

Multi-source Signal Detection with Arbitrary Noise Covariance

Lu Wei, *Member, IEEE*, Olav Tirkkonen, *Member, IEEE*, and Ying-Chang Liang, *Fellow, IEEE*

Abstract—Detecting the presence of signals in noise from multiple sources is a fundamental problem in statistical signal processing. In this paper, we consider multi-antenna signal detection when the noise covariance matrix is assumed to be arbitrary and unknown. We address this problem in the context of cognitive radio, where a multiple-primary-user detector is analyzed. This detector is known as Wilks’ detector in statistics literature, which was derived under the generalized likelihood ratio criterion. We calculate the moments of Wilks’ detector, which lead to simple and accurate approximate analytical formulae for the false alarm probability, the detection probability and the receiver operating characteristic. From the considered simulation settings, performance gain over existing detection algorithms is observed in scenarios with arbitrary and unknown noise correlation and multiple primary users.

Index Terms—Cognitive radio; moment-based approximation; signal detection; Wilks’ detector.

I. INTRODUCTION

A basic problem in statistical hypothesis testing is to infer the presence or absence of signals embedded in noise under different modeling assumptions. These assumptions, though, may not often be met in practical scenarios. Consequently, detection algorithms that are robust to deviations from the presumed assumptions are of particular interest. In this paper, we consider a deviation from the prevailing assumption of perfectly known noise covariance. Namely, we focus on the case when the noise covariance matrix is arbitrary and unknown. We choose to study this problem in the context of spectrum sensing in Cognitive Radio (CR) networks, although the formulation and analysis in this paper are valid for other applications such as sonar and radar detection.

In CR networks, dynamic spectrum access is implemented to mitigate spectrum scarcity. A secondary (unlicensed) user is allowed to utilize the spectrum resources when it does not cause intolerable interference to the primary (licensed) users. Spectrum sensing is the first key step towards this dynamic spectrum access scenario. Prior work on multi-antenna cooperative spectrum sensing predominately employ the assumption of known noise covariance matrix. Based on this assumption, several eigenvalue based spectrum sensing algorithms have been proposed recently [1–10]. In contrast

to feature detection, these eigenvalue based detectors do not require any signalling information of the primary users. The assumption of a perfectly known noise covariance matrix may not be realistic for practical systems due to the time-varying nature of the noise statistics. The time-varying noise can be induced, for example, from the unpredictable interferences. Using the existing detection algorithms [1–10] in such a scenario will induce performance loss. In this paper the noise covariance matrix is assumed to be arbitrary and unknown. The true noise covariance matrix is estimated by periodically updated noise-only observations. Detectors derived under this assumption, a.k.a. the blind-noise-statistics detectors, are robust¹ to modeling assumptions of the noise statistics. Despite the practical importance of the blind-noise-statistics detectors, results in this direction are rather limited. A heuristic detector based on Roy’s statistics [11] was considered in [12, 13]. This detector was proposed under the assumption of a single active primary user [12], see also (13b). The assumption of a single primary user may fail to reflect the situations in forthcoming CR networks, where the primary system could be a cellular network, and the existence of more than one primary users would be the prevailing condition. Moreover, since the spectrum resources also include vacant unlicensed bands, several unlicensed systems, such as Wi-Fi, Bluetooth, and DECT, may share the same band without coordination, giving a scenario where multiple primary users occupy the same band.

To address this challenge, in this paper we consider a detector for arbitrary and unknown noise covariance matrix in the presence of multiple primary users. Under the Generalized Likelihood Ratio (GLR) criterion, Wilks’ detector, which was first proposed in statistics literature in [14], turns out to be the blind-noise-statistics detector we are interested in. We investigate its detection performance by deriving closed-form approximations to the false alarm probability, the detection probability as well as the Receiver Operating Characteristic (ROC). The derived results are easily computable and reasonably accurate. Simulations show the robustness of Wilks’ detector for arbitrary noise covariance matrix and in the presence of multiple primary users. To the best of the authors’ knowledge the contributions of this paper, regarding the performance analysis of Wilks’ detector summarized in the four propositions including Lemma 1, are new.

The rest of this paper is organized as follows. After outlining the signal model in Section II, we introduce the considered

L. Wei is with the Department of Mathematics and Statistics, University of Helsinki, Finland (e-mail: lu.wei@helsinki.fi).

O. Tirkkonen is with the Department of Communications and Networking, Aalto University, Finland (e-mail: olav.tirkkonen@aalto.fi).

Y.-C. Liang is with the Institute for Infocomm Research, A*STAR, Singapore and also with the University of Electronic Science and Technology of China, Chengdu, China (e-mail: ycliang@i2r.a-star.edu.sg).

This work has been presented in part at IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2013.

¹Here, robustness refers to sustained performance in the absence of knowledge of noise covariance matrix. In statistics literature, however, robustness typically means countermeasures against non-Gaussianity.

blind-noise-statistics detector for multiple primary users in Section III. Performance analysis of this detector is addressed in Section IV. Section V presents numerical examples to verify the derived results and to study the detection performance in diverse scenarios. Finally in Section VI we conclude the main results of this paper.

II. SIGNAL MODEL

Consider the standard model for K -sensor cooperative detection² in the presence of P primary users,

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (1)$$

where $\mathbf{x} \in \mathbb{C}^K$ is the received data vector. The $K \times 1$ vector \mathbf{n} is the complex Gaussian noise with zero mean and covariance matrix $\mathbf{\Psi}$; the $K \times P$ matrix $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_P]$ represents the channels between the P primary users and the K sensors; and the $P \times 1$ vector $\mathbf{s} = [s_1, \dots, s_P]'$ denotes the transmitted signals from the primary users, which are commonly assumed to follow an i.i.d zero mean Gaussian distribution and is uncorrelated with the noise. This assumption, for instance, is nearly valid for an OFDM signal in which each carrier is modulated by independent data streams. We further consider deterministic channels, i.e., the channel matrix \mathbf{H} is assumed to be constant during sensing process. Note that the focus of this paper is performance analysis for a given channel realization \mathbf{H} and analyzing the average performance over the statistics of \mathbf{H} is beyond the scope of this work.

We collect N independent observations from model (1) to a $K \times N$ ($K \leq N$) received data matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$. By the above assumptions, the (un-normalized) sample covariance matrix³ $\mathbf{R} = \mathbf{X}\mathbf{X}^\dagger$ of the received data matrix follows a complex Wishart distribution, denoted by $\mathbf{R} \sim \mathcal{W}_K(N, \mathbf{\Sigma})$, with the corresponding population covariance matrix calculated in the absence of primary users, denoted by hypothesis \mathcal{H}_0 (null hypothesis), as

$$\mathcal{H}_0 : \quad \mathbf{\Sigma} = \mathbb{E}[\mathbf{X}\mathbf{X}^\dagger]/N = \mathbf{\Psi}, \quad (2)$$

and in the presence of primary users, denoted by hypothesis \mathcal{H}_1 , as

$$\mathcal{H}_1 : \quad \mathbf{\Sigma} = \mathbf{\Psi} + \sum_{i=1}^P \gamma_i \mathbf{h}_i \mathbf{h}_i^\dagger. \quad (3)$$

Here $\gamma_i = \mathbb{E}[s_i s_i^\dagger]$ defines the transmission power of the i -th primary user and the received SNR of primary user i across the K sensors is defined as⁴

$$\text{SNR}_i = \frac{\gamma_i \|\mathbf{h}_i\|^2}{\text{tr}(\mathbf{\Psi})/K}. \quad (4)$$

These characterize the interference level close to the primary transmitter from a transmission of the secondary system, the control of which is the target of dynamic spectrum management. We note that declaring wrongly \mathcal{H}_0 defines the false alarm probability P_{fa} , and declaring correctly \mathcal{H}_1 defines the

detection probability P_d . The hypothesis under \mathcal{H}_0 is often referred to as null hypothesis.

III. TEST STATISTICS

The differences between the population covariance matrices under \mathcal{H}_0 (2) and under \mathcal{H}_1 (3) can be explored to detect the presence of primary users. With different assumptions on the noise covariance matrix $\mathbf{\Psi}$, and the knowledge of the number of primary users P , various test statistics have been proposed in literature.

A. *When the noise covariance matrix is assumed to be $\mathbf{\Psi} = \sigma^2 \mathbf{I}_K$.*

Here the noise from the K antennas is assumed to be independent, and has a common power σ^2 . In this case, the sufficient statistics is the sample covariance matrix \mathbf{R} of the received data matrix [15] and we denote its ordered eigenvalues by $0 \leq \lambda_K \leq \dots \leq \lambda_1 < \infty$.

