2014

Spectrum sensing, spectrum monitoring, and security in cognitive radios

Erfan Soltanmohammadi
Louisiana State University and Agricultural and Mechanical College

SPECTRUM SENSING, SPECTRUM MONITORING, AND SECURITY IN COGNITIVE RADIOS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The School of Electrical Engineering and Computer Sciences

by
Erfan Soltanmohammadi
B.Sc., K.N. Toosi University of Technology, 2007
M.Sc., Amirkabir University of Technology, 2010
August 2014

To Ebrahim, Shirin, and Ehsan

Acknowledgments

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

I would like to express my deepest gratitude to my advisor, Dr. Mort Naraghi-Pour, for his excellent guidance, friendship, caring, patience, and providing me with an excellent atmosphere for doing research. He gave me the main idea of this work, and without his continuous encouragement next to his deep knowledge of communication and mathematics, this thesis could not have been accomplished.

I would like to acknowledge the financial, academic and technical support of Louisiana State University.

I want to express my special thanks to my thesis committee, Dr. Shuangqing Wei, Dr. Xue-bin Liang, Dr. Guoxiang Gu, and Dr. Karsten Thompson for attending my general and final exams and providing me with their thoughtful comments and suggestions.

Table of Contents

List of Tables

List of Figures

Abstract

Spectrum sensing is a key function of cognitive radios and is used to determine whether a primary user is present in the channel or not. In this dissertation, we formulate and solve the generalized likelihood ratio test (GLRT) for spectrum sensing when both primary user transmitter and the secondary user receiver are equipped with multiple antennas. We do not assume any prior information about the channel statistics or the primary user's signal structure. Two cases are considered when the secondary user is aware of the energy of the noise and when it is not. The final test statistics derived from GLRT are based on the eigenvalues of the sample covariance matrix.

In-band spectrum sensing in overlay cognitive radio networks requires that the secondary users (SU) periodically suspend their communication in order to determine whether the primary user (PU) has started to utilize the channel. In contrast, in spectrum monitoring the SU can detect the emergence of the PU from its own receiver statistics such as receiver error count (REC). We investigate the problem of spectrum monitoring in the presence of fading where the SU employs diversity combining to mitigate the channel fading effects. We show that a decision statistic based on the REC alone does not provide a good performance. Next we introduce new decision statistics based on the REC and the combiner coefficients. It is shown that the new decision statistic achieves significant improvement in the case of maximal ratio combining (MRC).

Next we consider the problem of cooperative spectrum sensing in cognitive radio networks (CRN) in the presence of misbehaving radios. We propose a novel approach based on the iterative expectation maximization (EM) algorithm to detect

the presence of the primary users, to classify the cognitive radios, and to compute their detection and false alarm probabilities.

We also consider the problem of centralized binary hypothesis testing in a cognitive radio network (CRN) consisting of multiple classes of cognitive radios, where the cognitive radios are classified according to the probability density function (PDF) of their received data (at the FC) under each hypotheses.

Chapter 1

Spectrum Sensing Over MIMO Channels Using Generalized Likelihood Ratio Tests

1.1 Introduction

Cognitive radio (CR) is the enabling technology for dynamic spectrum access allowing unlicensed secondary users (SU) to utilize a frequency band when it is vacant of the licensed primary users (PU) [1, 2]. Spectrum sensing (SS) is a key functionality of CR by which it detects whether the PU signal is present in the channel or not. Reliable SS is difficult to accomplish. To protect the PU against undue interference from SU, stringent requirements are established on the performance of SS (e.g., detection probability, detection delay) [3]. The challenge is that these requirements must be met for weak PU signals with unknown parameters and in the presence of unknown noise power or channel impairments.

Considering the PU transmitter and the SU receiver, we can categorize the sensing approaches based on the number of antennas at each side. Several methods have been proposed for single-input-single-output (SISO) case where the PU transmitter and the SU receiver each have a single antenna. These methods are different based on the amount of knowledge available at the SU receiver regarding the PU signal, the channel or the noise power. For a review of some of these methods we refer to [4, 5] and the references therein.

Using multiple antennas can improve the performance of SS. Recently, several algorithms have been proposed for single-input-multiple-output (SIMO) case where the SU receiver is equipped with multiple antennas [6, 7, 8, 9]. In [6, 7, 8], the authors derived test statistics using the generalized likelihood ratio test (GLRT) assuming that the PU signal is Gaussian. The authors in [9] proposed a blind

test statistic based on the channel path correlation among the signals received at different antennas.

In modern wireless communication multiple antenna systems are also used in the transmitter where space-time block coding (STBC), [10], can be employed to improve performance. Today multiple antenna transmitters have become an integral part of many standards such as Long Term Evolution (LTE) [11], IEEE 802.11n [12], and IEEE 802.16 [13]. In [14, 15] spectrum sensing methods are proposed for multiple-input-multiple-output (MIMO) channel where both the PU transmitter and the SU receiver are equipped with multiple antennas. In [14], the authors only considered fast fading channels where all channel coefficients are independent Gaussian random variables in both time and antenna. They also assumed that, in the presence of the PU, the received signal is Gaussian. These assumptions make it easy to find the test statistic based on GLRT since all the received symbols from different antennas are independent and identically distributed Gaussian variables. Clearly this is not always the case in practice. In [15] two test statistics based on the eigenvalues of the received sample covariance matrix were proposed for which the performance (in term of probabilities of false alarm and detection) was evaluated in [16, 17].

In this chapter, we find the solution of GLRT for blind spectrum sensing without assuming any prior information on the structure and statistics of the PU signal or the channel. We consider two cases where the SU is aware of the noise variance and when it is not. The solution of GLRT is a new test statistic based on the eigenvalues of the sample covariance matrix. Our results are reminiscent of the MUSIC algorithm in that when the PU signal is present, and when the number of SU receiver antennas, N, is larger than the number of PU transmitter antennas,

M, then the M largest eigenvalues correspond to the signal+noise subspace and the $N - M$ remaining eigenvalues correspond to the noise subspace.

1.2 Notations and problem formulation

In this dissertation, we denote matrices by capital bold letters, vectors by small bold letter and scalars by small letters. Let M and N denote the number of antennas in the PU transmitter and the SU receiver, respectively. Let T denote the number of samples at each antenna of the SU receiver (T is assumed to be greater than N and M). The $N \times T$ received data matrix $\mathbf{Y}$ in the presence ($\mathcal{H}_1$) or absence ($\mathcal{H}_0$) of the PU is modeled as

$$\begin{cases} \mathcal{H}_1 : \mathbf{Y} = \mathbf{HS} + \mathbf{W} \\ \mathcal{H}_0 : \mathbf{Y} = \mathbf{W} \end{cases} \tag{1.1}$$

where $\mathbf{H}$ is an $N \times M$ matrix denoting the channel coefficients, and $\mathbf{S}$ is the $M \times T$ matrix of transmitted symbols which is assumed to be full rank, and $\mathbf{W}$ denotes the noise. The components of $\mathbf{W}$ are assumed to be independent and identically distributed Gaussian random variables with mean zero and variance σ^2. We note that both $\mathbf{H}$ and $\mathbf{S}$ are unknown to the SU. In this dissertation we consider two cases where σ^2 may or may not be known at the SU.

The log-likelihood functions under $\mathcal{H}_i$ denoted by L^i, $i = 0, 1$ are given by

$$L^0(\sigma^2) = -NT \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathrm{tr}\left\{\mathbf{YY}^\dagger\right\} \tag{1.2}$$

$$L^1(\mathbf{H}, \mathbf{S}, \sigma^2)$$
$$= -NT \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathrm{tr}\left\{(\mathbf{Y} - \mathbf{HS})(\mathbf{Y} - \mathbf{HS})^\dagger\right\} \tag{1.3}$$

where superscript $\dagger$ is Hermitian transpose, and $\mathrm{tr}\{\mathbf{X}\}$ denotes the trace of matrix $\mathbf{X}$. Using GLRT for detection, we get the following test statistics

$$\max L^1(\mathbf{H}, \mathbf{S}, \sigma^2) - \max L^0(\sigma^2) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi \tag{1.4}$$

where maximization is over all the unknown parameters.

1.3 Generalized Likelihood Ratio Test Solution

In the following, we investigate (1.4) for two different cases, when the variance of the noise is or is not available at the SU receiver.

1.3.1 Case 1: Known Noise Variance

Since σ^2 is known, the maximization in (1.4) is over $\mathbf{H}$ and $\mathbf{S}$. The gradient with respect to a complex matrix $\mathbf{X}$ is given by $\nabla_{\mathbf{X}} \triangleq \nabla_{\Re\mathbf{X}} + j\nabla_{\Im\mathbf{X}}$, where $\Re\mathbf{X}$ and $\Im\mathbf{X}$ denote the real and imaginary parts of $\mathbf{X}$, respectively [18, 19]. Therefore, we can write,

$$\nabla_{\mathbf{H}} \mathrm{tr}\left\{\mathbf{H}\mathbf{S}\mathbf{Y}^{\dagger}\right\} = 0 \tag{1.5}$$

$$\nabla_{\mathbf{H}} \mathrm{tr}\left\{\mathbf{Y}\mathbf{S}^{\dagger}\mathbf{H}^{\dagger}\right\} = 2\mathbf{Y}\mathbf{S}^{\dagger} \tag{1.6}$$

$$\nabla_{\mathbf{H}} \mathrm{tr}\left\{\mathbf{H}\mathbf{S}\mathbf{S}^{\dagger}\mathbf{H}^{\dagger}\right\} = 2\mathbf{H}\mathbf{S}\mathbf{S}^{\dagger} \tag{1.7}$$

$$\nabla_{\mathbf{S}} \mathrm{tr}\left\{\mathbf{H}\mathbf{S}\mathbf{Y}^{\dagger}\right\} = 0 \tag{1.8}$$

$$\nabla_{\mathbf{S}} \mathrm{tr}\left\{\mathbf{Y}\mathbf{S}^{\dagger}\mathbf{H}^{\dagger}\right\} = 2\mathbf{H}^{\dagger}\mathbf{Y} \tag{1.9}$$

$$\nabla_{\mathbf{S}} \mathrm{tr}\left\{\mathbf{H}\mathbf{S}\mathbf{S}^{\dagger}\mathbf{H}^{\dagger}\right\} = 2\mathbf{H}^{\dagger}\mathbf{H}\mathbf{S} \tag{1.10}$$

Setting $\nabla_{\mathbf{H}} L^1(\mathbf{H}, \mathbf{S}, \sigma^2) = 0$, followed by (1.5), (1.6), and (1.7) implies

$$\mathbf{H} = \mathbf{Y}\mathbf{S}^{\dagger}(\mathbf{S}\mathbf{S}^{\dagger})^{-1} \tag{1.11}$$

Setting $\nabla_{\mathbf{S}} L^1(\mathbf{H}, \mathbf{S}, \sigma^2) = 0$ followed by (1.8), (1.9), and (1.10) implies

$$\mathbf{H}^{\dagger}\mathbf{Y} = \mathbf{H}^{\dagger}\mathbf{H}\mathbf{S} \tag{1.12}$$

Substituting (1.11) into (1.12) results in

$$\mathbf{Y}^{\dagger}\mathbf{Y}\mathbf{S}^{\dagger} = \mathbf{S}^{\dagger}(\mathbf{S}\mathbf{S}^{\dagger})^{-1}\mathbf{S}\mathbf{Y}^{\dagger}\mathbf{Y}\mathbf{S}^{\dagger} \tag{1.13}$$

Now $\mathbf{S}^{\dagger}(\mathbf{S}\mathbf{S}^{\dagger})^{-1}\mathbf{S}$ is an idempotent matrix. Therefore its eigenvalues are either zero or one [20]. This implies that any summation of its eigenvectors corresponding

to its nonzero eigenvalues is also an eigenvector. Since $\mathbf{S}^\dagger(\mathbf{S}\mathbf{S}^\dagger)^{-1}\mathbf{S}\mathbf{S}^\dagger = \mathbf{S}^\dagger$, any column of $\mathbf{S}^\dagger$ is an eigenvector of $\mathbf{S}^\dagger(\mathbf{S}\mathbf{S}^\dagger)^{-1}\mathbf{S}$ corresponding to an eigenvalue of one. As $\mathbf{S}^\dagger(\mathbf{S}\mathbf{S}^\dagger)^{-1}\mathbf{S}$ is idempotent then

$$\operatorname{rank}\left(\mathbf{S}^\dagger(\mathbf{S}\mathbf{S}^\dagger)^{-1}\mathbf{S}\right) = \operatorname{tr}\left(\mathbf{S}^\dagger(\mathbf{S}\mathbf{S}^\dagger)^{-1}\mathbf{S}\right) = M \tag{1.14}$$

Therefore the columns of $\mathbf{S}^\dagger$ form the set of all the eigenvectors of $\mathbf{S}^\dagger(\mathbf{S}\mathbf{S}^\dagger)^{-1}\mathbf{S}$ with eigenvalues equal to one. This together with the fact that (1.13) is an eigenvalue-eigenvector equation of $\mathbf{S}^\dagger(\mathbf{S}\mathbf{S}^\dagger)^{-1}\mathbf{S}$, gives

$$\mathbf{Y}^\dagger\mathbf{Y}\mathbf{S}^\dagger = \mathbf{S}^\dagger\mathbf{E} \tag{1.15}$$

where $\mathbf{E}$ is an $M \times M$ elementary matrix. This is a homogeneous Sylvester equation and has nonsingular solution if all of the eigenvalues of $\mathbf{E}$ belong to the set of eigenvalues of $\mathbf{Y}\mathbf{Y}^\dagger$ [21]. Moreover, substituting (1.11) and (1.15) into (1.3) results,

$$L^1(\mathbf{H}, \mathbf{S}, \sigma^2) = -NT\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\{\operatorname{tr}\left(\mathbf{Y}\mathbf{Y}^\dagger\right) - \operatorname{tr}\left(\mathbf{E}\right)\right\} \tag{1.16}$$

Note that a diagonal matrix with M of the eigenvalues of $\mathbf{Y}^\dagger\mathbf{Y}$ as its diagonal elements is a possible candidate solution for $\mathbf{E}$, and a matrix with the corresponding eigenvectors as columns is a possible solution for $\mathbf{S}^\dagger$. Since the trace of $\mathbf{E}$ is the sum of its eigenvalues and since all of the eigenvalues of $\mathbf{Y}^\dagger\mathbf{Y}$ are nonnegative, to maximize $L^1(\mathbf{H}, \mathbf{S}, \sigma^2)$, $\mathbf{E}$ should have the M largest eigenvalues of $\mathbf{Y}^\dagger\mathbf{Y}$ as its eigenvalues. We would like to point out that the other candidates for $\mathbf{E}$ constructed from other combinations of the eigenvalues are responsible for other critical points of $L^1(\mathbf{H}, \mathbf{S}, \sigma^2)$ corresponding to local maxima.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_T$ denote all of the eigenvalues of $Y^\dagger Y$, where at most N of them are nonzero. Thus,

$$\max_{\mathbf{H},\mathbf{S}} L^1(\mathbf{H}, \mathbf{S}, \sigma^2) = -NT\log(2\pi\sigma^2) - \frac{\sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{M}\lambda_i}{2\sigma^2} \tag{1.17}$$

Therefore, the test statistic in (1.4) is simplified to

$$\sum_{i=1}^{M} \lambda_i \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 2\sigma^2\xi \triangleq \mu_1 \tag{1.18}$$

Note if $M = N$, the test statistic in (1.18) reduces to the energy detector.

1.3.2 Case 2: Unknown Noise Variance

In this case we assume that there is no prior information available at the SU. Therefore $L^1(\mathbf{H}, \mathbf{S}, \sigma^2)$ and $L^0(\sigma^2)$ must be maximized with respect to σ^2. Taking derivative of $\max_{\mathbf{H},\mathbf{S}} L^1(\mathbf{H}, \mathbf{S}, \sigma^2)$ with respect to σ^2 and setting it to zero we get,

$$\sigma^2 = \frac{1}{2NT}\left\{\sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{M}\lambda_i\right\} \tag{1.19}$$

Note that under $\mathcal{H}_1$, the M largest eigenvalues correspond to the signal+noise subspace whereas the remaining $(N - M)$ eigenvalues correspond to the noise subspace. Using (1.19), the maximum of the likelihood is given by,

$$\max_{\mathbf{H},\mathbf{S},\sigma^2} L^1(\mathbf{H}, \mathbf{S}, \sigma^2) = -NT\log\left(\frac{\pi}{NT}\left\{\sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{M}\lambda_i\right\}\right) - NT \tag{1.20}$$

Similarly, setting the derivative of $L^0(\sigma^2)$ with respect to σ^2 to zero we get

$$\max_{\sigma^2} L^0(\sigma^2) = -NT\log\left(\frac{\pi}{NT}\left\{\sum_{i=1}^{N}\lambda_i\right\}\right) - NT \tag{1.21}$$

After some algebra, the GLRT in (1.4) is given by

$$\frac{\sum_{i=1}^{M}\lambda_i}{\sum_{i=1}^{N}\lambda_i} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 1 - \exp\left(-\frac{\xi}{NT}\right) \triangleq \mu_2 \tag{1.22}$$

Note that for $M = 1$, (1.18) and (1.22) reduce to the results in [6] found for a single transmitter antenna.

1.4 Numerical Results

In this section we evaluate the performance of the proposed decision statistics from simulations. Each result is obtained from at least 10^4 independent runs of the experiment. Each entry of the noise matrix, $\mathbf{W}$, and channel matrix, $\mathbf{H}$, is obtained from an independent zero-mean complex Gaussian distribution with variance 0.5. Signal-to-noise ratio is defined as SNR$= 10\log_{10}\frac{E\|S\|^2}{T\sigma^2}$, where E denotes expectation.

Fig. 1.1 shows the normalized eigenvalues (i.e., the ratio of each eigenvalue to the sum of the eigenvalues) when $M = 3$, $N = 8$, $T = 128$ and for SNR $= \infty$ dB, 10 dB, 5 dB, 0 dB, -5 dB, -10 dB, $-\infty$ dB and when the PU samples are independent, identically distributed circularly symmetric complex Gaussian random variables. It can be seen that, as SNR grows, the five smallest eigenvalue approach zero and the three larger eigenvalues become more prominent.

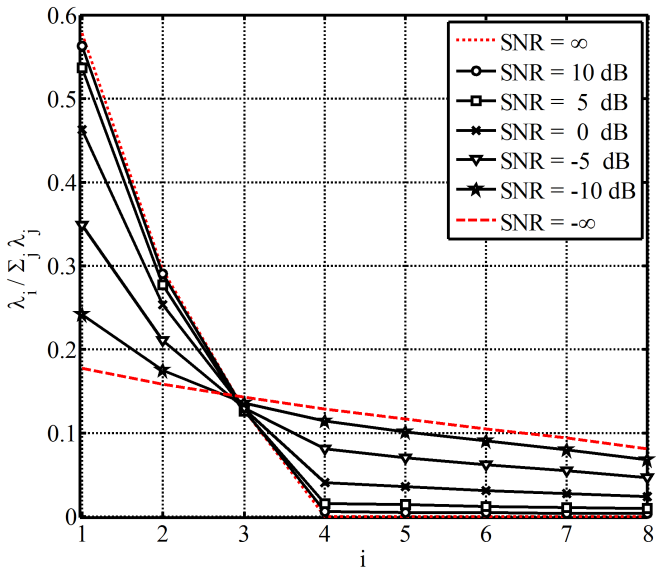

FIGURE 1.1. Normalized energy of eigenvalues when $M = 3$, $N = 8$, $T = 128$ and for SNR $= \infty$ dB, 10 dB, 5 dB, 0 dB, -5 dB, -10 dB, $-\infty$ dB and when the PU samples are independent identical circularly symmetric complex Gaussian random variables.

In the following examples we construct the transmitted signal, $\mathbf{S}$, using the unitary STBC in [22] which is widely used in practical MIMO systems. We also compare our test statistics with the maximum-to-minimum eigenvalue (MME) test statistic and energy to minimum eigenvalue (EME) test statistic, [15]. The performance of the proposed test statistics for Case 1, (1.18), and Case 2, (1.22), MME and EME are evaluated in Fig. 1.2, where we show the ROCs for $M = 2$, $N = 4$, and $T = 512$ and for SNR $= -8$ dB, -11 dB when PU employs the 2×2 Alamouti's block code and the modulation for the elements is 16-QAM. As expected, by increasing SNR the performance improves, and the performance of Case 1, where the SU knows the variance of the noise, is better than Case 2 which has to estimate the variance of the noise, and the proposed test statistics also outperform MME and EME.

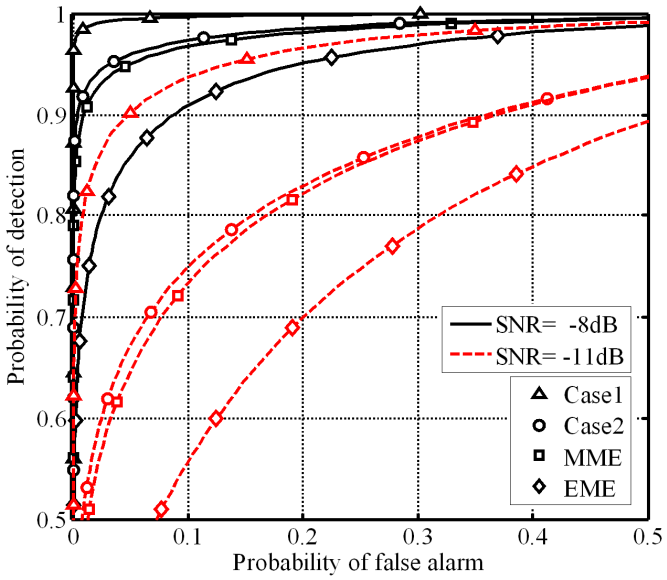

FIGURE 1.2. Comparison of the ROC curves for Case 1, Case 2, MME, and EME for $M = 2$, $N = 4$, $T = 512$ and for SNR $= -8$ dB, -11 dB, when the PU uses Alamouti's scheme with 16-QAM modulation.

Fig. 1.3 shows the probability of missed detection, $1 - p_d$, versus the probability of false alarm, p_f, for different test statistics, when the PU uses eigen-beamforming

for transmission, [22]. The parameters are $M = 5$, $N = 8$, and $T = 1620$, where the PU uses 256-QAM modulation.

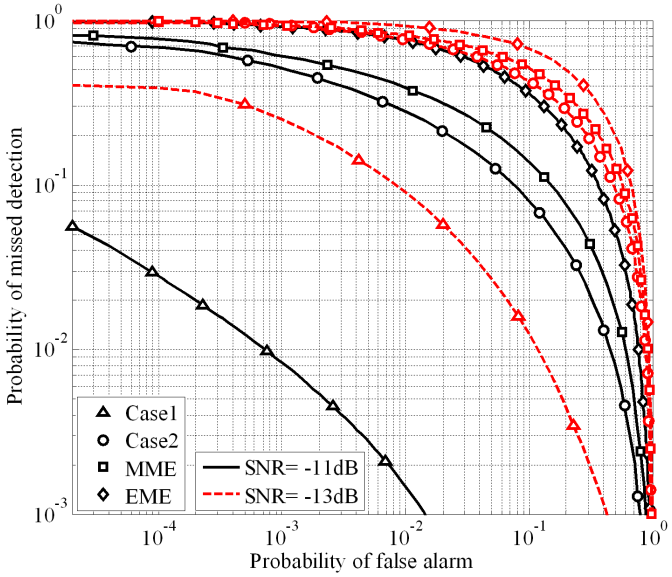

FIGURE 1.3. Comparison of the performance of Case 1, Case 2, MME, and EME for $M = 5$, $N = 8$, for $T = 1620$, and for SNR $= -11$ dB, -13 dB when the PU uses eigen-beamforming for transmission with 256-QAM modulation.

Fig. 1.4 compares the detection probability of Case 1, Case 2, MME, and EME for different number of SU receiver antennas, N, and different SNR values. The simulation was run for $M = 3$, $p_f = 0.1$, $T = 512$, and for SNR$= -8$ dB, -10 dB, when the PU uses unitary STBC with QPSK modulation. The probability of detection is increased by increasing the number of antennas and by increasing SNR. As always, the proposed test statistics outperform MME and EME.

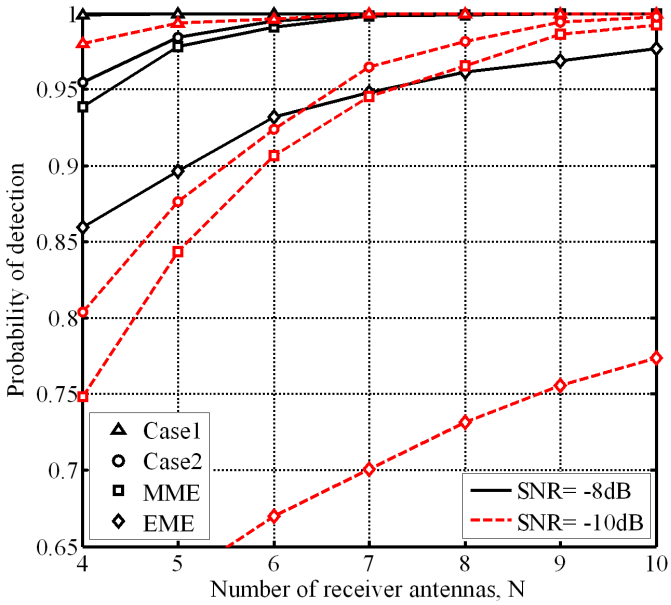

FIGURE 1.4. Probability of detection versus the number of receiver antennas, N, for Case 1 and Case 2, $M = 3$, $p_f = 0.1$, $T = 512$, and for SNR $= -8$ dB, -10 dB, when the PU uses unitary STBC with QPSK modulation.

Chapter 2

Improving the Sensing-Throughput Tradeoff for Cognitive Radios in Rayleigh Fading Channels

2.1 Introduction

Vehicular networks are expected to significantly improve safety and convenience of transportation systems and mitigate traffic congestion by improving road traffic flow. Dynamic spectrum access (DSA) has been proposed for vehicular ad-hoc networks (VANET) to allow access to licensed spectral bands such as TV white spaces [23, 24, 25, 26, 27]. In particular, in the European "DRiVE" project, DSA is the main focus for spectrum allocation in heterogeneous networks [28].

DSA allows unlicensed secondary users (SU) to utilize the licensed spectral bands that are not in use by the incumbent primary users (PU). Cognitive radio (CR), viewed as the enabling technology for DSA, relies on spectrum sensing (SS) to determine whether a given frequency band is vacant of the PU signal [29, 30, 31][1]. Since during their own communication the SUs do not sense the channel, they must periodically suspend their transmission and enter a sensing period so as to determine whether the PU has emerged or not. In order to protect the PU against undue interference from the SUs, stringent requirements are imposed on the detection probability and maximum detection delay of the SS algorithm (see for example [33]). Detection probability can be improved by increasing the duration of the sensing periods and detection delay can be reduced by decreasing the duration

[1]We should point out that this approach is referred to as overlay CR. In contrast, in underlay CR the SU can always access the licensed spectrum provided it can regulate its transmit power so as not to cause harmful interference to the PU. Overlay CR is considered to be more practical since, in contrast to underlay CR, it does not require instantaneous information on the interference channel [32].

of the SU's transmission periods. Both approaches, however, result in reduced throughput in the secondary network.

There is an intricate tradeoff between protection of the PU and the quality of service (QoS) of the SU, referred to as sensing-throughput tradeoff in [34]. In [35], Tang *et al.* evaluate the effect of PU traffic on the SU throughput. In [36], Akin *et al.* assume statistical QoS and maximize the throughput for the SU. To improve the SU's throughput, adaptive scheduling of spectrum sensing to the primary user activities has been investigated in [37] and [38]. These approaches, however, are mainly concerned with spectrum sensing and do not consider the possibility of sensing while the SU is communicating.

It is clear that during the SU's transmissions, the emergence of the PU increases the interference experienced by the SU. This in turn causes a drop in the SU's signal-to-noise (plus interference) ratio (SNR) and it may increase the number of errors in the SU packets. Therefore, while communicating, the SU may attempt to detect the emergence of the PU by monitoring the changes in the receiver's SNR or the number of errors in each received packet. Using this idea, in [39, 40] Boyd *et al.* introduced spectrum monitoring (SM) in which the SU utilizes its receiver statistics to detect the emergence of the PU during the SU's own communication. In [41] we proposed a decision statistic for SM based on the receiver error count (REC)[2] and the output of a cyclic redundancy check (CRC) code and show that for AWGN channels the proposed algorithm significantly improves the throughput[3] of the SU subject to a maximum PU detection delay.

Using receiver statistics to detect the emergence of the PU would be effective provided that the changes are mainly due to the emergence of the PU (e.g., in

[2]REC denotes the number of errors observed in a received packet and is more carefully defined in Section 2.2.1.

