(k) \tag{4.6}$$

where $T(k)$ is the energy in the kth FFT bin and $\tau(k)$ is its corresponding detection threshold. The wideband signal is divided into N_f frames and the output of the FFT for the mth frame is represented as $\hat{X}_m(k)$. $T(k)$ is approximately normally distributed according to the central limit theorem based on the assumption that M is large enough. Under this condition, the required number of averaging frames and the detection threshold to achieve a given target probability of false alarm, P_{FA}, and probability of detection, P_D, for a given SNR on the kth-bin with

measured noise power of $\sigma_\nu^2(k)$ can be derived from expressions in [QCS09] as

$$N_f = 2\left(\frac{Q^{-1}(P_{FA}) - Q^{-1}(P_D)\sqrt{1+2SNR}}{SNR}\right)^2, \tag{4.7}$$

and

$$\tau(k) = \left(Q^{-1}(P_{FA})\sqrt{2N_f} + N_f\right)\sigma_\nu^2(k). \tag{4.8}$$

Based on the binary detection results of (4.6) and by assuming that adjacent active bins are associated to a single PU, $\Delta\hat{\omega}_{BW_i}$ can be estimated as the number of consecutive active bins and $\hat{\omega}_{c_i}$ as the centroid of these consecutive active bins. However, such an approach is not robust when the bandwidth is large. If even one of the bins occupied by one PU is mis-detected, then the signal-of-interest is regarded as multiple signals resulting in failed band segmentation. To address this issue, a threshold can be used for the minimum number of contiguous empty bins before a band is considered empty. This exploits the high degree of correlation between adjacent bins.

A significant limitation of this approach is the inability to accurately distinguish between multiple signals when they are in adjacent frequency bands. Further, any spectral overlap, will cause this approach to fail. Finally, even if signals are not spectrally overlapped, the imperfect spectral mask due to roll-off and spectral leakage between FFT bins causes such a method to treat 2 or more adjacent signals as a single larger bandwidth signal.

4.3.2 Traffic-aware FPD-based Band Segmentation

In order to address the limitations of the FPD-based band segmentation method presented in the previous subsection, we propose to utilize the temporal variation in PU activity. In the FPD approach, a particular band is assumed to either be active, $A(k) = 1$, or inactive, $A(k) = 0$, for the entire duration of the sensing period. However, assuming that PUs are not active all the time, and by performing spectrum sensing L times, $(L > 1)$, over a given time interval, the on-off behavior

of each PU can be observed. As a result, in addition to being a function of the FFT bin k, the detection results are now also dependent on when sensing is performed. We represent the temporal dependence of both the true FFT bin activity and the computed test statistic explicitty by $A(k,l)$ and $T(k,l)$ respectively, where the second argument represents the fact that this is lth sensing period ($k,l \in \mathbb{Z}, k \leq K, l \leq L$). Effectively, we want to utilize the traffic behavior of individual PUs in order to distinguish between them even in scenarios where there is spectral overlap due to reasons described in the previous subsection. Due to the fact that it incorporates the actual changing activity of inidividual PUs, we refer to this approach as traffic-aware FPD-based band segmentation (TA-FPD).

We take advantage of the PU traffic in TA-FPD by performing multiple measurements of the FPD test statistic given in (4.6). Recall that this test statistic is an estimate of the average power in each FFT-bin. By assuming that signals from each PU are independent from each other, we can approximate the average power in each FFT-bin as the sum of the average power of all active PUs for that particular bin. More explicitly, the new test statistic can represented as

$$T(k,l) = \sum_{i=1}^{N_{\text{total}}} a_i(l) T_i(k) \tag{4.9}$$

where we define $a_i(l) \in \{0,1\}$ as the activity of the ith PU during the lth sensing period and $T_i(k)$ is the FPD test statistic if only the ith user was active. Thus $T_i(k)$ can be thought of as the PSD of the ith user when estimated using a K-point FFT. Thus if we can estimate $T_i(k) \forall i \in \{1,2,\ldots,N_{\text{total}}\}$ from $T(k,l)$, then we could apply an appropriate threshold to each $\hat{T}_i(k)$ similar to (4.6) to perform band segmentation. Since $\hat{T}_i(k)$ is now the PSD for a single signal, the band segmentation becomes trivial. This solves the problemn in FPD band segmentation we described in the previous subsection in which we are unable to distinguish adjacent bands or bands with partial spectral overlap. In effect what we want to do is to decompose $T(k,l)$ into each PUs contribution and perform detection

on each PU individually. We present one approach to achieve this in the follow subsection.

4.3.3 Proposed Band Segmentation Approach

We propose to estimate $\hat{T}_i(k)$ from $T(k,l)$ by formulating the decomposition of (4.9) into the corresponding $a_i(l)$ and $T_i(k)$ for each PU into a matrix factorization problem. Defining the activity matrix to be

$$\mathbf{A} \triangleq \begin{bmatrix} a_1(1) & \cdots & a_{N_{\text{total}}}(1) \\ \vdots & \ddots & \vdots \\ a_1(L) & \cdots & a_{N_{\text{total}}}(L) \end{bmatrix}. \tag{4.10}$$

Thus, $\mathbf{A}$ is a matrix whose ith column contains the temporal activity of the ith user. Let us also define the PSD matrix as

$$\mathbf{T} \triangleq \begin{bmatrix} T_1(1) & \cdots & T_{N_{\text{total}}}(1) \\ \vdots & \ddots & \vdots \\ T_1(K) & \cdots & T_{N_{\text{total}}}(K) \end{bmatrix}. \tag{4.11}$$

For this matrix the ith column contains the PSD of only the ith user defined earlier to be $T_i(k)$. Finally let us define a new matrix $\mathbf{R}$ whose elements are given by $\{\mathbf{R}\}_{kl} \triangleq T(k,l)$ which is formed from the L FPD sensing periods.

Using these newly defined matrices we can rewrite (4.9) as $\mathbf{R} = \mathbf{A}\mathbf{T}^T$. The problem of estimating $T_i(k)$ for all PUs is equivalent to estimating the matrix $\mathbf{T}$ which can be solved as a matrix factorization. However, a constraint in our problem is that all elements of both $\mathbf{A}$ and $\mathbf{T}$ need to be non-negative. Since $\mathbf{T}$ represents the PSD and $\mathbf{A}$ the activity of each individual PU, then it can easily be seen that their values must be non-negative.

Using (4.9) and the non-negativity constraint we can solve the band segmentation as an Approximate Non-Negative Matrix Factorization (NNMF) [LS00]. In an approximate NNMF solution, a matrix $\mathbf{R}$ can be decomposed as $\mathbf{R} = \hat{\mathbf{A}}\hat{\mathbf{T}}^T + \hat{\mathbf{V}}$,

where $\hat{\mathbf{A}}$ and $\hat{\mathbf{T}}$ are the appropriately sized matrices (i.e. $L \times N_{\text{total}}$ and $K \times N_{\text{total}}$ respectively) and $\hat{\mathbf{V}}$ is the residual. In order to solve this matrix factorization problem, the following cost function must be minimized:

$$F(\hat{\mathbf{A}}, \hat{\mathbf{T}}) = \|\mathbf{R} - \hat{\mathbf{A}}\hat{\mathbf{T}}^T\|_F^2 \tag{4.12}$$

where $\|\cdot\|_F$ is the Frobenius norm. Various other costs functions and more importantly, numerical algorithms for solving this optimization problem has been presented in [LS00]. The band segmentation results can then be found by applying a threshold on the elements of $\hat{\mathbf{T}}$.

In order to improve the results of the NNMF we apply a pre-processing stage based on energy detection in order to eliminate the unnecessary contribution from the receiver noise. In Fig. 4.3 we show an example of the result of applying NNMF to the time-frequency map $\mathbf{R}$. The time-frequency map is shown in the form of a spectrogram on the left and the resulting matrix $\hat{\mathbf{S}}$ after threshold is shown on the right as the NNMF Result. The PU occupy bands as shown in the diagram in Fig. 4.2. From this result we can then find an estimate of $\Delta\omega_{BW_k}$ with the number of consecutive active bins and ω_{c_k} as the centroid of these consecutive active bins. This process is similar to the naive band-segmentation approach described earlier. However, one key advantage is the ability to correctly identify adjacent or overlapped signals such as the one given shown in Fig. 4.2 and Fig. 4.3.

4.3.4 Results and Comparison

4.3.4.1 Simulation Setup

In this subsection we evaluate the performance of the TA-FPD band segmentation method under various scenarios. The configurable parameters of the proposed approach are as follows:

- N_f = Number of averaged FFT frames for FPD. This number determines

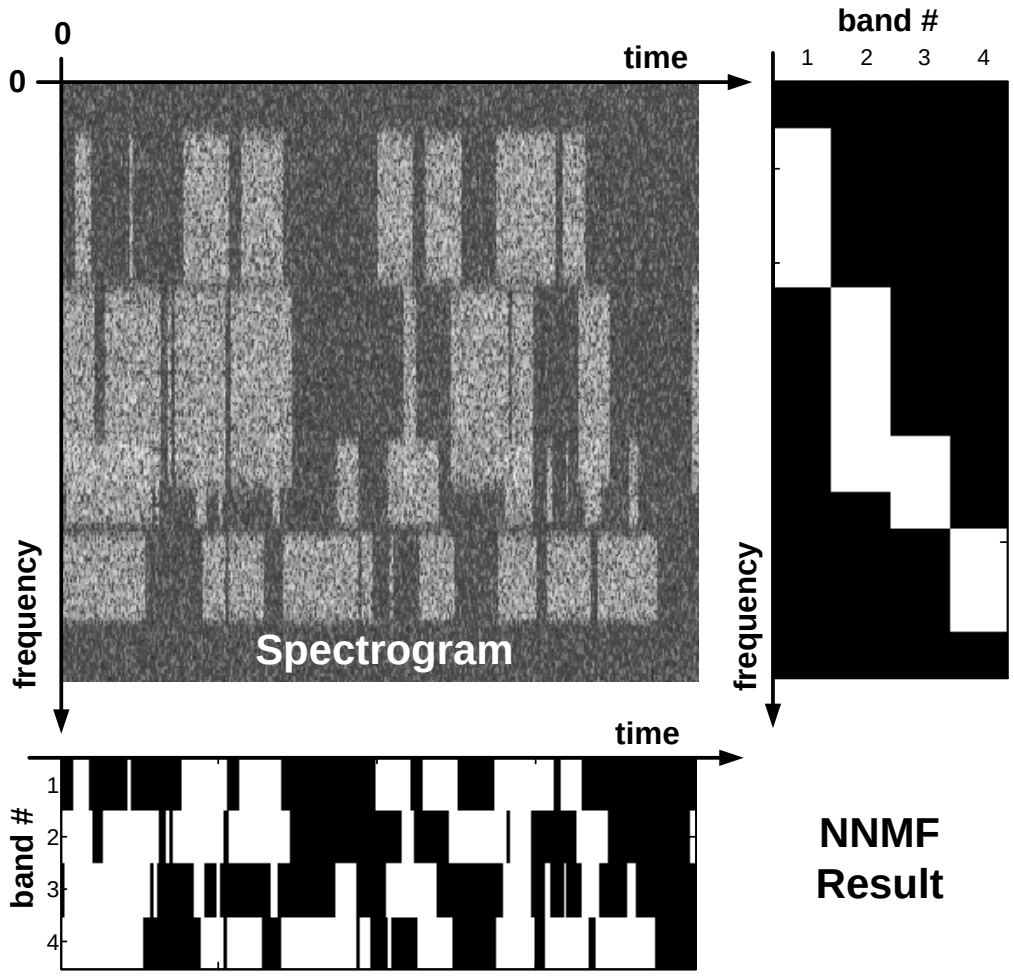


Figure 4.3: Illustration of Non-negative Matrix Factorization for band segmentation. The upper right image shows the spectrogram of 4 PUs, while the figure on the right shows the segmented bands and the bottom figure shows the estimated activity.

the achievable P_D and P_{FA} depending on the SNR according to (4.7).

- $L =$ Number of FPD sensing periods. This parameter determines, how much of the temporal variation in PU activity is observed.
- $K =$ Number of FFT points. This parameter determines the best resolution that can be achieved by both BW and center frequency estimation.