In the presence of a *single* primary user, $P = 1$, the hypothesis test is

$$\mathcal{H}_0 : \quad \mathbf{\Sigma} = \sigma^2 \mathbf{I}_K \quad (5a)$$

$$\mathcal{H}_1 : \quad \mathbf{\Sigma} = \sigma^2 \mathbf{I}_K + \gamma_1 \mathbf{h}_1 \mathbf{h}_1^\dagger. \quad (5b)$$

Under this hypothesis test and when σ^2 is assumed to be known, the Largest Eigenvalue based (LE) detector

$$T_{\text{LE}} = \lambda_1 \quad (6)$$

was derived under the GLR criterion [2]. The assumption of known noise power, besides being impractical, leads to detectors which may suffer severe performance degradation due to noise power uncertainty [7, 10]. Under the assumption of unknown noise power σ^2 , the test proposed under the GLR criterion is the Scaled Largest Eigenvalue based (SLE) detector

$$T_{\text{SLE}} = \frac{\lambda_1}{\sum_{i=1}^K \lambda_i}. \quad (7)$$

The SLE detector was first proposed in the context of spectrum sensing in [3] and further analyzed in [4, 5]. Detection without assuming any knowledge of certain parameter is often referred to as blind detection. For example the SLE detector belongs to the blind σ^2 detection, which is more robust than the LE detector to the noise power uncertainty.

In the presence of *multiple* primary users, when $P \geq 2$ but not known a priori i.e. blind P detection, the corresponding hypothesis test is expressed as

$$\mathcal{H}_0 : \quad \mathbf{\Sigma} = \sigma^2 \mathbf{I}_K \quad (8a)$$

$$\mathcal{H}_1 : \quad \mathbf{\Sigma} \succ \sigma^2 \mathbf{I}_K, \quad (8b)$$

where the symbol \succ denotes positive definite. Under this hypothesis test, the detector derived from the GLR criterion is the Spherical Test based (ST) detector⁵

$$T_{\text{ST}} = \frac{|\mathbf{R}|}{\left(\frac{1}{K} \text{tr}(\mathbf{R})\right)^K} = \frac{\prod_{i=1}^K \lambda_i}{\left(\frac{1}{K} \sum_{i=1}^K \lambda_i\right)^K}, \quad (9)$$

⁵ $|\cdot|$ denotes the matrix determinant operation.

²This scenario is more relevant when the K sensors are in one device. For distributed collaborating sensors, accurate time synchronization between devices and communications to the fusion center become an issue.

³ $(\cdot)^\dagger$ denotes the conjugate-transpose operation.

⁴ $\text{tr}(\cdot)$ denotes the matrix trace operation.

which is also a blind σ^2 detector [6]. The spherical test was formulated in [6] as a spectrum sensing algorithm and the detection performance has been analytically addressed in [7]. Although in general the ST detector achieves good performance, it is not an optimal one in the low SNR regime. A test statistics that is optimal in detecting small deviations from \mathcal{H}_0 is John's detector

$$T_J = \frac{\text{tr}(\mathbf{R}^2)}{(\text{tr}(\mathbf{R}))^2} = \frac{\sum_{i=1}^K \lambda_i^2}{\left(\sum_{i=1}^K \lambda_i\right)^2}, \quad (10)$$

which was studied in the context of spectrum sensing in [8]. The criterion under which John's detector is derived is known as the Locally Best Invariant (LBI) criterion. Unlike the GLR criterion, the LBI criterion often leads to detectors that perform particularly well in the low SNR regime. Besides the ST and John's detectors, another blind P detection is the Eigenvalue Ratio based (ER) detector [9, 10],

$$T_{\text{ER}} = \frac{\lambda_1}{\lambda_K}. \quad (11)$$

The ER detector is not constructed from any decision-theoretic consideration, such as the GLR or LBI criterion. In fact, it performs substantially worse than the ST and John's detectors in both single and multiple primary user scenarios [7]. We emphasize that the considered blind P detection i.e. the ST, John's and the ER detectors are also blind σ^2 detectors, which are robust not only to the number of primary users but also to the noise power uncertainty. To complete the story, we note that the cooperative Energy Detector

$$T_{\text{ED}} = \text{tr}(\mathbf{R}) = \sum_{i=1}^K \lambda_i, \quad (12)$$

which assumes σ^2 to be known, is also a blind P detector. The cooperative energy detector is often used as a benchmark detector for performance comparisons [10], whose performance is considerably degraded by a relatively small noise power uncertainty. Finally, we note that for arbitrary and *known* P , the corresponding GLR detectors have been derived in [16].

B. When the noise covariance matrix is assumed to be arbitrary and unknown.

In this case the blindness of the detection is extended in a new dimension. The resulting detectors belong to the so-called blind Ψ detection⁶, which are robust to any modeling assumptions on Ψ . This extension is partially motivated by the existence but usually unknown noise correlation due to e.g. antenna coupling or interferences in practical systems. Instead of a perfectly known Ψ , here we assume to have, in addition to the received data matrix \mathbf{X} , another independent noise-only observation matrix \mathbf{Z} consisting of M samples from the K sensors with $M \geq K$. This noise-only observation matrix \mathbf{Z} can be obtained e.g. when absence of the primary

⁶Strictly speaking, the concept of blindness here is different from those of blind σ^2 or blind P , since an estimate of Ψ has been constructed from the noise-only samples. In the context of this paper, blind Ψ refers to the fact that no assumption on the structure of Ψ is made.

users is declared from an initial coarse sensing period. The initial sensing can be performed in a database assisted manner. Moreover, when the signals of interest are narrow-band and located in a known frequency band, such as the case of TV primary systems, the noise-only samples collected at a frequency just outside this band can be justified as having the same noise covariance characteristics. The time-varying nature of the noise correlation can be coped with periodically updating the measurement \mathbf{Z} . The unknown noise population covariance matrix Ψ can be estimated via the (un-normalized) noise-only sample covariance matrix $\mathbf{E} = \mathbf{Z}\mathbf{Z}^\dagger$, which, by the assumptions in Section II, follows a complex Wishart distribution i.e. $\mathbf{E} \sim \mathcal{W}_K(M, \Psi)$. In this setting, the sufficient statistics is the 'whitened' sample covariance matrix of the form $\mathbf{E}^{-1}\mathbf{R}$ [15] and its ordered eigenvalues are denoted by $0 \leq \theta_K \leq \dots \leq \theta_1 < \infty$.

In the scenario of a *single* primary user, the corresponding hypothesis test is

$$\mathcal{H}_0 : \Sigma = \Psi \quad (13a)$$

$$\mathcal{H}_1 : \Sigma = \Psi + \gamma_1 \mathbf{h}_1 \mathbf{h}_1^\dagger. \quad (13b)$$

Essentially we are testing the equality of the two population covariance matrices Σ and Ψ against a rank-1 perturbation alternative (13b) based on the received data and noise-only observation matrices \mathbf{X} and \mathbf{Z} . In this scenario, a reasonable test statistics to choose is Roy's largest eigenvalue based detector [11]

$$T_R = \theta_1. \quad (14)$$

Nadler et al. [12, 13] were among the first to consider Roy's detector in the spectrum sensing application.

Although Roy's detector is a blind Ψ detector, it is not blind in P . When the actual number of primary users is more than one, Roy's detector is expected to suffer performance loss. Thus, it is of interest to extend the blindness of detection to the practical scenario of multiple primary users. Following the same line of reasoning in constructing (8), the hypothesis test under the assumptions of arbitrary and unknown Ψ and multiple primary users is

$$\mathcal{H}_0 : \Sigma = \Psi \quad (15a)$$

$$\mathcal{H}_1 : \Sigma \succ \Psi. \quad (15b)$$

For this hypothesis test, the so-called Wilks' detector [14] turns out to be the corresponding GLR detector [17]

$$T_W = \frac{|\mathbf{E}|}{|\mathbf{R} + \mathbf{E}|} = \prod_{i=1}^K \frac{1}{1 + \theta_i}, \quad (16)$$

where it can be verified that $T_W \in [0, 1]$. The test procedure for T_W and T_{ST} is

$$T \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\geq}} \zeta, \quad (17)$$

and for other detectors in Table I is

$$T \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \zeta, \quad (18)$$

where ζ is a threshold. Wilks' detector is blind to both Ψ and P , i.e., its performance is robust to the degree of

TABLE I
SUMMARY OF MULTI-ANTENNA SPECTRUM SENSING ALGORITHMS

assumptions	$\Psi = \sigma^2 \mathbf{I}_K$ with known σ^2	$\Psi = \sigma^2 \mathbf{I}_K$ with unknown σ^2	arbitrary and unknown Ψ
single primary user	T_{LE} [1, 2]	T_{SLE} [3–5]	T_R [12, 13]
multiple primary users	T_{ED} [10]	T_{ST} [6, 7], T_J [8], T_{ER} [9, 10]	T_W

noise correlation as well as to the number of primary users, which renders it the most robust detector under the framework developed in this paper. In a similar setting, random matrix theory based inference and estimation for generic deterministic matrices and Toeplitz matrices have been discussed in [18] and [19], respectively. These results do not rely on the existence of noise-only samples. Note that the applicability of Wilks' detection is not limited by the specific problem in CR networks considered in this paper. The analysis in the next Section can be also applied to areas such as sonar and radar detection, whenever Wilks' detector is involved.

For convenience, the detectors discussed within the framework of this paper are summarized in Table I according to their modeling assumptions. More detailed discussions regarding Table I can be found in [20].

Finally we note that for the case of arbitrary and *known* Ψ , the forms of the test statistics (6), (7), (9), (10) and (11) in Section III-A remain the same and are directly applicable. The only difference is that these test statistics are now functions of $\Psi^{-1}\mathbf{R}$ instead of \mathbf{R} .