[3]In this chapter we have used the notions of *throughput* and *channel utilization* interchangeably to refer to the average fraction of time that the SU is able to use the channel when the PU is absent.

the case of AWGN channels). However, this approach may not be effective in the presence of fading in the secondary channel as the changes in the receiver statistics may be due to the variations of the channel rather than the interference from the PU signal. In this chapter we investigate the problem of spectrum monitoring in the case that the secondary channel experiences flat Rayleigh fading. We first show that approaches which are based on the REC alone do not perform well. Next we consider the use of a multi-antenna system to improve the performance of spectrum monitoring. Multi-antenna systems in conjunction with diversity combining have been widely used in wireless communication to combat the deleterious effects of channel fading. Recently, multi-antenna systems have also been proposed for SS where it is shown that they can significantly improve the performance of SS techniques [42, 43, 44, 9, 45]. We assume that the SU uses a multi-antenna system along with one of three diversity combining techniques, namely maximal ratio combining (MRC), equal gain combining (EGC), or selective combining (SC). We introduce a new decision statistic based on the REC, a CRC code and the combiner statistics. The performance of this new decision statistic is evaluated in terms of detection and false alarm probabilities, channel utilization and detection delay. We also simulate the proposed system using two forward error correcting codes, namely a BCH code and a convolutional code. It is shown that the results from these simulations are closely matched with those from analysis.

2.2 System Model and Problem Formulation

The SU starts with a spectrum sensing interval (SSI) of duration T_s during which it senses the channel. If at the end of an SSI the channel is found to be occupied, another SSI begins[4] and this continues until the SU finds the channel to be vacant.

[4]We should point out that the results presented here will not change if the SU moves to another channel once it finds the current channel to be occupied.

At this time a spectrum monitoring interval (SMI) begins during which the SU transmits a maximum of K_M packets. After the reception of each packet the SU computes a decision statistic (described below) in order to detect whether the PU has emerged in the in-band channel. If it is decided that the PU has emerged, the SU terminates the SMI and enters the spectrum sensing phase. Otherwise the channel is deemed to be vacant and the SU continues its packet transmission. To allow for periodic sensing of the channel the SU terminates an SMI after the transmission of (at most) K_M packets and starts a new SSI.

Let H_η denote the hypothesis of interest where $\eta = 0$ and 1 correspond to the absence and the presence of the PU signal, respectively. We assume that the SU receiver is equipped with $L \geq 1$ identical antenna branches and that, as in [46], the L branches experience identically distributed, uncorrelated flat fading. The nth received symbol at the lth branch of the SU under H_η is given by,

$$r_{l,n} = s_n h_l + v_{l,n} + \eta u_{l,n}, \quad l = 1, 2, \cdots, L, \quad n = 1, 2, \cdots \tag{2.1}$$

where $\{s_n\}$ is the sequence of SU's transmitted symbols, $\{v_{l,n}\}_{l=1}^{L}$ denote L independent, identically distributed (i.i.d.) circularly symmetric Gaussian noise processes with zero mean and variance $\mathcal{E}_v$, and for $k \neq l$, $\{v_{k,n}\}$ and $\{v_{l,n}\}$ are independent, and $\{u_{l,n}\}$ denotes the sequence of primary user symbols at the lth branch of the SU receiver. We assume that the PU symbols $\{u_{l,n}\}$ have undergone independent flat fading which is not explicitly shown but is included in the symbols $\{u_{l,n}\}$. Finally, $\{h_l\}_{l=1}^{L}$, which denote the (secondary) channel fading coefficients, are i.i.d. circularly symmetric Gaussian random variables with mean zero and variance 1, i.e., $h_l \sim \mathcal{CN}(0, 1)$. Let $\alpha_l \triangleq |h_l|$ and let $\theta_l \triangleq \angle h_l$.

We assume linear combining in the SU receiver where the output of the combiner is given by,

$$r_n \triangleq \sum_{l=1}^{L} w_l r_{l,n} \tag{2.2}$$

and where w_l, $l = 1, 2, \cdots, L$ are the combiner weighting coefficients which are determined by the diversity combining technique [47]. Table 2.1 shows the values of w_l for the three combining techniques MRC, EGC, and SC.

2.2.1 Decision Statistic

At the SU transmitter the information sequence is first encoded using a CRC code (for error detection) followed by a forward error correction (FEC) scheme to obtain an N-bit packet. A block diagram of the receiver is shown in Fig. 2.1 (with the switch $\mathcal{S}$ open for now) where the received packet is demodulated and decoded. The decoded packet is then checked by the CRC and also encoded using a replica of the transmitter's encoder. The encoder output is compared to the output of the demodulator[5] to calculate the number of errors referred to as REC and denoted by e in the following. Note that the actual number of errors in a packet, subsequently denoted by k, is not always available in the receiver. In particular, when the packet is not decoded correctly, then $e \neq k$ and therefore the value of k is unknown to the receiver. However, if the packet is decoded correctly, then $k = e$.

Remark 1. *In today's communication systems FEC and CRC are in widespread use to combat channel errors and to verify whether the packet is correctly decoded or not, respectively [48]. Therefore there is no loss of throughput due to FEC if it is already in use by the SU; moreover, the throughput loss due to the addition of CRC is very small considering the number of CRC bits compared to the length of a packet.*

[5]If the decoder uses soft decision, then hard decision must be performed on the demodulator output before comparison with the encoder's output.

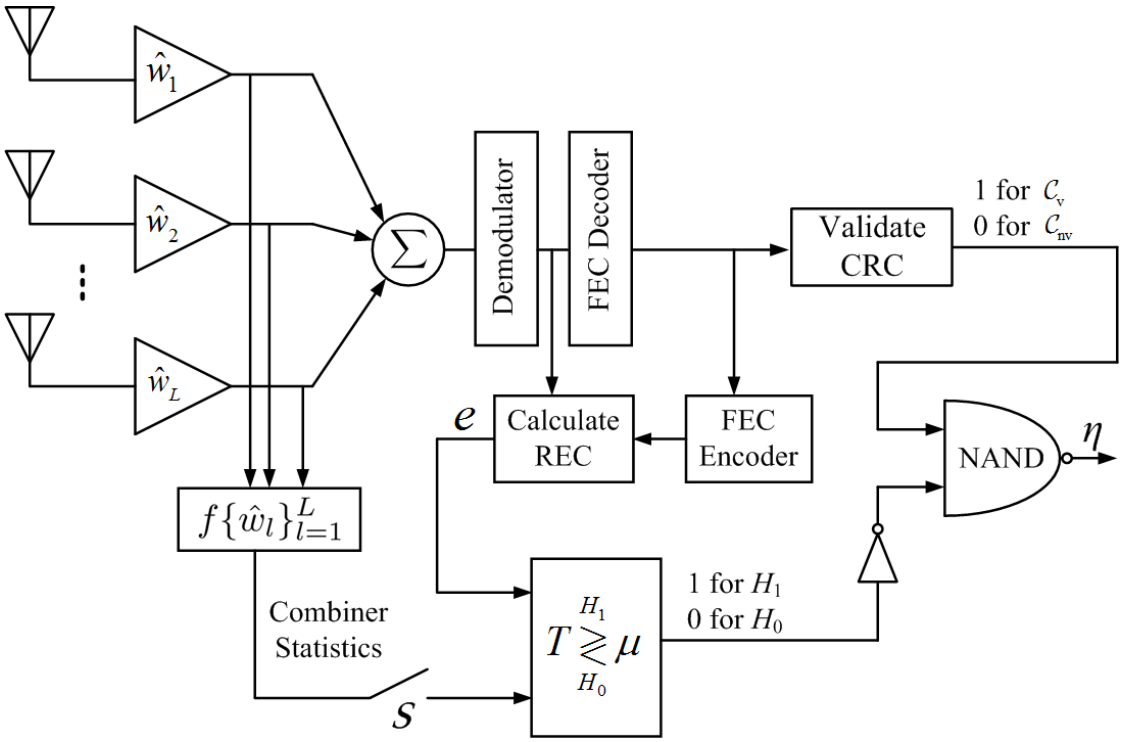

FIGURE 2.1. Proposed model using demodulator statistics and combiner statistics.

Denote by $\mathcal{C}_{\rm v}$ and $\mathcal{C}_{\rm nv}$ the events that the CRC is validated and not validated, respectively. The decision statistic in SMI is defined by,

$$T^{(\rm REC)} \sim \begin{cases} \left(\{e \geq \mu^{(\rm REC)}\} \cap \mathcal{C}_{\rm v}\right) \cup \mathcal{C}_{\rm nv}, & \text{Decide } H_1 \\ \text{Otherwise}, & \text{Decide } H_0 \end{cases} \tag{2.3}$$

where $\mu^{(\rm REC)}$ is the REC threshold which is chosen not to exceed $t^{(\rm FEC)}$, the maximum number of errors in a packet that the FEC is able to correct. The decision statistic in (2.3) indicates the emergence of the PU if the CRC is not validated, or if the CRC is validated and the REC e exceeds the threshold $\mu^{(\rm REC)}$. As (2.3) indicates, the decision statistic in not a function of the actual number of errors k.

If the packet is decoded correctly, then the CRC will correctly identify this event ($\mathcal{C}_{\rm v}$) and in this case $e = k$. On the other hand if the decoder fails, then either the CRC will identify this event ($\mathcal{C}_{\rm nv}$) or the CRC fails to identify the decoder failure ($\mathcal{C}_{\rm v}$). In the former case the SMI will be terminated. However, in the latter

case when the packet is not decoded correctly and the CRC also fails to identify this event, the proposed SM scheme may fail for the current packet. Consequently the SMI may be terminated when PU is not present (resulting in loss of channel utilization for the SU) or it may be continued when PU is present (resulting in increased detection delay). It is shown in [49] that for large packets (e.g. $N > 100$) the probability of failure for an $L-$bit CRC is approximately 2^{-L}. (Thus for the commonly used CRCs such as CCITT-16, CRC-32-Castagnoli and CRC-32-IEEE [50, 51, 52], the probability of CRC failure is around 1.5×10^{-5} and 2.3×10^{-10}, respectively.) It is shown in [41] that the increase in detection delay due to the CRC failure is less than $2^{-L} \times T_p$ seconds and the loss in channel utilization for the SU is less than $2^{-L} \times \frac{T_s}{T_s + K_M \times T_p} < 2^{-L}$, where T_p is the packet transmission time. In light of this, in the following we ignore the event of a decoder failure followed by a CRC failure.

If $\mathcal{C}_{\text{nv}}$ occurs, then the received packet is not correctly decoded (decoder failure). So, $\mathcal{C}_{\text{nv}}$ implies that $k \geq t^{(\text{FEC})} \geq \mu^{(\text{REC})}$. On the other hand (ignoring the event that the decoder and the CRC both fail), $\mathcal{C}_{\text{v}}$ implies that the packet is correctly decoded. Thus from (2.3) we get

$$\begin{aligned} & p\left(\left(\{e \geq \mu^{(\text{REC})}\} \cap \mathcal{C}_{\text{v}}\right) \cup \mathcal{C}_{\text{nv}}\right) \\ & = p\left(\left(\{k \geq \mu^{(\text{REC})}\} \cap \mathcal{C}_{\text{v}}\right) \cup \left(\{k \geq \mu^{(\text{REC})}\} \cap \mathcal{C}_{\text{nv}}\right)\right) \\ & = p(\{k \geq \mu^{(\text{REC})}\}) \end{aligned} \tag{2.4}$$

Therefore, (2.3) is equivalent to,

$$T^{(\text{REC})} = k \underset{H_0}{\overset{H_1}{\gtrless}} \mu^{(\text{REC})}. \tag{2.5}$$

The probabilities of false alarm and detection in SMI are given by $p_f = p(\{k \geq \mu^{(\text{REC})}\} \mid H_0)$ and $p_d = p(\{k \geq \mu^{(\text{REC})}\} \mid H_1)$, respectively.

It is shown in [9] that, if the modulation scheme of the PU is a constant modulus scheme such as MPSK, then after undergoing Rayleigh fading, the received PU sequences $\{u_{l,n}\}$, for $l = 1, 2, \cdots, L$ are i.i.d. zero-mean circularly symmetric complex Gaussian (CSCG) random processes. This model is also accurate if the PU uses orthogonal frequency division multiplexing (OFDM) [44]. For other modulation schemes with a large constellation this assumption is approximately true [53, 9]. Therefore we model $\{u_{l,n}\}$ as a zero-mean CSCG random process with variance $\mathcal{E}_u$. This model is assumed merely to make the analysis tractable and the proposed decision statistic does not depend on this assumption. Other articles using this model include [34] and [54].

From the SU receiver point of view, the PU signal during SMI is an additive noise. So the SNR for branch l under H_η is given by,

$$\gamma_\eta^{(l)} \triangleq \frac{|h_l|^2 \mathcal{E}_s}{\mathcal{E}_v + \eta \mathcal{E}_u} \tag{2.6}$$

where $\mathcal{E}_s$ is the energy of the SU transmitted signal. Note that for a given packet this SNR is fixed.

In the SU receiver, the signals from different branches are combined. Let $\gamma_\eta, \eta = 0, 1$, denote the SNR at the output of the combiner and let $p_{\gamma_\eta}(x)$ denotes its probability density function. Table 2.1 shows γ_η and $p_{\gamma_\eta}(x)$ for different combining techniques [55]. To find the number of bits in error in a packet of length N, we assume that the SU uses BPSK modulation. The results can be extended to other modulation scheme in an straightforward manner by substituting the probability of bit error corresponding to the modulation of interest. The probability of k errors in a packet of length N under H_η can now be written as

$$p_k(k|H_\eta) = E_{\gamma_\eta}[p_k(k|\gamma_\eta, H_\eta)] = \int_0^\infty \binom{N}{k} p_b(x)^k (1 - p_b(x))^{N-k} p_{\gamma_\eta}(x)\, dx \tag{2.7}$$

where $p_b(\gamma_\eta)$ is the bit error probability for SNR γ_η.

TABLE 2.1. Combiner-Weights, Instantaneous SNR, and its Distribution for Maximum Ratio Combining (MRC), Equal Gain Combining (EGC) and Selection Combining (SC) Techniques. In the table, $\gamma_{b,\eta}$ is the average SNR of the received signal per branch under H_η and $\mathcal{E}_r^{(l)}$ is the energy of the received packet for the l^{th} branch.

Diversity	w_l	$p_{\gamma_\eta}(x)$	γ_η
MRC	h_l^*	$\frac{1}{(L-1)!\gamma_{b,\eta}^L} x^{(L-1)} e^{\frac{-x}{\gamma_{b,\eta}}}$	$\sum_{l=1}^{L} \gamma_\eta^{(l)}$
EGC	$e^{-j\theta_l}$	L=2: $\frac{1}{\gamma_{b,\eta}} e^{\frac{-2x}{\gamma_{b,\eta}}} - \sqrt{\pi} e^{\frac{-x}{\gamma_{b,\eta}}}$ $\times \left(\frac{1}{2\sqrt{x\gamma_{b,\eta}}} - \sqrt{\frac{x}{\gamma_{b,\eta}^3}} \right)$ $\times \left(1 - 2Q\sqrt{2x/\gamma_{b,\eta}} \right)$,	$\frac{(\sum_{i=1}^{L} \alpha_l)^2 \gamma_{b,\eta}}{L}$
SC	$\begin{cases} 1 & \mathcal{E}_r^{(l)} > \mathcal{E}_r^{(k)}, \forall k \neq l \\ 0 & \text{otherwise} \end{cases}$	$\frac{L}{\gamma_{b,\eta}} \left(1 - e^{-x/\gamma_{b,\eta}} \right)^{(L-1)}$ $\times e^{-x/\gamma_{b,\eta}}$	$\max \left\{ \gamma_\eta^{(1)}, \ldots, \gamma_\eta^{(L)} \right\}$

Fig. 2.2 shows the receiver operating characteristic (ROC) curves (p_d vs. p_f) obtained from analysis as described above as well as from simulation for three cases of MRC, EGC and SC. The average SNR per branch under H_0 and H_1 is fixed and equal to 2 dB and 0.6 dB, respectively. In the case of EGC and for $L > 2$ branches, $p_{\gamma_\eta}(x)$ cannot be written in closed form [55]. Therefore, the performance is only evaluated from simulations. For comparison we also show the ROC curves for the decision statistic in (2.3) over AWGN channel and for the same SNR values. As Fig. 2.2 shows, while $T^{(\text{REC})}$ is effective in detecting the emergence of the PU in AWGN channels, in the case of fading channels its performance deteriorates significantly. This degradation is expected and is due to the fact that $T^{(\text{REC})}$ cannot determine whether an increase in the number of errors in a packet is due to the interference from the PU signal or is caused by channel fading. This result implies that for fading channels, using the REC alone as a test statistic may not provide acceptable performance even when diversity techniques are used. Hence, alternative decision statistics are needed.

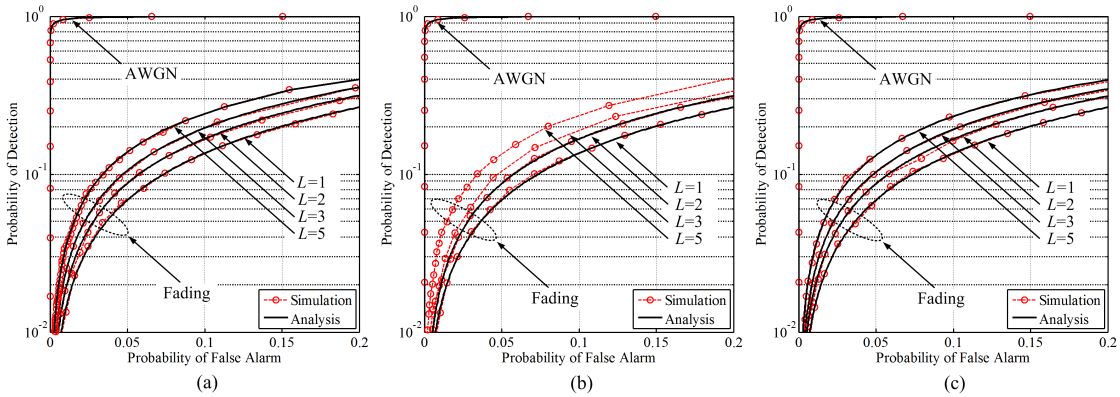

FIGURE 2.2. Comparison between the performance of $T^{(\mathrm{REC})}$ in AWGN channel and fading channel for different number of antennas, L, and different diversity techniques for $N = 1024$, $\gamma_{b,0} = 2$ dB and $\gamma_{b,1} = 0.6$ dB and BPSK signaling. (a): Maximum Ratio Combining, $T^{(\mathrm{REC})}_{\mathrm{MRC}}$, (b): Equal Gain Combining, $T^{(\mathrm{REC})}_{\mathrm{EGC}}$ (c): Selection Combining, $T^{(\mathrm{REC})}_{\mathrm{SC}}$.

2.2.2 Channel Estimation

The probability in (2.7) is derived assuming that the combiner weighting coefficients $w_l, l = 1, 2, \cdots, L$ are derived from precise knowledge of the channel coefficients $h_l, l = 1, 2, \cdots, L$. However, in practice the channel coefficients have to be estimated and there is always an error between the estimated channel coefficients and their actual values. In general, channel estimation error is caused by two distinct channel impairments [56]. One is due to the decorrelation of the pilots from the signal due to distinct distortions that the channel imparts on them because of their separation in time or frequency. The second is due to noise. It can be seen that the first phenomenon affects the channel estimation in the same manner whether the PU is present (H_1) or not (H_0). The estimation error due to noise, however, will be different as the SU experiences more noise when PU is present due to the interference from the PU signal.

Denote by $\hat{h}_l = \hat{\alpha}_l e^{j\hat{\theta}_l}$ the estimated channel coefficient corresponding to h_l. As in [46] we assume that the channel estimation errors, defined by $\ell_l \triangleq \hat{h}_l - h_l$, are independent of the channel coefficients h_l, and that $\{\ell_l\}_{l=1}^{L}$ are independent

and identically distributed (i.i.d.) circularly symmetric complex Gaussian random variables. Given the hypothesis H_η, the complex correlation coefficient ϱ_η between h_l and $\hat{h}_l$ and its magnitude denoted by ρ_η are defined by

$$\varrho_\eta \triangleq \frac{E[h_l \hat{h}_l^* | H_\eta]}{\sqrt{E[|h_l|^2] E[|\hat{h}_l|^2 | H_\eta]}} = \varrho_\eta^R + j\varrho_\eta^I \tag{2.8}$$

$$\rho_\eta^2 \triangleq |\varrho_\eta|^2 = (\varrho_\eta^R)^2 + (\varrho_\eta^I)^2, \quad \eta = 0, 1.$$

where here and subsequently, superscripts R and I represent the real and imaginary parts, respectively. From the assumptions on $\{h_l\}_{l=1}^L$ and $\{\ell_l\}_{l=1}^L$ we conclude that the estimated channel coefficients $\{\hat{h}_l\}_{l=1}^L$ are also i.i.d. circularly symmetric Gaussian random variables and conditioned on H_η,

$$\hat{h}_l | H_\eta \sim \mathcal{CN}(0, 2 - \rho_\eta^2) \quad \text{for} \quad l = 1, 2, \cdots, L, \tag{2.9}$$

where by $X|\Lambda \sim \mathcal{CN}(m, \sigma^2)$ we denote the conditional distribution of X given Λ. Finally we have $\text{var}(\ell_l | H_\eta) = \text{var}(h)(1 - \rho_\eta^2)$.

In the case of imperfect channel estimation, it is shown in [46] that the probability of observing a bit in error is identical to the case of perfect channel estimation with effective SNR $\gamma_\eta^{(\text{eff})} \triangleq (\varrho_\eta^R)^2 \gamma_\eta / (1 + \gamma_\eta(1 - \rho_\eta^2))$. Consequently, when weighting coefficients are not perfectly estimated, the performance of the proposed decision statistic will be equivalent to that of a system with a lower SNR.

2.3 Decision Statistics Using Error Counts and Combiner Coefficients

We saw in the previous section that the REC is not a good indicator of the presence or absence of the PU signal. Therefore, in our decision statistic we would like to augment the REC with the channel state information (CSI) that is available in the SU receiver in the form of the combiner coefficients. To emphasize the fact that the combiner coefficients are obtained from an estimate of the CSI (rather than the exact values), in the following we denote the combiner coefficients by

$\hat{w}_l, l = 1, 2, \cdots, L$ and let $\hat{\mathbf{w}} = (\hat{w}_1, \hat{w}_2, \cdots, \hat{w}_L)$. We define a new decision statistic as follows.

$$T \sim \begin{cases} \left(\left\{ \dfrac{p(e, f(\hat{\mathbf{w}}) \mid H_1)}{p(e, f(\hat{\mathbf{w}}) \mid H_0)} \geq \mu \right\} \cap \mathcal{C}_{\mathrm{v}} \right) \cup \mathcal{C}_{\mathrm{nv}}, & \text{Decide } H_1 \\ \text{Otherwise}, & \text{Decide } H_0 \end{cases} \tag{2.10}$$

where $f(\hat{\mathbf{w}})$ is a function of the combiner coefficients to be determined for each diversity scheme. Fig. 2.1 (with the switch $\mathcal{S}$ closed) shows the proposed model. Similar to the approach from (2.3) to (2.5), one can show that[6] when $t^{(\mathrm{FEC})} \geq \mu$,

$$T = \frac{p(k, f(\hat{\mathbf{w}}) \mid H_1)}{p(k, f(\hat{\mathbf{w}}) \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \mu \ . \tag{2.11}$$

While the receiver implements the decision rule in (2.10), for our analysis an in order to determine the function $f(.)$ for each combining method, we consider (2.11) in the following.

2.3.1 Maximal Ratio Combining

It is well known that MRC is the optimum diversity technique in the sense of maximizing the output SNR of the combiner [57]. In the case of imperfect channel estimation, the combiner coefficients are given by $\hat{w}_l = \hat{h}_l^*$. To evaluate the decision statistic in (2.11), we first find the joint probability of observing k errors and an estimated channel fading vector $\underline{\hat{h}} \triangleq (\hat{h}_1, \hat{h}_2, \cdots, \hat{h}_L)$ given H_η, i.e.,

$$p(k, \underline{\hat{h}} | H_\eta) = p(k | \underline{\hat{h}}, H_\eta) p(\underline{\hat{h}} | H_\eta) \tag{2.12}$$

We have

$$p(\underline{\hat{h}} | H_\eta) = \prod_{l=1}^{L} p(\hat{h}_l^R | H_\eta) p(\hat{h}_l^I | H_\eta), \tag{2.13}$$

[6] Ignoring the event that the decoder and the CRC both fail.

From (2.9) and the fact that $\hat{h}_l$'s are i.i.d., we get

$$p(\underline{\hat{h}}|H_\eta) = \frac{1}{\left[2\pi(1-\rho_\eta^2/2)\right]^L} \exp\left(-\frac{\sum_{l=1}^{L}|\hat{h}_l|^2}{2(1-\rho_\eta^2/2)}\right) \tag{2.14}$$

To find $p(k|\underline{\hat{h}}, H_\eta)$, let

$$\psi \triangleq \frac{\mathrm{Re}\left(\sum_{l=1}^{L} h_l \hat{h}_l^*\right)}{\sqrt{\sum_{l=1}^{L}|\hat{h}_l|^2}}. \tag{2.15}$$

Then,

$$\begin{aligned} p(k|\underline{\hat{h}}, H_\eta) &= \int_{-\infty}^{\infty} p(k|\psi, \underline{\hat{h}}, H_\eta) p(\psi|\underline{\hat{h}}, H_\eta) d\psi \qquad (2.16)\\ &= \int_{-\infty}^{\infty} \binom{N}{k} [P(E|\psi, \underline{\hat{h}}, H_\eta)]^k [1 - P(E|\psi, \underline{\hat{h}}, H_\eta)]^{N-k}\ p(\psi|\underline{\hat{h}}_\eta, H_\eta)\ d\psi \end{aligned}$$

where $P(E|\psi, \underline{\hat{h}}, H_\eta)$ is the bit error probability given $\psi, \underline{\hat{h}}$ and H_η and is given by, [46]

$$P(E|\psi, \underline{\hat{h}}, H_\eta) = Q\left(\psi\sqrt{2\gamma_\eta}\right) \tag{2.17}$$

Moreover, it is shown in Appendix 5 that,

$$\psi|\underline{\hat{h}}, H_\eta \sim \mathcal{N}\left(\frac{\sqrt{\sum_{l=1}^{L}|\hat{h}_l|^2}}{2-\rho_\eta^2},\ \frac{1-\rho_\eta^2}{2(2-\rho_\eta^2)}\right) \tag{2.18}$$

By substituting (2.17) and (2.18) into (2.16), we get,

$$\begin{aligned} p(k|\underline{\hat{h}}, H_\eta) = \int_{-\infty}^{\infty} &\binom{N}{k} Q^k(\psi\sqrt{2\gamma_\eta})(1 - Q(\psi\sqrt{2\gamma_\eta}))^{N-k} \\ &\times \frac{1}{\sqrt{\pi\left(\frac{1-\rho_\eta^2}{2-\rho_\eta^2}\right)}}\ e^{-\frac{(\hat{\mathcal{A}} - (2-\rho_\eta^2)\psi)^2}{(1-\rho_\eta^2)(2-\rho_\eta^2)}}\ d\psi \end{aligned} \tag{2.19}$$

where

$$\hat{\mathcal{A}} \triangleq \sqrt{\sum_{l=1}^{L}|\hat{h}_l|^2}. \tag{2.20}$$

Substituting (2.14), (2.19), and (2.20) into (2.12), we get $p(k, \underline{\hat{h}}|H_\eta)$.

From (2.14) and (2.19) it is evident that $p(k, \underline{\hat{h}}|H_\eta)$ depends only on $\hat{\mathcal{A}}$ and not the values of individual $\hat{h}_l$'s. All combinations of the estimated channel coefficients $\hat{h}_1, \hat{h}_2, \cdots, \hat{h}_L$ which result in the same value for $\hat{\mathcal{A}}$ are observed with equal probability at the SU. Consequently, in the case of MRC, instead of $(k, \underline{\hat{h}})$ it is sufficient to use the pair $(k, \hat{\mathcal{A}})$ in the decision statistic. Thus we let $f(\hat{\mathbf{w}}) \triangleq \hat{\mathcal{A}}$ and define our decision statistic by

$$T_{\text{MRC}} \triangleq \frac{p(k, \hat{\mathcal{A}} \mid H_1)}{p(k, \hat{\mathcal{A}} \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \mu^{(\text{MRC})} \tag{2.21}$$

where $\mu^{(\text{MRC})} \leq t^{(\text{FEC})}$ is the threshold in the case of MRC. Analysis of this rule requires $p(k, \hat{\mathcal{A}}|H_\eta)$ which can be obtained from $p(k, \underline{\hat{h}}|H_\eta)$. From (2.14) and (2.19) we see that $p(k, \underline{\hat{h}}|H_\eta)$ depends only on $||\underline{\hat{h}}||$. Letting $\Xi_\eta(k, ||\underline{\hat{h}}||) \triangleq p(k, \underline{\hat{h}}|H_\eta)$ we get

$$p(k, \hat{\mathcal{A}}|H_\eta) = \Xi_\eta(k, \hat{\mathcal{A}}) S_{2L}(\hat{\mathcal{A}}) \tag{2.22}$$

where

$$S_n(r) = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1} \tag{2.23}$$

is the surface area of the n-dimensional hyper-sphere of radius r[58]. Evaluation of (2.22) requires the computation of the integral in (2.19). In Appendix 5 an approximation for $p(k, \hat{\mathcal{A}}|H_\eta)$ is derived in closed form which does not involve any integration. The accuracy of this approximation is verified by comparing in Section 2.4 the performance results from analysis (using this approximation) with simulation results.