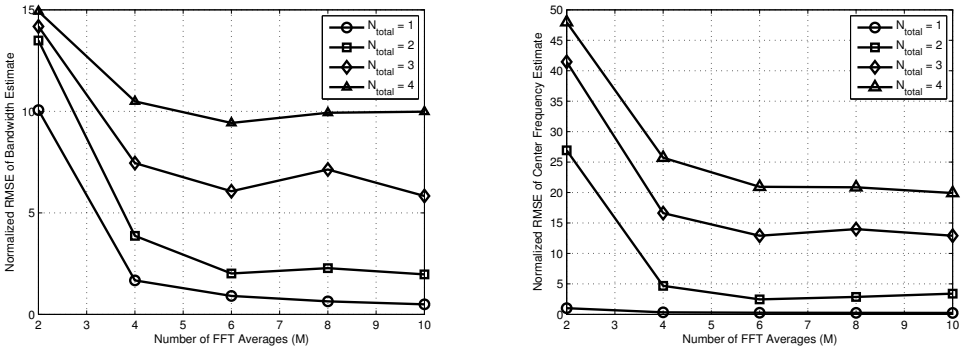
For our simulations we assume N_{total} PUs occupying a bandwdith of $\Delta\omega_{BW}$. We express all bandwidth and center frequency values in terms of normalized frequency (i.e. its ratio to the entire baseband signal bandwidth $\Delta\omega_{BW}$). For simplicity, all transmitters utilize a 4-QAM single-carrier modulation format. Note however, that since FPD is based only on the average power of each FFT bin, then the particular modulation class has negligible effect to the performance of our method. Further, since K determines the resolution of all frequency domain signal parameters (center frequency, bandwidth), we will fix $K = 256$ and normalize all average error measurement to the FFT resoltuion. Thus, the error of two particular estimates $\Delta\hat{\omega}_{BW_i}$ and $\hat{\omega}_{c_i}$ realative to their true values are given by

$$e(\Delta\hat{\omega}_{BW_i}) = \frac{K(\Delta\hat{\omega}_{BW_i} - \Delta\omega_{BW_i})}{\Delta\omega_{BW}}, \quad e(\Delta\hat{\omega}_{c_i}) = \frac{K(\Delta\hat{\omega}_{c_i} - \Delta\omega_{c_i})}{\Delta\omega_{BW}}. \tag{4.13}$$

Equivalently, we are measuring the average error as number of FFT bins. Finally, the activity of each PU is generated based on the model in Sec. 4.2.4.

4.3.4.2 Effect of Number of FFT averages (N_f)

In the first set of simulations, we study the effect of the number of FFT averaging frames on the accuracy of band segmentation. We maintain a fixed $P_{FA} = 0.005$ for the FPD. As such, the number of averages directly affects only the P_D. We vary N_{total} from 1 to 4, with all users transmitting with normalized bandwidth $\Delta\omega_{BW_i} = 1/8$ and whose center frequencies are uniformly distributed over the entire band, $\omega_{c_i} \sim U(-0.5, 0.5)$. The number of sensing periods $L = 20$ is also fixed and the traffic load is $G = 0.5$. The simulation results for this scenario is presented in Fig 4.4. As can be expected, the error in both bandwidth estimation and center frequency estimation decreases with increased FFT averages which in turn decreases the variance of the elements of $\mathbf{R}$. However, the accuracy quickly saturates after $N_f = 4$ which is equivalent to a $P_D = 99.5\%$. More importantly, we find that the saturated value of the error increases with increasing number of



(a) (b)

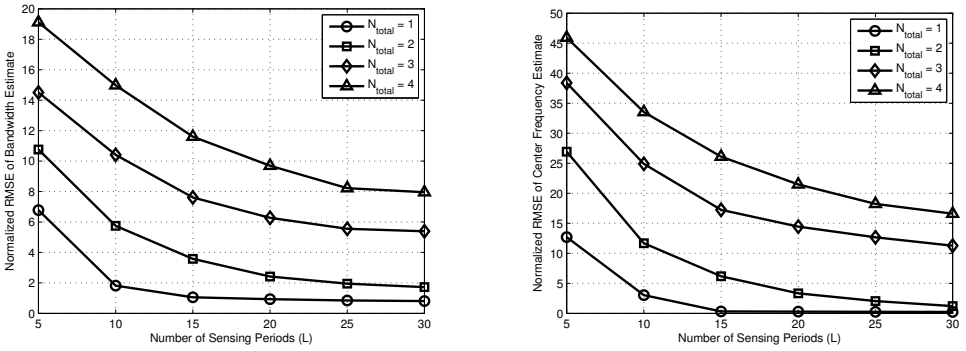
Figure 4.4: Effect of varying the number of FFT Averages, N_f, on TA-FPD band segmentation ($L = 20$, $K = 256$). (a) Bandwidth, $\Delta\omega_{BW_i}$, estimation error (b) Center frequncy, $\Delta\omega_{c_i}$ estimation error.

PU N_{total}.

In the case of a single user, the algorithm achieves the highest accuracy since there is no potential overlap between different bands. However, the center frequency accuracy is lower bounded by 0.5 bins which is minimum spectral resolution of the FFT. In the case of bandwdith estimation, both the FPD and TA-FPD methods overestimate the value because of the imperfect spectral mask and the pulse-shaping roll-off.

4.3.4.3 Effect of Varying Number of Sensing Periods (L)

As we have observed, the estimation accuracy can only be improved up to a certain level by increasing N_f. To further improve the accuracy of band segmentation we need to increase the number of sensing periods, L which allows the band segemtation algorithm to observe the random PU traffic further. In Fig. 4.5 we can confirm that the accuracy of TA-FPD can indeed be improved by increasing the number of sensing periods L which equates to increasing the number of columns in $\mathbf{R}$. This confirms our intuition that by increasing L the independence between



(a) (b)

Figure 4.5: Effect of varying the number of sensing periods, L, on TA-FPD band segmentation accuracy ($N_f = 6$, $K = 256$). (a) Bandwidth, $\Delta\omega_{BW_i}$, estimation error (b) Center frequncy, $\Delta\omega_{c_i}$ estimation error.

overlapping PU bands can be increased which is the basis of the NNMF and other matrix factorization methods. However, even with increased sensing period the error in band segmentation still increases with number of users.

4.3.4.4 Hybrid FPD/TA-FPD Algorithm

Since we have determined that more users contributing to $\mathbf{R}$, we can improve the accuracy of band segmentation significantly by only performing TA-FPD on signals which are actually overlapped. To achieve this, we propose a hybrid FPD and TA-FPD approach. We divide the band segmentation algorithm into two stages. Stage 1 involves performing the band segmentation based solely on the energy detection method presented in Sec. 4.3.1. The first stage segments all non-overlapping PU bands. Each of these band segments is expected to have one or more PUs. In Stage 2, we perform the NNMF-based (TA-FPD) band segmentation on each submatrix of $\mathbf{R}$ formed by the bins of the bands resulting from Stage 1. We can see from Fig. 4.6 that compared to TA-FPD, this hybrid approach achieves significantly better accuracy even with increasing number of

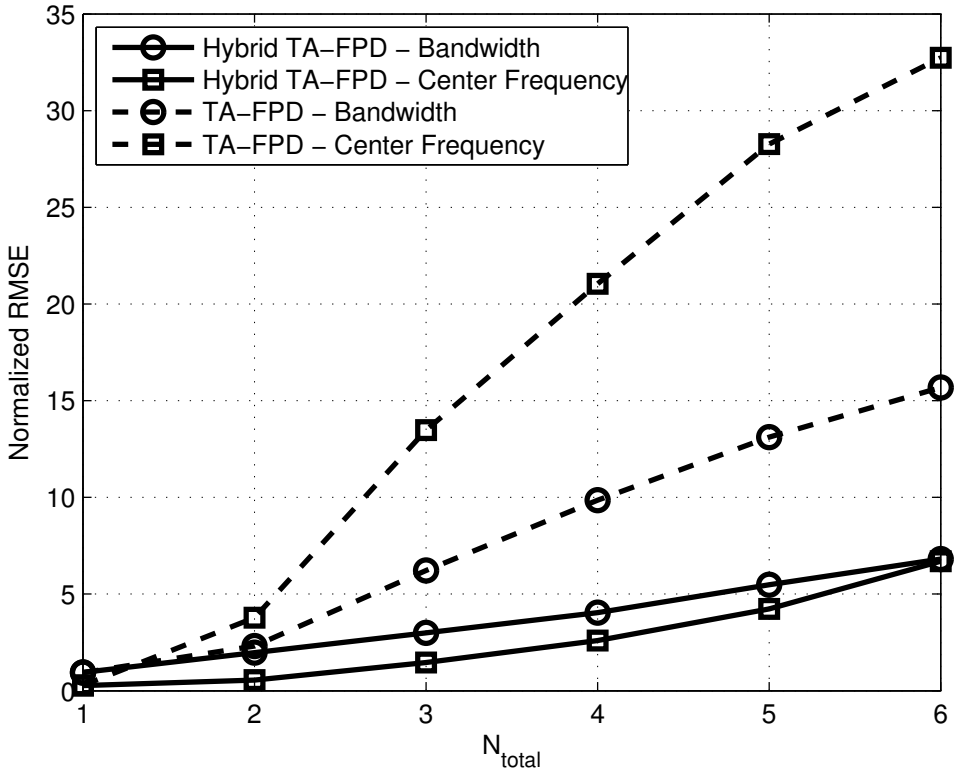


Figure 4.6: Accuracy comparison of TA-FPD and the Hybrid TA-FPD band segmenatation methods under varying number of users ($N_f = 6$, $L = 20$, $K = 256$)

.

users.

In summary, we have shown that NNMF can be a very effective means to perform band-segmentation. The key benefit of such an approach is that it requires only a single antenna and is able to segement bands that overlap. As a further output of the band segmentation procedure, we also obtain an estimate of traffic activity $\mathbf{A}$ which we later use in the channel access scheme identification discussed in the following section.

4.4 Identifying Channel Access Scheme

Once the individual bands occupied by the primary network have been identified, the next step is to distinguish between TDMA, OFDMA, CDMA, and contention-based channel access methods. TDMA, OFDMA, and CDMA have distinguishable features due to their signal structure. In particular, modulation used in TDMA, the effect of the inverse fourier transform in OFDMA, and the pseudonoise sequence used by users in CDMA. Contention-based methods have distinguishable features depending on the constellation used and the probability of collisions between users.

In this section, we start by describing normalized cumulants. Next, the sample normalized fourth-order cumulant is proposed to distinguish between TDMA, CDMA, and OFDMA. Its sample variance is proposed for identifying contention-based methods. We conclude this section with theoretical analysis of the proposed two-step method and simulation results for the performance of the method.

4.4.1 Normalized 4th-Order Cumulants and its Properties

Cumulants of multiple random variables are particular linear combinations of their joint higher order moments [Bri81]. For our purpose, we consider cumulants of the complex-valued zero mean signal $y(n)$ computed by considering multiple copies of the same signal as distinct random variables. We define kth-order cumulants C_{ki}, for $i \in \{0, \ldots, k\}$, of the signal $y(n)$ as cumulant of $k-i$ repeated copies of $y(n)$ and i repeated copies of $y^*(n)$.

In particular, the second-order cumulants are defined to be [SS00a],

$$C_{20} = E[y^2(n)] \quad \text{and} \quad C_{21} = E[|y(n)|^2]. \tag{4.14}$$

Note that there are two different 2nd-order cumulants. Fourth-order cumulants

are defined as [SS00a]

$$C_{40} = M_{40} - 3M_{20}^2, \text{ and} \tag{4.15}$$

$$C_{42} = M_{42} - |M_{20}|^2 - 2M_{21}^2. \tag{4.16}$$

Note that these expressions do not involve third-order moments of the signal because most linear modulations used in communication signals are zero-mean and have zero third-order moments. Similarly, the C_{41} cumulant is also zero for most linearly modulated communication signals. Sixth-order and eighth-order cumulants are defined in Appendix B.

To estimate the cumulants from J samples of a received signal, we use the unbiased maximum likelihood estimators [SS00a]

$$\hat{C}_{21} = \frac{1}{J}\sum_{n=1}^{J} |y(n)|^2, \tag{4.17}$$

$$\hat{C}_{20} = \frac{1}{J}\sum_{n=1}^{J} y^2(n), \tag{4.18}$$

$$\hat{C}_{40} = \frac{1}{J}\sum_{n=1}^{J} y^4(n) - 3\hat{C}_{20}^2, \text{ and} \tag{4.19}$$

$$\hat{C}_{42} = \frac{1}{J}\sum_{n=1}^{J} |y(n)|^4 - |\hat{C}_{20}|^2 - 2\hat{C}_{21}^2. \tag{4.20}$$

Note that these cumulant estimates depend on the signal's average power. Hence, similar to [SS00a], the cumulant estimates are normalized by the signal power $\hat{C}_{21} - \sigma_\nu^2$ where σ_ν^2 is the noise power. In particular, we focus on the fourth-order cumulant estimate

$$\tilde{C}_{42} = \frac{\hat{C}_{42}}{(\hat{C}_{21} - \sigma_\nu^2)^2} \tag{4.21}$$

There are four key properties of normalized fourth order cumulants that are crucial to channel access method classification:

1. *One-to-one correspondence with modulation type:* It is shown in [SS00a] that the normalized C_{42} for single user signals modulated with different

modulation types is distinct. This also holds true for TDMA systems since only one user transmits at any given time. We also show, in Appendices B.2 and B.3, that the normalized C_{42} for signals from OFDMA and CDMA systems are also distinct. The different values are illustrated in Fig. 4.7.