IV. PERFORMANCE ANALYSIS

In this section we derive analytical expressions for the moments of Wilks' detector under both hypotheses. Based on the derived results, we construct simple yet accurate closed-form approximations to the false alarm probability, the detection probability, as well as the receiver operating characteristic. These results are valid when the covariances of received data matrix and noise-only observation matrix are both non-singular.

A. False Alarm Probability

We first derive the exact moments of T_W under \mathcal{H}_0 , which is summarized in the following proposition.

Proposition 1. *Under \mathcal{H}_0 , the exact m -th moment of random variable T_W is*

$$\mathcal{M}_m = \frac{\Gamma_K(N+M)\Gamma_K(M+m)}{\Gamma_K(M)\Gamma_K(N+M+m)}, \quad (19)$$

where

$$\Gamma_K(N) = \pi^{\frac{1}{2}K(K-1)} \prod_{j=0}^{K-1} \Gamma(N-j), \quad (20)$$

and $\Gamma(\cdot)$ denotes the Gamma function.

The proof of Proposition 1 is in Appendix A. Note that the corresponding moments for the real case can be found, e.g. in [17, Eq. (14)]. The moments of T_W do not depend on the

unknown noise covariance matrix Ψ under the null hypothesis. Despite the fact that the problem of finding the exact T_W distribution under \mathcal{H}_0 has received much attention [17, 22–28], this problem is not completely settled. For example, the exact density representation [22] via the Meijer's G-function is, although of theoretical interest, too complicated for computational purposes. Similarly, the Beta-type density representations [17, 23] involve numerically determining a large number of unknown coefficients, which limits their usefulness in practice. A moment based heuristic curve fitting approach was proposed in [24]. Moreover, when some of the parameters are large an asymptotic expansion of Wilks' test was derived in [25]. In fact, explicit and exact density expressions are available in literature only for $K \leq 4$ [26]. Note that the above a priori results are for real Wishart matrices. For complex Wishart matrices, exact T_W densities were derived for a few limited cases, i.e. $K = 2$ and $K = 3$ in [27]. Since an exact and computable distribution of T_W seems intractable to obtain for arbitrary parameter values, we will construct a simple yet accurate approximative T_W distribution by the moment matching techniques [7, 29]. Contrary to the previously discussed results, the proposed closed-form approximation is valid for any K , N and M .

Motivated by the fact that the exact densities for $K = 2$ and $K = 3$ in [27] hold the same polynomial form $x^i(1-x)^j$ as a Beta density, we choose the Beta distribution to approximate the distribution of T_W for general parameter values. An additional motivation is due to the fact the Beta random variable has the same support as that of T_W . Accordingly we have

Proposition 2. *For any sensor size K , sample size N and noise-only sample size M , the Beta approximation to the CDF of T_W under \mathcal{H}_0 , based on the exact two first moments in (19), is*

$$F_W(y) \approx \frac{B(y; \alpha_0, \beta_0)}{B(\alpha_0, \beta_0)}, \quad y \in [0, 1], \quad (21)$$

where

$$\alpha_0 = \frac{\mathcal{M}_1(\mathcal{M}_1 - \mathcal{M}_2)}{\mathcal{M}_2 - (\mathcal{M}_1)^2}, \quad \beta_0 = \frac{(1 - \mathcal{M}_1)(\mathcal{M}_1 - \mathcal{M}_2)}{\mathcal{M}_2 - (\mathcal{M}_1)^2}. \quad (22)$$

Here, $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, $B(x; a, b) = \int_0^x z^{a-1}(1-z)^{b-1}dz$ define the Beta function and the lower incomplete Beta function, respectively.

The proof of Proposition 2 is in Appendix B. Note that an asymptotic T_W distribution for real Wishart matrices can be found in [28, Eq. (5.4)]. There may be a possibility to extend this result to the complex Wishart case using the tools in [28]. However, the corresponding complex analog

is expected to possess similar undesired properties such as being slowly converging and computationally intensive. On the contrary, in the proposed Beta approximation (21) we have established simple *closed-form* relations (22) between the parameters α_0, β_0 and K, N, M in the complex Wishart case.

From the test procedure (17) and Proposition 2, the resulting approximation to the false alarm probability, for a given threshold ζ , equals

$$P_{\text{fa}}(\zeta) = F_{\mathbf{W}}(\zeta) \approx \frac{B(\zeta; \alpha_0, \beta_0)}{B(\alpha_0, \beta_0)}, \quad (23)$$

where $\zeta \in [0, 1]$. Equivalently, for any P_{fa} requirement a threshold can be calculated by inverting $F_{\mathbf{W}}(\zeta)$,

$$\zeta = F_{\mathbf{W}}^{-1}(P_{\text{fa}}). \quad (24)$$

B. Detection Probability

We first study the moments of $T_{\mathbf{W}}$ under \mathcal{H}_1 . Unlike the case of \mathcal{H}_0 , a tractable and exact formula for the moments seems difficult to obtain. The existing exact moment representations are either in terms of the matrix-valued hypergeometric function [30, Eq. (1.2)] or the a sum over partition of Zonal polynomials [31, Eq. (2.1)], the evaluation of which is computationally prohibitive. For this reason, in the next proposition we propose a simple but accurate approximative moment expression of $T_{\mathbf{W}}$ under \mathcal{H}_1 . The key to establish this result relies on the following lemma.

Lemma 1. *For two independent $K \times K$ Wishart matrices $\mathbf{A} \sim \mathcal{W}_K(n, \mathbf{\Sigma})$ and $\mathbf{B} \sim \mathcal{W}_K(m, \mathbf{\Omega})$, the first two-moment matching Wishart approximation to the sum $\mathbf{A} + \mathbf{B} = \mathbf{C} \sim \mathcal{W}_K(f, \mathbf{\Theta})$ has the Degrees of Freedom (DoF) f and covariance matrix $\mathbf{\Theta}$ as*

$$f = \frac{(n\text{tr}(\mathbf{\Sigma}) + m\text{tr}(\mathbf{\Omega}))^2}{n\text{tr}^2(\mathbf{\Sigma}) + m\text{tr}^2(\mathbf{\Omega})}, \quad (25)$$

$$\mathbf{\Theta} = \frac{1}{f} (n\mathbf{\Sigma} + m\mathbf{\Omega}), \quad (26)$$

where the difference in the third moment is given by

$$\begin{aligned} e_3 = & f^3 \text{tr}(\mathbf{\Theta}^3) + 3f^2 \text{tr}(\mathbf{\Theta}) \text{tr}(\mathbf{\Theta}^2) + f \text{tr}^3(\mathbf{\Theta}) \\ & + f \text{tr}(\mathbf{\Theta}^3) - n^3 \text{tr}(\mathbf{\Sigma}^3) - 3n^2 \text{tr}(\mathbf{\Sigma}) \text{tr}(\mathbf{\Sigma}^2) \\ & - n \text{tr}^3(\mathbf{\Sigma}) - n \text{tr}(\mathbf{\Sigma}^3) - m^3 \text{tr}(\mathbf{\Omega}^3) \\ & - 3m^2 \text{tr}(\mathbf{\Omega}) \text{tr}(\mathbf{\Omega}^2) - m \text{tr}^3(\mathbf{\Omega}) - m \text{tr}(\mathbf{\Omega}^3) \\ & - 3n^2 m \text{tr}(\mathbf{\Sigma}^2 \mathbf{\Omega}) - 3nm \text{tr}(\mathbf{\Sigma}) \text{tr}(\mathbf{\Sigma} \mathbf{\Omega}) \\ & - 3m^2 n \text{tr}(\mathbf{\Omega}^2 \mathbf{\Sigma}) - 3mn \text{tr}(\mathbf{\Omega}) \text{tr}(\mathbf{\Omega} \mathbf{\Sigma}). \end{aligned} \quad (27)$$

The proof of Lemma 1 is in Appendix C. It can be easily seen that when $\mathbf{\Sigma} = \mathbf{\Omega}$ the above approximation reduces to the exact result [15, Th. 3.2.4], i.e. $\mathbf{C} \sim \mathcal{W}_K(f, \mathbf{\Theta})$ with $f = n + m$, $\mathbf{\Theta} = \mathbf{\Sigma}$, and $e_3 = 0$. Note that a similar result on approximating the distribution of the sum of Wishart matrices by a single Wishart matrix was proposed in [33] with the same equivalent covariance matrix (26) but a different degrees of freedom as

$$f^* = \frac{\text{tr}((n\mathbf{\Sigma} + m\mathbf{\Omega})^2) + (n\text{tr}(\mathbf{\Sigma}) + m\text{tr}(\mathbf{\Omega}))^2}{n(\text{tr}(\mathbf{\Sigma}^2) + \text{tr}^2(\mathbf{\Sigma})) + m(\text{tr}(\mathbf{\Omega}^2) + \text{tr}^2(\mathbf{\Omega}))}. \quad (28)$$

Simulations performed in Section V-B indicate that the proposed degrees of freedom (25) leads to more accurate moments estimation of $T_{\mathbf{W}}$ than (28) does. Using Lemma 1 we now derive the corresponding approximation to the moments of $T_{\mathbf{W}}$ under \mathcal{H}_1 , which is summarized in the following proposition.