For an intuitive explanation of the proposed decision rule in (2.21), consider the 2D space of $e \in \mathbb{N},\ 0 \leq e \leq N$, and $\hat{\mathcal{A}} \in \mathbb{R}^+$ which is split into two decision regions,

Ω_0 and Ω_1 associated with H_0 and H_1, respectively. Fig. 2.3 demonstrates two examples of these decision regions when $\gamma_0 = 6$ dB, $\gamma_1 = 0$ dB, $\rho_0 = 0.95$, $\rho_1 = 0.85$, $N = 256$ and $L = 2$ antennas are employed in the MRC combiner. The two decision boundaries are plotted for the false alarm probabilities of $p_f = 0.01$ and 0.05 and the corresponding detection probabilities, $p_d = 0.73$ and 0.86, respectively. In each case the area under the curve shows Ω_0 and the area above the curve shows Ω_1. Note that when the SU experiences large fades (small $\hat{\mathcal{A}}$), it expects to observe a large number of errors per packet due to fading alone. Therefore, as demonstrated in Fig. 2.3, in this case only for a very large number of errors a decision is made in favor of H_1. On the other hand when fading is small (large $\hat{\mathcal{A}}$), only a few errors per packet can be attributed to fading. As a result, in this case even for a small number of errors a decision is made in favor of H_1. This is how the inclusion of $\hat{\mathcal{A}}$ in the decision statistic improves the performance of spectrum monitoring over fading channels.

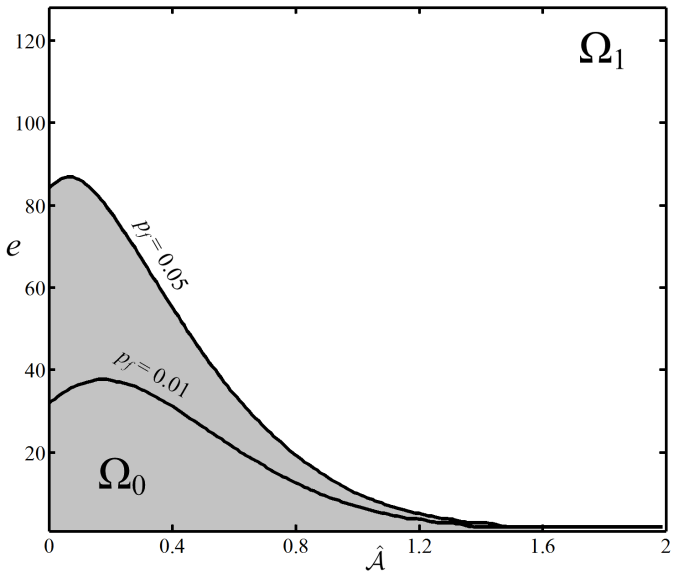

FIGURE 2.3. The decision regions for T_{MRC} when $N = 256$, $L = 2$, $\gamma_0 = 6$ dB, $\gamma_1 = 0$ dB, $\rho_0 = 0.95$, $\rho_1 = 0.85$, and $(p_f, p_d) = (0.01, 0.73)$ and $(0.05, 0.86)$.

By defining the decision regions (Ω_0 and Ω_1), and from (2.21), the probabilities of false alarm and detection in the case of MRC are given by,

$$p_f^{(\mathrm{MRC})} \triangleq \sum \int_{(k,\hat{\mathcal{A}})\in\Omega_1} p(k,\hat{\mathcal{A}}|H_0)\ d\hat{\mathcal{A}} \tag{2.24}$$

$$p_d^{(\mathrm{MRC})} \triangleq \sum \int_{(k,\hat{\mathcal{A}})\in\Omega_1} p(k,\hat{\mathcal{A}}|H_1)\ d\hat{\mathcal{A}} \tag{2.25}$$

2.3.2 Equal Gain Combining

In equal gain combining we first co-phase the signals on individual branches and then combine them with equal magnitude. Therefore in this case the combiner coefficients are given by $\hat{w}_l = e^{-j\hat{\theta}_l}$ where $\hat{\theta}_l$ is the estimated phase of the fading coefficient on branch l. Consequently, we define the decision statistic by

$$T_{\mathrm{EGC}} = \frac{p(k, f(\underline{\hat{\theta}})|H_1)}{p(k, f(\underline{\hat{\theta}})|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \mu^{(\mathrm{EGC})} \tag{2.26}$$

where $\underline{\hat{\theta}} \triangleq (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_L)$. Towards deriving the decision statistic we consider the following joint distribution.

$$p(k, \underline{\hat{\theta}}|H_\eta) = p(k|\underline{\hat{\theta}}, H_\eta)p(\underline{\hat{\theta}}|H_\eta) \tag{2.27}$$

It is well known that the estimated channel phases $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_L$ are i.i.d. and uniformly distributed over $[0, 2\pi)$, Therefore,

$$p(\underline{\hat{\theta}}|H_\eta) = \prod_{l=1}^{L} p(\hat{\theta}_l|H_\eta) = (\frac{1}{2\pi})^L, \quad 0 \le \hat{\theta}_l < 2\pi,\ l = 1, 2, \cdots, L. \tag{2.28}$$

To evaluate $p(k|\underline{\hat{\theta}}, H_\eta)$ let

$$\zeta \triangleq \frac{\mathrm{Re}\left(\sum_{l=1}^{L} h_l e^{-j\hat{\theta}_l}\right)}{\sqrt{L}}. \tag{2.29}$$

Then by conditioning on ζ, we get

$$p(k|\underline{\hat{\theta}}, H_\eta) = \int_{-\infty}^{\infty} p(k|\zeta, \underline{\hat{\theta}}, H_\eta)\ p(\zeta|\underline{\hat{\theta}}, H_\eta)\ d\zeta$$
$$= \binom{N}{k} \int_{-\infty}^{\infty} [P(E|\zeta, \underline{\hat{\theta}}, H_\eta)]^k\ [1 - P(E|\zeta, \underline{\hat{\theta}}, H_\eta)]^{N-k}\ p(\zeta|\underline{\hat{\theta}}, H_\eta)\ d\zeta \tag{2.30}$$

It is shown in [46] that,

$$P(E|\zeta, \underline{\hat{\theta}}, H_\eta) = Q(\zeta\sqrt{2\gamma_\eta}) \tag{2.31}$$

which is independent of $\underline{\hat{\theta}}$. Moreover, it is proven in Appendix 5 that given H_η, ζ is independent of $\underline{\hat{\theta}}$, i.e., $p(\zeta|\underline{\hat{\theta}}, H_\eta) = p(\zeta|H_\eta)$. From this we conclude that $p(k|\underline{\hat{\theta}}, H_\eta) = p(k|H_\eta)$. Finally from (2.27), (2.28) we get,

$$p(k, \underline{\hat{\theta}}|H_\eta) = p(k|H_\eta) \tag{2.32}$$

So, the decision statistic in the case of EGC is then given by

$$T_{\text{EGC}} = \frac{p(k, f(\underline{\hat{\theta}})|H_1)}{p(k, f(\underline{\hat{\theta}})|H_0)} = \frac{p(k|H_1)}{p(k|H_0)} \tag{2.33}$$

This shows that, in the case of EGC, the estimated phases cannot help us decide whether an increase in the REC at the SU is due to fading or the emergence of the PU signal. In light of (2.28), this result in fact makes intuitive sense. We conclude that in the presence of fading, EGC diversity technique is not a good option for spectrum monitoring in a fading channel. This is also demonstrated by the simulation results in Section 2.4.

2.3.3 Selection Combining

In selection combining, all the weighting coefficients are zero except for the branch with the highest SNR for which the coefficient is one. Therefore in this case the

weighting coefficients do not provide any information about the emergence of the primary user. In other words,

$$p(\{\hat{w}_l\}_{l=1}^{L} \mid H_0) = p(\{\hat{w}_l\}_{l=1}^{L} \mid H_1). \tag{2.34}$$

Therefore in this case

$$T_{\mathrm{SC}} = \frac{p(k, f(\{\hat{w}_l\}_{l=1}^{L}) \mid H_1)}{p(k, f(\{\hat{w}_l\}_{l=1}^{L}) \mid H_0)} = \frac{p(k \mid H_1)}{p(k \mid H_0)} \tag{2.35}$$

Similar to EGC, the CSI from SC combining method does not enhance the performance of spectrum monitoring over the REC alone.

Remark 2. *We need to discuss the complexity associated with the proposed spectrum monitoring method. As pointed out in [40], receiver statistics such as REC are useful in adaptive transmission protocols where modulation, coding or transmit power may be adjusted in order to mitigate the effects of time-varying channel and interference. If REC is already being collected by the receiver, then no significant additional hardware is required by the proposed method. If not, then the receiver is required to implement the CRC check, the FEC encoder and the hypothesis testing as shown in Fig. 2.1. The hardware and computational complexity of CRC and the FEC encoder is not very high particularly in comparison with the complexity of the rest of the SU receiver including multiple RF chains for diversity combining, the demodulator and the decoder. Hypothesis testing requires the computation of T in* (2.11) *for which $p(k, f(\hat{\boldsymbol{w}}) \mid H_\eta)$, $\eta = 0, 1$, is needed. In the case of MRC, we evaluate an accurate approximation for $p(k, f(\hat{\boldsymbol{w}}) \mid H_\eta)$ in Appendix 5 which alleviates the need for computation of integrals. In summary the incremental complexity of the proposed method for the SU receiver is not significant.*

2.4 Numerical Results

In this section we provide performance results from simulation and analysis to asses the effectiveness of the proposed spectrum monitoring methods. We should

point out that our goal here is to demonstrate the advantage of a hybrid spectrum sensing/spectrum monitoring system over a system that uses spectrum sensing alone. Therefore we are not concerned with the specific spectrum sensing method that is being used and only need to assume its probabilities of detection and false alarm which, subsequently, are denoted by $\hat{p}_d$ and $\hat{p}_f$, respectively.

Simulation results are obtained by running at least 10^4 independent trials, and analytical results of the proposed spectrum monitoring for MRC is obtained using the approximation in Appendix 5. The length of the spectrum sensing interval is identical to the length of a packet, and the transmitter uses BPSK modulation with rate 2 Mbps. In the simulations Jakes' model [59] with the sum of sinusoids is used to model a flat Rayleigh fading channel. In particular, we use 16 sinusoids for the Jakes' model with the maximum Doppler frequency of 90 Hz (corresponding to a mobile speed of 54 Km/h and the carrier frequency of 1.8 GHz).

Fig. 2.4 compares the ROC curve of the proposed decision statistic for MRC (T_{MRC}) to three other cases which use diversity combining but make their SM decisions based only on the REC alone. These three cases are MRC, EGC and SC denoted by $T_{\mathrm{MRC}}^{(\mathrm{REC})}$, $T_{\mathrm{EGC}}^{(\mathrm{REC})}$, and $T_{\mathrm{SC}}^{(\mathrm{REC})}$, respectively. A remark is in order here. Although the latter three cases use only REC to detect the presence of the PU signal, their performance is not identical as seen in Fig. 2.4. This is due to the fact that since the combining techniques are different, the REC's (under each hypothesis) are also different in these three cases.

As expected, T_{MRC} outperforms the other three test statistics. For example for probability of false alarm $p_f = 0.1$, the probability of detection for T_{MRC} is .97 whereas it is below .62 in the other cases. Due to the fact that the performance of $T_{\mathrm{MRC}}^{(\mathrm{REC})}$, $T_{\mathrm{EGC}}^{(\mathrm{REC})}$ and $T_{\mathrm{SC}}^{(\mathrm{REC})}$ are close to each other (see Figs. 2.2 and 2.4), in the following we only consider the decision statistic $T_{\mathrm{MRC}}^{(\mathrm{REC})}$ for our comparisons.

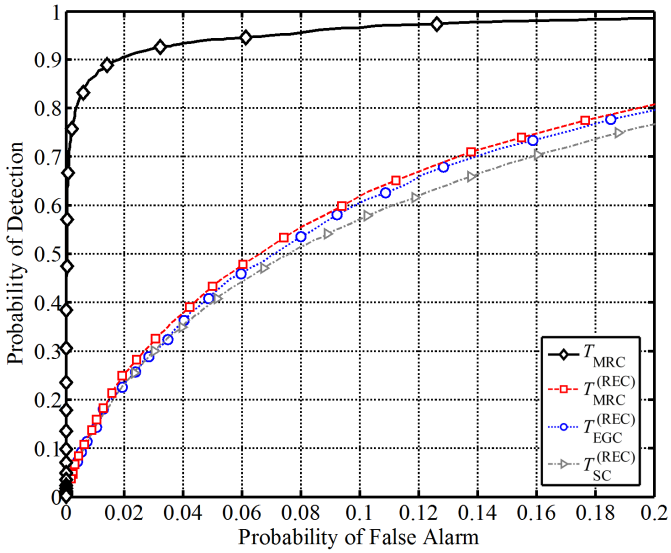

FIGURE 2.4. Comparison from simulations between T_{MRC}, $T_{\mathrm{MRC}}^{(\mathrm{REC})}$, $T_{\mathrm{EGC}}^{(\mathrm{REC})}$, and $T_{\mathrm{SC}}^{(\mathrm{REC})}$ for $N = 1024$, $\gamma_0 = 2$ dB, $\gamma_1 = -2$ dB, $\rho_0 = 0.95$, $\rho_1 = 0.85$, and $L = 4$.

Channel utilization for the SUs and detection delay of the PUs can be used to evaluate the efficacy of the hybrid spectrum sensing/spectrum monitoring systems. As described previously, channel utilization is defined as the average fraction of time that under hypothesis H_0 the SU communicates over the channel. Detection delay is the average time it takes to detect the presence of the primary user after it emerges in the channel. Using Markov chain models in [41] we evaluated channel utilization and detection delay for a hybrid spectrum sensing/spectrum monitoring technique for AWGN channel. For AWGN channels, the event of observing k errors is independent from packet to packet, and so Markov models can be employed. In contrast, in the case of fading channels, the fading coefficients affecting consecutive packets are correlated (in time). As a result the decision statistics are also correlated and the Markov model is not applicable. We have not been able to obtain closed form formulas for channel utilization and detection delay in the case of fading channels. The following results are obtained from extensive simulations.

Figs. 2.5 and 2.6 show channel utilization and detection delay of T_{MRC} and $T_{\mathrm{MRC}}^{(\mathrm{REC})}$ versus the detection probability p_d of spectrum monitoring for packet length

$N = 1024$, number of antennas $L = 2$, $\gamma_0 = 2$ dB, $\gamma_1 = -2$ dB, $\rho_0 = 0.95$, $\rho_1 = 0.85$, and the probabilities of false alarm and detection for the spectrum sensing method $(\hat{p}_f, \hat{p}_d) = (.1, .9)$, for different values of K_M.

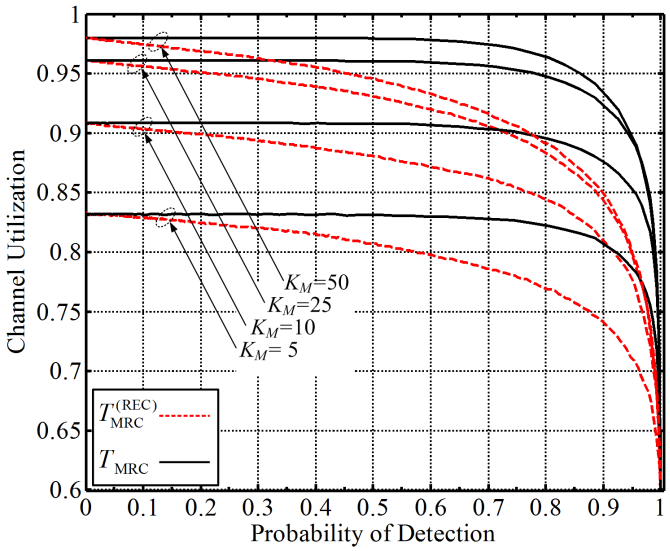

FIGURE 2.5. Channel utilization versus the probability of detection for T_{MRC} and $T_{\mathrm{MRC}}^{(\mathrm{REC})}$, $N = 1024$, $L = 2$, $\gamma_0 = 2$ dB, $\gamma_1 = -2$ dB, $\rho_0 = 0.95$, $\rho_1 = 0.85$, $f_{\mathrm{m}} = 90$Hz, $(\hat{p}_f, \hat{p}_d) = (0.1, 0.9)$ and K_M= 5, 10, 25, and 50.

Channel utilization is a decreasing function of false alarm probability p_f owing to the fact that the portion of time that the SU has a chance to access the channel decreases with p_f. Since p_d is an increasing function of p_f, channel utilization is also a decreasing function of p_d. Channel utilization increases with the duration of the spectrum monitoring interval (K_M). This is due to the fact that, for a fixed spectrum sensing interval, as K_M increases, the fraction of time that the SU is able to transmit also increases resulting in increased throughput for the SU. However, increasing K_M will also increase detection delay. The reason is that for equal probabilities of false alarm for spectrum monitoring and spectrum sensing, spectrum sensing has a higher probability of detection. Since increasing K_M (for fixed spectrum sensing intervals) reduces the fraction of time the SU spends in spectrum sensing, detection delay increases with K_M.

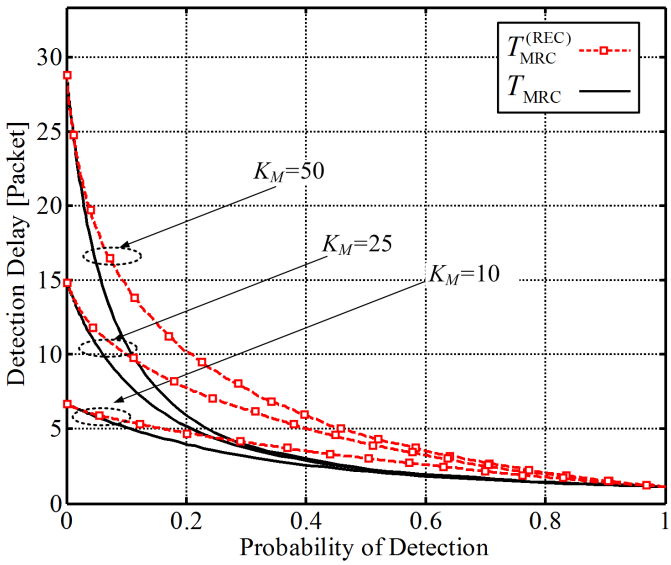

FIGURE 2.6. Detection delay versus the probability of detection for T_{MRC} and $T_{\text{MRC}}^{(\text{REC})}$, $N = 1024$, $L = 2$, $\gamma_0 = 2$ dB, $\gamma_1 = -2$ dB, $\rho_0 = 0.95$, $\rho_1 = 0.85$, $(\hat{p}_f, \hat{p}_d) = (0.1, 0.9)$ and $f_{\text{m}} = 90$Hz.

Figs. 2.5 and 2.6 also show that T_{MRC} significantly outperforms $T_{\text{MRC}}^{(\text{REC})}$ for the same value of p_d. One should note that, to obtain the same value of p_d, $T_{\text{MRC}}^{(\text{REC})}$ requires significantly higher SNR than T_{MRC} as evident from the ROC curves in Fig. 2.4.

In Fig. 2.7 we plot channel utilization versus detection delay p_d for the hybrid spectrum sensing/spectrum monitoring techniques using T_{MRC} and $T_{\text{MRC}}^{(\text{REC})}$ for different values of K_M and L. The performance of the spectrum sensing alone is also shown. It can be seen that the hybrid technique significantly outperform spectrum sensing alone. As illustrated by this figure, for any given channel utilization and fixed K_M, spectrum sensing is equivalent to the hybrid system with $p_d = p_f = 0$, and has the maximum detection delay. Moreover, the decision statistic T_{MRC}, outperforms $T_{\text{MRC}}^{(\text{REC})}$. For example for $K_M = 25$, channel utilization of 95% can be achieved by T_{MRC} and $T_{\text{MRC}}^{(\text{REC})}$ resulting in detection delays of 1.5 and 8.2 packets, respectively.

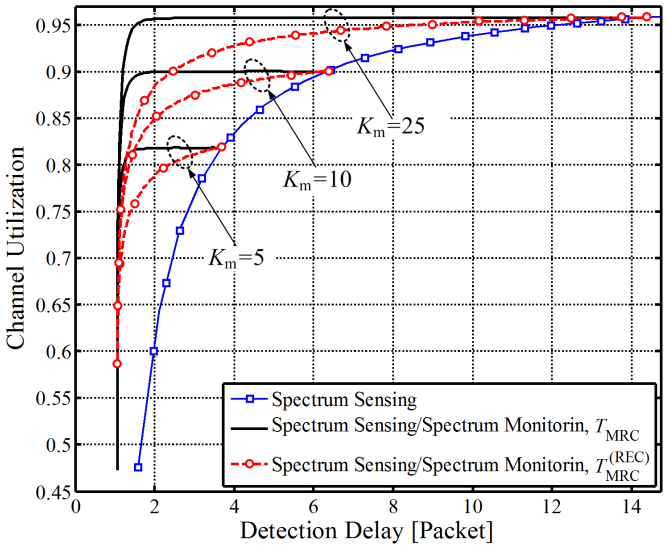

FIGURE 2.7. Channel utilization versus detection delay for T_{MRC} and $T_{\text{MRC}}^{(\text{REC})}$, $N = 1024$, $L = 2$, $\gamma_0 = 2$ dB, $\gamma_1 = -2$ dB, $\rho_0 = 0.95$, $\rho_1 = 0.85$, $(\hat{p}_f, \hat{p}_d) = (0.1, 0.9)$ and $f_{\text{m}} = 90$Hz .

In Fig. 2.8 we show the ROC curves for T_{MRC} for different number of branches, L. The simulation results are obtained for a SU which uses one of two error-correcting codes. The first is a rate 1/2 convolutional code with the generator matrix $[g^{(0)} = (716502)_8\ ;\ g^{(1)} = (514576)_8]$, [60]. The second code is a $(1023, 503)$ binary BCH code with rate $503/1023 \approx 1/2$. We also employed the CRC-8 code with the generator polynomial $x^8 + x^7 + x^6 + x^4 + x^2 + 1$. As the plots illustrate the simulation results using actual coding schemes closely match the results from analysis. Note that for $L = 1$, there is no diversity and $\hat{\mathcal{A}} = |\hat{h}|$. As L increases to 2, the performance improves. However, for $L = 3$ the performance starts to degrade and for $L = 10$ the ROC is close to the chance line, i.e. $p_f = p_d$. This behavior is due to the fact that as L increases the SNR at the output of the combiner improves and the REC is reduced. For very large values of L the emergence of the PU does not cause a significant change in the SNR or the REC. Therefore in such cases it is difficult for the decision statistic to detect the emergence of the PU. In Appendix 5 we present a method for the judicious selection of the number of

antennas L. For the parameters in Fig. 2.8, (5.39) results in $\tilde{L} = 2.56$ from which we get $L_{\text{opt}} = \left\lfloor \tilde{L} \right\rfloor = 2$. This matches the result in Fig. 2.8.

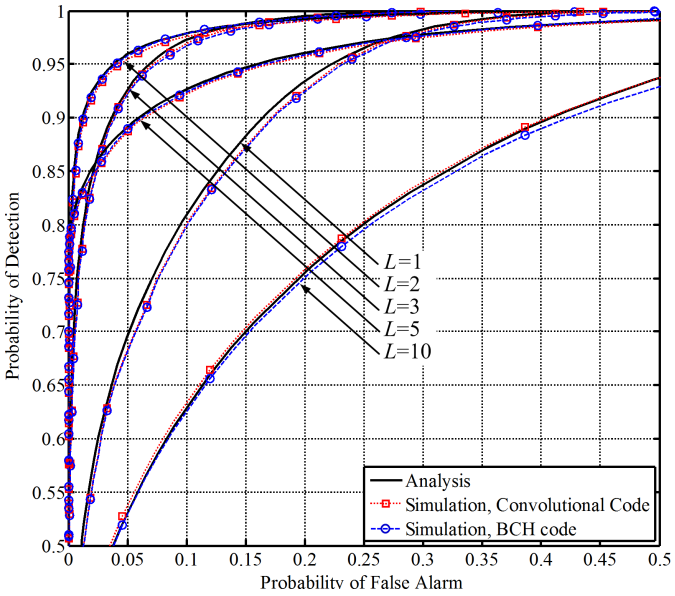

FIGURE 2.8. ROC for the proposed decision statistic T_{MRC} for $N = 1024$, $\gamma_0 = 6$ dB, $\gamma_1 = 0$ dB, $\rho_0 = 0.9$ and $\rho_1 = 0.8$ for $L = 1, 2, 3, 5, 10$.

Fig. 2.9 shows detection delay versus channel utilization for different number of branches and K_M when $N = 256$, $\gamma_0 = 4$ dB, $\gamma_1 = -1$ dB, $\rho_0 = 0.9$ and $\rho_1 = 0.8$. This figure also shows that the performance improves from $L = 1$ to $L = 4$ but it degrades as L increases to 10. From (5.39) we get $\tilde{L} = 3.2$ and $L_{\text{opt}} = \left\lceil \tilde{L} \right\rceil = 4$.

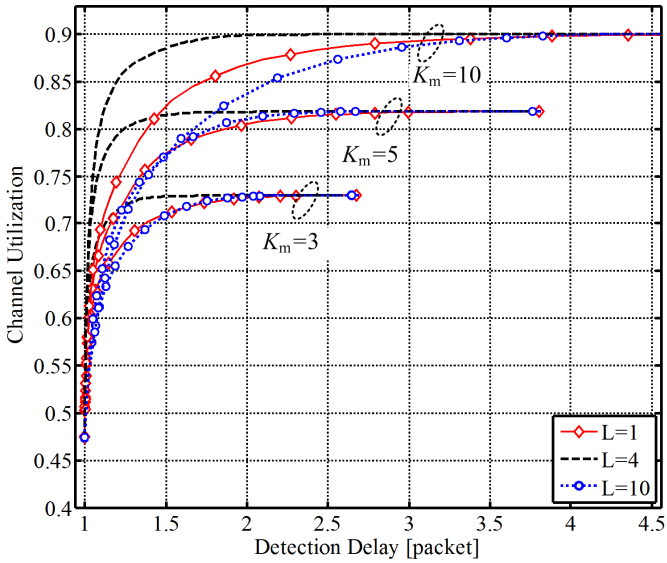

FIGURE 2.9. Channel utilization versus detection delay for T_{MRC}, $N = 256$, $\gamma_0 = 4$ dB, $\gamma_1 = -1$ dB, $\rho_0 = 0.9$, $\rho_1 = 0.8$, $K_M = 5, 10$ and $L = 1, 4, 10$.

Chapter 3

Fast Detection of Malicious Behavior in Cooperative Spectrum Sensing

3.1 Introduction

Dynamic spectrum access (DSA) is a new paradigm in spectrum sharing whereby unlicensed secondary users (SUs) equipped with cognitive radios can utilize a licensed frequency band when it is vacant of the primary users (PUs). SUs must, however, quickly vacate the channel once the primary users emerge [1, 2, 4, 61]. Cognitive radios (CRs) rely on spectrum sensing (SS) to detect the presence or emergence of the PU signal.

Reliable SS is difficult to accomplish. Stringent requirements are imposed on the probability of detection and detection delay in order to ensure that SUs do not cause undue interference to the PUs. In addition, low false alarm probability is required in order to increase channel utilization when the PU is not active [43]. The challenge is that these requirements must be ensured in scenarios where the PU signal is often very weak (at times well below the noise floor) with unknown parameters (e.g., signal power, pulse shape, modulation type, etc), and in the presence of unknown noise power and channel impairments. Several blind SS techniques proposed in recent years rely on multiple antennas to achieve acceptable performance in the face of unknown signal, noise and channel conditions [6, 62, 63, 9, 42].

Cooperative SS is an alternative approach to achieving spatial diversity in order to combat the channel fading effects and to overcome the hidden terminal problem. Here multiple receivers attempt to detect the PU signal and report their decisions to a fusion center (FC) which makes the final decision on the presence of the PU [4, 61, 64, 65]. Cooperative SS provides significant performance improvement over

single-user techniques as well as lower complexity for the individual SU terminals at the cost of a small increase in communication between the SU terminals and the FC. Advantages of cooperative sensing have been acknowledged by Federal Communication Commission's recent rules for unlicensed operation in licensed TV bands requiring that the SUs share their information on channel occupancy obtained through sensing [66].

Unfortunately cooperative SS is vulnerable to misbehavior by the individual radios [67]. Malicious users may send false sensing information in order to gain unfair access to the channel or to disrupt the spectrum sensing process. False information may also be conveyed to the FC due to the malfunctioning of a user's terminal. In [68], it is shown that the performance of cooperative SS can be severely degraded in the presence of unknown malicious radios. Several methods have been recently proposed to counter such malicious attacks. An outlier-based detection mechanism is proposed in [69] where it is assumed that energy detection is used at the CRs who transmit their measurements (without quantization) to the FC. The algorithm is based on the detection of the radios whose transmitted data is far from the underlying distribution. This work is further extended in [70]. In [71], the authors also consider energy detection at the SUs and assume data fusion in that the (unquantized) received signal strengths (RSSs) are transmitted to the FC. The FC combines the RSSs to decide on the presence or absence of the PU. The misbehaving CRs may transmit strong or weak RSS values in order to confuse the FC. The authors have developed an outlier detection scheme in order to eliminate the deleterious effects of the misbehaving CRs as follows. The CRs are grouped into clusters so that all the radios in a cluster experience correlated shadow-fading effects [72]. This correlation is then exploited to detect and filter out the abnormal sensing reports using a weighted gain combining (WGC) fusion

method. While WGC reduces the effects of outlier attacks on false alarm and misdetection probabilities, it cannot eliminate the effects of weak attacks which cannot be differentiated from normal sensing reports. The authors have then developed a sequential hypothesis testing scheme which along with WGC allows for effective detection of outlying reports from misbehaving CRs.