2. *Additivity:* The normalized C_{42} of the sum of independent signals, such as signals from colliding users, is the linear combination of the individual normalized C_{42}. Suppose in the fth frame, $N_{\text{colliding}}$ signals collide then the C_{42} estimate of the received signal is

$$\tilde{C}_{42} = \sum_{i=1}^{N_{\text{colliding}}} \frac{\left(\hat{C}_{21,i} - \sigma_\nu^2\right)^2}{\left(\hat{C}_{21} - \sigma_\nu^2\right)^2} \tilde{C}_{42,i} + \frac{\sigma_\nu^4}{\left(\hat{C}_{21} - \sigma_\nu^2\right)^2} \tilde{C}_{42,\nu} \tag{4.22}$$

where we define $\tilde{C}_{42,i}$ as the normalized cumulant estimate of the received signal if only the ith signal were active and all other PUs are not transmitting. $\tilde{C}_{42,\nu}$ represents the estimate of the C_{42} of the noise alone. The proof of (4.22) is given in Appendix B. The C_{42} of gaussian signals is zero [SS00a]. Hence, $E\left[\tilde{C}_{42,\nu}\right]$ is zero for additive white gaussian noise (AWGN).

3. *Effect of Collisions:* When $N_{colliding} > 1$, $\sum_{i=1}^{N_{\text{colliding}}} \left(C_{21,i} - \sigma_\nu^2\right)^2 < \left(C_{21} - \sigma_\nu^2\right)^2$ due to the concavity of the square function. Hence, $\tilde{C}_{42}$ reduces in magnitude with increasing number of colliding signals. In expectation, we get

$$\mathrm{E}\left[\tilde{C}_{42}\right] \xrightarrow{N_{\text{colliding}} \to \infty} 0. \tag{4.23}$$

Collisions increase the sample variance of the estimated $\tilde{C}_{42}$ because $\mathrm{E}[\tilde{C}_{42}]$ depends on the users whose signals collide. Different users collide at different times and hence, the C_{42} value changes accordingly as given by (4.37). This increases the variance of the estimator.

4. *Effect of Flat-Fading:* Since the $\tilde{C}_{42}$ is normalized to signal power, its mean value is independent of the attenuation due to flat-fading [SS00a]. It is also independent of the noise power because $E\left[\tilde{C}_{42,\nu}\right] = 0$.

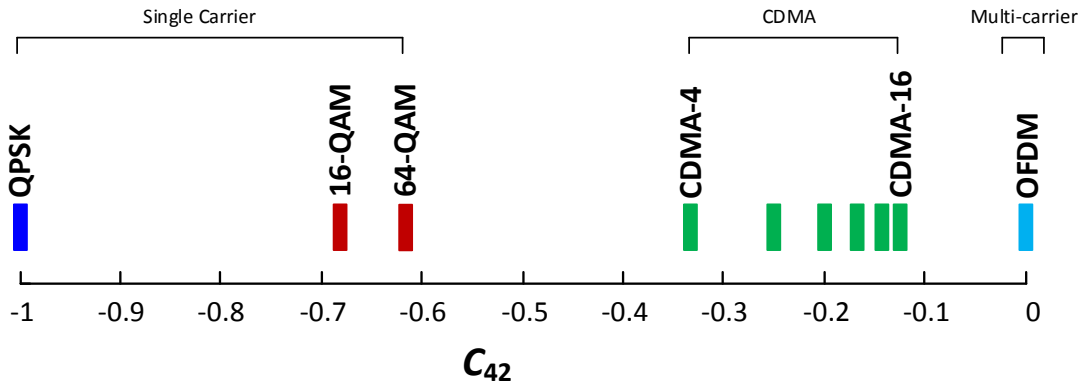


Figure 4.7: $E[C_{42}]$ values for various classes

4.4.2 Proposed Method

Received TDMA, OFDMA, and CDMA signals have distinct normalized fourth-order cumulant C_{42}. The value of the C_{42} for TDMA depends on the modulation type, that for OFDMA depends on the number of subcarriers, and that for CDMA depends on the number of users, the code in use, and the modulation type. The C_{42} values for OFDMA, CDMA, and some modulation types used in TDMA systems are illustrated in Fig. 4.7. However, when packets collide randomly, such as the case in contention-based schemes, then the C_{42} depends on the sets of users that have collided and the number of occurrences of each such collision. Since we cannot know this information in our system, we cannot identify properties of the received signal from the C_{42}. In other words, the C_{42} of the received signal can be used to label a signal as TDMA, OFDMA, or CDMA but the confidence in this inference will be low if collisions have occurred. We propose using the sample variance of the estimates of C_{42} to estimate the probability that collisions have occurred. If this probability is higher than a parameter $P_{\mathrm{C|T}}$ then we declare the channel access method to be a contention-based method. Otherwise, the channel access method inferred from the sample estimate of C_{42} is declared to be true.

Note that TDMA and slotted contention-based channel access methods, such

as slotted ALOHA, are not distinguishable using time-domain features such as channel idle and busy durations alone [HYY14]. Hence, we propose discarding noise-only samples, i.e., squelching the received signal, and using features dependent only on the structure of the transmitted signal and the occurrence of collisions.

The signal received in a band over a sufficiently long time consists of transmissions from multiple users. Since we need to estimate the sample variance of $\tilde{C}_{42}$, we start by dividing the squelched received signal $\{y(n)\}_{n\in\{1,\ldots,JF\}}$ into F frames of J samples each. Next, the C_{42} is estimated for each frame:

$$\tilde{C}_{42}(f) = \frac{1}{J}\sum_{n=1}^{J} |y(J(f-1)+n)|^4 - |\hat{C}_{20}(f)|^2 - 2\hat{C}_{21}^2(f) \quad \forall f \in \{1,\ldots,F\} \tag{4.24}$$

where $\hat{C}_{21}(f)$ and $\hat{C}_{20}(f)$ are similarly computed from the fth frame.

Let the set $\mathcal{U}(f)$ be the set users transmitting simultaneously in frame f be and their C_{42} estimate be $\tilde{C}_{42,\mathcal{U}(f)}(f)$. Then the estimated C_{42} can be represented as

$$\tilde{C}_{42}(f) = \sum_{U\subseteq\mathcal{U}} 1_{\{U=\mathcal{U}(f)\}}\tilde{C}_{42,U}(f). \tag{4.25}$$

Since each C_{42} estimate is gaussian in nature [SS00a], the $\tilde{C}_{42}(f)$ has a gaussian mixture distribution

$$\tilde{C}_{42}(f) \sim \sum_{U\subseteq\mathcal{U}} 1_{\{U=\mathcal{U}(f)\}}\mathcal{N}\left(C_{42,U}, \operatorname{var}\left[\tilde{C}_{42,U}\right]\right) \tag{4.26}$$

where $C_{42,U}$ is given by (4.22) and $\operatorname{var}[\tilde{C}_{42,U}]$ is computed from (B.9). Therefore, its mean and variance are

$$\mathrm{E}\left[\tilde{C}_{42}(f)\right] = \sum_{U\subseteq\mathcal{U}} P\left(U=\mathcal{U}(f)\right) C_{42,U} \text{ and} \tag{4.27}$$

$$\operatorname{var}\left[\tilde{C}_{42}(f)\right] = \sum_{U\subseteq\mathcal{U}} P\left(U=\mathcal{U}(f)\right)\left\{\left(C_{42,U} - E\left[\tilde{C}_{42}(f)\right]\right)^2 + \operatorname{var}\left[\tilde{C}_{42,U}\right] + \frac{\sigma_\nu^8}{\left(\sum_{u\in U} C_{21,u} - \sigma_\nu^2\right)^4}\operatorname{var}\left[\tilde{C}_{42,\nu}\right]\right\} \tag{4.28}$$

respectively.

Now, as we had described earlier, we will use the C_{42} estimate to distinguish between OFDMA, CDMA, and TDMA with different modulations. In TDMA systems, exactly one user transmits at a time, i.e., $|\mathcal{U}(f)| = 1$. We assume that all users in the system have the same modulation type. The C_{42} estimate then has the mean and variance as

$$\mathrm{E}\left[\tilde{C}_{42}(f)\right] = \sum_{u\in\mathcal{U}} P\left(u = \mathcal{U}(f)\right) C_{42,u} = C_{42,1}, \text{ and} \tag{4.29}$$

$$\mathrm{var}\left[\tilde{C}_{42}(f)\right] = \sum_{u\in\mathcal{U}} P\left(u = \mathcal{U}(f)\right) \left\{ \mathrm{var}\left[\tilde{C}_{42,u}\right] + \frac{\sigma_\nu^8}{\left(C_{21,u} - \sigma_\nu^2\right)^4} \mathrm{var}\left[\tilde{C}_{42,\nu}\right] \right\} \tag{4.30}$$

In OFDMA and CDMA systems, all users are active simultaneously at all times. Therefore, $\tilde{C}_{42}(f)$ is a gaussian random variable with mean and variance as derived in Appendices B.2 and B.3 respectively. Thus, $E\left[\tilde{C}_{42}(f)\right]$ can be used to distinguish between TDMA, OFDMA, and CDMA. We propose the test statistic

$$W = \frac{1}{F} \sum_{f=1}^{F} \tilde{C}_{42}(f) \tag{4.31}$$

and use the maximum likelihood estimator (4.32) to infer the tuple $\hat{M}$ of channel access method and the modulation type:

$$\hat{M} = \arg\max_{M'\in\mathcal{M}} P\left(C_{42} = W | M'\right). \tag{4.32}$$

Here, we search from the set $\mathcal{M}$ which consists of tuples of channel access method and modulation type and level. These classes are listed in Table 4.1.

Next, we estimate our confidence in this inference. The confidence is estimated from the variance of the $\tilde{C}_{42}$ value. From (4.28), it is easy to see that

$$\mathrm{var}\left[\tilde{C}_{42}(f)\right] \geq \sum_{U\subseteq\mathcal{U}} P\left(U = \mathcal{U}(f)\right) \left\{ \mathrm{var}\left[\tilde{C}_{42,U}\right] + \frac{\sigma_\nu^8}{\left(\sum_{u\in U} C_{21,u} - \sigma_\nu^2\right)^4} \mathrm{var}\left[\tilde{C}_{42,\nu}\right] \right\} \tag{4.33}$$

Table 4.1: List of channel access methods and modulation types & levels that we search from in (4.32).

Class label	Channel access method	Modulation type & levels
C1	TDMA	BPSK
C2		M-PSK ($M > 2$)
C3-C7		4/8/16/32/64-PAM
C8-C10		16/64/256-QAM
C11	OFDMA	QPSK
C12		16-QAM
C13	CDMA	BPSK

with equality only if $C_{42,\mathcal{U}(f)}$ is equal for all $f \in \{1, \ldots, F\}$. In other words, $\text{var}\left[\tilde{C}_{42}(f)\right]$ increases if collisions occur. Hence, we can estimate the confidence in the absence of collisions from the sample variance of $\tilde{C}_{42}(f)$. The confidence in the occurrence of collisions is hard to estimate because collisions may occur infrequently and the number of colliding users is not fixed.

Let $P_{C|T}$ be a parameter specifying the maximum permissible probability of inferring the occurrence of collisions when none have occurred. This is analogous to setting a desired false alarm probability in a two hypothesis test. We use $P_{C|T}$ to set a threshold on the sample estimate $\hat{\varsigma}^2$ of $\text{var}\left[\tilde{C}_{42}(f)\right]$. The sample variance is computed by using the modulation type and channel access method inferred earlier:

$$\hat{\varsigma}^2 = \frac{1}{F}\sum_{f=1}^{F}\left(\tilde{C}_{42}(f) - C_{42}\left(\hat{M}\right)\right)^2. \tag{4.34}$$

By Cochran's Theorem [Coc34], the sample variance is distributed as

$$\hat{\varsigma}^2 \sim \frac{\text{var}\left[\tilde{C}_{42}(f)\middle| M, \left\{\hat{C}_{21}(f)\right\}_f, \sigma_\nu^2\right]}{F}\chi_F^2 \tag{4.35}$$

where χ_F^2 is a χ^2 distribution with F degrees. We choose a threshold τ such that

$P\left(\varsigma^2 > \tau | \mathcal{H} \neq \mathcal{H}_{\text{contention}}\right) < P_{C|T}$:

$$\tau = \frac{\text{var}\left[\tilde{C}_{42}(f)\middle|\hat{M}, \{\hat{C}_{21}(f)\}_f, \sigma_\nu^2\right]}{F} \chi_F^{-2}\left(1 - P_{C|T}\right). \tag{4.36}$$

If $\hat{\varsigma}^2 < \tau$ then we declare the inference $\hat{M}$ from (4.32) to be correct. If not, then we infer that the channel access method is contention-based.

The algorithm is summarized as Algorithm 1.