Proposition 3. *Under \mathcal{H}_1 , an approximative expression for the m -th moment of random variable $T_{\mathbf{W}}$ is*

$$\mathcal{N}_m = \frac{\Gamma_K(M + m) \Gamma_K(d - m) |\mathbf{\Upsilon}|^{-m}}{\Gamma_K(M) \Gamma_K(d) |\mathbf{\Psi}|^{-m}}, \quad (29)$$

where

$$d = \frac{(N\text{tr}(\mathbf{\Sigma}) + (M + m)\text{tr}(\mathbf{\Psi}))^2}{N\text{tr}^2(\mathbf{\Sigma}) + (M + m)\text{tr}^2(\mathbf{\Psi})}, \quad (30)$$

$$\mathbf{\Upsilon} = \frac{1}{d} (N\mathbf{\Sigma} + (M + m)\mathbf{\Psi}). \quad (31)$$

The proof of Proposition 3 is in Appendix D. With the above moment expression, closed-form approximations to the distribution of $T_{\mathbf{W}}$ under \mathcal{H}_1 can be constructed by matching its moments to some known distributions. Motivated by the fact that for the univariate case, $K = 1$, the random variable $T_{\mathbf{W}}$ under \mathcal{H}_1 follows a Beta distribution [31], in the multivariate case we also choose the Beta distribution to model $T_{\mathbf{W}}$. Accordingly we have

Proposition 4. *For any sensor size K , sample size N and noise-only sample size M , the Beta approximation to the CDF of $T_{\mathbf{W}}$ under \mathcal{H}_1 , based on the two first moments in (29), is*

$$G_{\mathbf{W}}(y) \approx \frac{B(y; \alpha_1, \beta_1)}{B(\alpha_1, \beta_1)}, \quad y \in [0, 1], \quad (32)$$

where

$$\alpha_1 = \frac{\mathcal{N}_1(\mathcal{N}_1 - \mathcal{N}_2)}{\mathcal{N}_2 - (\mathcal{N}_1)^2}, \quad \beta_1 = \frac{(1 - \mathcal{N}_1)(\mathcal{N}_1 - \mathcal{N}_2)}{\mathcal{N}_2 - (\mathcal{N}_1)^2}. \quad (33)$$

The proof of Proposition 4 essentially follows that of Proposition 2, and is omitted here. Note that under \mathcal{H}_1 , asymptotic $T_{\mathbf{W}}$ distributions for real Wishart matrices are available in [30, Eq. (4.4)] and [31, Eq. (3.3)], which may be generalized to the complex Wishart case. However, simulations show that these asymptotic results converge slowly with respect to sample sizes N and M for a fixed sensor size K . These results may not be able to capture the performance of Wilks' detector in practical scenarios when the sample size N or the noise-only sample size M are relatively small. On the other hand, the proposed approximation (32) for complex Wishart matrices does not involve any asymptotic expansions in N or M , thus its accuracy is not expected to be affected much by these parameter values. Simulations in Section V-B support this argument.

From the test procedure (17) and Proposition 4, the resulting approximation to the detection probability reads

$$P_d(\zeta) = G_{\mathbf{W}}(\zeta) \approx \frac{B(\zeta; \alpha_1, \beta_1)}{B(\alpha_1, \beta_1)}, \quad (34)$$

where $\zeta \in [0, 1]$.

The mapping between the false alarm probability and the detection probability is referred to as the receiver operating

characteristic. As an immediate result of the closed-form false alarm probability (23) and detection probability (34), an analytical ROC expression for Wilks' detector is obtained as

$$P_d = G_W(F_W^{-1}(P_{fa})). \quad (35)$$

Note that if we further truncate the values of the parameters in (22) and (33) to their respective nearest integers, both the false alarm probability (23) and detection probability (34) reduce to a finite sum of polynomials in the threshold ζ . Thus the computational complexity of threshold calculation becomes quite affordable for on-line implementations.

C. A Note on Approximation Accuracy

The proposed first-two-moment-based Beta approximations in Propositions 2 and 4 correspond to the simplest form of a general Jacobi polynomial approximation [34]. In the general framework, up to any m -th degree of Jacobi polynomials matching the corresponding first m moments of T_W would be used. The first-two-moment-based approximation is often referred to as the initial approximation [29]. According to the general principle of moment-based approximation [29], the choice of the initial approximation is decided by the support of the test statistics of interest. Namely, when $x \in (-\infty, \infty)$, $x \in [a, \infty)$, and $x \in [a, b]$ (a, b being finite) the initial approximations are chosen to be Gaussian, Gamma, and Beta distributions, respectively. These are the simplest representative probability density functions for each support [29]. The associated classical orthogonal polynomials of these density functions are Hermite, Laguerre, and Jacobi orthogonal polynomials, respectively. An important property of moment-based approximation is that for test statistics of finite support $x \in [a, b]$, the approximation becomes exact as the number of polynomials m goes to infinity. This result is known as the Weierstrass approximation theorem [35], which is formally stated as any square integrable function on a finite interval can be expressed in an orthogonal Jacobi polynomial basis. According to the Weierstrass's theorem, the m moment based approximation to T_W becomes exact as the degree of polynomials involved goes to infinity.

Clearly, the error of the moment-based approximation is related to the higher order orthogonal polynomials left out from the approximation. The functional form of the error term can be found, e.g. in Eq. (6) of [34]. In light of this, the exact $P_{fa}(\zeta)$ and $P_d(\zeta)$ can be written as a sum of the proposed Beta approximation and an error term $e_{\alpha_i, \beta_i}(\zeta)$, $i = 0, 1$,

$$e_{\alpha_i, \beta_i}(\zeta) = \sum_{n=3}^m A_{i,n} \sum_{p=1}^n \sum_{q=1}^p B_{i,p,q,n} \zeta^{\alpha_i+n-q} (1-\zeta)^{\beta_i}. \quad (36)$$

The expressions for the constants $A_{i,n}$ and $B_{i,p,q,n}$ can be found below Eq. (6) in [34]. Here, α_0, β_0 are defined in (22) and α_1, β_1 are defined in (33). In the most interesting regions of low false alarm probability $P_{fa}(\zeta \rightarrow 0)$ and high detection probability $P_d(\zeta \rightarrow 1)$, the behavior of the error can be analyzed. Consider an ϵ fulfilling $0 < \epsilon \ll 1$, it follows from (36) that the leading order term in $e_{\alpha_0, \beta_0}(\epsilon)$ for low false alarm probability $P_{fa}(\epsilon)$ is proportional to ϵ^{α_0} (when $p = q = n$ in (36)) and the leading order error in

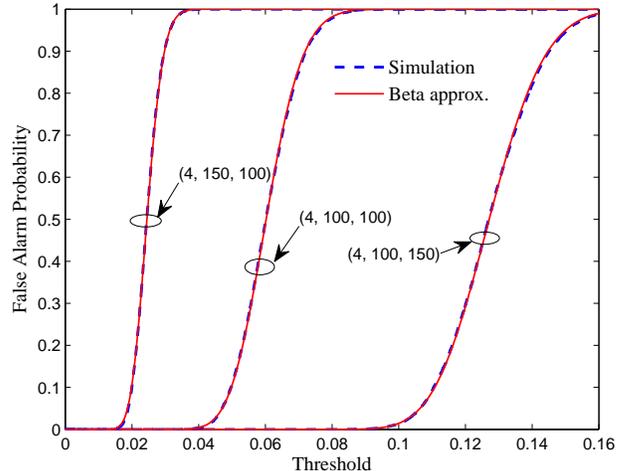


Fig. 1. False alarm probability: analytical (23) versus simulations for different parameter values (K, N, M) .

$e_{\alpha_1, \beta_1}(1 - \epsilon)$ for high detection probability $P_d(1 - \epsilon)$ is ϵ^{β_1} . Typically, the values α_0 and β_1 are positive and large. For example, (α_0, β_1) equals $(15.3, 75975.3)$, $(4162.6, 6.6 \times 10^6)$, $(28.3, 17395.8)$ and $(7469.1, 2.0 \times 10^6)$ for the scenarios considered in Figure 5 and Figure 6 for relatively low and low SNRs, respectively. Thus, the corresponding error for low P_{fa} and high P_d decreases quite fast. Moreover, according to Weierstrass' theorem [35], the approximation in Proposition 2 becomes exact as the degree of polynomials m goes to infinity, namely

$$F_W(y) = \frac{B(y; \alpha_0, \beta_0)}{B(\alpha_0, \beta_0)} + \lim_{m \rightarrow \infty} e_{\alpha_0, \beta_0}(y). \quad (37)$$

The approximation in Proposition 4 is not, however, asymptotically tight as m goes to infinity due to the additional approximation error from Lemma 1.

V. NUMERICAL RESULTS

In this section we first examine the accuracy of the derived approximative false alarm and detection probabilities via Monte-Carlo simulations. Then we compare the performance of Wilks' detector with several detectors in diverse scenarios. The considered values of the parameters K, N and M in this section reflect practical sensing situations. The sample sizes N and M can be as large as a couple of hundred whereas the number of sensors K is at most eight due to physical constraints of the device size.