Attack prevention in collaborative SS is also considered in [73] where the SUs report their binary decisions on the presence or absence of the PU to the FC which uses the 1-out-of-m (OR) fusion rule [74]. A single class of malicious radios is considered along with an aggressive attack strategy where the malicious CRs are able to hear the reports of the honest SUs and can cooperate in order to maximize their spectrum utilization. The key idea in [73] is that when the FC declares the channel to be busy, then no honest SU will subsequently transmit. Therefore, if a transmission occurs after the channel is declared to be busy, the FC can detect this transmission and impose a cost (punishment) on all the SUs. It is shown that if this cost is chosen to be large enough, the reward for the malicious CRs will be negative thereby discouraging them from such attacks. However, as noted in [73], for most practical systems this "direct punishment" scheme is difficult to implement. An indirect punishment mechanism is also developed in [73], where, upon the detection of malicious behavior, the FC declares an end to collaborative sensing and the SUs will no longer transmit their decisions to the FC. Indirect punishment is effective assuming that the attackers care about their future rewards, the so-called "stay-with-attack" strategy [73]. As noted in [75], this approach assumes a perfect control channel between the FC and the SUs, where the messages from the FC are always correctly received. Since indirect punishment can also be triggered due to control channel errors rather than attacks, a more practical implementation of this technique needs to be developed [75].

Reputation-based detection is another widely used approach [76, 77, 78, 79, 80, 81, 82]. Here each CR sends its binary decision regarding the presence of the PU to the FC which makes the final decision using a fusion rule. At the same time the FC constructs a reputation metric for each CR based on the similarity of CR's decisions to the final decisions of the FC. The reputation metric is updated over time and compared with a threshold in order to identify the radios. In [77] a weighted sequential probability ratio test (WSPRT) is proposed where the weights are derived from a radio's reputation metric. In [76, 78, 80], the detection of the hypothesis and the malicious radios is done in two steps. First based on the q-out-of-m rule, [74], the current hypothesis is detected and next, a CR is identified as malicious if its past decisions over a certain time period differ from the decisions of the FC by more than a threshold. In [79], it is assumed that the FC knows the probabilities of detection and false alarm of the honest CRs as well as the prior probability of each hypothesis. With these assumptions, the FC is aware of the expected behavior of an honest CR and the joint probability of decisions from two CRs. These parameters are estimated from the CRs' transmitted decisions and their deviation from the expected values is used to classify the radios.

In [82] the authors define a reputation metric based on two types of attackers: type-1, which report the channel to be busy when it is detected to be free, and type-0 which report the channel to be free when it is detected to be busy. Note that this model does not include malicious radios that attempt to confuse the FC under both hypotheses. In addition it is assumed that the subset of honest CRs is known to the FC.

A drawback of the reputation-based method is that the detection of the hypotheses is separated from the detection of the malicious CRs. This separation results in a loss of performance for this method. In addition, as shown in our numerical re-

sults in Section 3.4, for the reputation metric to be reliable, the FC must assemble a large number of decisions from each CR, and in fact the algorithm fails when the number of decisions is small. Finally, as explained in Section 3.4, in the absence of any prior information on the parameters of the network, the reputation-based method may fail even in cases where the honest radios are in majority.

In this chapter, we consider a network with several classes of CRs where all the radios in a class have the same detection and false alarm probabilities. In particular, there may be more than one class of misbehaving radios due to the presence of malicious as well as malfunctioning radios in the network. Moreover, there may be more than one class of honest CRs. This may arise when different SUs employ different SS techniques, or when, due to their geographic position with respect to the PU, they experience different paths loss and fading from the PU, resulting in different detection and false alarm probabilities. No parameters of the network are assumed to be known except that there is a class of honest CRs with more radios than any other class. We present a method for detecting the hypotheses and classifying the CRs based on the expectation maximization algorithm. In this approach we also compute the detection and false alarm probabilities of each class. Our results show that the proposed algorithm significantly outperforms the reputation-based method and even in cases where the reputation-based method fails, our proposed method is able to properly classify the CRs and detect the hypotheses.

3.2 System model and Notations

We consider a cognitive radio network (CRN) of L independent radios monitoring a spectral band in order to detect the presence (hypothesis $\mathcal{H}_1$) or absence (hypothesis $\mathcal{H}_0$) of a PU. We assume that there are K classes of CRs $c_1, c_2, \ldots c_K$,

where c_1 denotes the class of honest radios and the remaining classes represent honest, malicious or otherwise misbehaving radios.

At time $t \in \{1, 2, \cdots, T\}$ radio l makes a binary decision $r_{lt} \in \{0, 1\}$ regarding the presence of the PU where $r_{lt} = 0$ if $\mathcal{H}_0$ and $r_{lt} = 1$, otherwise. The hypothesis matrix H is defined by

$$H \triangleq \begin{bmatrix} h_{01} & h_{02} & \cdots & h_{0t} & \cdots & h_{0T} \\ h_{11} & h_{12} & \cdots & h_{1t} & \cdots & h_{1T} \end{bmatrix}_{2\times T} \tag{3.1}$$

where column t represents the state of the hypothesis at time t, $t = 1, 2, \cdots, T$. At each time t, one of the element in column t is 1 and the other is 0. If $h_{0t} = 0$, then $h_{1t} = 1$, indicating that at time t we have hypothesis $\mathcal{H}_1$. Similarly, if $h_{0t} = 1$ then $h_{1t} = 0$, indicating that at time t we have hypothesis $\mathcal{H}_0$.

At this point it is assumed that given H, the radios' decisions $\{r_{lt}, l = 1, 2, \cdots, L, t = 1, 2, \cdots, T\}$ are independent[1]. In Section 3.4.2 we discuss the implications of correlated decisions. The probabilities of detection and false alarm denoted by $\tilde{p}_{1k}$ and $\tilde{p}_{0k}$, respectively, for class c_k are given by

$$\tilde{p}_{\eta k} = Pr(r_{lt} = 1 | h_{1t} = \eta, l \in c_k), \quad \eta = 0, 1 \tag{3.2}$$

At time t, radio l transmits a single bit $d_{lt} \in \{0, 1\}$ to the FC. While for the honest radios we have $d_{lt} = r_{lt}$, radios in other classes may alter their decision before transmission to the FC. For $\eta = 0, 1$ let $\rho_\eta(k) \triangleq Pr(d_{lt} = 1 | r_{lt} = \eta, l \in c_k)$. Note that for the honest radios, $\rho_0(1) = 0$ and $\rho_1(1) = 1$. On the other hand, for a malicious radio in class, say ι, which flips its decisions before transmission, $\rho_0(\iota) = 1$ and $\rho_1(\iota) = 0$.

[1] As in [76, 80, 82] we assume that during the observation period the channel conditions (e.g., fading and shadowing) are constant. This assumption is then justified by the fact that, given H, the received signal samples at different radios are independent over time and from one radio to another.

The probabilities of detection and false alarm "perceived" by the FC for a radio in class c_k, and denoted by p_{1k} and p_{0k}, respectively, can be written as

$$p_{\eta k} \triangleq Pr(d_{lt} = 1 | h_{1t} = \eta, l \in c_k) \tag{3.3}$$
$$\rho_1(k)\tilde{p}_\eta(k) + \rho_0(k)(1 - \tilde{p}_\eta(k)), \quad \eta = 0, 1$$

It is assumed that for the malicious radios the values $\rho_0(k)$ and $\rho_1(k)$ do not change during the observation interval. As noted in [82], frequent changes of attack probabilities will not be meaningful even from the point of view of the adversary since it cannot produce predictable deleterious effects on the FC.

The FC collects T transmissions from each radio to form a decision matrix $D = [d_{lt}],\ l = 1, 2, \ldots, L,\ t = 1, 2, \ldots, T$ from which it tries to classify the radios and detect the hypotheses matrix H. In this process the FC will also determine the probabilities of detection and false alarm (p_{1k}, p_{0k}) for each class k. Hereafter we refer to the the pair (p_{1k}, p_{0k}) as the *operating point* of the radios in class c_k, and let $P \triangleq [p_{ik}]\ i = 0, 1,\ k = 1, 2, \ldots, K$.

Remark 3. *We have assumed an error free channel between the CRs and the FC. However, in the case of a noisy channel, the effect of a binary memoryless channel can be absorbed in the probabilities of detection and false alarm in* (3.3) *by adjusting the parameters $\rho_\eta(k)$. The only restriction is that the channels between all the radios in class c_k and the FC have the same transition probabilities.*

Let $z_{l,k} = 1$ if radio $l \in c_k$, and zero, otherwise and define the matrix $Z = [z_{lk}],\ l = 1, 2, \ldots, L,\ k = 1, 2, \ldots, K$ as the *radio identification matrix*. Note that row i corresponds to the ith radio and column j corresponds to class c_j. Thus if radio l belongs to class j, then $z_{l,j} = 1$ and $z_{l,k} = 0$ for all $k \neq j$. We assign a probability π_k for class c_k and define $\Pi = [\pi_k],\ k = 1, 2, \ldots, K$, where $\pi_k = Pr(z_{lk} = 1)$ for any $l \in \{1, 2, \ldots, L\}$. In order to formulate our detection problem

we allocate prior probabilities for the hypotheses defined by $\Phi = [\phi_{it}],\ i = 0, 1,\ t = 1, 2, \ldots, T$ where $\phi_{it} = Pr(h_{it} = 1)$ and $\phi_{0t} + \phi_{1t} = 1$. Finally the three-tuple $\Theta \triangleq \{P, \Pi, \Phi\}$ is defined as the parameter set. In the next section we propose an algorithm to estimate the parameter set from the received decision matrix which is then used in classifying the radios.

Our goal is to classify the radios and detect the hypothesis at each time $t = 1, 2, \cdots, T$. In other words we would like to detect the matrices Z and H from the decision matrix D. If the parameter set Θ were known, the maximum likelihood detection rule for (H, Z) is given by

$$(\hat{H}, \hat{Z}) = \arg\max_{H,Z} Pr(D|Z, H, \Theta) \tag{3.4}$$

However, since the parameter set is unknown, it must be first estimated from the received decision matrix D. A maximum likelihood estimate of Θ is given by $\tilde{\Theta} = \arg\max_{\Theta} Pr(D|\Theta)$. On the other hand, $Pr(D|\Theta)$ is not directly available and must be evaluated from

$$Pr(D|\Theta) = \sum_{H,Z} Pr(D, Z, H|\Theta). \tag{3.5}$$

Due to the complexity of the mixture model in (3.5), the estimate of Θ cannot be obtained in closed form. Therefore we employ the iterative EM algorithm [83] to estimate Θ with H and Z as latent variables. Each iteration of the EM algorithm is guaranteed to increase the likelihood function and the algorithm is guaranteed to converge to a (local) maximum of the likelihood function [84]. Once an estimate of Θ is obtained, H and Z can be obtained from the maximum likelihood rule in (3.4). Below in Section 3.3, we present the EM estimation of Θ and detection of Z and H.

A remark is in order here. Due to the presence of unknown parameters such as the radio identification matrix and the operating points of each class, the hypoth-

esis testing problem cannot be formulated with a Bayesian or Neyman-Pearson criterion. In particular the prior distribution matrix Φ is used only as an artifact in our detection problem and is not assumed to be the true prior of the hypotheses.

3.3 Parameter Estimation, Classification, and Hypotheses Testing

$$\begin{aligned} Pr(D, Z, H|\Theta) &= Pr(D|Z, H; \Theta) Pr(Z, H|\Theta) \\ &= \prod_{l=1}^{L} \prod_{k=1}^{K} \left[\prod_{t=1}^{T} \prod_{i=0}^{1} \pi_k^{\frac{1}{2T}} \left(p_{ik}^{d_{lt}} \left(1 - p_{ik}\right)^{(1-d_{lt})} \phi_{it}^{\frac{1}{L}} \right)^{h_{it}} \right]^{z_{lk}} \end{aligned} \tag{3.6}$$

$$\begin{aligned} L(\Theta; D, Z, H) &= \log Pr(D, Z, H|\Theta) \\ &\sum_{l=1}^{L} \sum_{k=1}^{K} z_{lk} \sum_{t=1}^{T} \sum_{i=0}^{1} \left\{ \frac{\log \pi_k}{2T} + h_{it} \left[d_{lt} \log p_{ik} + (1 - d_{lt}) \log \left(1 - p_{ik}\right) + \frac{1}{L} \log \phi_{it} \right] \right\} \end{aligned} \tag{3.7}$$

$$\begin{aligned} &Q(\Theta; \Theta^{\text{old}}) \triangleq E_{(Z,H)|D;\Theta^{\text{old}}}[L(\Theta; D, Z, H)] \\ &= \sum_{l=1}^{L} \sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{i=0}^{1} \Bigg\{ E_{(Z,H)|D;\Theta^{\text{old}}}[z_{lk}] \frac{1}{2T} \log \pi_k \\ &\quad + E_{(Z,H)|D;\Theta^{\text{old}}}[z_{lk} h_{it}] \left[d_{lt} \log p_{ik} + (1 - d_{lt}) \log \left(1 - p_{ik}\right) + \frac{1}{L} \log \phi_{it} \right] \Bigg\} \end{aligned} \tag{3.8}$$

3.3.1 Parameter Estimation

To estimate Θ from $\arg\max_{\Theta} Pr(D|\Theta) = \arg\max_{\Theta} \sum_{H,Z} Pr(D, Z, H|\Theta)$ using EM, we first need to evaluate $Pr(D, Z, H|\Theta)$ which is given in (3.6) and from which the log-likelihood function, denoted by $L(\Theta; D, Z, H)$, is obtained in (3.7). Note that in (3.6), in addition to Z, H is also considered as a latent variable which effectively makes the final process of detecting H easier and faster without the need for evaluating the likelihood function for all possible hypotheses.

An iteration of the EM algorithm involves the following two steps [84]. For a discussion of the convergence properties of the EM algorithm we refer to [83, 84].

1. *Expectation step:* In the first step the expectation of the log-likelihood function, denoted by $Q(\Theta;\Theta^{\text{old}})$ in (3.7), is evaluated with respect to the conditional distribution $P(Z,H|D,\Theta^{\text{old}})$ of the latent variables (Z,H), where Θ^{old} is the previous estimate for Θ. This is shown in (3.8).

2. *Maximization step:* In the second step $Q(\Theta;\Theta^{\text{old}})$ is maximized with respect to Θ.

To perform the expectation step, let $\beta(l,k) \triangleq E[z_{lk}|D;\Theta^{\text{old}}]$ and $\alpha(l,k,i,t) \triangleq E[z_{lk}h_{it}|D;\Theta^{\text{old}}]$. Then,

$$\begin{aligned}\beta(l,k) &= Pr(z_{lk}=1|D;\Theta^{\text{old}}) \qquad (3.9)\\ &= \frac{\pi_k^{\text{old}}\prod_{t=1}^{T} Pr\left(d_{lt}|z_{lk}=1;\Theta^{\text{old}}\right)}{\sum_{j=1}^{K}\pi_j^{\text{old}}\prod_{t=1}^{T} Pr\left(d_{lt}|z_{lj}=1;\Theta^{\text{old}}\right)}\\ &= \frac{\pi_k^{\text{old}}\prod_{t=1}^{T}\sum_{i=0}^{1}\phi_{it}^{\text{old}}\left((p_{ik}^{\text{old}})^{d_{lt}}\left(1-p_{ik}^{\text{old}}\right)^{(1-d_{lt})}\right)}{\sum_{k'=1}^{K}\pi_{k'}^{\text{old}}\prod_{t=1}^{T}\sum_{i=0}^{1}\phi_{it}^{\text{old}}\left((p_{ik'}^{\text{old}})^{d_{lt}}\left(1-p_{ik'}^{\text{old}}\right)^{(1-d_{lt})}\right)}.\end{aligned}$$

Moreover,

$$\begin{aligned}\alpha(l,k,i,t) &= Pr(z_{lk}=1,h_{it}=1|D;\Theta^{\text{old}}) \qquad (3.10)\\ &= Pr(h_{it}=1|z_{lk}=1,D;\Theta^{\text{old}})\times Pr(z_{lk}=1|D;\Theta^{\text{old}})\end{aligned}$$

where $Pr(h_{it}=1|z_{lk}=1,D;\Theta^{\text{old}})$ is derived in (3.11) in which $\mathbf{d}_t=[d_{1t},d_{2t},\cdots,d_{Lt}]^{\text{tr}}$ denotes the tth column of D, i.e., the decision vector received from all the radios at time t. Therefore from (3.8)-(3.10) and (3.11) $Q(\Theta;\Theta^{\text{old}})$ can be obtained as in (3.12).

$$Pr(h_{it}=1|z_{lk}=1,D;\Theta^{\text{old}}) = Pr(h_{it}=1|z_{lk}=1,\mathbf{d}_t;\Theta^{\text{old}}) \tag{3.11}$$
$$= \frac{Pr(\mathbf{d}_t|h_{it}=1,z_{lk}=1;\Theta^{\text{old}})Pr(h_{it}=1|z_{lk}=1;\Theta^{\text{old}})}{\sum_{j=0}^{1} Pr(\mathbf{d}_t|h_{jt}=1,z_{lk}=1;\Theta^{\text{old}})Pr(h_{jt}=1|z_{lk}=1;\Theta^{\text{old}})}$$
$$= \frac{Pr(d_{lt}|h_{it}=1,z_{lk}=1;\Theta^{\text{old}})\prod_{l'\neq l} Pr(d_{l't}|h_{it}=1;\Theta^{\text{old}}) \times \phi_{it}^{\text{old}}}{\sum_{j=0}^{1}\left\{Pr(d_{lt}|h_{jt}=1,z_{lk}=1;\Theta^{\text{old}})\prod_{l'\neq l} Pr(d_{l't}|h_{jt}=1;\Theta^{\text{old}}) \times \phi_{jt}^{\text{old}}\right\}}$$
$$= \frac{\phi_{it}^{\text{old}}\left((p_{ik}^{\text{old}})^{d_{lt}}\left(1-p_{ik}^{\text{old}}\right)^{(1-d_{lt})}\right)\prod_{l'\neq l}\left[\sum_{k'=1}^{K}\left((p_{ik'}^{\text{old}})^{d_{l't}}\left(1-p_{ik'}^{\text{old}}\right)^{(1-d_{l't})}\right)\pi_{k'}^{\text{old}}\right]}{\sum_{j=0}^{1}\left\{\phi_{jt}^{\text{old}}\left((p_{jk}^{\text{old}})^{d_{lt}}\left(1-p_{jk}^{\text{old}}\right)^{(1-d_{lt})}\right)\prod_{l'\neq l}\left[\sum_{k'=1}^{K}\left((p_{jk'}^{\text{old}})^{d_{l't}}\left(1-p_{jk'}^{\text{old}}\right)^{(1-d_{l't})}\right)\pi_{k'}^{\text{old}}\right]\right\}}$$

$$Q(\Theta;\Theta^{\text{old}}) = \sum_{l=1}^{L}\sum_{k=1}^{K}\beta(l,k)\log\pi_k + \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=0}^{1}\alpha(l,k,i,t)\log\phi_{it} \tag{3.12}$$
$$+\sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=0}^{1}\alpha(l,k,i,t)\left[d_{lt}\log p_{ik} + (1-d_{lt})\log(1-p_{ik})\right]$$

To perform the maximization step of the EM algorithm we need to maximize $Q(\Theta;\Theta^{\text{old}})$ with respect to the parameters Θ. Maximization with respect to the operating points is achieved from

$$\frac{\partial Q}{\partial p_{ik}} = \sum_{l=1}^{L}\sum_{t=1}^{T}\alpha(l,k,i,t)\left(\frac{d_{lt}}{p_{ik}} - \frac{(1-d_{lt})}{(1-p_{ik})}\right) = 0 \tag{3.13}$$

which results in

$$p_{ik}^{\text{new}} = \frac{\sum_{l=1}^{L}\sum_{t=1}^{T}\alpha(l,k,i,t)d_{lt}}{\sum_{l=1}^{L}\sum_{t=1}^{T}\alpha(l,k,i,t)} \tag{3.14}$$

To maximize $Q(\Theta;\Theta^{\text{old}})$ with respect to π_k we must also satisfy the constraint $\sum_{k=1}^{K}\pi_k = 1$. To this end we from the Lagrangian $\breve{Q}$ given by

$$\breve{Q}(\Theta,\lambda;\Theta^{\text{old}}) \triangleq Q(\Theta;\Theta^{\text{old}}) + \lambda\left\{\sum_{k=1}^{K}\pi_k - 1\right\} \tag{3.15}$$

Differentiating with respect to π_k results in

$$\frac{\partial\breve{Q}}{\partial\pi_k} = \sum_{l=1}^{L}\beta(l,k)\frac{1}{\pi_k} + \lambda = 0 \tag{3.16}$$

Now multiplying both sides by π_k and summing over k gives $\lambda = -L$, from which

$$\pi_k^{\text{new}} = \frac{1}{L}\sum_{l=1}^{L}\beta(l,k) \tag{3.17}$$

We should point out that since log(.) is a concave function and $\beta(l,k) \geq 0$ and since non-negative weighted sum of concave functions is still concave, $Q(\Theta;\Theta^{\text{old}})$ is a concave function of the π_k's. This followed by the fact that the constraint $\sum_{k=1}^{K}\pi_k = 1$ is linear implies that the Lagrange multiplier method achieves the optimal solution [85].

Similarly, to maximize $Q(\Theta;\Theta^{\text{old}})$ with respect to ϕ_{it} with the constraint that $\sum_{i=0}^{1}\phi_{it} = 1$, we maximize the Lagrangian $\tilde{Q}$ given by

$$\tilde{Q}(\Theta,\mu_t;\Theta^{\text{old}}) \triangleq Q(\Theta;\Theta^{\text{old}}) + \mu_t\left\{\sum_{i=0}^{1}\phi_{it} - 1\right\} \tag{3.18}$$

from which we get

$$\frac{\partial\tilde{Q}}{\partial\phi_{it}} = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\alpha(l,k,i,t)\frac{1}{\phi_{it}} + \mu_t = 0 \tag{3.19}$$

Multiplying both sides by ϕ_{it} and summing over i gives $\mu_t = -1$, which leads to

$$\phi_{it}^{\text{new}} = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\alpha(l,k,i,t) \tag{3.20}$$

By the same argument as in the case of π_k's, the Lagrange method above achieves the optimal solution for the ϕ_{it}'s.

We would like to comment on the complexity of the EM algorithm. At each iteration of EM, we should evaluate (3.9), (3.10), (3.14), (3.17), and (3.20) for updating $\beta(l,k)$, $\alpha(l,k,i,t)$, p_{ik}, π_k, and ϕ_{it}. To do these calculations, the required number of additions and multiplication is given by $15TKL + 6K^2L^2T - 6K^2L - 2L^2KT + 2KLT - 2K + LK - 2T$, and $11KTL + L + 4K^2L^2T - 4LK^2T + 2L^2TK + LK + 3K + 2T$, respectively. Therefore, the order of computational complexity for each EM iteration for both addition and multiplication is $\mathcal{O}(K^2L^2T)$. Furthermore, in all of our numerical results the EM algorithm converged in five or fewer iterations.

3.3.2 Resolving the Ambiguity in Parameter Estimation

It can be verified that the probability of receiving a one (or zero) from a radio with the operating point (p_{0k}, p_{1k}) under $\mathcal{H}_\eta$, $\eta \in \{0, 1\}$, is the same as from a radio with operating point (p_{1k}, p_{0k}) under $\mathcal{H}_{1-\eta}$. More specifically let $\Theta^{\mathrm{c}} \triangleq \{P^{\mathrm{c}}, \Pi, \Phi^{\mathrm{c}}\}$ be the counterpart of the parameter set Θ, where P^{c} and Φ^{c} are given by $p^{\mathrm{c}}_{ik} = p_{(1-i)k}$ and $\phi^{\mathrm{c}}_{it} = \phi_{(1-i)t}$, $i = 0, 1$. It then follows that $Pr(D|\Theta) = Pr(D|\Theta^{\mathrm{c}})$. Therefore, we always have two possible solutions Θ^{c} and Θ for the parameter set. Note that this ambiguity is not specific to the our proposed method and is inherent in any estimation procedure for the parameter set. The proposed EM algorithm converges to one of these two possible solutions. To resolve this ambiguity, we assume that the class of honest radios has the highest population among all the classes and the operating point of the honest radios is above the chance line, i.e., $p_{11} > p_{01}$[2]. Therefore, when the estimated probability of false alarm is higher than the probability of detection for the class with the highest population then the counterpart of the parameter set is the true solution.

3.3.3 Radio Identification and Hypotheses Testing

After the parameters are estimated from received decision matrix D, the identification and hypotheses matrices can be estimated using the MAP rule. Let $\hat{\Theta}$ denote the parameter set estimated by the EM algorithm. Then

$$\begin{aligned}(\hat{Z}, \hat{H}) &= \arg\max_{Z,H} Pr(Z, H|D; \hat{\Theta}) \\ &= \arg\max_{Z,H} \log Pr(D, Z, H|\hat{\Theta})\end{aligned} \tag{3.21}$$

Using (3.7) we get,

[2]We note that for practical CRNs, these assumption are not unrealistic.

$$(\hat{Z},\hat{H}) = \arg\max_{Z,H} \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=0}^{1} \left\{ \frac{z_{lk}}{2T}\log\hat{\pi}_k \right. \tag{3.22}$$
$$\left. + z_{lk}h_{it}\left[d_{lt}\log\hat{p}_{ik} + (1-d_{lt})\log(1-\hat{p}_{ik}) + \frac{1}{L}\log\hat{\phi}_{it} \right] \right\}$$

Unfortunately the complexity of the optimization in (3.22) is prohibitive. Therefore we employ a suboptimal algorithm for the estimation of Z and H as follows. Let

$$\hat{h}_{0t} = \begin{cases} 1, & \hat{\phi}_{0t} > \hat{\phi}_{1t} \\ 0, & \text{otherwise.} \end{cases} \tag{3.23}$$

We note that the above is equivalent to setting $\hat{H} = \arg\max_H Pr(H|\hat{\Theta})$. Next, let $N \triangleq \sum_{t=1}^{T}\hat{h}_{0t}$, $M \triangleq \sum_{t=1}^{T}\hat{h}_{1t}$, $n_l \triangleq \sum_{t=1}^{T}\hat{h}_{0t}d_{lt}$, and $m_l \triangleq \sum_{t=1}^{T}\hat{h}_{1t}d_{lt}$. Also let $\tilde{\mathbf{d}}_l \triangleq [d_{l1}, d_{l2}, \ldots, d_{lT}]$ denote the vector of decisions received form radio l. Then, the class of radio l is estimated as c_{k^*} (i.e., $\hat{z}_{lk^*} = 1$), where

$$\begin{aligned} k^* &= \arg\max_k Pr(z_{lk} = 1|\mathbf{d}_l, \hat{H}; \hat{\Theta}) \\ &= \arg\max_k Pr(\mathbf{d}_l|z_{lk} = 1, \hat{H}; \hat{\Theta}) \\ &= \arg\max_k \hat{p}_{0k}^{n_l}(1-\hat{p}_{0k})^{(N-n_l)}\hat{p}_{1k}^{m_l}(1-\hat{p}_{1k})^{(M-m_l)} \end{aligned} \tag{3.24}$$

The entire procedure of estimating the network parameters and the hypotheses matrix along with the classification of the radios is summarized in Algorithm 1 referred to as EM-based Classifier (EMC).

3.4 Numerical Results

In this section we compare the results obtained from the proposed algorithm with those from the reputation-based classifier (RBC) [76, 77, 78, 79, 80, 81] which can be used to classify the radios into two groups of honest and Byzantine (malicious) radios. In RBC, the hypothesis is first estimated at each time through a voting

Data: Decision matrix, D
Result: Estimation of the parameter set, Θ, detection of the hypotheses matrix, H, and classification of the radios, Z.

begin
  Estimating matrix of class parameters, Θ, using EM-Algorithm:
  Assume an initial estimation for Θ^{old};
  while *Convergence criterion is not satisfied* **do**
    E Step: Find $\alpha(l,k,i,t)$ and $\beta(l,k)$, using (3.10) and (3.9);
    M Step: Re-estimate p_{ik}, π_k, and ϕ_{it} using (3.14), (3.17), and (3.20);
  end
  Resolve the ambiguity between $\hat{\Theta}$ and $\hat{\Theta}^{\text{c}}$;
  if $\hat{p}_{01} > \hat{p}_{11}$ **then**
    $\hat{\Theta} = \hat{\Theta}^{\text{c}}$
  end
  Detect hypotheses, $\hat{H}$, using (3.23) and classify the radios, $\hat{Z}$, using (3.24);
end

Algorithm 1: Estimating the parameter set, detection of hypotheses matrix, and classificatin of radios using the EM algorithm.

scheme[3] as $\hat{h}_{1t} = 1$ if $\sum_{l=1}^{L} d_{lt} > q$ and $\hat{h}_{1t} = 0$, otherwise for $t = 1, 2, \ldots, T$, where q is a threshold. Then, the operating points are obtained from $\hat{p}_{ik} = \frac{\sum_{t=1}^{T} \hat{h}_{it} d_{lt}}{\sum_{t=1}^{T} \hat{h}_{it}}$. Finally radio classification is performed by comparing the reputation R_l of radio l with a threshold where

$$R_l \triangleq \sum_{t=1}^{T} (1 - d_{lt}) \hat{h}_{0t} + d_{lt} \hat{h}_{1t} \underset{\text{Byzantine}}{\overset{\text{Honest}}{\gtrless}} \lambda \tag{3.25}$$

The value of λ affects the probability of misclassifying an honest radio as Byzantine and vice verse. In the following results we set $\lambda = 0.5$ so the probability of misclassifying an honest radio as Byzantine is the same as misclassifying a Byzantine as honest.