Algorithm 1 MAC Classification Algorithm

Input: $y \in \mathbb{C}^{FJ}$, σ_ν^2

for $f = 1 \ldots F$ **do**

Compute $\tilde{C}_{42}(f)$ from (4.24)

end for

$W \leftarrow \frac{1}{F}\sum_{f=1}^{F} \tilde{C}_{42}(f)$

for all $M' \in \mathcal{M}$ **do**

Look up $\text{E}\left[\tilde{C}_{42}(f)\middle|M', \tilde{C}_{21}(f), \sigma_\nu^2\right] = \text{E}\left[\tilde{C}_{42}(f)\middle|M'\right]$ from Table B.1

Compute $\text{var}\left[\tilde{C}_{42}(f)\middle|M', \tilde{C}_{21}(f), \sigma_\nu^2\right]$ by (B.13), (B.14)

Compute $P(C_{42} = T_1|M')$

end for

$\hat{M} \leftarrow \arg\max_{M' \in \mathcal{M}} P(C_{42} = T_1|M')$

Compute $\hat{\varsigma}^2$ by (4.34)

Compute τ by (4.36)

Output: If $\hat{\varsigma}^2 \leq \tau$ then M' else Contention-based

4.4.3 Results and Comparisons

We have proposed a two step procedure to distinguish channel access methods between TDMA, OFDMA, and CDMA. First, we use the average $\tilde{C}_{42}(f)$ to classify the channel access method as either TDMA, OFDMA, and CDMA. At this stage, we also infer the modulation type in the case of TDMA. Next, we estimate the

probability that collisions have occurred by thresholding the sample variance of $\tilde{C}_{42}(f)$. The threshold depends on the inferred channel access method, modulation type, and the signal power received in each frame.

4.4.3.1 Identifying Channel Access Method for Non-Contention-based Methods

Our method can have errors at two stages: the maximum likelihood estimator to distinguish between TDMA, CDMA, OFDMA and identifying the occurrence of collisions.

The first stage is multi-hypothesis problem with the possible hypotheses listed in Table 4.1. The probability of correctly inferring M by the maximum likelihood estimator (4.32) can be computed from the theoretical distribution of $\tilde{C}_{42}(f)$ for each of these classes. To verify the accuracy of our derivations, we provide the theoretical and empirical probability of accuracy for these classes in Table 4.2 for particular values of SNR.

Table 4.2: Verification of theoretical analysis of distributions of $\tilde{C}_{42}(f)$ for different modulation types. System has 4 users, at average SNR of 10dB SNR, and 0.6 load.

	Theory	**Simulation**
BPSK	0.938	0.938
M-PSK	0.942	0.935
M-PAM	0.840	0.905
QAM	0.934	0.926

Lastly, we discuss the detection of modulation type and level for a TDMA system. From Table B.1 in Appendix B, we see that C_{42} values of different levels of the same modulation type are very close to each other. In the case of PAM, for example, the C_{42} of 32-PAM differs from that of 64-PAM by 0.0018. However, achieving such an accurate estimate from $\tilde{C}_{42}$ requires samples of the order of 10^3

Table 4.3: Confusion matrix for modulation type detection in TDMA systems. Parameters: 4 users, average SNR 10dB, 0.6 offered load.

	BPSK	**M-PSK**	**PAM**	**QAM**
BPSK	0.957	0.025	0.000	0.018
M-PSK	0.004	0.957	0.018	0.021
PAM	0.010	0.047	0.924	0.018
QAM	0.005	0.035	0.013	0.946

or higher. Hence, in our simulations, we observe that our method is unable to identify the modulation level correctly. However, our proposed algorithm has a high probability of detecting the correct modulation type of a TDMA system. An example confusion matrix is shown in Table 4.3.

4.4.3.2 Identifying Contention-Based Channel Access Methods

Our proposed method thresholds the sample variance of the $\tilde{C}_{42}(f)$ using a specified probability $P_{C|T}$ of identifying a non-contention channel access method as a contention-based channel access system. Simulation results in Section 4.4.3.3 show that this probability is achieved exactly for a variety of system parameters.

Next, we analyze the probability of correctly detecting a contention-based channel access method. The normalized sample cumulant can be written as

$$\tilde{C}_{42}(f) = \sum_{\mathcal{C}\subseteq\mathcal{U}} \left[I_{\{\mathcal{U}(f)=\mathcal{C}\}} \frac{\sum_{u\in\mathcal{C}} C_{21,u}^2 \tilde{C}_{42,u}(f) + \sigma_\nu^4 \tilde{C}_{42,\nu}(f)}{\left(\sum_{u\in\mathcal{C}} C_{21,u} - \sigma_\nu^2\right)^2} \right] \tag{4.37}$$

where $\tilde{C}_{42,u}(f) \triangleq \hat{C}_{42,u}(f)/C_{21,u}^2$. Explicitly factoring the normalization separates the signal property ($\tilde{C}_{42,u}$) from the effect of collisions ($C_{21} = \sum_{u\in\mathcal{C}} C_{21,u}$). Its

mean and variance are

$$E\left[\tilde{C}_{42}(f)\right] = \sum_{\mathcal{C}\subseteq\mathcal{U}} \left[P\left(\mathcal{U}(f)=\mathcal{C}\right) \frac{\sum_{u\in\mathcal{C}} C_{21,u}^2 C_{42,u}}{\left(\sum_{u\in\mathcal{C}} C_{21,u} - \sigma_\nu^2\right)^2} \right] \tag{4.38}$$

$$\text{var}\left[\tilde{C}_{42}(f)\right] = \sum_{\mathcal{C}\subseteq\mathcal{U}} \left[P\left(\mathcal{U}(f)=\mathcal{C}\right) \frac{\sum_{u\in\mathcal{C}} C_{21,u}^4 \,\text{var}\left[\tilde{C}_{42,u}(f)\right] + \sigma_\nu^8 \,\text{var}[\tilde{C}_{42,\nu}(f)]}{\left(\sum_{u\in\mathcal{C}} C_{21,u} - \sigma_\nu^2\right)^4} \right] \tag{4.39}$$

Now, note that we define T as the second-order moment of $\tilde{C}_{42}(f)$ around $C_{42}(\hat{M})$ and not around $E\left[\tilde{C}_{42}(f)\right]$. Since the variance is the minimizer of such a second-order moment, $E[T] \geq \text{var}[\tilde{C}_{42}(f)]$. Instead, if we consider the probability that the sample variance of $\tilde{C}_{42}(f)$ exceeds our threshold τ, then we get a lower bound on the probability of detecting contention-based channel access methods. Let T' be the sample variance of $\tilde{C}_{42}(f)$. Since $\tilde{C}_{42}(f)$ is a mixture model of distributions with finite fourth-order moments, under hypothesis $\mathcal{H}_C$, T' is distributed as

$$T' \sim \mathcal{N}\left(\text{var}\left[\tilde{C}_{42}(f)\right], \frac{1}{F}\,\text{var}\left[\left(\tilde{C}_{42}(f) - E\left[\tilde{C}_{42}(f)\right]\right)^2\right]\right). \tag{4.40}$$

The mean of T is given by (4.39) and its variance can be computed as

$$\begin{aligned}
&\text{var}\left[\left(\tilde{C}_{42}(f) - E\left[\tilde{C}_{42}(f)\right]\right)^2\right] \\
&= E\left[\left(\tilde{C}_{42}(f) - E\left[\tilde{C}_{42}(f)\right]\right)^4\right] - E^2\left[\left(\tilde{C}_{42}(f) - E\left[\tilde{C}_{42}(f)\right]\right)^2\right] \\
&= \sum_{\mathcal{C}\subseteq\mathcal{U}} \sum_{k=0}^{4} \Bigg\{ \binom{4}{k} \left(\frac{\sum_{u\in\mathcal{C}} C_{21,u}^2 C_{42,u}}{(\sum_{u\in\mathcal{C}} C_{21,u} - \sigma_\nu^2)^2} - E\left[\tilde{C}_{42}(f)\right] \right)^{4-k} \\
&\qquad \times P(\mathcal{U}(f)=\mathcal{C}) E\left[\left(\frac{\sum_{u\in\mathcal{C}} C_{21,u}^2 (\tilde{C}_{42,u}(f) - C_{42,u}) + \sigma_\nu^4 \tilde{C}_{42,\nu}(f)}{(\sum_{u\in\mathcal{C}} C_{21,u} - \sigma_\nu^2)^2} \right)^k \right] \Bigg\} \\
&\quad - \text{var}^2\left[\tilde{C}_{42}(f)\right].
\end{aligned} \tag{4.41}$$

Note that

$$E\left[\left(\frac{\sum_{u\in\mathcal{C}} C_{21,u}^2 (\tilde{C}_{42,u}(f) - C_{42,u}) + \sigma_\nu^4 \tilde{C}_{42,\nu}(f)}{(\sum_{u\in\mathcal{C}} C_{21,u} - \sigma_\nu^2)^2} \right)^k \right] = E\left[\left(\tilde{C}_{42,\mathcal{C}}(f) - C_{42,\mathcal{C}} \right)^k \right] \tag{4.42}$$

is the kth central moment of the sample normalized C_{42} of the received signal when the $\mathcal{C}$ users collide. $\tilde{C}_{42,\mathcal{C}}(f)$ is a gaussian random variable with mean given by (4.22) and variance computable from (B.9). The gaussian nature of this estimate implies that the above term is zero for $k = 3$. The distribution of $\mathcal{U}(f)$ depends on the traffic parameter λ_i of all the users:

$$P\left(\mathcal{U}(f) = \mathcal{C}\right) = \prod_{u \in \mathcal{C}} \lambda_u \prod_{u \in \mathcal{U} \setminus \mathcal{C}} (1 - \lambda_u). \tag{4.43}$$

Now, we can compute a lower bound on the probability of detecting contention-based channel access methods as

$$P\left(\hat{\mathcal{H}} = \mathcal{H}_C \middle| \mathcal{H}_C\right) \geq 1 - \Phi^{-1}\left(\frac{\tau_T - E[T']}{\text{var}[T']}\right) \tag{4.44}$$

where $E[T']$ and $\text{var}[T']$ are computed as shown above.

The primary factors that affect the detection of contention are the received powers of each primary user and the probability of collisions, i.e., $P(|\mathcal{U}(f)| > 1)$. The mean of T increases monotonically with the received power of each primary user. Therefore, the probability of detecting contention increases with an increase in the received power of each primary user.

The probability of users colliding, i.e., $P(|\mathcal{U}(f)| > 1)$, increases monotonically with increasing λ_u and also with increasing number of users.

4.4.3.3 Simulation Results

Our channel access method identification algorithm operates on a single frequency band used by a single primary user network. Our simulation system consists of N primary users communicating by a TDMA, CDMA, OFDMA, or contention-based channel access method. In TDMA and contention-based systems, all users use the same modulation scheme. We assume a Rayleigh flat fading channel between the primary users and our sensor node. The received SNR of each individual user is exponentially distributed. The offered load by the system is varied from 0.2

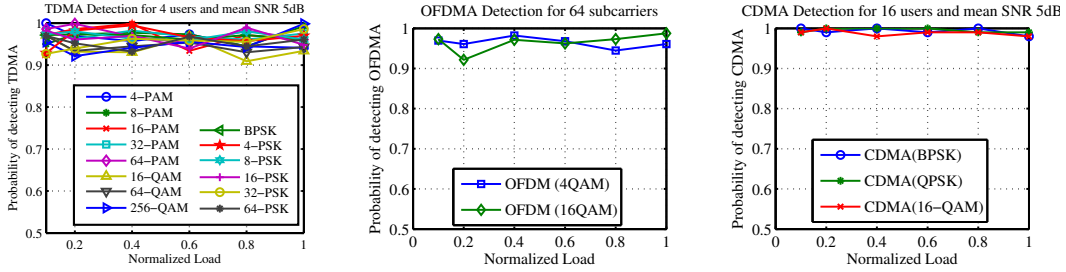
to 1 and is divided equally among all users. Our metrics are the probability of correctly detecting TDMA and contention-based channel methods. We test our proposed algorithm by transmitting the same packets once from users through TDMA and once more through a contention-based channel access scheme.

We compare our collision detection algorithm with the SVM-based algorithm proposed in [HYY14]. Their algorithm trains an SVM uses a feature vector consisting of the mean and variance of the received power; the minimum, median, and maximum lengths of the idle periods; and the minimum, median, and maximum lengths of busy periods. Unlike their algorithm, we do not need labeled training data.

The performance of our classifier depends on the modulation type, offered load, number of users and individual user's SNR.

First, we consider the probability of identifying between TDMA, OFDMA, and CDMA. Since we have chosen an algorithm based on bounding the probability of wrongly labeling a system as using a contention-based method, the probability of correctly detecting a system as TDMA, OFDMA, or CDMA is approximately independent of the offered load, the number of users, and the average SNR. For example, Fig. 4.8 shows the probability of correctly detecting TDMA, OFDMA, and CDMA for a system having an average SNR of 5dB. We have chosen the parameter $P_{C|T}$ as 0.05.