A. False Alarm Probability

In Figure 1 we plot the approximative (23) and simulated false alarm probabilities as a function of the threshold. To quantitatively show the approximation accuracy we tabulated in Table II the numerical values of approximation error, measured by the Cramér-von Mises goodness-of-fit criterion

$$\int_{\zeta} \left| \widetilde{P}_{fa}(\zeta) - P_{fa}(\zeta) \right|^2 d\widetilde{P}_{fa}(\zeta), \quad (38)$$

TABLE II
APPROXIMATION ERRORS OF FALSE ALARM AND DETECTION PROBABILITIES

(K, N, M)	(4, 150, 100)	(4, 100, 100)	(4, 100, 150)
P_{fa} overall	7.51×10^{-9}	7.88×10^{-9}	2.99×10^{-9}
$P_{fa} \leq 0.1$	9.13×10^{-9}	3.24×10^{-9}	1.26×10^{-9}
P_d overall	9.78×10^{-9}	1.19×10^{-8}	2.12×10^{-8}
$P_d \geq 0.9$	7.20×10^{-10}	3.85×10^{-9}	3.01×10^{-8}

TABLE III
COMPARISON OF APPROXIMATION ERRORS: THE FIRST TWO MOMENTS

SNRs in dB	$\rho = 0.3$		$\rho = 0.7$	
	-2, -3, -4	2, 3, 4	-2, -3, -4	2, 3, 4
when using the proposed DoF (25)	(0.86h, 1.12h)	(1.92h, 3.93h)	(0.93h, 2.81h)	(0.61h, 4.51h)
when using Nel's [33] DoF (28)	(2.04h, 3.95h)	(4.89h, 11.07h)	(5.61h, 12.93h)	(4.90h, 8.72h)

of the proposed false alarm probability (23) with respect to the simulated one \widehat{P}_{fa} . In addition to the overall error in the support $\zeta \in [0, 0.16]$, we also calculated the error for low false alarm probability $P_{fa} \leq 0.1$ as motivated by IEEE 802.22 standard [36]. In the numerical evaluation of (38), we assume a sampling size $Ns = 10^6$ for the overall error. For better illustration, the sampling size for low false alarm probability is counted as a proportion of Ns used up to $P_{fa} = 0.1$. Figure 1 and Table II show that the derived analytical false alarm probability (23) matches the simulations well and the approximation errors are of the same order of magnitude for different parameters. Moreover, it is seen that the overall errors and the errors of the low false alarm probability are also of the same order of magnitude, implying rather uniformly distributed error across the support.

B. Detection Probability

Here we first examine the accuracy of the proposed moment expression (29). In Table III we calculated the relative approximation errors⁷ of the particularly interesting first two moments, where each entry (\cdot, \cdot) denotes the relative errors in per mill of the first and the second moment, respectively. As a comparison, we also tabulated the relative errors of the first two moments when using Nel's DoF estimate (28). In Table III the parameters are set $(K, N, M) = (4, 100, 150)$ and the number of primary users is assumed to be three. The entries of the channel matrix \mathbf{H} were independently drawn from a standard complex Gaussian distribution. The exponential correlation model [37]

$$\Psi_{i,j} = \rho^{|i-j|}, \quad \rho \in [0, 1), \quad (39)$$

is chosen for the noise covariance matrix, where ρ specifies the degree of noise correlation. Different combinations of degrees

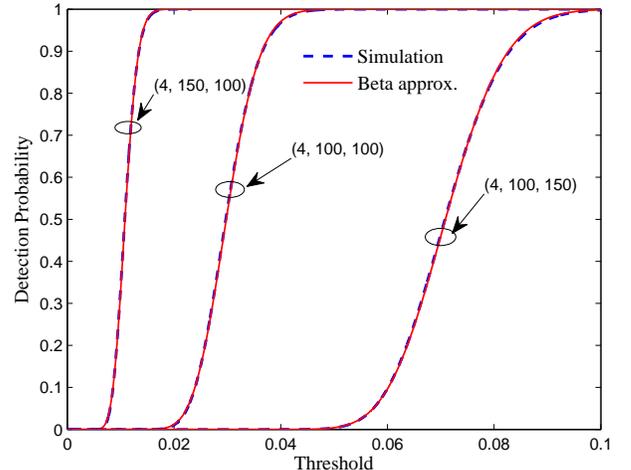


Fig. 2. Detection probability (assuming three primary users with $\text{SNR}_1 = -2$ dB, $\text{SNR}_2 = -3$ dB, $\text{SNR}_3 = -4$ dB and the degree of noise correlation $\rho = 0.3$): analytical (34) versus simulations for different parameter values (K, N, M) .

of noise correlation and SNR values are considered in Table III, where we observe that the approximative moments (29) are more accurate when using the proposed DoF (25) than Nel's DoF (28).

We now study the accuracy of the derived approximative detection probability (34). In Figure 2 we plot the approximative (34) and simulated detection probabilities in a scenario of three primary users with $\text{SNR}_1 = -2$ dB, $\text{SNR}_2 = -3$ dB, $\text{SNR}_3 = -4$ dB and the degree of noise correlation $\rho = 0.3$. For the specific channel realization considered in Figure 2, the eigenvalues⁸ of $\Psi^{-1}\Sigma$ are $[2.31, 1.41, 1.03, 1.00]$. The resulting approximation errors, calculated by the Cramér-von

⁷For a quantity a and its estimate \tilde{a} , the relative error is defined as the absolute value of $(a - \tilde{a})/a$.

⁸Roy's and Wilks' detectors depend on the population covariance matrix $\Psi^{-1}\Sigma$, induced from the sample covariance matrix $\mathbf{E}^{-1}\mathbf{R}$, only through its eigenvalues [15].

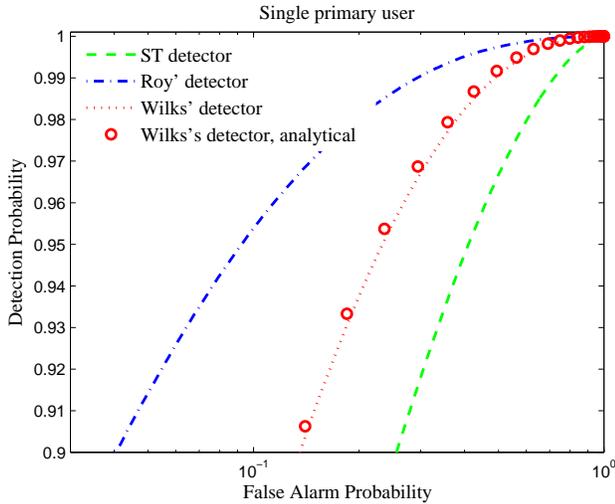


Fig. 3. Performance comparisons: assuming one primary user with $\text{SNR}_1 = -2$ dB and $(K, N, M) = (4, 100, 200)$. The noise correlation is set at $\rho = 0.2$.

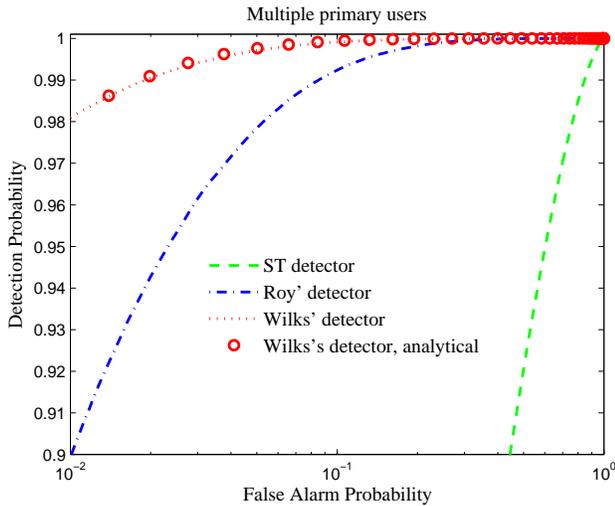


Fig. 4. Performance comparisons: assuming three primary users with $\text{SNRs} = (-2, -3, -4)$ dB and $(K, N, M) = (4, 100, 100)$. The noise correlation is set at $\rho = 0.2$.

Mises goodness-of-fit criterion (38) with $\zeta \in [0, 0.10]$ and a sampling size 10^6 , are summarized in Table II. Similarly, we consider both the overall error and the error in interesting region of high detection probability $P_d \geq 0.9$. The latter consideration is motivated by IEEE 802.22 standard [36], which requires that a secondary device must be able to detect, with at least 90% probability, the presence of primary user. From Table II we can see that the derived analytical detection probability (34) is reasonably accurate and errors in the high detection probability are not significantly different from that of the overall cases.

C. Performance Comparisons

In this subsection the performance of Wilks' detector is compared with some existing detectors by means of ROC curves. Since a ROC curve shows the achieved detection

probability as a function of the false alarm probability, it reflects the overall performance for a given detector. As the focus of this paper is blind Ψ detection, we consider for comparison Roy's detector (14) proposed in [12, 13]. In addition, we consider the ST detector (9) derived from the GLR criterion [6, 7], which is a candidate detector in the presence of multiple primary users. Comparisons with other non-blind Ψ detectors in Section III-A are excluded. For the relative performance among these detectors, the readers are referred to [4, 7, 8, 20]. Here we also choose the exponential correlation model (39) for the noise covariance matrix Ψ . Towards a fair comparison⁹, the ST detector also needs to utilize the available noise-only observations. To this end, we replace \mathbf{R} in the ST detector by the 'whitened' sample covariance matrix $\mathbf{E}^{-1}\mathbf{R}$. This modification is motivated by the fact that \mathbf{E} is the maximum likelihood estimate of Ψ , and for a known Ψ the ST detector becomes a function of $\Psi^{-1}\mathbf{R}$ [15]. We assume that the entries of the channel matrix \mathbf{H} , which are fixed during sensing, are drawn from a standard complex Gaussian distribution. For each ROC curve, 10^6 realizations of data matrix \mathbf{X} are generated to construct the empirical test statistics distributions under both hypotheses.