We evaluate the performance of the classifiers using the *discriminability* and *reliability* metrics [58, 86]. Discriminability, denoted by Δ_Z, is a measure of the

[3]Voting or q-out-of-L rule is the only available rule for RBC when the FC does not have any prior information regarding the radios' parameters [82].

misclassification rate of the classifier and is given by

$$\Delta_Z \triangleq \frac{1}{2L} \sum_{l=1}^{L} \sum_{k=1}^{K} |z_{l,k} - \hat{z}_{l,k}| \tag{3.26}$$

The performance of the classifiers in the estimation of the hypothese is measured by the *hypothesis discriminability* given by

$$\Delta_H \triangleq \frac{1}{2T} \sum_{i=0}^{1} \sum_{t=1}^{T} |h_{it} - \hat{h}_{it}| \tag{3.27}$$

Finally, to measure the reliability of the classifiers, we define the estimation error for the radios' operating points as

$$\Delta_{OP} \triangleq \frac{1}{\sqrt{2}} \sum_{k=1}^{K} \pi_k \sqrt{(p_{0k} - \hat{p}_{0k})^2 + (p_{1k} - \hat{p}_{1k})^2} \tag{3.28}$$

These three measures are appropriately normalized so that they are in the interval $[0, 1]$ where the smaller the values, the better the estimates are.

We use four sets of the operating points in the simulations as shown in table 3.1. In the first set denoted OP1, there are two classes of honest (c_1) and Byzantine radios (c_2), where Byzantines try to confuse the FC under both hypotheses. In the second set, OP2, there are also two classes of honest and Byzantine radios but with different populations from OP1. Moreover, the honest radios are not as effective in detecting the hypotheses. In the case of OP3 and OP4, c_3 represents the class of "almost-always-no" radios and c_4 represents the class of "almost-always-yes" radios.

The performance of the classifiers with respect to the number of decisions, T, is shown in Fig. 3.1 and Fig. 3.2 for the estimation error Δ_{OP} and the misclassification rate Δ_Z for $L = 20$ radios. It can be seen that the accuracy of estimation and radio classification improves with T and that the proposed method outperforms RBC. Note that RBC is not capable of classifying the radios into more than two classes; thus for RBC, Δ_Z is not defined for OP3, and OP4. As shown in Figs. 3.1 and 3.2,

TABLE 3.1. Class parameters of each operating point.

Set	c_k	p_{0k}	p_{1k}	π_k
OP1	c_1	0.1	0.9	0.6
	c_2	0.9	0.4	0.4
OP2	c_1	0.2	0.7	0.7
	c_2	0.9	0.2	0.3
OP3	c_1	0.2	0.9	0.5
	c_2	0.8	0.3	0.3
	c_3	0.05	0.05	0.2
OP4	c_1	0.2	0.8	0.4
	c_2	0.9	0.2	0.15
	c_3	0.05	0.05	0.2
	c_4	0.95	0.95	0.25

the performance of the classifiers for OP1 is better than the others cases since the malicious radios are collectively weaker in the case of OP1. By this we mean that given the fraction of Byzantines and their operating points, the average number of radios that provide false information to the FC under each hypothesis is smaller in the case of OP1.

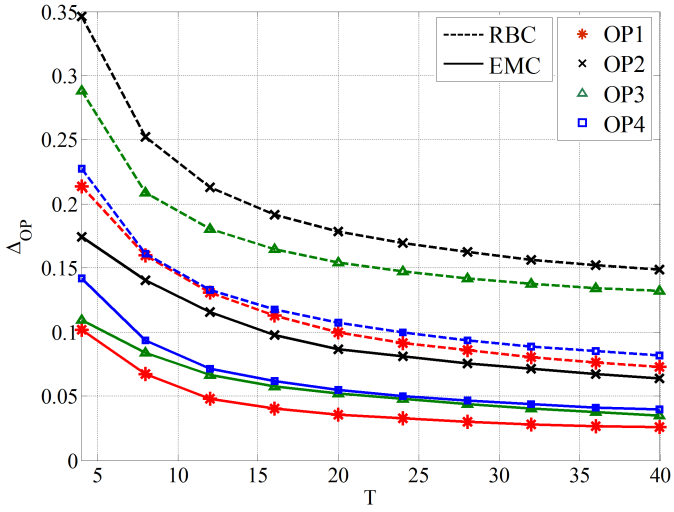

FIGURE 3.1. Estimation error of the operating point vs. T for $L = 20$.

To evaluate the performance of the classifier with respect to the fraction of honest radios in the network, we consider OP1 and OP2 and evaluate the estimation error and misclassification rate as a function of π_1 for $T = 18$ and $L = 20$. The

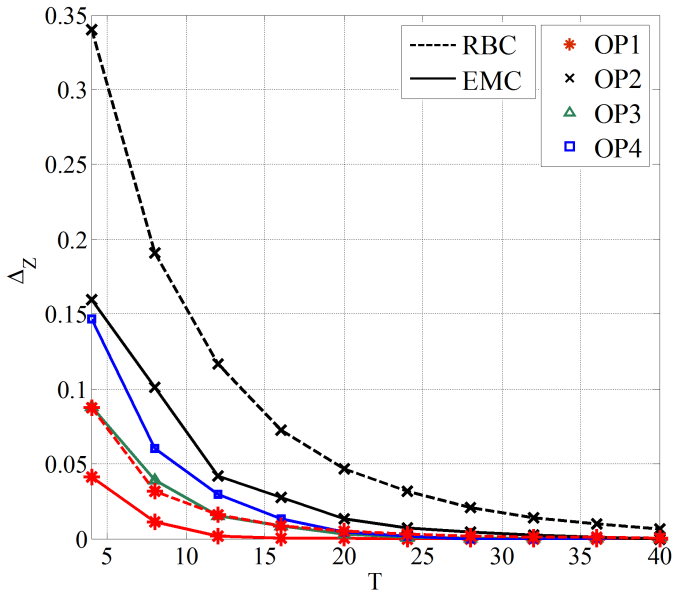

FIGURE 3.2. Misclassification rate vs. T for $L = 20$.

results are shown in Figs. 3.3 and 3.4. From Fig. 3.4 it can be seen that for OP2, the performance of RBC is not acceptable when $\pi_1 \leq 0.6$. In general RBC fails to classify the malicious radios correctly when the class of malicious radios is collectively stronger than the class of honest radios. In such cases even though the honest CRs are in majority, their performance in terms of detection and false alarm probability is not very good (i.e., it is close to the chance line). On the other hand the malicious CRs have good detection and false alarm probability and try to completely mislead the FC. As a result the FC receives more decisions in favor of the alternative hypothesis $H_{1-\eta}$ when H_η is the true hypothesis.

Performance of the classifier with respect to the number of radios in the network, L, is evaluated by the estimation error, misclassification rate, and hypothesis discriminability and the results are shown in Figs. 3.5-3.7. As expected, in all the cases the performance improves with L. However, it can be seen from these figures that for small number of decisions T, the performance of RBC is not acceptable for OP2 for any values of L.

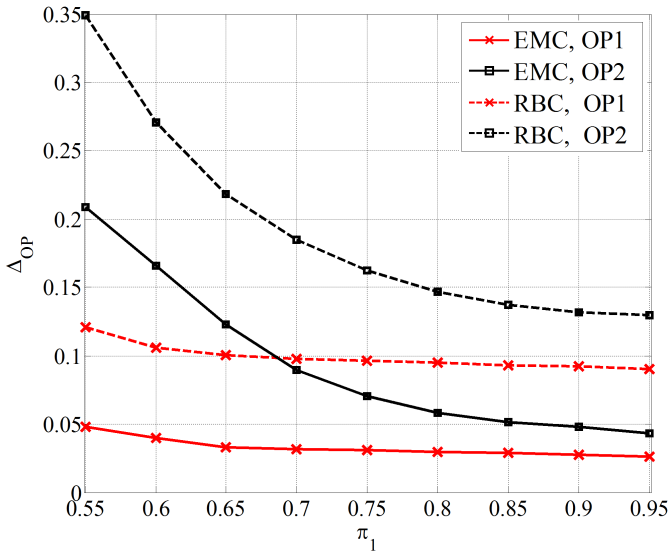

FIGURE 3.3. Estimation error of the operating point vs. π_1 for $T = 18$ and $L = 20$.

Remark 4. *In spectrum sensing, misdetection results in secondary user transmission when the primary user is present, causing undue interference to the primary user. Also false alarm results in wasted opportunity for the secondary user to utilize the channel, resulting in loss of throughput in the secondary network. In addition, channel is not utilized during the detection process. Therefore, in the operation of cognitive radios, fast and reliable detection of the primary user is critical. As it is shown in Fig.'s 1, 2, 5, 6, and 7, even with a few decisions from the radios (e.g., $T = 4$), the proposed method achieves a good performance, which is also significantly better than the performance of RBC.*

3.4.1 Variance of the Estimator vs. Cramer-Rao Lower Bound

To evaluate the efficacy of the proposed algorithm for the estimation of the operating points we compare our results with the Cramer-Rao lower bound (CRLB). Assume that the FC knows the class of each radio (i.e., the radio identification matrix Z) and the hypotheses matrix H. Let $\zeta_k =\triangleq \sum_{l=1}^{L} z_{lk}$ and $\mathcal{N}_i \triangleq \sum_{t=1}^{T} h_{it}$. Then it can be shown that the variance of any unbiased estimator can be lower bounded as follows.

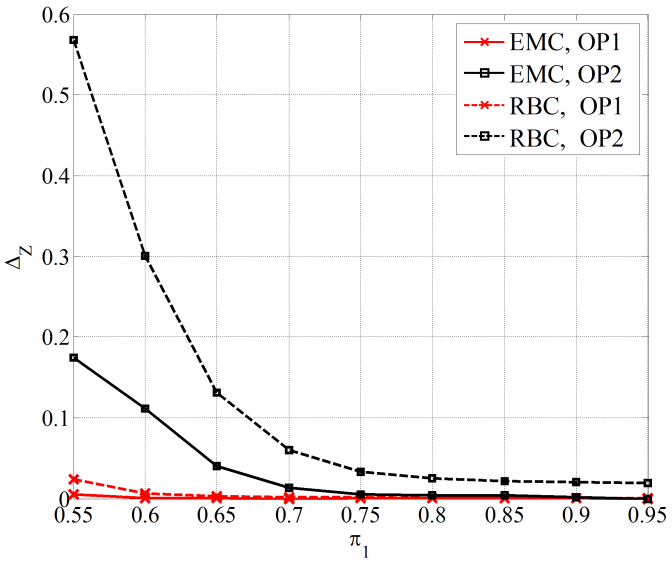

FIGURE 3.4. Misclassification rate vs. π_1 for $T = 18$ and $L = 20$.

$$\text{var}\{\hat{p}_{ik}|p_{ik}\} \geq \frac{p_{ik}(1-p_{ik})}{\zeta_k \mathcal{N}_i} \tag{3.29}$$

Fig. 3.8 shows the variances of the probability of false alarm estimated by EMC for the first class of OP1, OP2, OP3, and OP4 when $L = 20$. These results are also compared with the lower bound in (3.29) denoted by CLRB. It can be seen that the variance of the estimation decreases with the number of decisions T and approaches the bound as T increases.

3.4.2 Correlated Decisions

In some sensing scenarios the secondary radios may experience similar shadowing [72]. In this case, under hypothesis $\mathcal{H}_1$, the observations of the SUs will be correlated. Consequently, the decisions of the radios will also be correlated. To investigate the effects of such correlation on our proposed algorithm, as in [72], we consider a one-dimensional distribution of the SUs. We model the effects of correlated shadowing on the decisions of the radios by a Markov chain. Specifically, we assume that for each time t, the decisions, $\{r_{lt}, l \in c_k\}$, of the radios in each class c_k form a Markov chain with probability transition matrix $P = \left[\begin{smallmatrix} a & 1-a \\ b & 1-b \end{smallmatrix}\right]$.

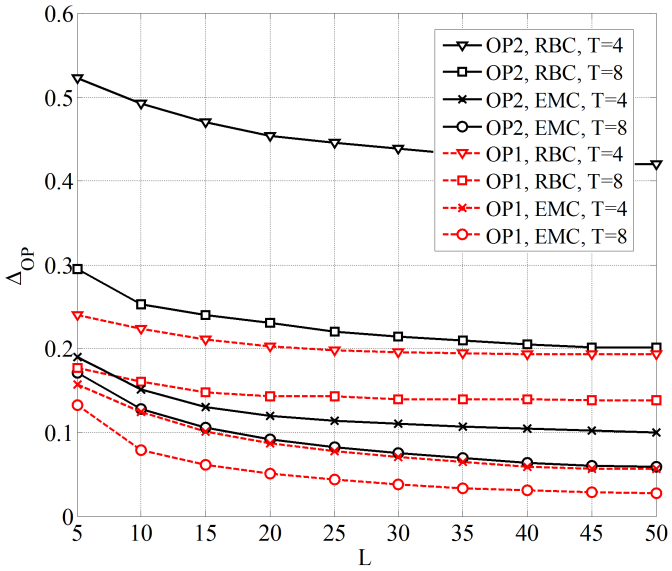

FIGURE 3.5. Estimation error vs. L for $T = 4, 8$.

Then $\tilde{p}_{1k} = P(r_{lt} = 1|h_{1t} = 1) = \frac{b}{1+b-a}$. On the other hand, the correlation of the decisions between adjacent SUs is given by $\zeta = E\{r_{lt}r_{(l+1)t} \mid h_{1t} = 1\} = a\tilde{p}_{1k}$. Therefore, from the given values of detection probabilities $\tilde{p}_{1k}$ and correlation ζ, one can determine the parameters a and b.

Simulation results are obtained for two classes of radios where the SUs in each class experience correlated shadowing. To be able to compare the results with those from independent decisions, we choose the parameters so that the operating points of the correlated case are the same as those in the independent case, namely OP1 and OP2 in Table 3.1. In particular, we set $a = .99$ and $b = .09$, which correspond to $\tilde{p}_{1k} = .9$ and $\zeta = .891$. For the SUs in c_1 (honest radios), we set $\rho_1(1) = 1$ and $\rho_0(1) = 0$. This results in $p_{01} = 0.1$ and $p_{11} = 0.9$. For the SUs in c_2 (Byzantine radios), we set $\rho_1(2) = 27/80$ and $\rho_0(2) = 77/80$. This results in $p_{02} = 0.9$ and $p_{12} = 0.4$. Note that these values are the same as those for OP1 in Table 3.1. For a second example, for the SUs in c_1, we set $\rho_1(1) = 61/80$ and $\rho_0(1) = 11/80$. This results in $p_{01} = 0.2$ and $p_{11} = 0.7$. For the SUs in c_2, we set $\rho_1(2) = 9/80$ and

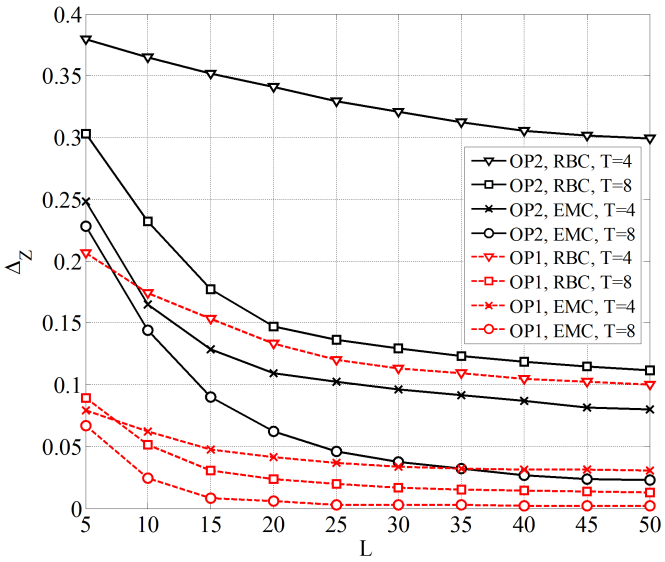

FIGURE 3.6. Misclassification rate vs. L for $T = 4, 8$.

$\rho_0(2) = 79/80$. This results in $p_{02} = 0.9$ and $p_{12} = 0.2$. Note that these values are the same as those for OP2 in Table 3.1.

In Figs. 3.9 and 3.10 we compare the simulation results from the proposed algorithm with those from RBC for the cases of dependent and independent observations. Fig. 3.9 shows Δ_Z vs. L and Fig. 3.10 shows Δ_H vs. L. It can be seen that the performance of both algorithms is slightly worse for dependent observations, with the proposed algorithm achieving significantly better performance than RBC. The fact that the performance of dependent and independent decisions are close is noteworthy, particularly given the large correlation value of $\zeta = .891$. This indicates that, although under correlated shadowing, the mis-detection probability is asymptotically lower bounded [72], in the case of a finite number of radios, the effect of correlated observations is not significant.

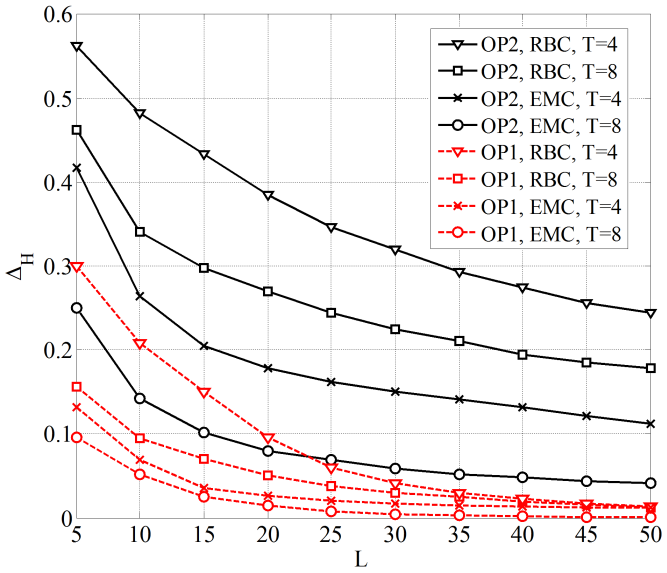

FIGURE 3.7. Hypothesis discriminability vs. L for $T = 4, 8$.

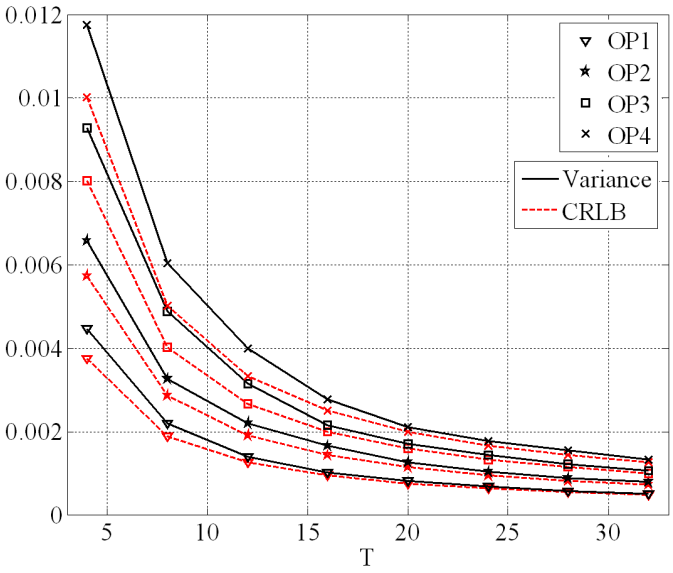

FIGURE 3.8. Comparison of the variances of the $\hat{p}_{00}$ with the CRLB when $L = 20$.

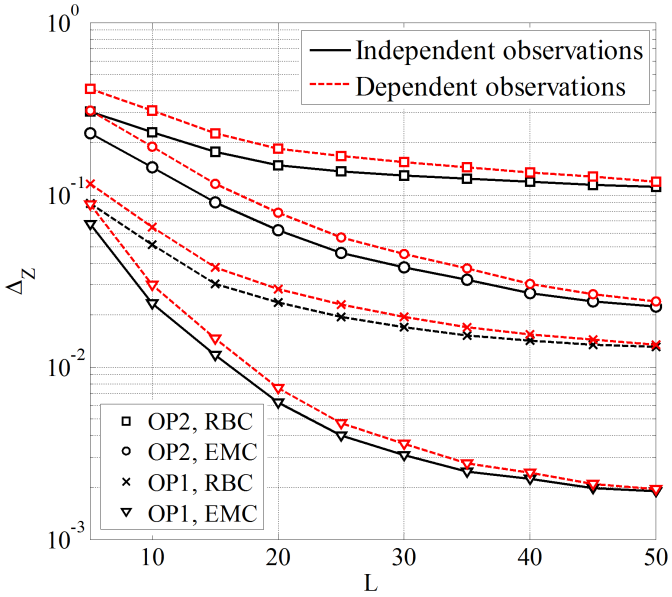

FIGURE 3.9. Misclassification rate vs. L for $T = 8$ for dependent and independent observations.

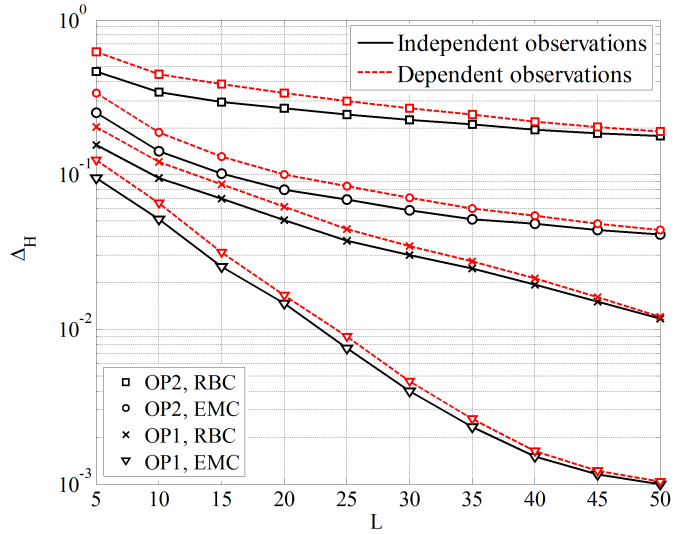

FIGURE 3.10. Hypothesis discriminability vs. L for $T = 8$ for dependent and independent observations.

Chapter 4

Nonparametric Density Estimation, Hypotheses Testing, and Radio Classification in Centralized Detection

4.1 Introduction

Detection with distributed cognitive radios is often used to improve the performance of systems such as radar and sonar, [87], cognitive radio networks (CRNs), [88], and cooperative spectrum sensing in cognitive radio networks [65], to name a few. The detection strategy can be categorized as centralized or decentralized. In centralized detection, the cognitive radio transmit their actual (raw) measurements without any pre-processing to the fusion center (FC) which fuses the received messages to detect the hypothesis. On the other hand, in decentralized detection, each radio quantizes its data before transmission to the FC. In the case of binary hypothesis testing and binary quantization, the radios in fact make a local decision regarding the state of the hypothesis and send their decisions to the FC [74, 89, 90, 91].

Performance of decentralized detection depends on the quantization rule used in the cognitive radios as well as the fusion rule used in the FC. Several authors have investigated the optimal design of the FC's fusion rule as well as the optimal design of the quantization rule for the cognitive radios' measurements, where it is assumed that the underlying distribution of the cognitive radios' data is known. [92, 93, 94, 95]. However, in many applications the statistical model of the cognitive radios' local measurements is unknown or difficult to find [96]. Moreover, the distribution of the cognitive radios' measurements may vary among the cognitive radios and over time. For example, due to their geographic distribution in the cognitive radio field, the cognitive radios' received signals may have different

statistics; the cognitive radios may experience different types of noise; the channels between the cognitive radios and the FC may not be identical; or the network may be operating in a time-varying environment. Another scenario is when a set of misbehaving radios, which may be faulty or under the control of an adversary, transmit false data to the FC. Decentralized detection in cognitive radio networks in the presence of Byzantine attacks has been the subject of many studies in recent years (see [97, 80, 98] and the references therein). Another example is in collaborative spectrum sensing where individual radios employ, say, an energy detector and transmit their measurements to the FC which decides on the presence or absence of a primary user (PU). Malicious radios may send high energy values when PU is not present or low energy values when it is present. In the former case malicious radios intend to reduce the throughput of honest radios in order to increase their own bandwidth utilization, while in the latter case, they reduce the detection probability, and thereby increase interference to the PU. Several authors have investigated outlier detection strategies in order to detect and mitigate the deleterious effects of such malicious attacks on cooperative spectrum sensing [69, 99, 71].

In this dissertation we consider centralized detection in a cognitive radio network consisting of multiple classes of cognitive radios, where each class consists of all the cognitive radios whose received data, under each hypothesis, are drawn from the same probability density function (PDF) which are assumed to be unknown a priori. The goal of the FC is to detect the state of nature and to classify the radios. However, for the optimal fusion of the received data and classification of the cognitive radios, the underlying PDFs must also be estimated. We develop a method based on the expectation maximization algorithm for nonparametric estimation of the underlying PDFs, classification of the cognitive radios and the detection of the hypotheses. In scenarios where the PDFs are time-varying, we

develop a two phase approach for the detection of the hypothesis and estimation of the PDFs so as to reduce the amount of data transmitted by the cognitive radios. Although we do not explicitly formulate our problem as such, the proposed method may also be viewed as an outlier detection strategy for collaborative spectrum sensing operating in the presence of misbehaving radios.

As mentioned previously, in scenarios where the underlying distributions of the cognitive radios' measurements is unknown a priori or is time-varying, the optimal design of the quantizer for decentralized detection is not feasible. Our proposed method here can be employed to alleviate this difficulty in decentralized detection as follows. Periodically the cognitive radios send several samples of their unquantized measurements to the FC. Using our proposed method the FC can estimate the PDFs and optimally design the quantizers for each class of cognitive radios[1]. The FC transmits the quantizer thresholds to the cognitive radios which can then revert back to decentralized detection using these thresholds.

Density estimation using EM algorithm has been investigated in a number of dissertations. In [96] the PDFs are modeled as a Gaussian mixture and the parameters are estimated assuming that the cognitive radios transmit only a binary decision to the FC. A Gaussian mixture model whose parameters vary in time is also assumed in [100] and the parameters are estimated at the FC from the cognitive radios received data. Following the pioneering work of [101], several authors have developed distributed EM algorithms for density estimation using the consensus algorithm [102, 103]. Our approach here is different in two regards. Our density estimation is in the context of a hypothesis testing problem. This has not

[1]Note that this step does not disrupt the operation of the network since the hypothesis detection is also performed in the proposed method.

been studied before. Moreover, our PDF estimation is nonparametric, whereas, previously, Gaussian mixture models were assumed.

4.2 Problem Formulation, System Model, and Notations

We consider a cognitive radio network consisting of L radios employed to detect the state of nature $\mathcal{H} \in \{\mathcal{H}_0, \mathcal{H}_1\}$. It is assumed that there are K classes of cognitive radios with π_k denoting the fraction of radios in class k. During each time slot, each cognitive radio transmits its measurement regarding the state of nature to the FC. It is assumed that under hypothesis $\mathcal{H}_i$, the measurements of the radios in class k are derived from PDF $f_i^{(k)}(x)$, $k = 1, 2, \cdots, K$, $i = 0, 1$. After T time slots, the FC has received $L \times T$ measurements from all of the radios which we denote by the $L \times T$ matrix $D = [d_{lt}]$, where $d_{lt} \in \Re$ is the measurement received from the lth radio at time t.

We assume that the cognitive radios do not collaborate and that their observations are contaminated by independent noise processes. Consequently, given the hypothesis $\mathcal{H}_\eta$, $\eta = 0, 1$, the cognitive radio observations are conditionally independent. In addition, given the hypothesis $\mathcal{H}_\eta$, the cognitive radio observations are assumed to be conditionally independent over time. This assumption is justified when the cognitive radio observations are contaminated with white noise processes.