On the other hand, the probability of detecting contention-based channel access methods is highly dependent on the modulation type, the average SNR, the offered load, and the number of users. The effect of the modulation type and the offered load can be seen in Fig. 4.9. BPSK has the least $\text{var}[\tilde{C}_{42}]$ and hence, the probability of identifying contention-based systems using BPSK is the least. PAM-modulated signals have the highest $\text{var}[\tilde{C}_{42}(f)]$ and hence, the probability of identifying contention among PAM signals is high. Increasing the offered load increases the number of collisions as can be seen from (4.43). This improves the



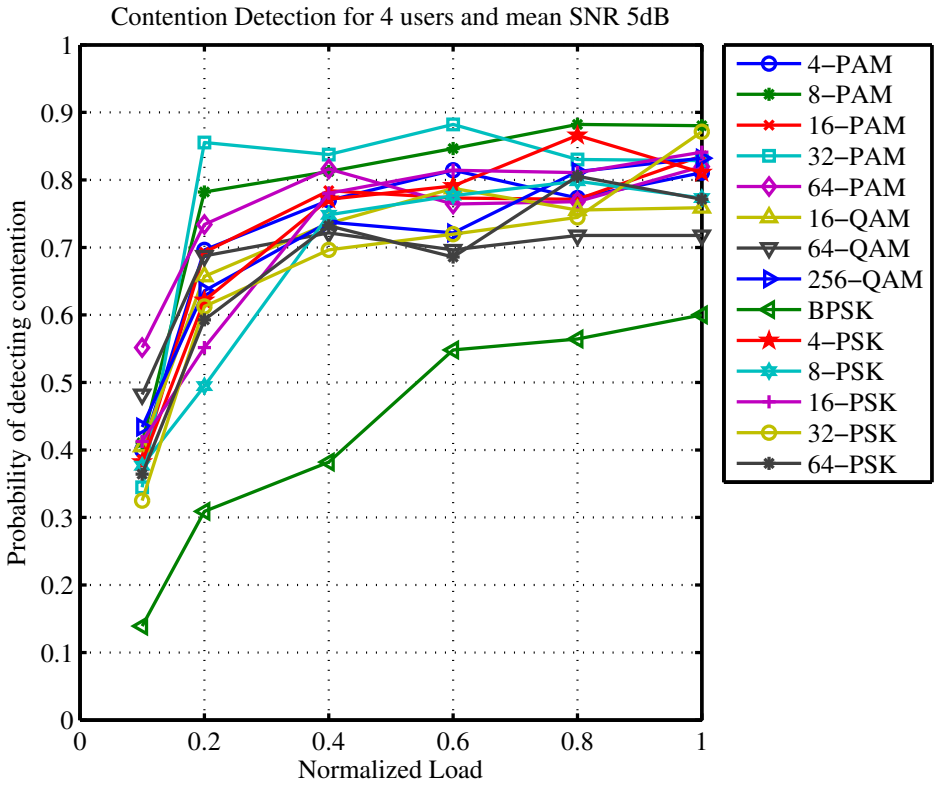
(a) Detecting 4 user TDMA (b) Detecting 64 subcarrier OFDMA (c) Detecting 16 user CDMA

Figure 4.8: Probability of correctly identifying the channel access scheme at 5dB average SNR

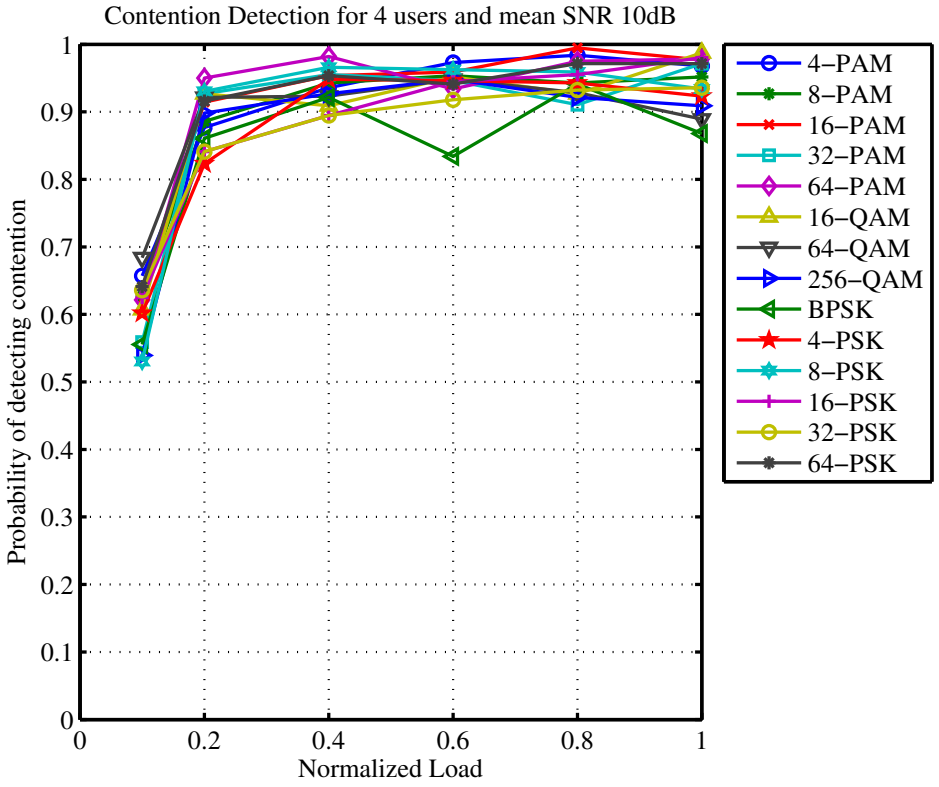
probability of correctly detecting contention-based algorithm.

To study the effect of SNR on the detection of contention-based channel access methods we illustrate, in Fig. 4.10, the probability of detecting a BPSK modulated contention-based system as a contention-based channel access method. We choose BPSK modulation because it has the worst probability of detecting contention-based channel access methods due to the low var $\left[\tilde{C}_{42}\right]$. We observe that the probability of detecting it as contention-based increases with increasing SNR of individual users. This is explained by the fact that $E[T]$, as described by (4.39), is monotonically decreasing with the individual signal powers $C_{21,u}$. Counter-intuitively, Fig. 4.10 also shows that as the number of users increases, the probability of detecting contention-based channel access method does not change significantly.

In summary, our proposed algorithm is able to successfully classify signals received from a primary user network as being either TDMA, OFDMA, CDMA, or contention-based. For TDMA systems, it is able to identify the modulation type with high accuracy.



(a) Average SNR 5dB



(b) Average SNR 10dB

Figure 4.9: Contention Detection of 4 users.

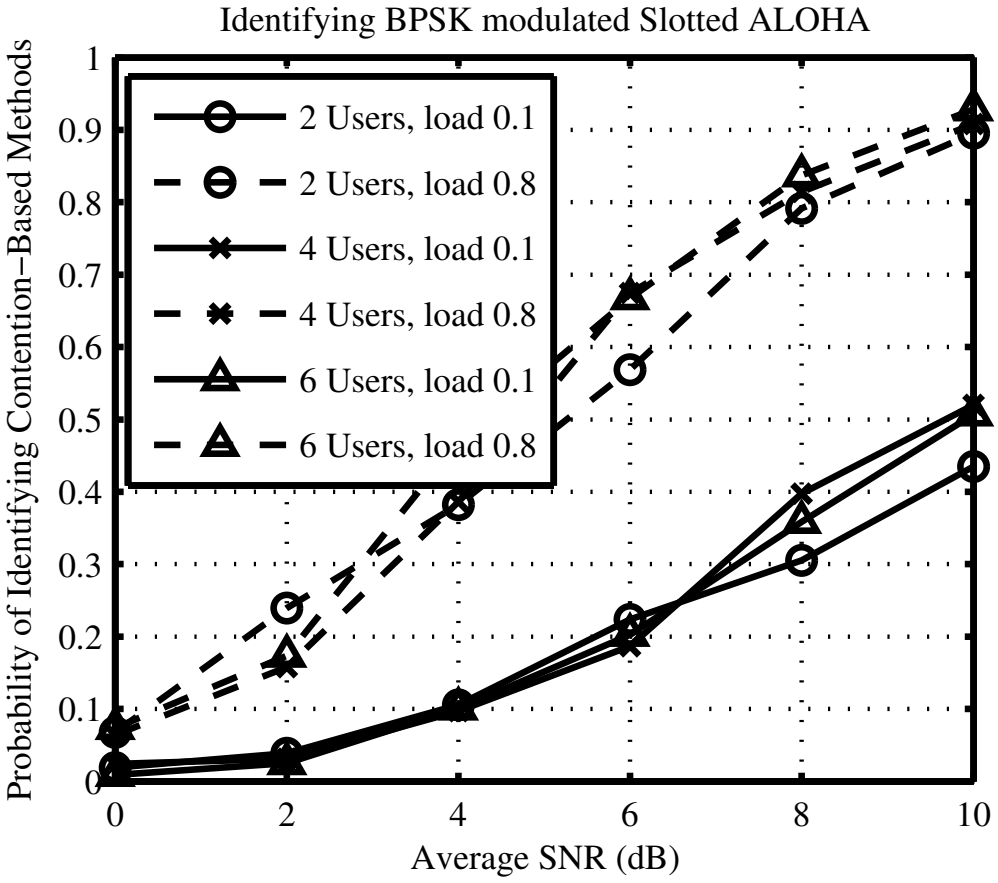


Figure 4.10: Probability of detecting contention-based channel access methods as average SNR and number of users increases

4.5 Conclusion and Future Work

We have presented in this chapter a new integrated approach for achieving channel access method classification. In particular we have developed a new method for band-segmentation that could be used in a single antenna system based on a frequency domain power detection method coupled with NNMF. Unlike other single-antenna band-segmentation appraoch, this method is able to successfully segment bands with partial spectral overlap by observing the temporal activity of each PU. We have also presented a hybrid approach that combines this matrix factorization method with a more robust energy detection based method in order to further improve the accuracy of both center freqeuncy and bandwdith estiamtion.

We have also shown in this chapter two key algorithms that can be used to identify the channel access method utilized by a primary network. In particular, we extended a cumulant-based modulation type classification technique to differentiate between OFDMA, CDMA and TDMA. A novel method based on the variance of the sample cumulant estimator has been proposed to identify contention-based channel access methods such as CSMA. The goal of channel access method classification is to further improve our knowledge of the behavior of PUs which is also the key goal of advanced spectrum sensing. We have also done comparisons with existing approaches and showed that our technique provides higher classification accuracy given the same target false alarm probability.

CHAPTER 5

Conclusion

5.1 Summary of Contributions

This dissertation has put forth the idea of advance spectrum sensing. Through several key enabling aspects of such as system, which include activity detection (conventional spectrum sensing), modulation classification and MAC-layer classification we have investigated the challenges faced by future cognitive radio systems. In particular we focused on the development, analysis, and evaluation of algorithms that solve three key tasks crucial to advance spectrum sensing.

We proposed a cyclostationary spectrum sensing method specifically targeted for a multiple antenna receivers which is highly robust to noise uncertainty because the detection threshold is not dependent on SNR. . We have shown that this method has lower computational complexity than existing methods of comparable performance because it only computes the cyclic covariance matrix once and performs the correlation in time which eliminates the need for taking a high frequency resolution FFT with large number of samples. More importantly we have derived asymptotic theoretical expressions for the probability of detection and false alarm. Using these expressions aided with simulations we evaluated the performance of this method under various scenarios including Rayleigh fading channel, correlated noise environments, and in the presence of a strong interferer. The method was shown to outperform existing techniques in all considered scenarios.

We developed a computationally efficient method for modulation level classifi-

cation based on distribution distance functions. Specifically, we proposed to use a metric based on Kolmogorov-Smirnov and Kuiper distances which exploits the distance properties between CDFs corresponding to different modulation levels. The proposed method results in faster modulation level classification than the commonly used cumulant-based method, by reducing the number of samples needed. It also results in lower computational complexity than the KS-GoF method, by eliminating the need for a sorting operation and using only a limited set of test points, instead of the entire CDF. We have further verified the practicality of these approach by implementing two modulation level classifiers in a practical real-time hardware platform with radio front-ends and a DSP engine (BEE2), and evaluated their classification performance under realistic impairments such as quantization errors and timing synchronization errors. Both classifiers were shown to follow similar trends with regard to these impairments, and their hardware computational complexity were proven to be similar. Finally we have developed the concept of distribution distance based classification. We also derived the optimal discriminant functions for classifying modulation schemes using the sampled distribution distance. This method was shown to provide substantial gains compared to other existing approaches. The performance of this method is also shown to be close to the maximum likelihood classifier but at significantly lower computational complexity. The same classifier can be generalized to any classification problem where the CDF of each class is available.