We start by studying the simple scenario of a single primary user in Figure 3, where we set $\text{SNR}_1 = -2$ dB, $(K, N, M) = (4, 100, 200)$ and $\rho = 0.2$. For the specific channel realization considered in Figure 3, the eigenvalues of the induced population covariance matrix $\Psi^{-1}\Sigma$ are $[1.69, 1.00, 1.00, 1.00]$. We emphasize that the performance of Roy's and Wilks' detectors depend on the parameter space, i.e. the values of K, N, M, \mathbf{H} , SNRs and ρ , only through the eigenvalues of $\Psi^{-1}\Sigma$ [15]. It is seen from this figure that Roy's detector performs best, and indeed its usefulness in detecting a single primary user in the case of arbitrary and unknown Ψ has been justified in [12, 13].

We now investigate the interesting case of multiple primary users. We first consider a scenario of three primary users in Figure 4, where $\text{SNRs} = (-2, -3, -4)$ dB with $(K, N, M) = (4, 100, 100)$ and $\rho = 0.2$. For the specific channel realization in Figure 4, the eigenvalues of $\Psi^{-1}\Sigma$ are $[1.82, 1.53, 1.21, 1.00]$. It is seen from Figure 4 that Wilks' detector performs better than Roy's detector. This is as expected since the former is designed for multiple P detection when Ψ is arbitrary. We further consider a case of five primary users with relatively low and low SNRs in Figures 5 and 6. The consideration of low SNRs scenario is motivated by the fact that the recent Federal Communications Commission regulations require that the secondary devices must be able to detect signals with SNR as low as -18 dB [38, 39]. In both figures, we set $K = 8$ and $\rho = 0.2$. In the upper subplots in Figures 5 and 6, we choose $\text{SNRs} = (-2, -3, -4, -5, -6)$ dB with $(N, M) = (120, 80)$ and $(N, M) = (120, 120)$, respectively. In this case, the eigenvalues of $\Psi^{-1}\Sigma$ are $[2.06, 1.56, 1.36, 1.15, 1.09, 1.00, 1.00, 1.00]$. In the lower subplots in Figures 5 and 6, we choose $\text{SNRs} =$

⁹Strictly speaking, comparing the performance of blind-noise detectors with the ST detector may not be fair since the latter does not require the noise-only samples. However, this comparison does help quantify the performance loss when the noise-only samples are utilized in a heuristic manner.

TABLE IV
IMPACT OF N AND M ON DETECTION PROBABILITY OF WILKS' DETECTOR

	$N + M = 10^5$	$N = 1 \times 10^4$	$N = 3 \times 10^4$	$N = 5 \times 10^4$	$N = 7 \times 10^4$	$N = 9 \times 10^4$
P_d assuming $P_{fa} = 0.1$		0.9690	0.9998	0.9999	0.9998	0.9678
P_d assuming $P_{fa} = 0.01$		0.7956	0.9933	0.9982	0.9930	0.7893

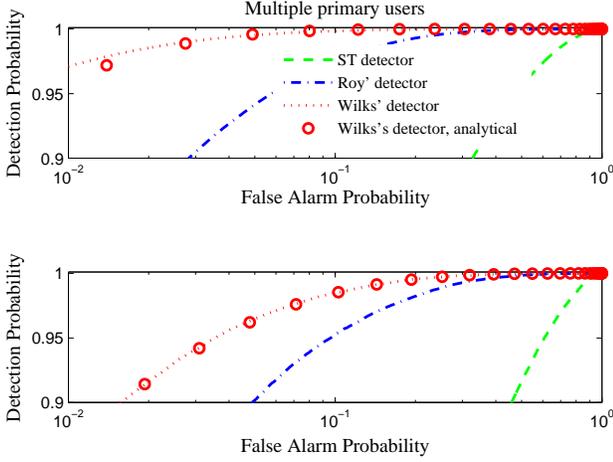


Fig. 5. Performance comparisons: assuming five primary users with SNRs = $(-2, -3, -4, -5, -6)$ dB and $(K, N, M) = (8, 120, 80)$ in the upper subplot, and SNRs = $(-16, -17, -18, -19, -20)$ dB and $(K, N, M) = (8, 3 \times 10^4, 2 \times 10^4)$ in the lower subplot. The noise correlation is set at $\rho = 0.2$.

$(-16, -17, -18, -19, -20)$ dB with $(N, M) = (3 \times 10^4, 2 \times 10^4)$ and $(N, M) = (3 \times 10^4, 3 \times 10^4)$, respectively. In this case, the eigenvalues of $\Psi^{-1}\Sigma$ are $[1.04, 1.02, 1.01, 1.01, 1.01, 1.00, 1.00, 1.00]$. From Figures 5 and 6, we see that Wilks' detector outperforms Roy's detector, as expected. Note that the observations in Figure 3 to Figure 6 regarding the relative performance of Roy's and Wilks' detectors are in line with those of [40, 41]. Comparing Figure 5 with Figure 6, we see that an increase of noise-only samples M enlarges the performance gap between Wilks' and the ST detectors, which indicates that the former is more efficient in using the noise-only samples than the latter does. This is intuitively clear since Wilks' detector was derived from a decision-theoretic criterion i.e. the GLR criterion whereas the modified ST detector utilizes the noise-only samples in a heuristic manner. Note that our intensive simulations show that both Roy's and Wilks' detectors perform substantially better than the ST detector when $\rho > 0.2$.

Finally, we study the relation between N and M on the performance of Wilks' detector in Table IV. A low SNR scenario with three active primary users is considered, where we set SNRs = $(-16, -17, -18)$ dB, $K = 4$ and $\rho = 0.2$. The total number of samples is assumed to be 10^5 , i.e. $N + M = 10^5$. For the specific channel realization considered in Table IV, the eigenvalues of $\Psi^{-1}\Sigma$ are $[1.04, 1.02, 1.00, 1.00]$. Two cases corresponding to $P_{fa} = 0.1$ and $P_{fa} = 0.01$ with N varying from 1×10^4 to 9×10^4 have been considered. The numerical

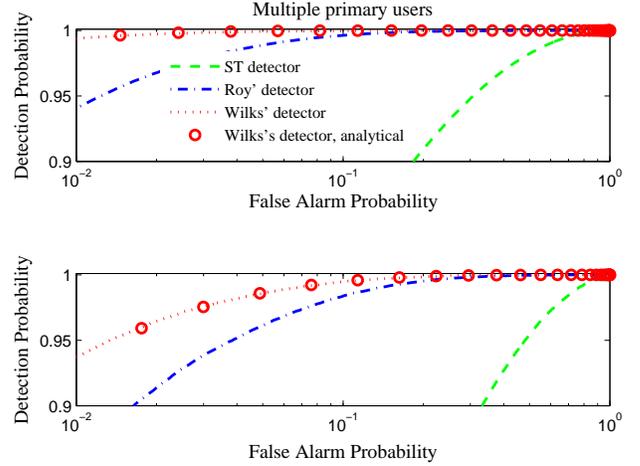


Fig. 6. Performance comparisons: assuming five primary users with SNRs = $(-2, -3, -4, -5, -6)$ dB and $(K, N, M) = (8, 120, 120)$ in the upper subplot, and SNRs = $(-16, -17, -18, -19, -20)$ dB and $(K, N, M) = (8, 3 \times 10^4, 3 \times 10^4)$ in the lower subplot. The noise correlation is set at $\rho = 0.2$.

results obtained in both cases seem to indicate that equally divided received data samples and noise-only observations lead to a higher detection probability for Wilks' detector.

VI. CONCLUSION

In this paper, we studied the performance of Wilks' detector, which is a blind-noise-statistics detector in the presence of multiple primary users. Using the moment matching techniques, simple and accurate closed-form expressions have been derived for its key performance metrics. Simulations show the robustness of Wilks' detector in scenarios with multiple primary users and arbitrary and unknown noise correlation. In such scenarios, Wilks' detector is a viable choice for spectrum sensing.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for the constructive comments which significantly improve this paper. This work has been supported in part by the Academy of Finland under grant 254299.

APPENDIX A PROOF OF PROPOSITION 1

Before proving Proposition 1 we need the following definition.

Definition 1. The density function of a Wishart matrix $\mathbf{R} \sim \mathcal{W}_K(N, \mathbf{\Sigma})$ reads

$$f(\mathbf{R}) = \frac{|\mathbf{\Sigma}|^{-N}}{\Gamma_K(N)} |\mathbf{R}|^{N-K} e^{-\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{R})}. \quad (40)$$

We now prove Proposition 1.