Having received the measurement matrix D, the FC desires to classify the radios, estimate the PDFs $f_i^{(k)}(x)$ $k = 1, 2, \cdots, K$, $i = 0, 1$ corresponding to each class and to detect the state of nature during each slot of the observation period T. To this end we assign probabilities ϕ_{0t} and $\phi_{1t} = 1 - \phi_{0t}$ to the states $\mathcal{H}_0$ and $\mathcal{H}_1$ at time t, respectively. Note that these are not prior probabilities and are only used as a tool to help us decide on the state of $\mathcal{H}$ at time t. In the following we estimate ϕ_{0t} and ϕ_{1t} and decide the state of nature at time t to be $\mathcal{H}_0$ if $\phi_{0t} > \phi_{1t}$, and $\mathcal{H}_1$, otherwise. In this way we convert the hypothesis detection problem into an

estimation problem for the parameters ϕ_{0t} and ϕ_{1t}. Let

$$\Phi \triangleq \begin{bmatrix} \phi_{01} & \phi_{02} & \cdots & \phi_{0T} \\ \phi_{11} & \phi_{12} & \cdots & \phi_{1T} \end{bmatrix} \tag{4.1}$$

denote the hypothesis probability matrix and define the PDF matrix, F as

$$F \triangleq \begin{bmatrix} f_0^{(1)}(x) & f_0^{(2)}(x) & \cdots & f_0^{(K)}(x) \\ f_1^{(1)}(x) & f_1^{(2)}(x) & \cdots & f_1^{(K)}(x) \end{bmatrix} \tag{4.2}$$

Finally we let

$$\Pi \triangleq [\pi_1 \quad \pi_2 \quad \cdots \quad \pi_K] \tag{4.3}$$

The unknown parameter set is now defined by $\Theta \triangleq \{F, \Pi, \Phi\}$. The maximum likelihood estimation of Θ from D is given by $\tilde{\Theta} = \arg\max_{\Theta} P(D|\Theta)$. However, since the identity of the radios is not known, $P(D|\Theta)$ is not directly available. Therefore to estimate the parameter set, we employ the well-known expectation-maximization (EM) algorithm in which the identity of the radios and the state of nature during the observation period constitute the latent variables. We end this section by defining two matrices, the radio identification matrix, Z,

$$Z \triangleq \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1K} \\ z_{21} & z_{22} & \cdots & z_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ z_{L1} & z_{L2} & \cdots & z_{LK} \end{bmatrix} \tag{4.4}$$

where the binary variable z_{lk} is one if radio l belongs to class k, and zero, otherwise, and the hypotheses matrix H as

$$H \triangleq \begin{bmatrix} h_{01} & h_{02} & \cdots & h_{0T} \\ h_{11} & h_{12} & \cdots & h_{1T} \end{bmatrix} \tag{4.5}$$

where column t represents the state of the hypothesis at time t in that $h_{it} \in \{0, 1\}$ for $i = 0, 1$ where $h_{0t} = 1 - h_{1t}$ and $h_{0t} = 1$ denotes $\mathcal{H}_0$ at time t and $h_{1t} = 1$ denotes $\mathcal{H}_1$ at time t.

4.3 Parameter Estimation

To establish the EM algorithm for estimating the parameter set we need to find the joint conditional PDF of the measurement matrix, the radio identification matrix, and the hypotheses matrix, given the parameter set, as

$$\begin{aligned} P(D,Z,H|\Theta) &= P(D|Z,H;\Theta)P(Z,H|\Theta) \\ &= \prod_{l=1}^{L}\prod_{k=1}^{K}\pi_k^{z_{lk}}\left[\prod_{t=1}^{T}\prod_{i=0}^{1}\left(f_i^{(k)}(d_{lt})\right)^{h_{it}}\phi_{it}^{\frac{h_{it}}{L}}\right]^{z_{lk}} \\ &= \prod_{l=1}^{L}\prod_{k=1}^{K}\left[\prod_{t=1}^{T}\prod_{i=0}^{1}\pi_k^{\frac{1}{2T}}\left(f_i^{(k)}(d_{lt})\phi_{it}^{\frac{1}{L}}\right)^{h_{it}}\right]^{z_{lk}} \end{aligned} \tag{4.6}$$

Therefore the log-likelihood function is given by

$$\begin{aligned} L(\Theta;D,Z,H) = \log P(D,Z,H|\Theta) = \sum_{l=1}^{L}\sum_{k=1}^{K} z_{lk}\times \\ \sum_{t=1}^{T}\sum_{i=0}^{1}\left\{\frac{1}{2T}\log\pi_k + h_{it}\left[\log f_i^{(k)}(d_{lt}) + \frac{1}{L}\log\phi_{it}\right]\right\} \end{aligned} \tag{4.7}$$

In the *expectation step* of EM, we compute the expectation of the log-likelihood function with respect to the latent variables (Z,H), given the measurement matrix, D, and the current estimate of the parameter set Θ^{old}. This is given by

$$\begin{aligned} &Q(\Theta;\Theta^{\text{old}}) \triangleq E_{(Z,H)|D;\Theta^{\text{old}}}[L(\Theta;D,Z,H)] \\ &= \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=0}^{1}\left\{E_{(Z,H)|D;\Theta^{\text{old}}}[z_{lk}]\frac{1}{2T}\log\pi_k + \right. \\ &\qquad \left. E_{(Z,H)|D;\Theta^{\text{old}}}[z_{lk}h_{it}]\left[\log f_i^{(k)}(d_{lt}) + \frac{1}{L}\log\phi_{it}\right]\right\} \end{aligned} \tag{4.8}$$

Defining $\alpha(l,k,i,t) \triangleq E_{(Z,H)|D;\Theta^{\text{old}}}[z_{lk}h_{it}]$ and $\beta(l,k) \triangleq E_{(Z,H)|D;\Theta^{\text{old}}}[z_{lk}] = P(z_{lk} = 1|D;\Theta^{\text{old}})$, we have

$$\begin{aligned} &\alpha(l,k,i,t) = \\ &P(h_{it}=1|z_{lk}=1,D;\Theta^{\text{old}})\times P(z_{lk}=1|D;\Theta^{\text{old}}), \end{aligned} \tag{4.9}$$

$$\beta(l,k) = P(z_{lk} = 1 | D; \Theta^{\text{old}}) \tag{4.10}$$

The probabilities in (4.9) and (4.10) are calculated in (4.11) and (4.12), where $\mathbf{d}^t_{\text{col}}$ denotes the tth column of D, i.e., the received vector of measurements at time t, and where $\mathbf{d}^l_{\text{row}}$ denotes the lth row of D, i.e., the received vector of measurements from radio l.

$$\begin{aligned}
&P(h_{it} = 1 | z_{lk} = 1, D; \Theta^{\text{old}}) = P(h_{it} = 1 | z_{lk} = 1, \mathbf{d}^t_{\text{col}}; \Theta^{\text{old}}) \qquad (4.11)\\
&= \frac{P(\mathbf{d}^t_{\text{col}} | h_{it} = 1, z_{lk} = 1; \Theta^{\text{old}}) \times P(h_{it} = 1 | z_{lk} = 1; \Theta^{\text{old}})}{\sum_{j=0}^{1} P(\mathbf{d}^t_{\text{col}} | h_{jt} = 1, z_{lk} = 1; \Theta^{\text{old}}) \times P(h_{jt} = 1 | z_{lk} = 1; \Theta^{\text{old}})}\\
&= \frac{P(d_{lt} | h_{it} = 1, z_{lk} = 1; \Theta^{\text{old}}) \prod_{l' \neq l} P(d_{l't} | h_{it} = 1; \Theta^{\text{old}}) \times \phi_{it}}{\sum_{j=0}^{1} \left\{ P(d_{lt} | h_{jt} = 1, z_{lk} = 1; \Theta^{\text{old}}) \prod_{l' \neq l} P(d_{l't} | h_{jt} = 1; \Theta^{\text{old}}) \times \phi_{jt} \right\}}\\
&= \frac{\phi_{it}\ f_i^{(k)}(d_{lt})\ \prod_{l' \neq l} \left[\sum_{k'=1}^{K} f_i^{(k')}(d_{l't}) \pi_{k'} \right]}{\sum_{j=0}^{1} \left\{ \phi_{jt}\ f_j^{(k)}(d_{lt}) \prod_{l' \neq l} \left[\sum_{k'=1}^{K} f_j^{(k')}(d_{l't}) \right] \pi_{k'} \right\}}
\end{aligned}$$

$$\begin{aligned}
&P(z_{lk} = 1 | D; \Theta^{\text{old}}) = P(z_{lk} = 1 | \mathbf{d}^l_{\text{row}}; \Theta^{\text{old}}) \qquad (4.12)\\
&= \frac{\pi_k Pr\left(\mathbf{d}^l_{\text{row}} | z_{lk} = 1; \Theta^{\text{old}}\right)}{\sum_{j=1}^{K} \pi_j Pr\left(\mathbf{d}^l_{\text{row}} | z_{lj} = 1; \Theta^{\text{old}}\right)}\\
&= \frac{\pi_k \prod_{t=1}^{T} Pr\left(d_{lt}) | z_{lk} = 1; \Theta^{\text{old}}\right)}{\sum_{j=1}^{K} \pi_j \prod_{t=1}^{T} Pr\left(d_{lt} | z_{lj} = 1; \Theta^{\text{old}}\right)}\\
&= \frac{\pi_k \prod_{t=1}^{T} \sum_{i=0}^{1} \phi_{it}\ f_i^{(k)}(d_{lt})}{\sum_{k'=1}^{K} \pi_{k'} \prod_{t=1}^{T} \sum_{i=0}^{1} \phi_{it}\ f_i^{(k')}(d_{lt})}
\end{aligned}$$

So we can rewrite $Q(\Theta; \Theta^{\text{old}})$ as

$$\begin{aligned}
Q(\Theta; \Theta^{\text{old}}) = &\sum_{l=1}^{L} \sum_{k=1}^{K} \beta(l,k) \log \pi_k \qquad (4.13)\\
&+ \frac{1}{L} \sum_{l=1}^{L} \sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{i=0}^{1} \alpha(l,k,i,t) \log \phi_{it}\\
&+ \sum_{l=1}^{L} \sum_{k=1}^{K} \sum_{t=1}^{T} \sum_{i=0}^{1} \alpha(l,k,i,t) \left[\log f_i^{(k)}(d_{lt}) \right]
\end{aligned}$$

In the *maximization step* of EM algorithm we maximize $Q(\Theta;\Theta^{\text{old}})$ with respect to the parameter set Θ so as to compute the next parameter set. To maximize $Q(\Theta;\Theta^{\text{old}})$ subject to the constraint $\sum_{k=1}^{K}\pi_k = 1$ we use the Lagrangian, $\mathcal{L}_\pi$, defined as

$$\mathcal{L}_\pi \triangleq Q(\Theta;\Theta^{\text{old}}) + \lambda\left\{\sum_{k=1}^{K}\pi_k - 1\right\} \tag{4.14}$$

The derivative of $\mathcal{L}_\pi$ is computed as

$$\frac{\partial}{\partial\pi_k}\mathcal{L}_\pi = \sum_{l=1}^{L}\beta(l,k)\frac{1}{\pi_k} + \lambda = 0 \tag{4.15}$$

Multiplying both sides by π_k and summing over k gives $\lambda = -L$. Hence,

$$\pi_k = \frac{1}{L}\sum_{l=1}^{L}\beta(l,k) \tag{4.16}$$

Similarly, to maximize $Q(\Theta;\Theta^{\text{old}})$ subject to the constraint $\sum_{i=0}^{1}\phi_{it} = 1$ we use the Lagrangian, $\mathcal{L}_{\phi_i}$, given by

$$\mathcal{L}_{\phi_i} \triangleq Q(\Theta;\Theta^{\text{old}}) + \mu_t\left\{\sum_{i=0}^{1}\phi_{it} - 1\right\} \tag{4.17}$$

whose derivative is obtained as

$$\frac{\partial}{\partial\phi_{it}}\mathcal{L}_\phi = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\alpha(l,k,i,t)\frac{1}{\phi_{it}} + \mu_t = 0 \tag{4.18}$$

Multiplying both sides by ϕ_{it} and summing over i results in $\mu_t = -1$. Hence,

$$\phi_{it} = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}\alpha(l,k,i,t) \tag{4.19}$$

Remark 5. *The Lagrangian methods above result in the optimal solution for* Π *and* ϕ_{it}*,* $i = 0, 1$*. This is due to the fact that* $Q(\Theta;\Theta^{old})$ *is concave with respect to the variables* π_k*,* $k = 1, 2, \cdots, K$ *and* ϕ_{it}*,* $i = 0, 1$ *and the constraints in both cases are linear.*

Next, we aim to maximize $Q(\Theta;\Theta^{\text{old}})$ with respect to the PDFs $f_i^{(k)}(x)$ for $i = 0,1$ and $k = 1,2,\cdots,K$. When the FC does not have any prior information about these PDFs, a nonparametric estimation method must be employed [84, 58]. Among the nonparametric density estimation methods histogram, kernel and orthogonal series based methods are most prevalent. In the next section, we consider the first tow methods to estimate the PDFs which maximize $Q(\Theta;\Theta^{\text{old}})$.

4.3.1 Histogram-Based Approach

In this approach, the cognitive radios' measurement sample space is partitioned into a number of bins with duration δ. Suppose the range of measurements is covered by N_{Bin} consecutive bins with the first bin starting at point x_0. So, $f_i^{(k)}(x)$ is stepwise and the constraint for maximizing $Q(\Theta;\Theta^{\text{old}})$ is $\delta\sum_{n=1}^{N_{\text{Bin}}} f_i^{(k)}(x_0+n\delta)=1$ which results in the following Lagrangian function, $\mathcal{L}_f$,

$$\mathcal{L}_f \triangleq Q(\Theta;\Theta^{\text{old}}) + \nu\left\{\delta\sum_{n=1}^{N_{\text{Bin}}} f_i^{(k)}(x_0+n\delta)-1\right\} \tag{4.20}$$

Taking derivative of the Lagrangian we get

$$\frac{\partial}{\partial f_i^{(k)}(x_0+m\delta)}\mathcal{L}_f = \sum_{(l,t)\in\mathcal{B}_m}\frac{\alpha(l,k,i,t)}{f_i^{(k)}(x_0+m\delta)} + \delta\nu = 0 \tag{4.21}$$

where $\mathcal{B}_m \triangleq \{(l,t) \mid x_0+m\delta \le d_{lt} < x_0+(m+1)\delta\}$. Multiplying both sides by $f_i^{(k)}(x_0+m\delta)$ and summing over m gives $\nu = -\sum_{l=1}^{L}\sum_{t=1}^{T}\alpha(l,k,i,t)$. Thus,

$$f_i^{(k)}(x_0+m\delta) = \frac{\sum_{\mathcal{B}_m}\alpha(l,k,i,t)}{\delta\sum_{l=1}^{L}\sum_{t=1}^{T}\alpha(l,k,i,t)} \tag{4.22}$$

4.3.2 Kernel-Based Approach

In this approach, the PDF $f_i^{(k)}(x)$ is modeled as

$$f_i^{(k)}(x) = \frac{1}{LT\sigma}\sum_{l=1}^{L}\sum_{t=1}^{T} w_i^{(k)}(l,t)\Psi\left(\frac{x-d_{lt}}{\sigma}\right) \tag{4.23}$$

where $\Psi(\cdot)$ is a kernel function, σ is the bandwidth, and $w_i^{(k)}(l,t)$ is the weight of the kernel placed at d_{lt} for the k-th class and hypothesis $\mathcal{H}_i$. Examples of kernel

functions include the Gaussian kernel, the Epanechnikov Kernel, the Laplacian Kernel, and the Quartic kernel [104].

Maximization of $Q(\Theta;\Theta^{\text{old}})$ with respect to $f_i^{(k)}(x)$ must be performed subject to the constraints that $\int f_i^{(k)}(x)dx = 1$ and $f_i^{(k)}(x) \geq 0, \quad \forall x$. From (4.8) it can be seen that this optimization problem is equivalent to evaluating $w_i^{(k)}(l,t), \forall\, l,t$, and each pair of (i,k) to maximize

$$\sum_{l=1}^{L}\sum_{t=1}^{T} \alpha(l,k,i,t)\times \\ \log\left[\frac{1}{LT\sigma}\sum_{l'=1}^{L}\sum_{t'=1}^{T} w_i^{(k)}(l',t')\Psi\left(\frac{d_{lt}-d_{l't'}}{\sigma}\right)\right] \tag{4.24}$$

subject to the constraints:

$$\sum_{l=1}^{L}\sum_{t=1}^{T} w_i^{(k)}(l,t) = 1 \tag{4.25}$$

$$w_i^{(k)}(l,t) \geq 0 \quad \forall\, t,l \tag{4.26}$$

The objective function in (4.24) is a concave function of the variables $w_i^{(k)}(l,t)$. This is due to the fact that log(.) is a concave function, $\alpha(l,k,i,t)$ is nonnegative for all l, k, i, and t followed by the fact that a nonnegative weighted sum of concave functions is concave. Therefore a convex optimization algorithm such as the interior-point method, [85], could be used to solve (4.24) subject to (4.25) and (4.26).

Due to this final optimization step, the kernel-based approach is slower than histogram-based method; however, the final estimations of the PDFs are smoother.

Remark 6. *The bin duration in histogram-based approach and the bandwidth in the kernel-based approach should be determined based on received data [84]. However, in the absence of any information on the underline densities, no optimal rule can be devised for the selection of these parameters. One option is to determine them*

based on partitioning the range of the received data into a specific number of bins. Other methods based on Doane's formula, [105], or Scott's rule, [106] can also be employed.

Let us define $\Theta^{\mathrm{c}} \triangleq \{F^{\mathrm{c}}, \Pi, \Phi^{\mathrm{c}}\}$ as the counterpart of the parameter set Θ, where F^{c} and Φ^{c} are given by

$$F^{\mathrm{c}} \triangleq \begin{bmatrix} f_1^{(1)}(x) & f_1^{(2)}(x) & \cdots & f_1^{(K)}(x) \\ f_0^{(1)}(x) & f_0^{(2)}(x) & \cdots & f_0^{(K)}(x) \end{bmatrix}, \tag{4.27}$$

$$\Phi^{\mathrm{c}} \triangleq \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1T} \\ \phi_{01} & \phi_{02} & \cdots & \phi_{0T} \end{bmatrix} \tag{4.28}$$

It can be verified that $P(D|\Theta) = P(D|\Theta^{\mathrm{c}})$. Therefore, there are always two possible solutions Θ^{c} and Θ for the parameter set to one of which the EM algorithm converges. This ambiguity can be resolved with some minimal information on the cognitive radio measurement data. For example in this dissertation we assume that for the class of radios with the largest population, the mean of the PDF under $\mathcal{H}_1$ is greater than the mean of the PDF under $\mathcal{H}_0$. Therefore, if the final solution does not satisfy this condition, then the counterpart parameter set should be chosen as the real solution. In the following we denote by $\hat{\Theta} \triangleq \left\{\hat{F}, \hat{\Pi}, \hat{\Phi}\right\}$ the final estimate of the parameter set after resolving the ambiguity.

4.4 Hypothesis Detection and Radio Classification

To detect the hypotheses matrix, we use the estimated $\hat{\Phi}$ which maximizes the likelihood of observing the current measurement matrix D. Therefore, we let

$$\hat{h}_{0t} = \begin{cases} 1, & \hat{\phi}_{0t} > \hat{\phi}_{1t} \\ 0, & \text{otherwise.} \end{cases} \tag{4.29}$$

Radios classification can be achieved using maximum a posterior rule as

$$\hat{Z} = \arg\max_{Z} P(Z|D,H;\hat{\Theta}) = \arg\max_{Z} \frac{P(D,Z|H;\hat{\Theta})}{P(D|H;\hat{\Theta})} =$$

$$\arg\max_{Z} P(D,Z|H;\hat{\Theta}) = \arg\max_{Z} \log P(D,Z|H;\hat{\Theta}) =$$

$$\arg\max_{Z} \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=0}^{1}\left\{\frac{z_{lk}}{2T}\log\hat{\pi}_k + z_{lk}h_{it}\log\hat{f}_i^{(k)}(d_{lt})\right\}$$

$$= \arg\max_{Z} \sum_{l=1}^{L}\sum_{k=1}^{K} z_{lk}\left\{\log\hat{\pi}_k + \sum_{t=1}^{T}\sum_{i=0}^{1} h_{it}\log\hat{f}_i^{(k)}(d_{lt})\right\} \tag{4.30}$$

Then, the class of radio l is detected as k^*, i.e., $\hat{z}_{lk^*} = 1$, and $\hat{z}_{lk} = 0$ for $k \neq k^*$, if

$$k^* = \arg\max_{k} \ \log\hat{\pi}_k + \sum_{t=1}^{T}\sum_{i=0}^{1} h_{it}\log\hat{f}_i^{(k)}(d_{lt}) \tag{4.31}$$

The entire procedure for the estimation of the parameter set, the detection of the hypotheses and the classification of the radios is summarized in Algorithm 2.

Data: Decision matrix, D

Result: Estimation of the parameter set, Θ, detection of the hypotheses, H, and classification of the radios, Z.

---

begin

Estimating parameters set, Θ, using the EM Algorithm:

Assume an initial value for Θ;

while *practical convergence criterion is not satisfied* **do**

E Step: Find $\alpha(l,k,i,t)$ and $\beta(l,k)$, using (4.9) and (4.10);

M Step: Reestimate π_k, and ϕ_{it} using (4.16), and (4.19), and reestimating $f_i^{(k)}(x)$ using (4.22) for histogram based approach or (4.23) and solving the optimization problem in (4.24) for kernel based approach;

end

Resolve the ambiguity between $\hat{\Theta}$ and $\hat{\Theta}^{\mathrm{c}}$;

Detect the hypotheses matrix, $\hat{H}$, using (4.29) and the classes of radios, $\hat{Z}$, using (4.31);

end

Algorithm 2: Estimating the parameter set, detecting hypotheses, and classifying the radios.

Remark 7. *In this dissertation we have assumed that the FC is aware of the number of classes of cognitive radios. This is a well-known problem in classification and is referred to as model order selection. Akaike information criterion (AIC) and Bayesian information criterion (BIC) have been proposed for model order selection. However, they do not always work satisfactorily and tend to favor overly simple models [84]. The main difficulty in model order selection is under- or overfitting the data. However, in large cognitive radio networks this may not be an issue as the number of cognitive radios L is significantly larger than the number of classes K. Therefore even if the number of classes is overestimated, it is unlikely that it will approach the number of cognitive radios and result in overfitting. In the case of our proposed algorithm, if the assumed number of classes K is larger than the actual number of classes, then the algorithm will not assign any radios to the fictitious classes.*

A question arises as to how many measurements from the cognitive radios are needed for the FC to obtain an accurate estimate of the underlying PDFs. Due to the complexity of the problem formulated in this dissertation, including the EM algorithm, the hypothesis detection and the radio classification, the analytical evaluation of the required number of measurements is not tractable. In Appendix 5 we evaluate the number of measurement samples needed in order to estimate a given PDF with its histogram so as to achieve a desired discriminability $\Delta_{\text{dist.}}$.

4.5 Numerical Results

We asses the discriminability and reliability, [86, 58], of the proposed approach in classifying the radios, detecting the hypotheses and estimating the PDFs using the following measures.

$$\Delta_H \triangleq \frac{1}{2T} \sum_{i=0}^{1} \sum_{t=1}^{T} |h_{it} - \hat{h}_{it}| \tag{4.32}$$

$$\Delta_Z \triangleq \frac{1}{2L}\sum_{l=1}^{L}\sum_{k=1}^{K}|z_{l,k}-\hat{z}_{l,k}| \tag{4.33}$$

$$\Delta_{\text{dist.}} \triangleq \frac{1}{4K}\sum_{k=1}^{K}\sum_{i=0}^{1}\int_{-\infty}^{\infty}\left|f_i^{(k)}(x)-\hat{f}_i^{(k)}(x)\right|^2 dx \tag{4.34}$$

where Δ_H measures hypotheses discriminability, Δ_Z measures the misclassification rate, and $\Delta_{\text{dist.}}$ measures densities discriminability. These three measures are normalized to be in in the interval [0 1] and the smaller values indicate better performances.

We use two sets of PDFs to evaluate the performance of the proposed method. Set 1 consists of four PDFs, as shown in Fig. 4.1, to describe two classes of cognitive radios. Let $g(\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. For Set 1 we have chosen $f_0^{(1)}(x) = .7g(-10,225)+.3g(30,225)$, $f_1^{(1)}(x) = .6g(70,100)+.4g(105,225)$, $f_0^{(2)}(x) = g(90,100)$ and $f_1^{(2)}(x) = g(150,225)$.

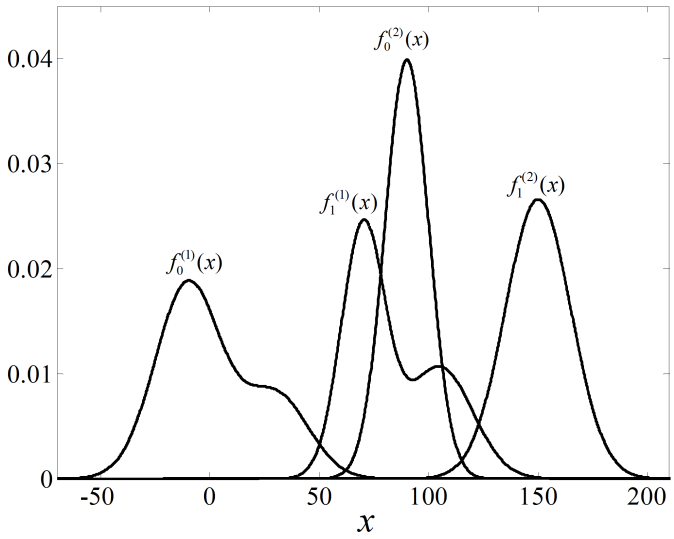

FIGURE 4.1. Set 1 of the probability density functions.

In many applications the observation of the cognitive radios is the sum of a constant signal component and a noise component which is often modeled as a Gaussian random variable. As a result the observations of the cognitive radios

are modeled as Gaussian random variables. A more general model, the Gaussian mixture model has also been used in many applications where the data may be viewed as arising from one of several populations. Therefore for set 1 we have chosen the PDFs from the Gaussian and the Gaussian mixture families.

Set 2 consists of six PDFs, as shown in Fig. 4.2, corresponding to three classes of cognitive radios. Here we have $f_0^{(1)}(x) = g(-10, 225)$, $f_1^{(1)}(x) = g(70, 100)$, $f_0^{(2)}(x)$ is a non-central chi-squared density function with 3 degrees of freedom and non-centrality parameter equal to 55, $f_1^{(2)}(x)$ is also a non-central chi-squared density with 4 degrees of freedom and non-centrality parameter equal to 5. Finally $f_0^{(3)}(x)$ is uniform on $[35, 60]$ and $f_1^{(3)}(x)$ is uniform on $[120, 145]$[2]. Set 2 may be considered to arise from a class of honest radios and two classes of misbehaving radios.

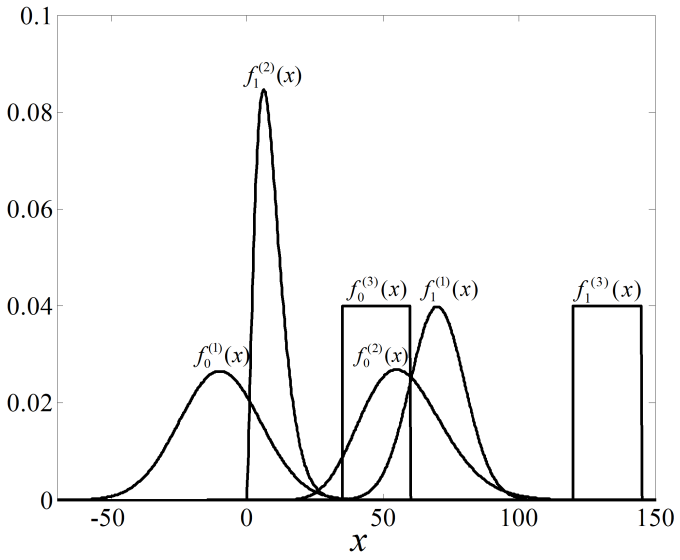

FIGURE 4.2. Set 2 of the probability density functions.

For kernel-based method, we use the kernel function $\Psi(x) = \frac{1}{2\pi}\exp(-\frac{x^2}{2})$ and use the interior point method to solve the optimization problem in (4.24).

[2]Note that for the second class, the PDF under $\mathcal{H}_1$ is on the left-hand-side of the PDF under $\mathcal{H}_0$. This class may be viewed as a class of malicious cognitive radios which try to fool the FC under both hypotheses.

The initial distributions are chosen to be K pairs of uniform PDFs which partition the entire range of received data. For each example, the results are obtained using at least 10^3 runs of the simulation. In addition, for all the examples, the maximum number of iterations of the EM algorithm is set to three.

In the first example, we evaluate the densities discriminability versus σ and δ for both kernel-based and histogram-based approaches for the PDF Set 1. The result is shown in Fig. 4.3 for $L = 15$, $T = 20$ and $\Pi = [0.6\ 0.4]$. It can be seen that the best value for δ is larger than the best value for σ. In general, when a Gaussian kernel is used, in order for the histogram- and kernel-based methods to cover the same range of data, we should set δ around 2 or 3 times σ. In the following examples we set $\delta = 3\sigma = (\mathcal{Q}_3 - \mathcal{Q}_1)/20$, where $\mathcal{Q}_1$ and $\mathcal{Q}_3$ are the first and third quartiles of the received data.