Finally, we have contributed to the problem of MAC layer classification through the development of algorithms to perform channel access method classification. In particular, we presented a novel scheme for acquiring this information comprising of three main stages 1) Band Segmentation, 2) Channel Access Method and Modulation Type Classification, and finally 3) Collision detection. We proposed and evaluated a novel method of performing band-segmentation on a wideband spectrum that is able to successfully distinguish between spectrally overlapped signals

without the need for multiple antennas as prior methods do. This is achieved by exploiting the temporal independence in primary user activity through the Non-Negative Matrix Factorization (NNMF) method. We extend existing fourth-order cumulant-based methods of modulation type classification to distinguish between multiplexing methods including TDMA, OFDMA, and CDMA. We proposed a novel method based on the sample variance of the cumulant estimator in order to identify contention-based systems. Through analysis and simulations we show that our scheme is capable of distinguishing between TDMA, OFDMA, CDMA, and contention-based channel access methods with a high probability. Further, our proposed method is capable of identifying the modulation type of non-contention-based primary users with high accuracy.

5.2 Furture Work

There are many aspects of the advance spectrum sensing system that need to be solved before it finally becomes practical. In the area of cyclostationary spectrum sensing, the approach we have presented is only able to detect a single cyclic frequency due to the limitations of the covariance matrix approach. Several work have shown that each man-made signal actually exhibits multiple cycle frequencies. There is therefore a need to adapt the eigenvalue-based approach into a multiple cycle frequency detector. Some promising ideas include applying a complex exponential that is a sum of the multiple cycle frequencies present in the signal of interest.

As for distribution distance based classification, we have already presented the optimal form of the classifier. However, the implementation of such a classifier in hardware is still an interesting challenge. As we have shown in our own hardware implementation, problems with the improper scaling of signals prior to level classification require very accurate noise power estimation algorithms. Techniques

that allow classification in the presence of interferers are also of great interest.

Lastly, the area of MAC layer classification is yet a new and barely explored problem in the context of cognitive radios. Therefore there are still many problems facing its practical application. Most importantly, the effect of large variation in received power of signals remains a big challenge in collision detection. The approaches we have shown perform best whenever collisions of almost equal received power are being observed. This is not necessarily the case in a large geographical area. There is also a need to solve the problem of frequency hopping MAC schemes which cannot be directly addressed using our technique.

APPENDIX A

Derivation of True Correlation under $\mathcal{H}_1$

In this appendix, we find the true correlation, ρ, under the non-null hypothesis $\mathcal{H}_1$ given a particular instance of the channel, $\mathbf{h}$, and noise covariance matrix, $\mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \sigma_\eta \mathbf{A}\mathbf{A}^H$. We begin by repeating (2.22) and (2.25) since these covariance matrices completely determine the test statistic.

$$\mathbf{R}_{\mathbf{xx}} = \mathbf{h}\mathbf{h}^H + \mathbf{R}_{\boldsymbol{\eta}\boldsymbol{\eta}} \tag{A.1}$$

$$\mathbf{R}_{\mathbf{xx}}^{\alpha_0} = \mathbf{h}\mathbf{h}^H R_{ss}^{\alpha}. \tag{A.2}$$

From (2.14) and under the assumption of only one SOI ($\mu_i = 0$ for $i > 1$) we have

$$1 - \mu_1^2 = \frac{1}{|\mathbf{R}_{\mathbf{xx}}|^2} \begin{vmatrix} \mathbf{R}_{\mathbf{xx}} & \mathbf{R}_{\mathbf{xx}}^{\alpha_0} \\ \mathbf{R}_{\mathbf{xx}}^{\alpha_0} & \mathbf{R}_{\mathbf{xx}} \end{vmatrix}, \tag{A.3}$$

where we recognize that both covariance matrices are Hermitian symmetric. Using the determinant for block 2×2 block matrices, we have

$$1 - \mu_1^2 = |\mathbf{R}_{\mathbf{xx}}| \left| \mathbf{R}_{\mathbf{xx}} - \mathbf{R}_{\mathbf{xx}}^{\alpha_0} \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{R}_{\mathbf{xx}}^{\alpha_0} \right| / |\hat{\mathbf{R}}_{\mathbf{xx}}|^2 \tag{A.4}$$

$$= \left| \mathbf{R}_{\mathbf{xx}} - \mathbf{R}_{\mathbf{xx}}^{\alpha_0} \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{R}_{\mathbf{xx}}^{\alpha_0} \right| / |\mathbf{R}_{\mathbf{xx}}| \tag{A.5}$$

Substituting $\mathbf{R}_{\mathbf{xx}}^{\alpha_0}$ and grouping together scalar terms

$$1 - \mu_1^2 = \left| \mathbf{R}_{\mathbf{xx}} - \left(|R_{ss}^{\alpha}|^2 \, \mathbf{h}^H \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{h} \right) \mathbf{h}\mathbf{h}^H \right| / |\mathbf{R}_{\mathbf{xx}}| \tag{A.6}$$

Using Sylvester's determinant theorem for the sum of a full-rank and rank-1 matrix

$$1 - \mu_1^2 = |\mathbf{R}_{\mathbf{xx}}| \left(1 - |R_{ss}^{\alpha}|^2 \left(\mathbf{h}^H \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{h} \right)^2 \right) / |\mathbf{R}_{\mathbf{xx}}| \tag{A.7}$$

$$\mu_1 = |R_{ss}^{\alpha}| \, \mathbf{h}^H \mathbf{R}_{\mathbf{xx}}^{-1} \mathbf{h} \tag{A.8}$$

To proceed further, we need to evaluate the inverse of (A.1) which is a sum of a full-rank matrix and a rank-1 matrix. Using the results in [Mil81, Eqn. 1] the inverse becomes

$$\mathbf{R}_{\mathbf{xx}}^{-1} = \mathbf{R}_{\boldsymbol{\eta\eta}}^{-1} - \frac{\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1}\mathbf{h}\mathbf{h}^H\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1}}{1+\mathbf{h}^H\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1}\mathbf{h}}. \tag{A.9}$$

Thus we have,

$$\mu_1 = |R_{ss}^{\alpha}|\,\mathbf{h}^H\left(\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1} - \frac{\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1}\mathbf{h}\mathbf{h}^H\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1}}{1+\mathbf{h}^H\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1}\mathbf{h}}\right)\mathbf{h} \tag{A.10}$$

$$= \frac{|R_{ss}^{\alpha}|\,\mathbf{h}^H\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1}\mathbf{h}}{1+\mathbf{h}^H\mathbf{R}_{\boldsymbol{\eta\eta}}^{-1}\mathbf{h}}. \tag{A.11}$$

Recalling that $\mathbf{R}_{\boldsymbol{\eta\eta}} = \sigma_\eta\mathbf{A}\mathbf{A}^H$, we arrive at the final expression

$$\mu_1 = |R_{ss}^{\alpha}|\,\mathbf{h}^H\mathbf{A}^{-H}\mathbf{A}^{-1}\mathbf{h}/(\sigma_\eta^2+\mathbf{h}^H\mathbf{A}^{-H}\mathbf{A}^{-1}\mathbf{h}). \tag{A.12}$$

APPENDIX B

Statistics of Sample Estimates of Cumulants

Consider J samples of a complex signal $r(n)$ as given by (4.5). Define $y(n) = r(n) - \frac{1}{J}\sum_{n=1}^{J} r(n)$ so as to obtain a zero-mean form of the received signal. We wish to derive the mean and variance of the estimator $\tilde{C}_{42,y}$ where the subscript y indicates that the cumulant is computed from the signal $y(n)$.

We begin by studying the unbiased estimator for the (unnormalized) $\hat{C}_{42,y}$ cumulant:

$$\hat{C}_{42,y} = M_{42,y} - |\hat{C}_{20,y}|^2 - 2\hat{C}_{21,y}^2. \tag{B.1}$$

Since these terms are correlated, we have

$$\begin{aligned}\operatorname{var}[\hat{C}_{42,y}] =& \operatorname{var}[\hat{M}_{42,y}] + \operatorname{var}[|\hat{M}_{20,y}|^2] + 4\operatorname{var}[\hat{M}_{21,y}^2] - 2\operatorname{cov}[\hat{M}_{42,y}, |\hat{M}_{20,y}|^2] \\ & - 4\operatorname{cov}[\hat{M}_{42,y}, \hat{M}_{21,y}^2] + 4\operatorname{cov}[|\hat{M}_{20,y}|^2, \hat{M}_{21,y}^2].\end{aligned} \tag{B.2}$$

From the Appendix in [SS00a] we have

$$\operatorname{var}[\hat{M}_{42,y}] = \frac{1}{J}(M_{84,y} - M_{42,y}^2) \tag{B.3}$$

The asymptotic analysis of both $\operatorname{var}[\hat{M}_{21,y}^2]$ and $\operatorname{cov}[\hat{M}_{42,y}, \hat{M}_{21,y}^2]$ as derived in [SS00a] are incorrect since they fail to take into account some $O(1/J)$ terms as shown in the following derivation. Using the definition $\alpha \triangleq M_{42,y} - M_{21,y}^2$,

$$\begin{aligned}\operatorname{var}[\hat{M}_{21,y}^2] =& \frac{(J-1)(J-2)(J-3)}{J^3} M_{21,y}^4 + \frac{6(J-1)(J-2)}{J^3} M_{21,y}^2 M_{42,y}^2 \\ & - \left(M_{21,y}^2 + \frac{\alpha}{J}\right)^2 + O(1/J^2) \\ =& \left(1 - \frac{6}{J}\right) M_{21,y}^4 + \frac{6}{J} M_{21,y}^2 M_{42,y} - \left(M_{21,y}^2 + \frac{\alpha}{J}\right)^2 + O(1/J^2) \\ \approx& \frac{4}{J} M_{21,y}^2 (M_{42,y} - M_{21,y}^2)\end{aligned} \tag{B.4}$$

The error is in failing to take into account the $-\frac{6}{J}M_{21,y}^4$ part of the first term. A similar derivation also gives the corrected expression

$$\operatorname{cov}[\hat{M}_{42,y}, \hat{M}_{21,y}^2] = \frac{2}{J} M_{21,y}(M_{63,y} - M_{42,y}M_{21,y}) \tag{B.5}$$

Following similar derivations we can find the rest of the terms in (B.2) as

$$\operatorname{var}[|\hat{M}_{20,y}|^2] = \frac{2}{J} M_{42,y}|M_{20,y}|^2 + \frac{2}{J}\operatorname{Re}\{M_{40,y}M_{20,y}^{*2}\} - \frac{4}{J}|M_{20,y}|^4 \tag{B.6}$$

$$\operatorname{cov}[\hat{M}_{42,y}, |\hat{M}_{20,y}|^2] = \frac{2}{J}\operatorname{Re}\{M_{62,y}M_{20,y}^*\} - \frac{2}{J}M_{42,y}|M_{20,y}|^2 \tag{B.7}$$

$$\operatorname{cov}[|\hat{M}_{20,y}|^2, \hat{M}_{21,y}^2] = \frac{4}{J} M_{21,y}\operatorname{Re}\{M_{41,y}M_{20,y}^*\} - \frac{4}{J}M_{21,y}^2|M_{20,y}|^2 \tag{B.8}$$

Substituting (B.3)–(B.8) into (B.2) we find the general expression for the asymptotic variance of the $C_{42,y}$ estimate as follows:

$$\begin{aligned} J\operatorname{var}[\hat{C}_{42,y}] \approx & M_{84,y} - M_{42,y}^2 + 8M_{21,y}\left[2M_{21,y}(M_{42,y} - M_{21,y}^2 - |M_{20,y}|^2) + 2\operatorname{Re}\{M_{41,y}M_{20,y}^*\}\right. \\ & \left. -M_{63,y} + M_{21,y}M_{42,y}\right] + 2\operatorname{Re}\{M_{20,y}^*(M_{40,y}M_{20,y}^* - 2M_{62,y})\} \\ & + 2|M_{20,y}|^2(3M_{42,y} - 2|M_{20,y}|^2). \end{aligned} \tag{B.9}$$

Now, $y(n)$ is a noisy signal. We will rewrite (B.9) in terms of cumulants so that we can quantify the effect of noise using the additive property of cumulants. The following moment–cumulant equivalence relations are easy to derive:

$$\begin{aligned} M_{84,y} = & C_{84,y} + 16C_{63,y}C_{21,y} + 12\operatorname{Re}\{C_{64,y}C_{20,y}\} + 72C_{21,y}^2C_{42,y} + 18C_{42,y}^2 + 16|C_{41,y}|^2 \\ & + |C_{40,y}|^2 + 6\operatorname{Re}\{C_{40,y}^*C_{20,y}^2\} + 96\operatorname{Re}\{C_{41,y}^*C_{20,y}\}C_{21,y} + 36|C_{20,y}|^2C_{42,y} \\ & + 72|C_{20,y}|^2C_{21,y}^2 + 24C_{21,y}^4 + 9|C_{20,y}|^4 \end{aligned} \tag{B.10}$$

$$M_{63,y} = C_{63,y} + 6\operatorname{Re}[C_{20,y}C_{43,y}] + 9|C_{20,y}|^2C_{21,y} + 6C_{21,y}^3 + 9C_{21,y}C_{42,y} \tag{B.11}$$

$$M_{42,y} = C_{42,y} + |C_{20,y}|^2 + 2C_{21,y}^2, \quad M_{40,y} = C_{40,y} + 3C_{20,y}^2, \quad M_{21,y} = C_{21,y}. \tag{B.12}$$

Gaussian noise has all the relevant cumulants zero except for $C_{21,\nu} = \sigma_\nu^2$. By slight abuse of notation, let $C_{ki,x}$ be the C_{ki} cumulant of the noiseless signal component

of $y(n)$. Then, except for $C_{21,y}$, we can rewrite all the relevant cumulants as $C_{ki,y} = C_{ki,x}$. $C_{21,y}$ can be rewritten as $C_{21,y} = C_{21,x} + \sigma_\nu^2$. Using these relations and (B.10)-(B.12), we can rewrite (B.9) in terms of cumulants.