Proof: Recall the definition of random variable

$$T_W = \frac{|\mathbf{E}|}{|\mathbf{R} + \mathbf{E}|} \in [0, 1], \quad (41)$$

where $\mathbf{R} \sim \mathcal{W}_K(N, \mathbf{\Sigma})$ and $\mathbf{E} \sim \mathcal{W}_K(M, \mathbf{\Psi})$. Under \mathcal{H}_0 it holds that $\mathbf{\Sigma} = \mathbf{\Psi}$. Using Definition 1 and the fact that \mathbf{R} and \mathbf{E} are independent, the m -th moment of T_W is calculated as shown on top of the next page, where now $\mathbf{E}' \sim \mathcal{W}_K(M+m, \mathbf{\Psi})$. Since the Wishart matrices \mathbf{R} and \mathbf{E}' have the same covariance matrix $\mathbf{\Psi}$ and dimension K , the sum $\mathbf{R} + \mathbf{E}'$ also follows a Wishart distribution [15, Th. 3.2.4], i.e. $\mathbf{R} + \mathbf{E}' \sim \mathcal{W}_K(N+M+m, \mathbf{\Psi})$. Now using the result for moments of the determinant of complex Wishart matrices [21] we have

$$\mathbb{E}[|\mathbf{R} + \mathbf{E}'|^{-m}] = \frac{|\mathbf{\Psi}|^{-m} \Gamma_K(N+M)}{\Gamma_K(N+M+m)}. \quad (45)$$

Inserting (45) into (44) completes the proof of Proposition 1. ■

APPENDIX B PROOF OF PROPOSITION 2

Proof: For a Beta random variable with density function

$$f_B(x) = \frac{1}{B(\alpha_0, \beta_0)} x^{\alpha_0-1} (1-x)^{\beta_0-1}, \quad x \in [0, 1], \quad (46)$$

the CDF and the m -th moment are given by

$$F_B(y) = \int_0^y f_B(x) dx = \frac{B(y; \alpha_0, \beta_0)}{B(\alpha_0, \beta_0)}, \quad (47)$$

and

$$\mathbb{E}[x^m] = \int_0^1 x^m f_B(x) dx = \frac{(\alpha_0)_m}{(\alpha_0 + \beta_0)_m}, \quad (48)$$

respectively. Here, $(\alpha)_m = \Gamma(\alpha+m)/\Gamma(\alpha)$ defines the Pochhammer symbol. In particular, by matching the first two moments of a Beta random variable in (48) to those of T_W in (19) we have

$$\mathcal{M}_1 = \frac{\alpha_0}{\alpha_0 + \beta_0}, \quad \mathcal{M}_2 = \frac{\alpha_0(\alpha_0 + 1)}{(\alpha_0 + \beta_0)(\alpha_0 + \beta_0 + 1)}. \quad (49)$$

From the above equations the parameters α_0 and β_0 are solved as in (22). This completes the proof. ■

APPENDIX C PROOF OF LEMMA 1

Proof: The idea of the proof is to solve the unknown parameters f and $\mathbf{\Theta}$ of the Wishart matrix \mathbf{C} as functions of the known parameters $n, m, \mathbf{\Sigma}$ and $\mathbf{\Omega}$ of the Wishart matrices \mathbf{A} and \mathbf{B} by moment matching. We start with matching the first moment. By definition, for any Wishart matrix $\mathbf{A} \sim \mathcal{W}_K(n, \mathbf{\Sigma})$ we have $\mathbf{\Sigma} = \mathbb{E}[\mathbf{A}]/n$, thus

$$f\mathbf{\Theta} = \mathbb{E}[\mathbf{C}] = \mathbb{E}[\mathbf{A}] + \mathbb{E}[\mathbf{B}] = n\mathbf{\Sigma} + m\mathbf{\Omega}, \quad (50)$$

from which the relation (26) is established. Now we move on solving the degrees of freedom f by matching the second moment $\mathbb{E}[\mathbf{C}^2]$. Since we need a scalar equation in f , we naturally consider the matrix trace operation on $\mathbb{E}[\mathbf{C}^2]$. Using the fact that $\text{tr}(\mathbb{E}[\mathbf{C}^2]) = \mathbb{E}[\text{tr}(\mathbf{C}^2)]$ and the following result [32] for moments of a complex Wishart matrix $\mathbf{A} \sim \mathcal{W}_K(n, \mathbf{\Sigma})$,

$$\mathbb{E}[\mathbf{A}_{i,j}\mathbf{A}_{k,l}] = n^2\mathbf{\Sigma}_{i,j}\mathbf{\Sigma}_{k,l} + n\mathbf{\Sigma}_{i,l}\mathbf{\Sigma}_{k,j}, \quad (51)$$

we have

$$\mathbb{E}[\text{tr}(\mathbf{C}^2)] = \sum_{i,j \in \{1, \dots, K\}} \mathbb{E}[\mathbf{C}_{i,j}\mathbf{C}_{j,i}] \quad (52)$$

$$= \sum_{i,j \in \{1, \dots, K\}} (f^2\mathbf{\Theta}_{i,j}\mathbf{\Theta}_{j,i} + f\mathbf{\Theta}_{i,i}\mathbf{\Theta}_{j,j}) \quad (53)$$

$$= f^2\text{tr}(\mathbf{\Theta}^2) + f\text{tr}^2(\mathbf{\Theta}). \quad (54)$$

On the other hand, by the independence of \mathbf{A}, \mathbf{B} and invoking (51) again, the right-hand-side of (52) is calculated as shown on the next page. Inserting the established relation (26) into (54) and equating it with (57), after some manipulations, the degrees of freedom f is solved as in (25). In a similar manner, the difference in the third moment (27) is established by repeatedly use of (51) and the following result [32] for moments of a complex Wishart matrix $\mathbf{A} \sim \mathcal{W}_K(n, \mathbf{\Sigma})$,

$$\begin{aligned} \mathbb{E}[\mathbf{A}_{i,j}\mathbf{A}_{k,l}\mathbf{A}_{m,n}] &= n^3\mathbf{\Sigma}_{i,j}\mathbf{\Sigma}_{k,l}\mathbf{\Sigma}_{m,n} + n^2(\mathbf{\Sigma}_{i,n}\mathbf{\Sigma}_{k,l}\mathbf{\Sigma}_{m,j} \\ &\quad + \mathbf{\Sigma}_{i,j}\mathbf{\Sigma}_{k,n}\mathbf{\Sigma}_{m,l} + \mathbf{\Sigma}_{i,l}\mathbf{\Sigma}_{k,j}\mathbf{\Sigma}_{m,n}) \\ &\quad + n(\mathbf{\Sigma}_{k,j}\mathbf{\Sigma}_{i,n}\mathbf{\Sigma}_{m,l} + \mathbf{\Sigma}_{k,n}\mathbf{\Sigma}_{m,j}\mathbf{\Sigma}_{i,l}). \end{aligned}$$

This completes the proof. ■

APPENDIX D PROOF OF PROPOSITION 3

Proof: Recall that under \mathcal{H}_1 we have $\mathbf{R} \sim \mathcal{W}_K(N, \mathbf{\Sigma})$, $\mathbf{E} \sim \mathcal{W}_K(M, \mathbf{\Psi})$ with $\mathbf{\Sigma} \neq \mathbf{\Psi}$. By Definition 1 and the independence of \mathbf{R}, \mathbf{E} , the m -th moment of T_W is calculated as

$$\begin{aligned} \mathbb{E}[T_W^m] &= \int_{\mathbf{R}, \mathbf{E} > 0} |\mathbf{R} + \mathbf{E}|^{-m} \frac{|\mathbf{\Sigma}|^{-N} |\mathbf{\Psi}|^{-M}}{\Gamma_K(N)\Gamma_K(M)} |\mathbf{R}|^{N-K} \times \\ &\quad |\mathbf{E}|^{M+m-K} e^{-\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{R})} e^{-\text{tr}(\mathbf{\Psi}^{-1}\mathbf{E})} d\mathbf{R}d\mathbf{E} \\ &= \frac{\Gamma_K(M+m)}{\Gamma_K(M)} \mathbb{E}[|\mathbf{R} + \mathbf{E}'|^{-m}], \end{aligned} \quad (58)$$

where now $\mathbf{E}' \sim \mathcal{W}_K(M+m, \mathbf{\Psi})$. By Lemma 1 the sum of \mathbf{R} and \mathbf{E}' can be approximated by a single Wishart as

$$\mathbf{R} + \mathbf{E}' \dot{\sim} \mathcal{W}_K(d, \mathbf{\Upsilon}), \quad (59)$$

where d and $\mathbf{\Upsilon}$ are calculated as in (30) and (31), respectively. Using the result for moments of determinant of complex Wishart matrices [21] we have

$$\mathbb{E}[|\mathbf{R} + \mathbf{E}'|^{-m}] \approx \frac{|\mathbf{\Upsilon}|^{-m} \Gamma_K(d-m)}{\Gamma_K(d)}. \quad (60)$$