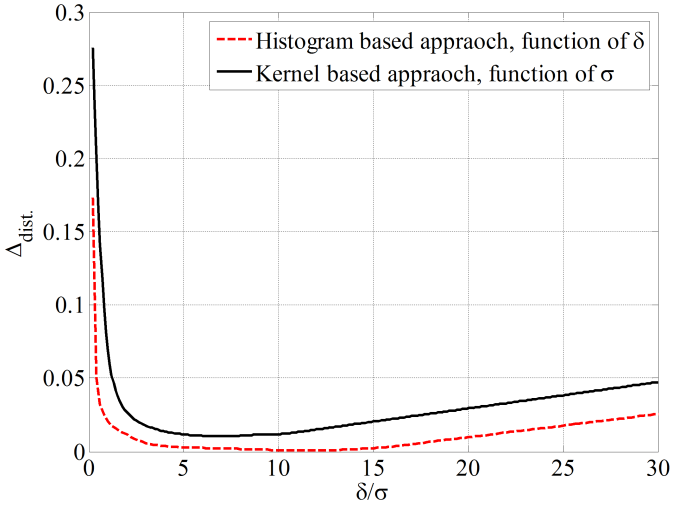

FIGURE 4.3. Discriminability of the PDFs, $\Delta_{\text{dens.}}$, versus δ and σ for histogram based and kernel based approaches when $K = 2$, $L = 15$, $T = 20$, $\Pi = [0.6\ 0.4]$ and for densities set 1.

Fig. 4.4 shows the underlying PDFs from Set 1 and the estimated densities from the proposed algorithm. The system parameters are $K = 2$, $T = 20$ $L = 20$, and $\Pi = [0.7\ 0.3]$. As it can be seen that even for $T = 20$ samples from each cognitive

radio, the estimated densities well approximate the original ones for both classes and under both approaches.

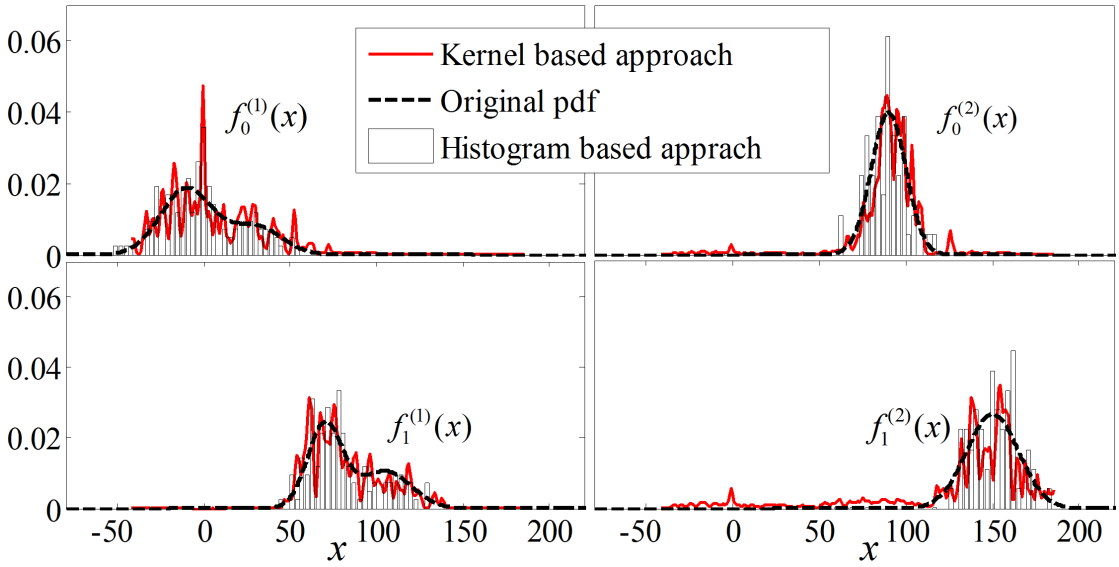

FIGURE 4.4. Estimated densities for histogram based and kernel based approaches and original densities when $K = 2$, $T = 20$ $L = 20$, $\Pi = [0.7\ 0.3]$ and for densities set 1.

In this example, we evaluate the performance of the proposed method using $\Delta_{\text{dist.}}$, Δ_Z, and Δ_H as a function of T. The results are shown in Figs. 4.5, 4.6, and 4.7 for both histogram- and kernel-based methods for PDF Set 1 with $\Pi = [0.6\ 0.4]$ and for PDF Set 2 with $\Pi = [0.4\ 0.4\ 0.2]$ when $L = 15$. As expected, $\Delta_{\text{dist.}}$, Δ_Z, and Δ_H all decrease with the number of data samples. Moreover, it can be seen that given a reasonable data sample size, the proposed method performs well in detecting the hypotheses, classifying the radios and estimating the underlying PDFs.

4.5.1 Computational complexity

The computational complexity is evaluated for both histogram- and kernel-based methods in terms of the time needed to run the MATLAB code for three iterations of the proposed EM algorithm on a Core-2-Duo, 1.8GHz PC, with 1GB RAM. The results vs the number of data samples T are shown in Fig. 4.8 for PDF Set 1 when

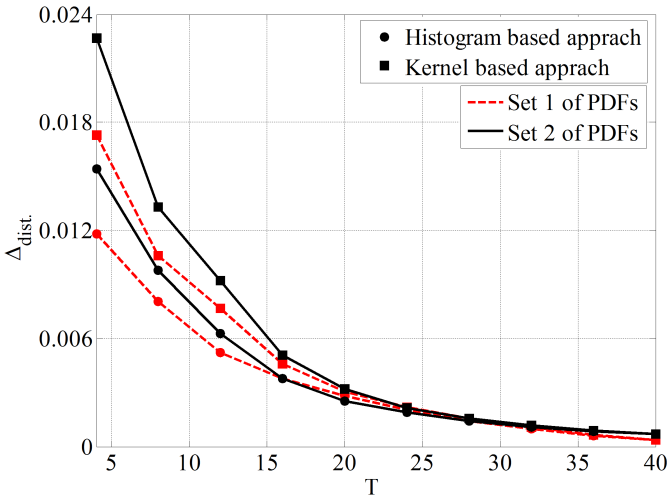

FIGURE 4.5. Discriminability of the PDFs, $\Delta_{\text{dens.}}$, versus T for histogram based and kernel based methods for densities set 1 with $\Pi = [0.6\ 0.4]$ and for densities set 2 with $\Pi = [0.4\ 0.4\ 0.2]$ when $L = 15$.

$L = 20$ and $\Pi = [0.6,\ 0.4]$. As the figure shows the complexity for the histogram-based approach is almost linear in the number of samples while it is exponential for the kernel-based method as a result of the additional optimization algorithms such as the interior point method.

4.6 Time-varying environment

In many practical scenarios the environment is time-varying. As a result the PDFs of the cognitive radio measurements will also vary with time.

Given the limited resources of the cognitive radio networks, we would like to develop a method for the FC to detect the hypotheses and update the PDFs while reducing the data transfer from the cognitive radios. This results in savings in cognitive radios' energy expenditure as well as reduce the required network bandwidth.

To this end we propose a two-phase method referred to as *identification phase* and *tracking phase* in the sequel. Time is divided into frames where each frame consists of an identification phase followed by a tracking phase which consists of Γ time slots as shown in Fig. 4.9.

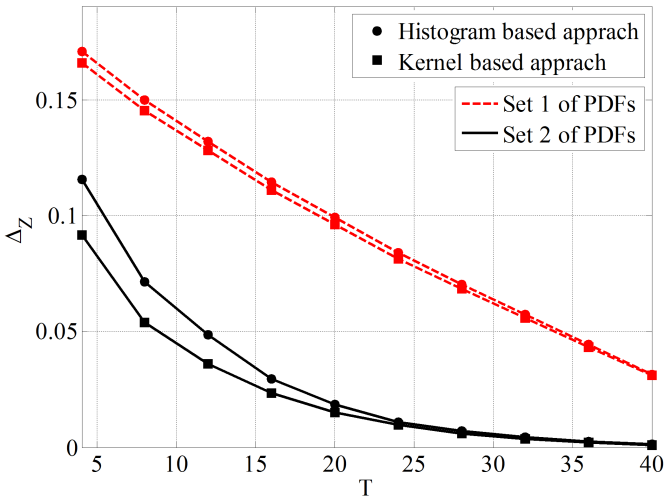

FIGURE 4.6. Misclassification rate, Δ_Z, versus T for histogram based and kernel based methods for densities set 1 with $\Pi = [0.6\ 0.4]$ and for densities set 2 with $\Pi = [0.4\ 0.4\ 0.2]$ when $L = 15$.

We assume that during the identification phase, the sampling rate is high compared to the time variations of the environment so that the PDFs are almost constant during this time. During this phase each radio transmits T samples to the FC. Using the total of LT samples, the FC employs Algorithm 1 to accurately estimate the PDFs, identify the radios, and detect the hypotheses. The PDF estimated for class k under hypothesis $\mathcal{H}_i$ at the end of the identification phase is denoted by $f_i^{(k)}(x;T)$.

The identification phase is followed by the tracking phase. During each slot of the tracking phase every radio sends a single sample to the FC. The FC uses the L samples from the radios to detect the hypothesis and update the PDFs. It is assumed that the identity of the radios do not change during the tracking phase. Therefore, the FC uses the radio identification matrix $\hat{Z}$ and the probability vector Π which were estimated during the identification phase. In the following the index $t = 1, 2, \cdots, \Gamma$ denotes the current slot of the tracking phase, and $d_{j(T+t)}, j = 1, 2, \cdots, L$, $t = 1, 2, \cdots, \Gamma$ denotes the sample received from cognitive radio j during this slot. In addition, we denote by $f_i^{(k)}(x;T+t)$ the PDF of class

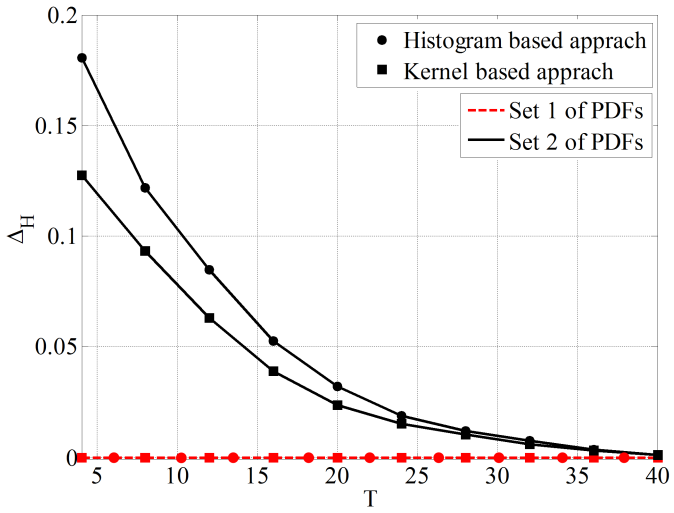

FIGURE 4.7. Hypotheses discriminability, Δ_H, versus T for histogram based and kernel based methods for densities set 1 with $\Pi = [0.6\ 0.4]$ and for densities set 2 with $\Pi = [0.4\ 0.4\ 0.2]$ when $L = 15$.

k under hypothesis i estimated during slot t. The likelihood ratio test is used for the detection of the hypothesis as follows.

$$\frac{P(d_{1(T+t)}, d_{2(T+t)}, \cdots, d_{L(T+t)}|\hat{Z}, \hat{\Theta}, \mathcal{H}_1)}{P(d_{1(T+t)}, d_{2(T+t)}, \cdots, d_{L(T+t)}|\hat{Z}, \hat{\Theta}, \mathcal{H}_0)} = \frac{\prod_{l=1}^{L}\prod_{k=1}^{K}\left(f_1^{(k)}(d_{l(T+t)}; T+t-1)\right)^{z_{lk}}}{\prod_{l=1}^{L}\prod_{k=1}^{K}\left(f_0^{(k)}(d_{l(T+t)}; T+t-1)\right)^{z_{lk}}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \xi \tag{4.35}$$

Note that the estimated PDFs during slot $t-1$ of the tracking phase are used in the hypothesis detection during slot t.

After detecting the hypothesis for slot t, the PDFs for this slot are updated according to the following.

$$f_i^{(k)}(x; T+t) = \frac{T+t-1}{T+t} f_i^{(k)}(x; T+t-1) + \frac{1}{T+t} C_i^{(k)}(x; t) \tag{4.36}$$

where $C_i^{(k)}(x; t)$ is a PDF correction term for class k and hypothesis $\mathcal{H}_i$, and is obtained from the following equations for the histogram- and kernel-based ap-

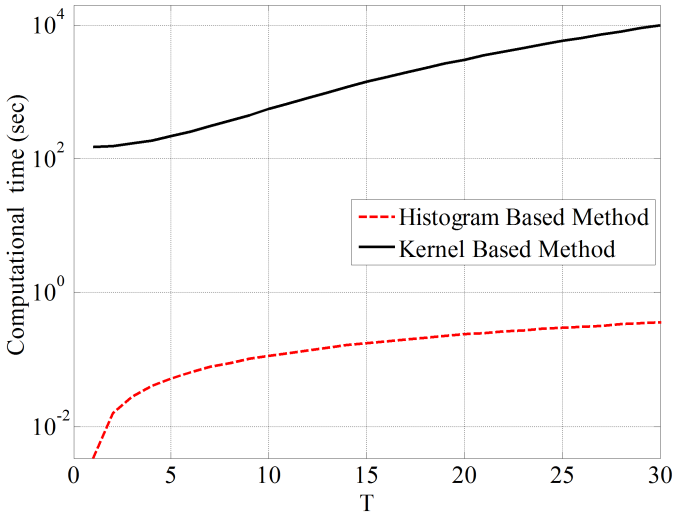

FIGURE 4.8. Computational time versus T, for histogram based and kernel based approaches for densities set 1 when $L = 20$, $\Pi = [0.6,\ 0.4]$.

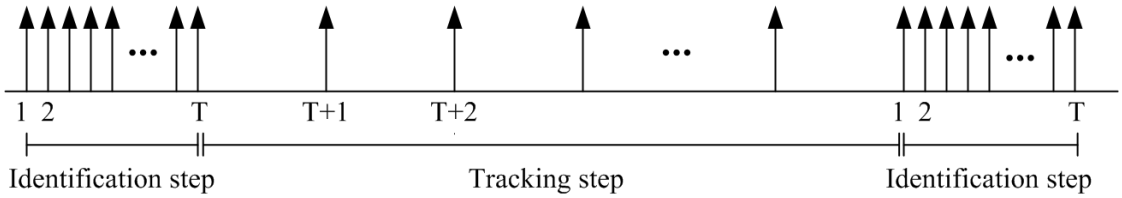

FIGURE 4.9. Sampling rate in the identification step and tracking step. The radios decrease the sampling rate after identification step.

proaches, respectively.

$$C_i^{(k)}(x_0 + m\delta; t) = \frac{1}{\delta \sum_{l=1}^{L} z_{lk}} \sum_{\{l:\ (l,T+t)\in\mathcal{B}_m\}} z_{lk} \tag{4.37}$$

$$C_i^{(k)}(x; t) = \frac{1}{\sigma \sum_{l=1}^{L} z_{lk}} \sum_{l=1}^{L} z_{lk} \Psi\left(\frac{x - d_{l(T+t)}}{\sigma}\right) \tag{4.38}$$

At the end of the tracking phase, a new time frame starts with its own identification phase and tracking phase. The duration of the time frame, the identification and tracking phases should be chosen in accordance with the rate of change in the environment and the requirements of hypothesis detection. However, data transfer rate during the tracking phase is much lower than that in the identification phase.

To evaluate the performance of this method, we modify the densities in set 1 using an autoregressive (AR) model as follows:

$$f_i^{(k)}(x;T+t) = (1-\rho) f_i^{(k)}(x;T+t-1) + \rho w_i^{(k)}(x),$$
$$t = 1, 2, \cdots, \Gamma, \tag{4.39}$$

where $f_i^{(k)}(x;T) = f_i^{(k)}(x)$, $w_0^{(1)}(x) = g(10, 225)$, $w_1^{(1)}(x) = g(90, 150)$, $w_0^{(2)}(x) = g(100, 100)$, and $w_1^{(2)}(x) = g(150, 150)$. Fig. 4.10 shows the result for different values of ρ when $T = 40$, $L = 15$ and $\Pi = [0.6\ 0.4]$. It can be seen that for values of $\rho = 0.02, 0.05$, the estimation of the PDFs improves during the tracking phase as $\Delta_{\text{dist.}}$ deceases with t. For a highly time-varying environment when $\rho = 0.1$, the PDFs estimation error $\Delta_{\text{dist.}}$ increases. However, even in this case the increase in the estimation error is modest. Therefore if the estimation from the identification phase is good (small value of $\Delta_{\text{dist.}}$ at $t = 0$), then the tracking phase may be applied for a number of slots and still maintain acceptable performance in terms of PDF estimation and hypothesis detection. We should point out that for this example, the Hypotheses discriminability, $\Delta_H = 0$ for all values of $t = 1, 2, \cdots, 20$ and therefore it is not plotted here.

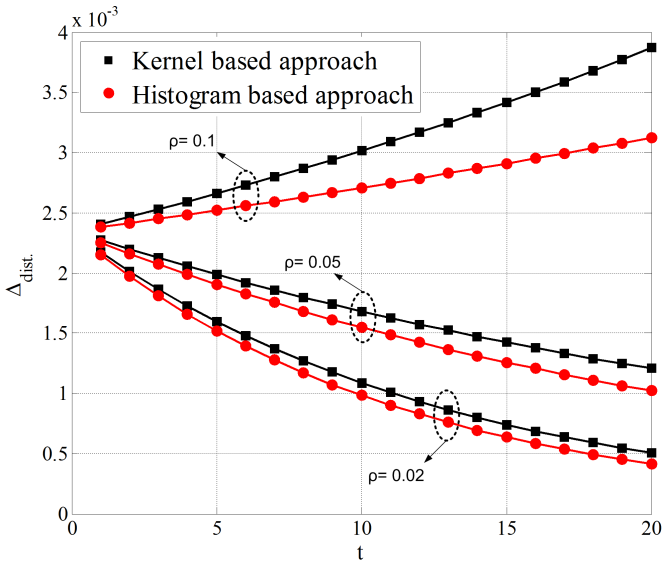

FIGURE 4.10. $\Delta_{\text{dist.}}$ vs. t for different values of ρ when $T = 40$, $L = 15$ and $\Pi = [0.6\ 0.4]$ and when the PDFs from set 1 are modified using the AR model in (4.39).

Chapter 5
Conclusion

In Chapter 1, a GLRT-based spectrum sensing technique is proposed when both the PU transmitter and the SU receiver use multiple antennas. No assumptions are made on the PU's signal structure and statistics and about the channel coefficients. Two test statistics are derived when the receiver knows the energy of the noise and when it does not. The analysis shows that the M largest eigenvalues of sample covariance matrix correspond to the signal+noise subspace (where M is the number of PU antennas) and the remaining eigenvalues correspond to the noise subspace. The performance of the proposed method is presented in terms of the receiver operating characteristics and detection probabilities for several cases of interest and compared to two recently proposed techniques in the literature.

In Chapter 2, we investigate the problem of spectrum monitoring over Rayleigh fading channels. It is assumed that the secondary user is equipped with multiple antennas and uses diversity combining to mitigate the effects of fading. We consider maximal ratio combining, equal gain combining and selection combining. It is shown that spectrum monitoring using REC alone is not as effective in the case of fading channels as it is in the case of AWGN channels. Next we introduce new decision statistics based on the REC and the combiner coefficients for the three combining schemes. It is shown that only in the case of MRC the combiner coefficients improve the decision statistic over the REC alone. Numerical results are presented to compare the performance of the hybrid spectrum sensing/spectrum monitoring technique with spectrum sensing alone. The results show that the proposed decision statistic significantly outperforms the decision statistic using REC

alone. Moreover the hybrid spectrum sensing/spectrum monitoring significantly outperforms spectrum sensing alone.

In Chapter 3, we have addressed the problem of cooperative spectrum sensing in cognitive radio networks (CRNs) in the presence of misbehaving cognitive radios (CRs). No prior information on the parameters of the network is assumed except that a class of honest radios is in majority. The iterative expectation maximization (EM) algorithm is formulated and solved for the detection of the hypotheses and the classification of the radios. In contrast to other recently proposed methods, our approach can classify the radios into more than two classes. This applies in cases where the honest CRs may employ different spectrum sensing techniques or encounter dissimilar channel and noise conditions resulting in different detection and false alarm probabilities. Another case is when the CRN includes more than one type of misbehaving CRs such as both malicious and malfunctioning radios. Our numerical results show significant improvements over the widely popular reputation-based classifier (RBC). In particular, with only a few decisions from the radios the proposed algorithm can quickly and efficiently classify the CRs whereas the RBC method fails. The numerical results are also compared with the Cramer-Rao lower bound and show a close match.

In Chapter 4, we consider the problem of centralized binary hypothesis testing in a cognitive radio network (CRN) consisting of multiple classes of cognitive radios, where each class consists of all the cognitive radios whose received data, under each hypothesis, is drawn from the same probability density function (PDF). We proposed a method based on the expectation maximization (EM) algorithm to estimate the probability density function of cognitive radios' data, to classify the cognitive radios and to detect the hypotheses. To estimate the PDFs we propose two nonparametric approaches; a histogram-based approach and a kernel-based ap-

proach. Numerical results obtained using only three iterations of the EM algorithm show that the proposed method is effective in estimating the PDFs, classifying the cognitive radios and detecting the hypotheses.

References

[1] I. Mitola, J. and J. Maguire, G.Q., "Cognitive radio: making software radios more personal," *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13 –18, aug 1999.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 23, no. 2, pp. 201 – 220, feb. 2005.

[3] M. Naraghi-Pour and T. Ikuma, "Autocorrelation-based spectrum sensing for cognitive radios," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 2, pp. 718 –733, 2010.

[4] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *Communications Surveys Tutorials, IEEE*, vol. 11, no. 1, pp. 116 –130, quarter 2009.

[5] D. Ariananda, M. Lakshmanan, and H. Nikookar, "A survey on spectrum sensing techniques for cognitive radio," in *Cognitive Radio and Advanced Spectrum Management, 2009. CogART 2009. Second International Workshop on*, may 2009, pp. 74 –79.

[6] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 2, pp. 814 –823, february 2010.

[7] R. Zhang, T. Lim, Y.-C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A glrt approach," *Communications, IEEE Transactions on*, vol. 58, no. 1, pp. 84 –88, january 2010.

[8] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 4, pp. 1791 –1800, may 2010.

[9] M. Orooji, R. Soosahabi, and M. Naraghi-Pour, "Blind spectrum sensing using antenna arrays and path correlation," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 8, pp. 3758 –3767, oct. 2011.

[10] S. Alamouti, "A simple transmit diversity technique for wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1451 –1458, oct 1998.

[11] 3GPP, LTE Release 9. 2009.

[12] IEEE Standard 802.11n-2009. Part 11, 2009.

[13] IEEE Computer Society and the IEEE Microwave Theory and Techniques Society, IEEE Standard 802.16-2009. Part 16: Air Interface for Broadband Wireless Access Systems 2009.

[14] J. Font-Segura and X. Wang, "Glrt-based spectrum sensing for cognitive radio with prior information," *Communications, IEEE Transactions on*, vol. 58, no. 7, pp. 2137 –2146, july 2010.

[15] Y. Zeng and Y. chang Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *Communications, IEEE Transactions on*, vol. 57, no. 6, pp. 1784 –1793, june 2009.

[16] A. Kortun, T. Ratnarajah, M. Sellathurai, and C. Zhong, "On the performance of eigenvalue-based spectrum sensing for cognitive radio," in *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*, april 2010, pp. 1 –6.

[17] T. Ratnarajah, C. Zhong, A. Kortun, M. Sellathurai, and C. B. Papadias, "Complex random matrices and multiple-antenna spectrum sensing," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, may 2011, pp. 3848 –3851.

[18] S. Haykin, *Adaptive filter theory*, 4th ed. Upper Saddle River, NJ: Prentice Hall, 2002.

[19] D. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Microwaves, Optics and Antennas, IEE Proceedings H*, vol. 130, no. 1, pp. 11 –16, february 1983.

[20] R. Horn and C. R. Johnson, *Matrix Analysis*, 1985.

[21] J. W. Demmel, *Applied Numerical Linear Algebra*. SIAM, 1997.

[22] H. Jafarkhani, *Space-time coding: theory and practice*. Cambridge Univ Pr, 2005.

[23] S. Chen, A. Wyglinski, R. Vuyyuru, and O. Altintas, "Feasibility analysis of vehicular dynamic spectrum access via queueing theory model," in *Vehicular Networking Conference (VNC), 2010 IEEE*, dec. 2010, pp. 223 –230.

[24] K. Tsukamoto, Y. Omori, O. Altintas, M. Tsuru, and Y. Oie, "On spatially-aware channel selection in dynamic spectrum access multi-hop inter-vehicle communications," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, sept. 2009, pp. 1 –7.

[25] H. Li and D. Irick, "Collaborative spectrum sensing in cognitive radio vehicular ad hoc networks: Belief propagation on highway," in *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, may 2010, pp. 1 –5.

[26] J.-H. Chu, K.-T. Feng, C.-N. Chuah, and C.-F. Liu, "Cognitive radio enabled multi-channel access for vehicular communications," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, sept. 2010, pp. 1 –5.

[27] X. Y. Wang and P.-H. Ho, "A novel sensing coordination framework for cr-vanets," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 4, pp. 1936 –1948, may 2010.

[28] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *COMPUTER NETWORKS JOURNAL (ELSEVIER*, vol. 50, pp. 2127–2159, 2006.

[29] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *Communications Surveys Tutorials, IEEE*, vol. 11, no. 1, pp. 116 –130, quarter 2009.

[30] D. Ariananda, M. Lakshmanan, and H. Nikoo, "A survey on spectrum sensing techniques for cognitive radio," in *Cognitive Radio and Advanced Spectrum Management, 2009. CogART 2009. Second International Workshop on*, may 2009, pp. 74 –79.

[31] B. Wang and K. Liu, "Advances in cognitive radio networks: A survey," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 1, pp. 5 –23, feb. 2011.

[32] J. Oh and W. Choi, "A hybrid cognitive radio system: A combination of underlay and overlay approaches," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, sept. 2010, pp. 1 –5.

[33] "Draft standard for wireless regional area networks," Mar. 2008.

[34] Y.-C. Liang, Y. Zeng, E. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 4, pp. 1326 –1337, april 2008.

[35] L. Tang, Y. Chen, E. Hines, and M.-S. Alouini, "Effect of primary user traffic on sensing-throughput tradeoff for cognitive radios," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 4, pp. 1063 –1068, april 2011.

[36] S. Akin and M. Gursoy, "Effective capacity analysis of cognitive radio channels for quality of service provisioning," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3354 –3364, november 2010.

[37] A. Hoang, Y.-C. Liang, and Y. Zeng, "Adaptive joint scheduling of spectrum sensing and data transmission in cognitive radio networks," *Communications, IEEE Transactions on*, vol. 58, no. 1, pp. 235 –246, january 2010.

[38] Q. Zhao, S. Geirhofer, L. Tong, and B. Sadler, "Opportunistic spectrum access via periodic channel sensing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 2, pp. 785 –796, feb. 2008.

[39] S. Boyd and M. Pursley, "Enhanced spectrum sensing techniques for dynamic spectrum access cognitive radio networks," in *MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010*, 31 2010-nov. 3 2010, pp. 317 –322.

[40] S. Boyd, J. Frye, M. Pursley, and T. Royster IV, "Spectrum monitoring during reception in dynamic spectrum access cognitive radio networks," *Communications, IEEE Transactions on*, vol. PP, no. 99, pp. 1 –12, 2011.

[41] M. Orooji, E. Soltanmohammadi, and M. Naraghi-Pour, "Enhancing sensing-throughput tradeoff in cognitive radios using receiver statistics," in *Submitted to Vehicular Technology, IEEE Transactions on.*

[42] M. Naraghi-Pour and T. Ikuma, "Diversity techniques for spectrum sensing in fading environments," in *Military Communications Conference, 2008. MILCOM 2008. IEEE*, nov. 2008, pp. 1 –7.

[43] ——, "Autocorrelation-based spectrum sensing for cognitive radios," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 2, pp. 718 –733, 2010.

[44] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 2, pp. 814 –823, february 2010.

[45] R. Soosahabi, M. Orooji, and M. Naraghi-Pour, "Multi-antenna blind spectrum sensing for cognitive radios using path correlations," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, dec. 2011, pp. 1 –5.

[46] R. Annavajjala and L. Milstein, "Performance analysis of linear diversity-combining schemes on rayleigh fading channels with binary signaling and gaussian weighting errors," *Wireless Communications, IEEE Transactions on*, vol. 4, no. 5, pp. 2267 – 2278, sept. 2005.

[47] D. Brennan, "Linear diversity combining techniques," *Proceedings of the IEEE*, vol. 91, no. 2, pp. 331 – 356, feb 2003.

[48] A. S. Tanenbaum and D. J. Wetherall, *Computer Networks*, 5th ed. Prentice Hall, 2010.

[49] K. Witzke and C. Leung, "A comparison of some error detecting crc code standards," *Communications, IEEE Transactions on*, vol. 33, no. 9, pp. 996 – 998, sep 1985.

[50] A. Leon-Garcia and I. Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures*, 2nd ed. New York: McGraw Hill, 2004.

[51] G. Castagnoli, S. Brauer, and M. Herrmann, "Optimization of cyclic redundancy-check codes with 24 and 32 parity bits," *Communications, IEEE Transactions on*, vol. 41, no. 6, pp. 883 –892, jun 1993.

[52] P. Koopman, "32-bit cyclic redundancy codes for internet applications," in *Dependable Systems and Networks, 2002. DSN 2002. Proceedings. International Conference on*, 2002, pp. 459 – 468.

[53] R. Zhang, T. Lim, Y.-C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A glrt approach," *Communications, IEEE Transactions on*, vol. 58, no. 1, pp. 84 –88, january 2010.