B.1 Single User Signals

If it is known that the received signal $r(n)$ consists of a single user's signal, i.e., no collisions have occurred and it is not a CDMA or OFDMA signal, then we can use the modulation type M of the signal to describe $\text{var}\left[\tilde{C}_{42,y}\right]$. We do this by assuming that the normalizing factor $\left(\hat{C}_{21,y} - \sigma_\nu^2\right)^2$ is perfectly estimated. Then, after normalization, $C_{ki,x}$ are replaced by $C_{ki}(M)$ where $C_{ki}(M)$ is the C_{ki} cumulant of a unit power signal having modulation M. After normalization, $C_{21,y}$ would be replaced by $C_{21,y}/(C_{21,y} - \sigma_\nu^2)$. Using these relations and (B.10)-(B.12), we can rewrite (B.9) in terms of cumulants for signals modulated by real constellations as:

$$\begin{aligned} J\,\text{var}[\tilde{C}_{42,y}] =& C_{84}(M) + 4C_{63}(M)\left[\frac{C_{21,y}}{C_{21,y}-\sigma_\nu^2}\right] + 12\,\text{Re}\{C_{62}^*(M)C_{20}(M)\} + 17C_{42}(M)^2 \\ & - 8\left[\frac{C_{21,y}}{C_{21,y}-\sigma_\nu^2}\right]^2 C_{42}(M) + 34|C_{20}(M)|^2 C_{42}(M) + 16|C_{41}(M)|^2 \\ & + 24\,\text{Re}\{C_{41}(M)^* C_{20}(M)\}\left[\frac{C_{21,y}}{C_{21,y}-\sigma_\nu^2}\right] + |C_{40}(M)|^2 \\ & + 6\,\text{Re}\{C_{40}(M)^* C_{20}(M)^2\} + 24\left[\frac{C_{21,y}}{C_{21,y}-\sigma_\nu^2}\right]^4 \end{aligned} \tag{B.13}$$

where we use the fact that real constellations have all real moments, i.e., $M_{20} = M_{21}$, $M_{40} = M_{41} = M_{42}$, and $M_{62} = M_{63}$. Similarly, for signals modulated by constellations having four-fold symmetry, such as QAM, we use $C_{20} = 0 = C_{21}$ to get

$$\begin{aligned} J\,\text{var}[\tilde{C}_{42,y}] =& C_{84}(M) + |C_{40}(M)|^2 + 8\left[\frac{C_{21,y}}{C_{21,y}-\sigma_\nu^2}\right] C_{63}(M) + 20\left[\frac{C_{21,y}}{C_{21,y}-\sigma_\nu^2}\right]^2 C_{42}(M) \\ & + 4\left[\frac{C_{21,y}}{C_{21,y}-\sigma_\nu^2}\right]^4 + 17C_{42}(M)^2 \end{aligned} \tag{B.14}$$

This is the corrected form of [SS00a, Eqns. 13].

Table B.1 lists the statistics of the C_{42} estimation for unit power noise-less signals having different modulation types.

B.2 Statistics for OFDMA Signals

The use of fourth order cumulants for distinguishing OFDM signals from single carrier signals is proposed and analyzed in [SLB08]. The moments for OFDM are found to be

$$M_{84} \approx 24, \quad M_{63} \approx 6, \quad M_{40} \approx 0, \quad M_{42} \approx 2, \quad M_{21} = 1. \tag{B.15}$$

Using these moments in the corrected expression in (B.9) gives that $J \operatorname{var}[\hat{C}_{42}] \approx 4$ which coincidentally also matches the result derived in [SLB08] from the incorrect expression of [SS00a]. Note however that this result only applies for OFDM with subcarrier modulations that satisfy the fourfold symmetry such as QPSK. Using these moments the variance of the fourth order cumulant for a noisy OFDM signal can be found using (B.14).

B.3 Statistics for CDMA Signals

A CDMA signal which uses BPSK chips can be viewed as the sum of N_{total} BPSK signals given as

$$x(n) = \sum_{i=1}^{N_{\text{total}}} s_i(n) \tag{B.16}$$

where $s_i(n)$ is a BPSK formed by spreading the data to be transmitted with the particular code assigned to that user and is given by (4.4). Thus we can find the mean of the normalized fourth-order cumulant of a noiseless CDMA signal as $E[\hat{C}_{42}] = -2/N_{\text{total}}$ where we invoke the additivity property as given in (4.22). In effect, the mean normalized fourth order cumulant approaches 0 as the number of

Table B.1: Theoretical Cumulant Statistics C_{42} for Various Constellation Types, and Variances of Their Sample Estimates

Constellation	C_{42}	$Nvar_1(\hat{C}_{42})$ from [SS00a]	$Nvar_1(\hat{C}_{42})$
BPSK	-2.0000	36.00	0.00
PAM(4)	-1.3600	34.72	10.24
PAM(8)	-1.2381	32.27	9.98
PAM(16)	-1.2094	31.67	9.90
PAM(32)	-1.2024	31.52	9.88
PAM(64)	-1.2006	31.49	9.88
PAM(∞)	-1.2000	31.47	9.87
PSK($\geq$4)	-1.0000	12.00	0.00
V32	-0.6900	9.70	1.42
V29	-0.5816	8.75	1.77
QAM(4,4)	-0.6800	9.54	1.38
QAM(8,8)	-0.6191	8.82	1.39
QAM(16,16)	-0.6047	8.65	1.39
QAM(32,32)	-0.6012	8.61	1.39
QAM(∞)	-0.6000	8.59	1.39
BPSK-OFDM	0	–	$\sim$8
QPSK-OFDM	0	–	$\sim$4

users increases.

As for the variance of $\hat{C}_{42}$ we can use the general expression in (B.9) once the moments are found. Due to the blind nature of our classification problem, we do not have knowledge of the true spreading code used by each user. As a result, the correlations from one chip to another within the same symbol period cannot be known. However, unlike the single carrier signals presented in this appendix, these correlations are clearly non-zero and are dependent on the codes used. To circumvent this issue we will assume that such correlations are negligible. Note that with this assumption we are treating CDMA signals to be similar to a sum of BPSK signals in which symbols from different symbol periods and different users are regarded as i.i.d. This is clearly an approximation, but we have found through simulations that the discrepancy is negligible in practice.

With this assumption we can proceed to derive the moments of a sum of N_{total} BPSK signals to be

$$M_{42} = \binom{4}{2}\binom{N_{\text{total}}}{2}\frac{1}{N_{\text{total}}^2} + \frac{1}{N_{\text{total}}} \tag{B.17}$$

$$M_{63} = \binom{6}{2}\binom{4}{2}\binom{N_{\text{total}}}{3}\frac{1}{N_{\text{total}}^3} + \binom{6}{4}\binom{N_{\text{total}}}{2}\frac{2}{N_{\text{total}}^3} + \frac{1}{N_{\text{total}}^2} \tag{B.18}$$

$$M_{84} = \binom{8}{2}\binom{6}{2}\binom{4}{2}\binom{N_{\text{total}}}{4}\frac{1}{N_{\text{total}}^4} + \binom{8}{4}\binom{4}{2}\binom{N_{\text{total}}}{3}\frac{3}{J^4} \tag{B.19}$$

$$+\left(\binom{8}{6} + \frac{1}{2}\binom{8}{4}\right)\binom{N_{\text{total}}}{2}\frac{2}{N_{\text{total}}^4} + \frac{1}{N_{\text{total}}^3} \tag{B.20}$$

The detailed derivations of theses expressions have been left out. However, all expressions can be derived using combinatorics. The final variance with noise can then be found through (B.12).

References

[20111a] "Guest Editorial Advances in Cognitive Radio Networking and Communications (II)." *IEEE J. Sel. Areas Commun.*, **29**(4):673 –675, april 2011.

[20111b] "IEEE Standard for Information Technology–Telecommunications and information exchange between systems Wireless Regional Area Networks (WRAN)–Specific requirements Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Procedures for Operation in the TV Bands." *IEEE Std 802.22-2011*, pp. 1 –680, 1 2011.

[ALL12] E. Axell, G. Leus, E. G. Larsson, and H.V. Poor. "Spectrum Sensing for Cognitive Radio : State-of-the-Art and Recent Advances." *IEEE Signal Process. Mag.*, **29**(3):101–116, May 2012.

[And03] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis.* Wiley, 3rd edition, 2003.

[Are77] N. Arenbaev. "Asymptotic Behavior of the Multinomial Dstribution." *Theory of Probability & Its Applications*, **21**(4):805–810, 1977.

[BAP09] F. Bernardo, R. Agusti, J. Perez-Romero, and O. Sallent. "A Self-Organized Spectrum Assignment Strategy in Next Generation OFDMA Networks Providing Secondary Spectrum Access." In *Communications, 2009. ICC '09. IEEE International Conference on*, pp. 1–5, June 2009.

[Bar38] M. S. Bartlett. "Further aspects of the theory of multiple regression." *Mathematical Proceedings of the Cambridge Philosophical Society*, **34**(01):33–40, 1938.

[Bar47] M. S. Bartlett. "The General Canonical Correlation Distribution." *The Annals of Mathematical Statistics*, **18**(1):1–17, 1947.

[Bri81] D.R. Brillinger. *Time Series: Data Analysis an Theory.* Holden-Day Series in Time Series Analysis. Holden-Day, 1981.

[CDG04] G. A. P. Cirrone, S. Donadio, S. Guatelli, A. Mantero, B. Masciliano, S. Parlati, M. G. Pia, A. Pfeiffer, A. Ribon, and P. Viarengo. "A Goodness of Fit Statistical Toolkit." *IEEE Trans. Nucl. Sci.*, **51**(5):2056–2063, October 2004.

[CGD07] Hou-Shin Chen, Wen Gao, and D.G. Daut. "Spectrum Sensing Using Cyclostationary Properties and Application to IEEE 802.22 WRAN." In *Proc. IEEE GLOBECOM*, November 26–30, 2007.

[CGG10] H. Celebi, I. Guvenc, S. Gezici, and H. Arslan. "Cognitive-Radio Systems for Spectrum, Location, and Environmental Awareness." *IEEE Trans. Antennas Propag.*, **52**(4):41 –61, aug. 2010.

[CMB04] D. Cabric, S.M. Mishra, and R.W. Brodersen. "Implementation issues in spectrum sensing for cognitive radios." In *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on*, volume 1, pp. 772 – 776 Vol.1, nov. 2004.

[Coc34] W. G. Cochran. "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance." *Mathematical Proceedings of the Cambridge Philosophical Society*, **30**(02):178–191, April 1934.

[CWB05] C. Chang, J. Wawrzynek, and R.W. Brodersen. "BEE2: a high-end reconfigurable computing system." *IEEE Des. Test. Comput.*, **22**(2):114 – 125, March–April 2005.

[CXH08] Xing Chen, Wenjun Xu, Zhiqiang He, and Xiaofeng Tao. "Spectral Correlation-Based Multi-Antenna Spectrum Sensing Technique." In *Proc. IEEE WCNC*, Las Vegas, NV, USA, March 31–April 3, 2008.

[DAB07] Octavia A. Dobre, Ali Abdi, Yeheskel Bar-Ness, and Wei Su. "Survey of Automatic Modulation Classification Techniques: Classical Approaches and New Trends." *IET Communications*, **1**(2):137–156, April 2007.

[DBM10] M. Di Benedetto, S. Boldrini, C.J.M. Martin, and J.R. Diaz. "Automatic network recognition by feature extraction: A case study in the ISM band." In *Cognitive Radio Oriented Wireless Networks Communications (CROWNCOM), 2010 Proceedings of the Fifth International Conference on*, pp. 1–5, June 2010.

[DHS01] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification.* John Wiley & Sons, Inc., 2001.

[ES99] M. Eric and M. Skender. "Automatic band segmentation based on spatio-frequency processing using MUSIC algorithm." In *Proc. IEEE VTC*, Amsterdam, Netherlands, September 19–22, 1999.

[Fed05] Federal Communications Commission. "Notice of proposed rule making and order: Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies." ET Docket No. 03-108, February 2005.

[GBC87] William A. Gardner, William A. Brown, and Chih-Kang Chen. "Spectral Correlation of Modulated Signals: Part II–Digital Modulation." *IEEE Trans. Commun.*, **35**(6):595–601, June 1987.

[GPC11] W. Gabran, P. Pawelczak, and D. Cabric. "Throughput and Collision Analysis of Multichannel Multistage Spectrum Sensing Algorithms." *IEEE Trans. Veh. Technol.*, **60**(7):3309–3323, September 2011.

[GS08] A. Ghasemi and E.S. Sousa. "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs." *IEEE Commun. Mag.*, **46**(4):32 –39, april 2008.

[Hay05] Simon Haykin. "Cognitive radio: brain-empowered wireless communications." *Selected Areas in Communications, IEEE Journal on*, **23**(2):201–220, Feb 2005.

[HPM07] Rachid Hachemani, Jacques Palicot, Christophe Moy, et al. "A new standard recognition sensor for cognitive radio terminal." *EURASIP, Kessariani, Greece*, 2007.

[HYY14] Sanqing Hu, Yu-Dong Yao, and Zhuo Yang. "MAC protocol identification using support vector machines for cognitive radio networks." *Wireless Communications, IEEE*, **21**(1):52–60, February 2014.

[JLZ10] K. Jitvanichphaibool, Y.-C. Liang, and Yonghong Zeng. "Spectrum Sensing Using Multiple Antennas for Spatially and Temporally Correlated Noise Environments." In *Proc. IEEE DySPAN*, Singapore, April 6–9, 2010.

[Jon85] F. Jondral. "Automatic classification of high frequency signals." *Signal Processing*, **9**(3):177–190, October 1985. Jondral1985.

[Kai67] Thomas Kailath. "The Divergence and Bhattacharyya Distance Measures in Signal Selection." *IEEE Trans. Commun. Technol.*, **15**(1):52–60, February 1967.

[KYY12] V. Kone, L. Yang, X. Yang, B. Y. Zhao, and H. Zheng. "The Effectiveness of Opportunistic Spectrum Access: A Measurement Study." *Networking, IEEE/ACM Transactions on*, **20**(6):2005–2016, December 2012.

[Law59] D. N. Lawley. "Tests of Significance in Canonical Analysis." *Biometrika*, **46**(1/2):59–66, June 1959.

[LKK10] J. Lundén, S. A. Kassam, and V. Koivunen. "Robust Nonparametric Cyclic Correlation-Based Spectrum Sensing for Cognitive Radio." *IEEE Trans. Signal Process.*, **58**(1):38–52, 2010.

[LKT14] Yingxi Liu, N. Kundargi, and A. Tewfik. "Channel Idle Time Statistics Based Spectrum Accessing Strategies With CSMA Based Primary Networks." *Signal Processing, IEEE Transactions on*, **62**(3):572–582, February 2014.

[LS00] Daniel D. Lee and H. Sebastian Seung. "Algorithms for Non-negative Matrix Factorization." In *Proc. NIPS*, Denver, CO, USA, November 28–30, 2000. MIT Press.

[MB08] Ali Motamedi and Ahmad Bahai. "Optimal Channel Selection for Spectrum-Agile Low-Power Wireless Packet Switched Networks in Unlicensed Band." *EURASIP Journal on Wireless Communications and Networking*, **2008**(1):896420, 2008.

[Mil81] Kenneth S. Miller. "On the Inverse of the Sum of Matrices." *Mathematics Magazine*, **54**(2):67–72, 1981.

[Min09] Daniel Minoli. *Satellite Systems Engineering in an IPv6 Environment*. Auerbach Publications, 2009.

[Mis05] S. Mishra et. al. "A real time cognitive radio testbed for physical and link layer experiments." In *Proc. DySPAN 2005*, pp. 562 –567, nov. 2005.

[MM99] III Mitola, J. and Jr. Maguire, G.Q. "Cognitive radio: making software radios more personal." *Personal Communications, IEEE*, **6**(4):13 –18, aug 1999.

[MS09] A.W. Min and K.G. Shin. "An Optimal Sensing Framework Based on Spatial RSS-profile in Cognitive Radio Networks." In *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON '09. 6th Annual IEEE Communications Society Conference on*, pp. 1 –9, june 2009.

[MTM06] Mark A. McHenry, Peter A. Tenhula, Dan McCloskey, Dennis A. Roberson, and Cynthia S. Hood. "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal." In *Proceedings of the first international workshop on Technology and policy for accessing spectrum*, TAPAS '06, New York, NY, USA, 2006. ACM.

[MW80] Robb J. Muirhead and Christine M. Waternaux. "Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations." *Biometrika*, **67**(1):31–43, 1980.

[Nee06] J. O. Neel. *Analysis and design of cognitive radio networks and distributed radio resource management algorithms*. PhD thesis, Virginia Polytechnic Institute and State University, September 2006.

[Oga07] Haruhiko Ogasawara. "Asymptotic expansions of the distributions of estimators in canonical correlation analysis under nonnormality." *Journal of Multivariate Analysis*, **98**(9):1726–1750, 2007.

[PSM05] P. Papadimitratos, S. Sankaranarayanan, and A. Mishra. "A bandwidth sharing approach to improve licensed spectrum utilization." *Communications Magazine, IEEE*, **43**(12):supl.10–supl.14, Dec 2005.

[QCP08] Zhi Quan, Shuguang Cui, H. Poor, and A. Sayed. "Collaborative wideband sensing for cognitive radios." *IEEE Signal Process. Mag.*, **25**(6):60–73, November 2008.

[QCS08] Zhi Quan, Shuguang Cui, A.H. Sayed, and H.V. Poor. "Wideband Spectrum Sensing in Cognitive Radio Networks." In *Communications, 2008. ICC '08. IEEE International Conference on*, pp. 901–906, May 2008.

[QCS09] Zhi Quan, Shuguang Cui, A.H. Sayed, and H.V. Poor. "Optimal Multiband Joint Detection for Spectrum Sensing in Cognitive Radio Networks." *IEEE Transactions on Signal Processing*, **57**(3):1128–1140, 2009.

[RC11] Eric Rebeiz and Danijela Cabric. "Low Complexity Feature-Based Modulation Classifier and Its Non-Asymptotic Analysis." In *Proc. IEEE GLOBECOM*, Houston, TX, USA, December 5–9, 2011.

[SA08] H. Sadeghi and P. Azmi. "A novel primary user detection method for multiple-antenna cognitive radio." In *Proc. International Symposium on Telecommunications*, pp. 188 –192, August 2008.

[Sch86] Ralph O. Schmidt. "Multiple emitter location and signal parameter estimation." *IEEE Trans. Antennas Propag.*, **34**(3):276 – 280, March 1986.

[SCZ10] Chengqi Song, Dawei Chen, and Qian Zhang. "Understand the Predictability of Wireless Spectrum: A Large-Scale Empirical Study." In *IEEE ICC*, pp. 1 –5, May 2010.

[SF69] Nariaki Sugiura and Yasunori Fujikoshi. "Asymptotic Expansions of the Non-Null Distributions of the Likelihood Ratio Criteria for Multivariate Linear Hypothesis and Independence." *The Annals of Mathematical Statistics*, **40**(3):942–952, 1969.

[SFT11] T. Suzuki, Takeo Fujii, and O. Takyu. "An intelligent secondary MAC protocol using shared spectrum with IEEE802.11 wireless LAN." In *Ubiquitous and Future Networks (ICUFN), 2011 Third International Conference on*, pp. 358–362, June 2011.

[SG90] S.V. Schell and W.A. Gardner. "Detection of the Number of Cyclostationary Signals in Unknown Interference and Noise." In *Proc. ACSSC*, Pacific Grove, CA, USA, November 5–7, 1990.

[SLB08] Miao Shi, A. Laufer, Y. Bar-Ness, and Wei Su. "Fourth order cumulants in distinguishing single carrier from OFDM signals." In *IEEE MILCOM*, pp. 1–6, Nov 2008.

[SS97] T.E. Shrimpton and S.V. Schell. "Source enumeration using a signal-selective information theoretic criterion." In *Proc. MILCOM*, November 2–5, 1997.

[SS00a] A. Swami and B.M. Sadler. "Hierarchical digital modulation classification using cumulants." *IEEE Trans. Commun.*, **48**(3):416–429, March 2000.

[SS00b] Ananthram Swami and Brian M. Sadler. "Hierarchical digital modulation classification using cumulants." *IEEE Trans. Commun.*, **48**(3):416–429, March 2000.

[Ste74] Michael A. Stephens. "EDF Statistics for Goodness of Fit and Some Comparisons." *Journal of the American Statistical Association*, **69**(347):730–737, September 1974.

[Sug73] Nariaki Sugiura. "Asymptotic Non-Null Distributions of the Likelihood Ratio Criteria for Covariance Matrix Under Local Alternatives." *The Annals of Statistics*, **1**(4):718–728, 1973.

[SZ09] Hang Su and Xi Zhang. "Secondary User Friendly TDMA Scheduling for Primary Users in Cognitive Radio Networks." In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pp. 1–6, Nov 2009.

[Tan05] H. Tang. "Some physical layer issues of wide-band cognitive radio systems." In *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, pp. 151–159, Nov 2005.

[TCP09] Sheng-Yuan Tu, Kwang-Cheng Chen, and R. Prasad. "Spectrum Sensing of OFDMA Systems for Cognitive Radio Networks." *Vehicular Technology, IEEE Transactions on*, **58**(7):3410–3425, September 2009.

[TNG10] A. Taherpour, M. Nasiri-Kenari, and S. Gazor. "Multiple antenna spectrum sensing in cognitive radios." *IEEE Trans. Wireless Commun.*, **9**(2):814–823, February 2010.

[TS06] R. Tandra and A. Sahai. "Performance of the Power Detector with Noise Uncertainty." *IEEE 802.22-06/0075r0*, July 2006.

[TS08] R. Tandra and A. Sahai. "SNR Walls for Signal Detection." *IEEE J. Sel. Topics Signal Process.*, **2**(1):4–17, February 2008.

[Tug12] J. K. Tugnait. "On Multiple Antenna Spectrum Sensing Under Noise Variance Uncertainty and Flat Fading." *IEEE Trans. Signal Process.*, **60**(4):1823 –1832, April 2012.

[URP11] P. Urriza, E. Rebeiz, P. Pawelczak, and D. Cabric. "Computationally Efficient Modulation Level Classification Based on Probability Distribution Distance Functions." *IEEE Commun. Lett.*, **PP**(99):1 –3, 2011.

[WB12] W.D. Wellisch and AN. Barreto. "Spectrum sensing of TETRA systems through timefrequency analysis." In *Communications (LATINCOM), 2012 IEEE Latin-America Conference on*, pp. 1–6, November 2012.

[WC12] Fanggang Wang and Chung Chan. "Variational-Distance-Based Modulation Classifier." In *Proc. IEEE ICC*, Ottawa, Canada, June 10-15, 2012.

[WRP09] Matthias Wellens, Janne Riihijrvi, and Petri Mhnen. "Empirical time and frequency domain models of spectrum use." *Physical Communication*, **2**(1-2):10 – 32, 2009.

[WW10] Fanggang Wang and Xiaodong Wang. "Fast and robust modulation classification via Kolmogorov-Smirnov test." *IEEE Trans. Wireless Commun.*, **58**(8):2324–2332, August 2010.

[YA09] Tevfik Yücek and Hüseyin Arslan. "A survey of spectrum sensing algorithms for cognitive radio applications." *IEEE Commun. Surveys Tuts.*, **11**(1):116–130, March 2009.

[Yan05] J. Yang. "Spatial Channel Characterization for Cognitive Radios." Technical Report UCB/ERL M05/8, EECS Department, University of California, Berkeley, Jan 2005.

[YSR11] Tsung-Han Yu, O. Sekkat, S. Rodriguez-Parera, D. Markovic, and D. Cabric. "A Wideband Spectrum-Sensing Processor With Adaptive Detection Threshold and Sensing Time." *IEEE Transactions on Circuits and Systems I: Regular Papers*, **58**(11):2765–2775, 2011.

[Yu13] T. H. Yu. *Energy-Efficient VLSI Signal Processing for Wideband Spectrum Sensing in Cognitive Radios.* PhD thesis, UCLA, Los Angeles, CA, USA, 2013.

[ZL09] Yonghong Zeng and Y-C Liang. "Eigenvalue-based spectrum sensing algorithms for cognitive radio." *IEEE Trans. Commun.*, **57**(6):1784–793, June 2009.

[ZTS07] Qing Zhao, Lang Tong, A. Swami, and Yunxia Chen. "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A

POMDP framework." *Selected Areas in Communications, IEEE Journal on*, **25**(3):589–600, April 2007.