[54] D. Ramirez, G. Vazquez-Vilar, R. Lopez-Valcarce, J. Via, and I. Santamaria, "Detection of rank- signals in cognitive radio networks with uncalibrated multiple antennas," *Signal Processing, IEEE Transactions on*, vol. 59, no. 8, pp. 3764 –3774, aug. 2011.

[55] G. L. Stuber, "Principles of mobile communication," New York, 2002.

[56] M. Gans, "The effect of gaussian error in maximal ratio combiners," *Communication Technology, IEEE Transactions on*, vol. 19, no. 4, pp. 492 –500, august 1971.

[57] A. Goldsmith, "Wireless communications," New York, 2005.

[58] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. England: John Wiley & Sons, 2001.

[59] W. C. Jakes, "Microwave mobile communications," New York, 1994.

[60] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*, 1st ed. Upper Saddle River, New Jersey: Prentice-Hall, 1995.

[61] J. Ma, G. Li, and B. H. Juang, "Signal processing in cognitive radio," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 805 –823, may 2009.

[62] R. Zhang, T. Lim, Y.-C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A GLRT approach," *Communications, IEEE Transactions on*, vol. 58, no. 1, pp. 84 –88, jan. 2010.

[63] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *Vehicular Technology, IEEE Transactions on*, vol. 59, no. 4, pp. 1791 –1800, May 2010.

[64] M. Gandetto and C. Regazzoni, "Spectrum sensing: A distributed approach for cognitive terminals," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 3, pp. 546 –557, april 2007.

[65] G. Ganesan and L. Ye, "Cooperative spectrum sensing in cognitive radio, part ii: Multiuser networks," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 6, pp. 2214 –2222, june 2007.

[66] FCC, "Second Report and Order and Memorandum Opinion and Order, ET Docket No. 04-186," *Federal Communications Commission*, Nov. 2008.

[67] A. Fragkiadakis, E. Tragos, and I. Askoxylakis, "A survey on security threats and detection techniques in cognitive radio networks," *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1 –18, 2012.

[68] S. Mishra, A. Sahai, and R. Brodersen, "Cooperative sensing among cognitive radios," in *Communications, 2006. ICC '06. IEEE International Conference on*, vol. 4, june 2006, pp. 1658 –1663.

[69] P. Kaligineedi, M. Khabbazian, and V. Bhargava, "Secure cooperative sensing techniques for cognitive radio systems," in *Communications, 2008. ICC '08. IEEE International Conference on*, may 2008, pp. 3406 –3410.

[70] ——, "Malicious user detection in a cognitive radio cooperative sensing system," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 8, pp. 2488 –2497, august 2010.

[71] A. Min, K. Shin, and X. Hu, "Secure cooperative sensing in ieee 802.22 wrans using shadow fading correlation," *Mobile Computing, IEEE Transactions on*, vol. 10, no. 10, pp. 1434–1447, 2011.

[72] A. Ghasemi and E. Sousa, "Asymptotic performance of collaborative spectrum sensing under correlated log-normal shadowing," *Communications Letters, IEEE*, vol. 11, no. 1, pp. 34–36, 2007.

[73] L. Duan, A. Min, J. Huang, and K. Shin, "Attack prevention for collaborative spectrum sensing in cognitive radio networks," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 9, pp. 1658–1665, 2012.

[74] P. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag New York, Inc., 1997.

[75] L. Duan, A. Min, J. Huang, and K. Shin, "Attack prevention for collaborative spectrum sensing in cognitive radio networks," http://arxiv.org/abs/1109.1021, 2011, [Online; Technical Report].

[76] A. Rawat, P. Anand, H. Chen, and P. Varshney, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks," *Signal Processing, IEEE Transactions on*, vol. 59, no. 2, pp. 774 –786, feb. 2011.

[77] R. Chen, J.-M. Park, and K. Bian, "Robust distributed spectrum sensing in cognitive radio networks," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, april 2008, pp. 1876 –1884.

[78] H. Chen, X. Jin, and L. Xie, "Reputation-based collaborative spectrum sensing algorithm in cognitive radio networks," in *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, sept. 2009, pp. 582 –587.

[79] A. Vempaty, K. Agrawal, H. Chen, and P. Varshney, "Adaptive learning of byzantines' behavior in cooperative spectrum sensing," in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*, march 2011, pp. 1310 –1315.

[80] A. Rawat, P. Anand, H. Chen, and P. Varshney, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks," *Signal Processing, IEEE Transactions on*, vol. 59, no. 2, pp. 774 –786, feb. 2011.

[81] F. Adelantado and C. Verikoukis, "A non-parametric statistical approach for malicious users detection in cognitive wireless ad-hoc networks," in *Communications (ICC), 2011 IEEE International Conference on*, june 2011, pp. 1 –5.

[82] F. Penna, Y. Sun, L. Dolecek, and D. Cabric, "Detecting and counteracting statistical attacks in cooperative spectrum sensing," *Signal Processing, IEEE Transactions on*, vol. 60, no. 4, pp. 1806 –1822, april 2012.

[83] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

[84] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[85] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[86] D. J. Hand, "Construction and assessment of classification rules," Chichester, West Sussex, England: John Wiley & Sons, 1997.

[87] R. Srinivasan, "Distributed radar detection theory," *Communications, Radar and Signal Processing, IEE Proceedings F*, vol. 133, no. 1, pp. 55 –60, february 1986.

[88] Q. Tian and E. Coyle, "Optimal distributed detection in clustered wireless sensor networks," *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3892 –3904, july 2007.

[89] J. N. Tsitsiklis, "Decentralized detection," in *In Advances in Statistical Signal Processing*. JAI Press, 1993, pp. 297–344.

[90] J.-F. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *Signal Processing, IEEE Transactions on*, vol. 51, no. 2, pp. 407–416, 2003.

[91] B. Liu and B. Chen, "Channel-optimized quantizers for decentralized detection in sensor networks," *Information Theory, IEEE Transactions on*, vol. 52, no. 7, pp. 3349–3358, 2006.

[92] R. Tenney and N. Sandell, "Detection with distributed sensors," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-17, no. 4, pp. 501 –510, july 1981.

[93] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-22, no. 1, pp. 98 –101, jan. 1986.

[94] A. Reibman and L. Nolte, "Optimal detection and performance of distributed sensor systems," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-23, no. 1, pp. 24 –30, jan. 1987.

[95] Q. Zhang, P. Varshney, and R. Wesel, "Optimal bi-level quantization of i.i.d. sensor observations for binary hypothesis testing," *Information Theory, IEEE Transactions on*, vol. 48, no. 7, pp. 2105–2111, 2002.

[96] A. Dogandzic and B. Zhang, "Nonparametric probability density estimation for sensor networks using quantized measurements," in *Information Sciences and Systems, 2007. CISS '07. 41st Annual Conference on*, 2007, pp. 759–764.

[97] E. Soltanmohammadi, M. Orooji, and M. Naraghi-Pour, "Decentralized hypothesis testing in wireless sensor networks in the presence of misbehaving nodes," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 1, pp. 205–215, 2013.

[98] P. Anand, A. Rawat, H. Chen, and P. Varshney, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks," in *Communication Systems and Networks (COMSNETS), 2010 Second International Conference on*, jan. 2010, pp. 1 –9.

[99] P. Kaligineedi, M. Khabbazian, and V. Bhargava, "Malicious user detection in a cognitive radio cooperative sensing system," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 8, pp. 2488–2497, 2010.

[100] A. Mukherjee and D. Mukherjee, "Distributed probability density function estimation of environmental function from sensor network data," in *Signal Processing Image Processing Pattern Recognition (ICSIPR), 2013 International Conference on*, 2013, pp. 346–350.

[101] R. Nowak, "Distributed em algorithms for density estimation and clustering in sensor networks," *Signal Processing, IEEE Transactions on*, vol. 51, no. 8, pp. 2245–2253, 2003.

[102] P. Forero, A. Cano, and G. Giannakis, "Consensus-based distributed expectation-maximization algorithm for density estimation and classification using wireless sensor networks," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 1989–1992.

[103] B. Safarinejadian, M.-B. Menhaj, and M. Karrari, "Distributed unsupervised gaussian mixture learning for density estimation in sensor networks," *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 9, pp. 2250–2260, 2010.

[104] B. Silverman, *Density Estimation for Statistics and Data Analysis*. New York, NY, USA: Chapman and Hall, 1986.

[105] D. Doane, "Aesthetic frequency classification," *American Statistician*, vol. 30, no. 4, pp. 181 – 183, Nov. 1976.

[106] D. W. Scott, "On optimal and data-based histograms," vol. 66, 1979, pp. 605–610. [Online]. Available: http://www.mendeley.com/research/on-optimal-and-databased-histograms/

[107] I. S. Gradshteyn and I. M. Ryshik, *Table of Integrals, Series, and Products*, 5th ed. London: Academic Press, 1994.

Appendix A: Evaluation of $p(\psi|\underline{\hat{h}}, H_\eta)$

First let us find $p_{h_l^R|\hat{h}_l^R}(x|y, H_\eta)$. We have

$$p_{h_l^R|\hat{h}_l^R}(x|y, H_\eta) = \frac{p_{\hat{h}_l^R|h_l^R}(y|x, H_\eta)p_{h_l^R}(x)}{p_{\hat{h}_l^R}(y, H_\eta)} = \frac{p_{\ell_l^R}(y-x|H_\eta)p_{h_l^R}(x)}{p_{\hat{h}_l^R}(y|H_\eta)} \tag{5.1}$$

Note that h_l^R is independent of the hypothesis H_η. It is discussed in section 2.2 that, ℓ_l and h_l are two independent zero-mean circular Gaussian random variables. Thus

$$h_l^R \text{ and } h_l^I \quad \sim \mathcal{N}(0, 1/2) \tag{5.2}$$

$$\ell_l^R|H_\eta \text{ and } \ell_l^I|H_\eta \sim \mathcal{N}(0, 1/2 - \rho_\eta^2/2) \tag{5.3}$$

$$\hat{h}_l^R|H_\eta \text{ and } \hat{h}_l^I|H_\eta \sim \mathcal{N}(0, 1 - \rho_\eta^2) \tag{5.4}$$

By substituting (5.2), (5.3), and (5.4) into (5.1) and after some manipulations one can show that,

$$h_l^R|\hat{h}_l^R, H_\eta \sim \mathcal{N}\left(\frac{\hat{h}_l^R}{2-\rho_\eta^2}\ ,\ \frac{1-\rho_\eta^2}{2(2-\rho_\eta^2)}\right) \tag{5.5}$$

Similarly the distribution of $h_l^I|\hat{h}_l^I, H_\eta$ can be derived. From (5.5) we can write,

$$\frac{\hat{h}_l^R h_l^R}{\sqrt{\sum_{l=1}^L |\hat{h}_l|^2}}\ \Bigg|\ \hat{h}_l, H_\eta \tag{5.6}$$

$$\sim \mathcal{N}\left(\frac{(\hat{h}_l^R)^2}{(2-\rho_\eta^2)(\sqrt{\sum_{l=1}^L |\hat{h}_l|^2})}, \frac{(1-\rho_\eta^2)(\hat{h}_l^R)^2}{2(2-\rho_\eta^2)(\sum_{l=1}^L |\hat{h}_l|^2)}\right)$$

and in the same way, one can rewrite (5.6) for the imaginary parts. Let us rewrite ψ in (2.15) as,

$$\psi = \frac{\mathrm{Re}\left(\sum_{l=1}^L h_l \hat{h}_l^*\right)}{\sqrt{\sum_{l=1}^L |\hat{h}_l|^2}} = \frac{\sum_{l=1}^L \left(h_l^R \hat{h}_l^R + h_l^I \hat{h}_l^I\right)}{\sqrt{\sum_{l=1}^L |\hat{h}_l|^2}} \tag{5.7}$$

Then (5.6) and (5.7) imply that,

$$\psi|\underline{\hat{h}}, H_\eta \sim \mathcal{N}\left(\frac{\sqrt{\sum_{l=1}^L |\hat{h}_l|^2}}{2-\rho_\eta^2}\ ,\ \frac{1-\rho_\eta^2}{2(2-\rho_\eta^2)}\right) \tag{5.8}$$

Appendix B:
The approximation of $p(k, \hat{\mathcal{A}}|H_\eta)$

To make a decision on the hypothesis, the integral in (2.16) should be evaluated and multiplied by (2.14) and (2.23).

For $0 < k < N$, let us approximate,

$$\binom{N}{k} Q^k(x)(1-Q(x))^{N-k} \approx a e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{5.9}$$

where m, a and σ are the solutions of following equations,

$$\binom{N}{k} Q^k(x)(1-Q(x))^{N-k}\Big|_{x=m} = a \tag{5.10}$$

$$\binom{N}{k} Q^k(x)(1-Q(x))^{N-k}\Big|_{x=m} = 0 \tag{5.11}$$

$$\binom{N}{k} Q^k(x)(1-Q(x))^{N-k}\Big|_{x=m} = -\frac{a}{\sigma^2} \tag{5.12}$$

Equations (5.10), (5.11) and (5.12) give

$$m = Q^{-1}(\frac{k}{N}) \tag{5.13}$$

$$a = \binom{N}{k} \frac{k^k (N-k)^{N-k}}{N^N} \tag{5.14}$$

$$\sigma^2 = 2\pi \frac{k^2}{N^2} (\frac{1}{k} - \frac{1}{N}) \exp\left((Q^{-1}(\frac{k}{N}))^2 \right) \tag{5.15}$$

where Q^{-1} is inverse Q-function.

For cases $k = 0$ and $k = N$, the left hand-side of (5.9) is equal to $(1-Q(x))^N$ and $Q^N(x)$, respectively, which are approximated by $U(x-m_0)$ and $1-U(x-m_N)$, respectively, where $U(.)$ is the unit step-function, and m_0 and m_N are the solutions of following equations,

$$(1-Q(x))^N\big|_{x=m_0} = \frac{1}{2} \tag{5.16}$$

$$Q^N(x)\big|_{x=m_N} = \frac{1}{2} \tag{5.17}$$

This gives,

$$(1-Q(x))^N \approx U(x - Q^{-1}(1 - \sqrt[N]{1/2})) \tag{5.18}$$

$$Q^N(x) \approx 1 - U(x - Q^{-1}(\sqrt[N]{1/2})) \tag{5.19}$$

By substituting (5.9), (5.18), and (5.19) into (2.19), $p(k|\underline{\hat{h}}, H_\eta)$ is approximated by

$$p(k|\underline{\hat{h}}, H_\eta) \approx \begin{cases} \frac{1}{\sqrt{2\pi\acute{\sigma}_\eta^2}} \int_{m_0}^{\infty} e^{\frac{-(\acute{m}_\eta - x)^2}{2\acute{\sigma}_\eta^2}}\, dx, & k = 0 \\ \frac{a}{\sqrt{2\pi\acute{\sigma}_\eta^2}} \int_{-\infty}^{\infty} e^{-\frac{(\acute{m}_\eta - x)^2}{2\acute{\sigma}_\eta^2} - \frac{(m-x)^2}{2\sigma^2}}\, dx, & 0 < k < N \\ \frac{1}{\sqrt{2\pi\acute{\sigma}_\eta^2}} \int_{-\infty}^{m_N} e^{\frac{-(\acute{m}_\eta - x)^2}{2\acute{\sigma}_\eta^2}}\, dx, & k = N \end{cases} \tag{5.20}$$

where $\acute{m}_\eta \triangleq \frac{\hat{\mathcal{A}}\sqrt{2\gamma_\eta}}{2-\rho_\eta^2}$ and $\acute{\sigma}_\eta^2 \triangleq \gamma_\eta \frac{1-\rho_\eta^2}{2-\rho_\eta^2}$. Approximation of $p(k|\underline{\hat{h}}, H_\eta)$ in (5.20) is found by considering that,

$$\int_x^{\infty} e^{-t^2/2}\, dt = \sqrt{2\pi}\, Q(x) \tag{5.21}$$

$$\int_{-\infty}^{\infty} e^{-(b_2 t^2 + b_1 t + b_0)}\, dt = \sqrt{\frac{\pi}{b_2}}\, e^{\frac{b_1^2 - 4b_2 b_0}{4b_2}} \tag{5.22}$$

Finally by substituting the approximation of $p(k|\underline{\hat{h}}, H_\eta)$ into (2.12), $p(k, \hat{\mathcal{A}}|H_\eta)$ is approximated by (5.23). Comparision between the simulation results and analysis results (see Fig. 2.8) shows the accuracy of this approximation.

$$p(k, \hat{\mathcal{A}}|H_\eta) \approx \begin{cases} Q\left(\frac{Q^{-1}(1 - \sqrt[N]{1/2}) - \acute{m}_\eta}{\acute{\sigma}_\eta}\right) \frac{2\hat{\mathcal{A}}^{(2L-1)} e^{-\hat{\mathcal{A}}^2/(2-\rho_\eta^2)}}{(2-\rho_\eta^2)^L (L-1)!}, & k = 0 \\ \frac{a\acute{\sigma}_\eta}{\sqrt{\sigma^2 + \acute{\sigma}_\eta^2}} \exp\left(\frac{-(m - \acute{m}_\eta)^2}{2(\sigma^2 + \acute{\sigma}_\eta^2)}\right) \frac{2\hat{\mathcal{A}}^{(2L-1)} e^{-\hat{\mathcal{A}}^2/(2-\rho_\eta^2)}}{(2-\rho_\eta^2)^L (L-1)!}, & 0 < k < N \\ \left(1 - Q\left(\frac{Q^{-1}(\sqrt[N]{1/2}) - \acute{m}_\eta}{\acute{\sigma}_\eta}\right)\right) \frac{2\hat{\mathcal{A}}^{(2L-1)} e^{-\hat{\mathcal{A}}^2/(2-\rho_\eta^2)}}{(2-\rho_\eta^2)^L (L-1)!}, & k = N \end{cases} \tag{5.23}$$

Appendix C:
Proof of Independence of $\zeta|H_\eta$ and $\underline{\hat{\theta}}|H_\eta$

The goal is to prove $p(\zeta|\underline{\hat{\theta}}, H_\eta) = p(\zeta|H_\eta)$. Let us rewrite

$$\zeta = \sum_{l=1}^{L} z_l \tag{5.24}$$

where

$$z_l \triangleq \frac{\mathrm{Re}\left(h_l e^{-j\hat{\theta}_l}\right)}{\sqrt{L}} = \frac{h_l^R \cos\hat{\theta}_l + h_l^I \sin\hat{\theta}_l}{\sqrt{L}} \tag{5.25}$$

Since $\{h_l\}_{l=1}^{L}$ are i.i.d. random variables, given $\underline{\hat{\theta}}$ and H_η the distribution of ζ is the convolution of the distribution of $z_1, z_2, \ldots, z_l$. Besides,

$$p(z_l|\hat{\theta}_l, H_\eta) = \int_{-\infty}^{\infty} p(z_l|\hat{\theta}_l, \hat{\alpha}_l, H_\eta)\ p(\hat{\alpha}_l|\hat{\theta}_l, H_\eta)\ d\hat{\alpha}_l \tag{5.26}$$

Since $\hat{h}$ is circularly symmetric Gaussian, its magnitude, $\hat{\alpha}$, is Rayleigh distributed and is independent of its angle, $\hat{\theta}$. Thus

$$p(\hat{\alpha}_l = x|\hat{\theta}_l, H_\eta) = p(\hat{\alpha}_l = x|H_\eta) = \frac{2x}{2-\rho_\eta^2} e^{-\frac{x^2}{2-\rho_\eta^2}} \tag{5.27}$$

Moreover, from

$$\begin{aligned} h_l^R &= \hat{\alpha}_l \cos\hat{\theta}_l - \ell_l^R \\ h_l^I &= \hat{\alpha}_l \sin\hat{\theta}_l - \ell_l^I, \end{aligned} \tag{5.28}$$

we get

$$\begin{aligned} h_l^R|\hat{\alpha}, \hat{\theta}, H_\eta &\sim \mathcal{N}\left(\frac{\hat{\alpha}\cos\hat{\theta}}{2-\rho_\eta^2}, \frac{1-\rho_\eta^2}{2(2-\rho_\eta^2)}\right) \\ h_l^I|\hat{\alpha}, \hat{\theta}, H_\eta &\sim \mathcal{N}\left(\frac{\hat{\alpha}\sin\hat{\theta}}{2-\rho_\eta^2}, \frac{1-\rho_\eta^2}{2(2-\rho_\eta^2)}\right), \end{aligned} \tag{5.29}$$

which implies that,

$$z_l|\hat{\theta}_l, \hat{\alpha}_l, H_\eta \sim \mathcal{N}\left(\frac{\hat{\alpha}_l}{\sqrt{L}(2-\rho_\eta^2)}, \frac{1-\rho_\eta^2}{2L(2-\rho_\eta^2)}\right). \tag{5.30}$$

It can be seen that for any given $\hat{\theta}$ the last distribution is Gaussian with mean and variance which are independent of $\hat{\theta}$. Substituting (5.27) and (5.30) into (5.26) and after some manipulations we get

$$\begin{aligned} p_{z_l|\hat{\theta}_l}(x|y, H_\eta) &= p_{z_l}(x|H_\eta) \\ &= \frac{1}{\sqrt{\pi\left(\frac{2-\rho_\eta^2}{1-\rho_\eta^2}\right)}}\, e^{\frac{-x^2(2-\rho_\eta^2)}{1-\rho_\eta^2}} + \frac{2xe^{-x^2}}{\sqrt{2-\rho_\eta^2}}\, Q\left(-\sqrt{\frac{2x^2}{1-\rho_\eta^2}}\right) \end{aligned} \tag{5.31}$$

Finally,

$$\begin{aligned} p(\zeta|\underline{\hat{\theta}}, H_\eta) &= p_{z_1}(.|H_\eta) \otimes p_{z_2}(.|H_\eta) \otimes \cdots \otimes p_{z_L}(.|H_\eta) \\ &= p(\zeta|H_\eta) \end{aligned} \tag{5.32}$$

where $\otimes$ is convolution. The last equation implies that given H_η, ζ and $\underline{\hat{\theta}}$ are independent.

Appendix D: Optimum Number of Diversity Branches in MRC

In this appendix we derive the optimal number of diversity branches for spectrum monitoring for the MRC receiver. It is clear that the efficacy of the proposed method relies on the statistics of the REC in the presence or absence of the PU. In particular, the performance of the algorithm improves if the emergence of the primary user causes a higher number of errors in each packet. Therefore we choose the number of diversity branches so as to maximize the difference between the average symbol error probabilities under H_1 and H_0. More specifically let

$$L_{\text{opt}} = \arg\max_{L} \mathfrak{D}(L) \tag{5.33}$$

where $\mathfrak{D}(L) \triangleq \bar{p}_1(L) - \bar{p}_0(L)$, and where $\bar{p}_\eta(L)$ for $\eta = 0, 1$ is the average (with respect to the channel coefficients) symbol error probability under H_η. Using the bound $Q(x) \leq 1/2e^{-x^2/2}$ we have,

$$\begin{aligned} \bar{p}_\eta(L) &= \int_0^\infty Q\left(\sqrt{2\gamma_\eta^{(\text{eff})} x^2}\right) \frac{2e^{-x^2} x^{2L-1}}{(L-1)!} dx \\ &\leq \frac{1}{(L-1)!} \int_0^\infty e^{-\gamma_\eta^{(\text{eff})} x^2} e^{-x^2} x^{2L-1} dx \\ &= \frac{1}{(L-1)!} \int_0^\infty x^{2L-1} \exp\left(-(1+\gamma_\eta^{(\text{eff})}) x^2\right) dx \end{aligned} \tag{5.34}$$

Note that, [107],

$$\int_0^\infty x^m \exp(-\beta x^n) dx = \frac{\Gamma(\frac{m+1}{n})}{n\beta^{\frac{m+1}{n}}} \tag{5.35}$$

Thus we get

$$\bar{p}_\eta(L) \leq \frac{1}{(L-1)!} \frac{\Gamma(L)}{2(1+\gamma_\eta^{(\text{eff})})^L} = \frac{1}{2}(1+\gamma_\eta^{(\text{eff})})^{-L}. \tag{5.36}$$

Using the upper bound in (5.36) we get an approximation for $\mathfrak{D}(L)$. Assuming this approximation to be exact we get

$$\mathfrak{D}(L) = \frac{1}{2}\left((1+\gamma_1^{(\text{eff})})^{-L} - (1+\gamma_0^{(\text{eff})})^{-L}\right) \tag{5.37}$$

Treating L as a continuous variable, the derivative of $\mathfrak{D}$ with respect to L is given by

$$\frac{\partial \mathfrak{D}}{\partial L} = \frac{1}{2}\left(-(1+\gamma_1^{(\text{eff})})^{-L} \log(1+\gamma_1^{(\text{eff})}) + (1+\gamma_0^{(\text{eff})})^{-L} \log(1+\gamma_0^{(\text{eff})})\right) \tag{5.38}$$

Setting the derivative to zero and solving for L we get

$$\tilde{L} = \log\left(\frac{\log\left(1+\gamma_0^{(\mathrm{eff})}\right)}{\log\left(1+\gamma_1^{(\mathrm{eff})}\right)}\right) \Big/ \log\left(\frac{1+\gamma_0^{(\mathrm{eff})}}{1+\gamma_1^{(\mathrm{eff})}}\right) \tag{5.39}$$

Therefore, L_{opt} is obtained as either $\left\lfloor \tilde{L} \right\rfloor$ or $\left\lceil \tilde{L} \right\rceil$.

Appendix E:
Number of samples for PDF estimation

Suppose that we use a histogram to estimate the PDF for a single class and for a given hypothesis. Let $f(x)$ be the actual PDF and denote by $\hat{f}(x)$ its estimate based on its histogram. Then

$$\begin{aligned} 2\Delta_{\text{dist.}} &= \int_{-\infty}^{\infty} \left| f(x) - \hat{f}(x) \right|^2 dx \\ &= \sum_i \int_{x_i}^{x_i+\delta} \left| f(x) - \frac{n_i}{N\delta} \right|^2 dx \\ &= \sum_i \int_{x_i}^{x_i+\delta} \left(f^2(x) - 2f(x)\frac{n_i}{N\delta} + \frac{n_i^2}{N^2\delta^2} \right) dx \end{aligned} \tag{5.40}$$

where n_i is the number of measurement samples that are in the interval $(x_i, x_i+\delta]$. It can be seen that n_i is a Binomial random variable with parameters N and $\vartheta_i = \int_{x_i}^{x_i+\delta} f(x)dx$, i.e., $n_i \sim B(N, \vartheta_i)$. Moreover, $\sum_i n_i = N$, $E[n_i] = N\vartheta_i$ and $E[n_i^2] = N\vartheta_i + N(N-1)\vartheta_i^2$. Therefore, we can find the expected value of $\Delta_{\text{dist.}}$ as follows.

$$\begin{aligned} &2E\left\{\Delta_{\text{dist.}}\right\} \\ &= \int_{-\infty}^{\infty} f^2(x)dx - \frac{2}{N\delta}\sum_i \vartheta_i E[n_i] + \frac{1}{N^2\delta}\sum_i E[n_i^2] \\ &= \int_{-\infty}^{\infty} f^2(x)dx - \frac{2}{\delta}\sum_i \vartheta_i^2 \\ &\quad + \frac{1}{N^2\delta}\left[N\sum_i \vartheta_i + N(N-1)\sum_i \vartheta_i^2 \right] \\ &\approx \frac{1}{N\delta} - \frac{1}{N}\int_{-\infty}^{\infty} f^2(x)dx \end{aligned} \tag{5.41}$$

where we have used the approximation $\sum_i \vartheta_i^2 \approx \delta \int_{-\infty}^{\infty} f^2(x)dx$.

Note that since the PDF is integrable, if the PDF is bounded, then the integral $\int_{-\infty}^{\infty} f^2(x)\, dx$ is finite. Ignoring the second term in (5.41), a single PDF can be estimated with its histogram with the accuracy of $E\left\{\Delta_{\text{dist.}}\right\} < \epsilon$, if we have approximately $N = \frac{1}{2\delta\epsilon}$ samples. It follows that for K classes of radios and for the two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, in order to have the expected discriminability $E\left\{\Delta_{\text{dist.}}\right\} \leq \epsilon$, we need to have a total of $LT = \frac{2K}{2\delta\epsilon}$ samples which results in $T = \frac{K}{\delta\epsilon L}$ time slots. For example for the case of $K = 2$, $L = 15$, $\delta = 10$ and $\epsilon = .001$, we need $T = 14$ samples. This number is of course very optimistic as it does not consider the problems of PDF estimation for several PDFs simultaneously, classification of the radios and the detection of the hypotheses which is

studied in this dissertation. For the given parameters above, Fig. 4.5 shows that the required number of samples is in fact around 35. Finally we should point out that for the kernel-based method, this approach is not tractable. However, a rule of thumb is to set $\delta = 3\sigma$ and and then to use the same formula for T as calculated above.

Vita

Erfan Soltanmohammadi was born in Karaj, Iran, in 1984. He received the B.Sc. in electrical engineering from K.N. Toosi University of Technology (KNTU), Tehran, Iran, in 2007, and the M.Sc. from Amirkabir University of Technology (AUT), Tehran, Iran, in 2010. Erfan worked as a graduate assistant at LSU from September 2010 to May 2014. He has been working under the supervision of Dr. Mort Naraghi Pour. His research interests are wireless communication and signal processing.