Inserting (60) into (58) completes the proof of Proposition 3. ■

$$\mathbb{E}[T_W^m] = \int_{\mathbf{R}, \mathbf{E} > 0} |\mathbf{R} + \mathbf{E}|^{-m} \frac{|\Psi|^{-N} |\Psi|^{-M}}{\Gamma_K(N) \Gamma_K(M)} |\mathbf{R}|^{N-K} |\mathbf{E}|^{M+m-K} e^{-\text{tr}(\Psi^{-1}\mathbf{R})} e^{-\text{tr}(\Psi^{-1}\mathbf{E})} d\mathbf{R} d\mathbf{E} \quad (42)$$

$$= \frac{\Gamma_K(M+m)}{\Gamma_K(M) |\Psi|^{-m}} \int_{\mathbf{R}, \mathbf{E} > 0} |\mathbf{R} + \mathbf{E}|^{-m} \frac{|\Psi|^{-N} |\Psi|^{-M-m}}{\Gamma_K(N) \Gamma_K(M+m)} |\mathbf{R}|^{N-K} |\mathbf{E}|^{M+m-K} \times e^{-\text{tr}(\Psi^{-1}\mathbf{R})} e^{-\text{tr}(\Psi^{-1}\mathbf{E})} d\mathbf{R} d\mathbf{E} \quad (43)$$

$$= \frac{\Gamma_K(M+m)}{\Gamma_K(M) |\Psi|^{-m}} \mathbb{E}[|\mathbf{R} + \mathbf{E}'|^{-m}], \quad (44)$$

$$\sum_{i,j \in \{1, \dots, K\}} \mathbb{E}[\mathbf{C}_{i,j} \mathbf{C}_{j,i}] = \sum_{i,j \in \{1, \dots, K\}} (\mathbb{E}[\mathbf{A}_{i,j} \mathbf{A}_{j,i}] + \mathbb{E}[\mathbf{A}_{i,j}] \mathbb{E}[\mathbf{B}_{j,i}] + \mathbb{E}[\mathbf{B}_{i,j}] \mathbb{E}[\mathbf{A}_{j,i}] + \mathbb{E}[\mathbf{B}_{i,j} \mathbf{B}_{j,i}]) \quad (55)$$

$$= \sum_{i,j \in \{1, \dots, K\}} (n^2 \Sigma_{i,j} \Sigma_{j,i} + n \Sigma_{i,i} \Sigma_{j,j} + nm \Sigma_{i,j} \Omega_{j,i} + mn \Omega_{i,j} \Sigma_{j,i} + m^2 \Omega_{i,j} \Omega_{j,i} + m \Omega_{i,i} \Omega_{j,j}) \quad (56)$$

$$= n^2 \text{tr}(\Sigma^2) + n \text{tr}^2(\Sigma) + 2nm \text{tr}(\Sigma \Omega) + m^2 \text{tr}(\Omega^2) + m \text{tr}^2(\Omega). \quad (57)$$

REFERENCES

- [1] Y. Zeng, C. L. Koh, and Y.-C. Liang, "Maximum eigenvalue detection: Theory and application," in *Proc. IEEE Int. Conf. Commun.*, May 2008.
- [2] A. Taherpour, M. N. Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 814-823, Feb. 2010.
- [3] Y. Zeng, Y.-C. Liang, and R. Zhang, "Blindly combined energy detection for spectrum sensing in cognitive radio," *IEEE Signal Process. Lett.*, vol. 15, pp. 649-652, 2008.
- [4] P. Bianchi, M. Debbah, M. Maida, and J. Najim, "Performance of statistical tests for single-source detection using random matrix theory," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2400-2419, Apr. 2011.
- [5] B. Nadler, F. Penna, and R. Garello, "Performance of eigenvalue-based signal detectors with known and unknown noise power," in *Proc. IEEE Int. Conf. Commun.*, June 2011.
- [6] R. Zhang, T. J. Lim, Y.-C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: a GLRT approach," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 84-88, Jan. 2010.
- [7] L. Wei and O. Tirkkonen, "Spectrum sensing in the presence of multiple primary users," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1268-1277, May 2012.
- [8] L. Wei, P. Dharmawansa, and O. Tirkkonen, "Multiple primary user spectrum sensing in the low SNR regime," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1720-1731, May 2013.
- [9] Y. Zeng and Y.-C. Liang, "Eigenvalue based spectrum sensing algorithms for cognitive radio," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1784-1793, June 2009.
- [10] F. Penna, R. Garello, and M. A. Spirito, "Cooperative spectrum sensing based on the limiting eigenvalue ratio distribution in Wishart matrices," *IEEE Commun. Lett.*, vol. 13, issue 7, pp. 507-509, July 2009.
- [11] S. N. Roy, "On a heuristic method of test construction and its use in multivariate analysis," *Ann. Math. Statist.*, vol. 24, no. 2, pp. 220-238, Jun. 1953.
- [12] B. Nadler and I. M. Johnstone, "Detection performance of Roy's largest root test when the noise covariance matrix is arbitrary," in *Proc. IEEE Stat. Signal Process. Conf.*, June 2011.
- [13] B. Nadler and I. M. Johnstone, "On the distribution of Roy's largest root test in MANOVA and in signal detection in noise," Technical Report (No. 2011-04), Department of Statistics, Stanford University, 2011.
- [14] S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika*, vol. 24, no. 3/4, pp. 471-494, 1932.
- [15] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
- [16] D. Ramírez, G. Vazquez-Vilar, R. López-Valcarce, J. Vía, and I. Santamaría, "Detection of rank-P signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764-3774, Aug. 2011.
- [17] Y.-S. Lee, "Some results on the distribution of Wilks's likelihood-ratio criterion," *Biometrika*, vol. 59, no. 3, pp. 649-664, 1972.
- [18] J. Vinogradova, R. Couillet, and W. Hachem, "Statistical inference in large antenna arrays under unknown noise pattern," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5633-5645, Nov. 2013.
- [19] J. Vinogradova, R. Couillet, and W. Hachem, "Estimation of Toeplitz covariance matrices in large dimensional regime with application to source detection," <http://arxiv.org/abs/1403.1243>
- [20] L. Wei, *Towards Robust Spectrum Sensing in Cognitive Radio Networks*. Ph.D thesis, Aalto University, 2013. <http://lib.tkk.fi/Diss/2013/isbn9789526053356/isbn9789526053356.pdf>
- [21] N. R. Goodman, "The distribution of the determinant of a complex Wishart distributed matrix," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 178-180, 1963.
- [22] T. Pham-Gia, "Exact distribution of the generalized Wilks's statistic and applications," *J. Multivariate Anal.*, vol. 99, pp. 1698-1716, 2008.
- [23] M. Schatzoff, "Exact distributions of Wilks's likelihood ratio criterion," *Biometrika*, vol. 53, no. 3/4, pp. 347-358, 1966.
- [24] R. P. Alberto and C. A. Coelho, "Study of the quality of several asymptotic and near-exact approximations based on moments for the distribution of the Wilks Lambda statistic," *J. Statist. Plann. Inference*, vol. 137, pp. 1612-1626, 2007.
- [25] H. Wakaki, "Edgeworth expansion of Wilks' lambda statistic," *J. Multivariate Anal.*, vol. 97, pp. 1958-1964, 2006.
- [26] K. C. S. Pillai and A. K. Gupta, "On the exact distribution of Wilks's criterion," *Biometrika*, vol. 56, no. 1, pp. 109-118, 1969.
- [27] A. K. Gupta, "Distribution of Wilks' likelihood-ratio criterion in the complex case," *Ann. Inst. Statist. Math.*, vol. 23, no. 1, pp. 77-87, 1971.
- [28] B. N. Nagarsenker and J. Suniaga, "Distributions of a class of statistics useful in multivariate analysis," *J. Amer. Statist. Assoc.*, vol. 78, no. 382, pp. 472-475, 1983.
- [29] H. T. Ha, *Advances in Moment-Based Density Approximation Methods*. Ph.D thesis, University of Western Ontario, 2006.
- [30] K. Subrahmaniam, "On the asymptotic distributions of some statistics used for testing $\Sigma_1 = \Sigma_2$," *Ann. Statist.*, vol. 3, no. 4, pp. 916-925, 1975.
- [31] R. W. Kulp and B. N. Nagarsenker, "An asymptotic expansion of the nonnull distribution of Wilks criterion for testing the multivariate linear hypothesis," *Ann. Statist.*, vol. 12, no. 4, pp. 1576-1583, 1984.
- [32] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEE Proc.-Radar, Sonar Navig.*, vol. 147, no. 4, pp. 162-168, Aug. 2000.
- [33] D. G. Nel and C. A. van der Merwe, "A solution to the multivariate Behrens-Fisher problem," *Comm. Statist. Theory Methods*, 15(12), pp. 3719-3735, 1986.

- [34] R. J. Boik, "Algorithm AS 284: Null distribution of a statistics for testing sphericity and additivity: a Jacobi polynomial expansion," *J. Roy. Statist. Soc. Ser. C*, vol. 42, no. 3, pp. 567-576, 1993.
- [35] H. Hochstadt, *Special Functions of Mathematical Physics*. New York: Holt, Rinehart and Winston, 1961.
- [36] IEEE Standard for Information Technology - Telecommunications and information exchange between systems Wireless Regional Area Networks (WRAN) - Specific requirements, Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Procedures for Operation in the TV Bands, IEEE Std 802.22, 2011.
- [37] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502-513, Mar. 2000.
- [38] FCC, "In the matter of unlicensed operation in the TV broadcast bands: second memorandum opinion and order," *Federal Communications Commission*, FCC 10-174, Sept. 2010.
- [39] H.-S. Chen and W. Gao, "Spectrum sensing for TV white space in north America," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 316-326, Feb. 2011.
- [40] K. C. S. Pillai and K. Jayachandran, "Power comparisons of tests of equality of two covariance matrices based on four criteria," *Biometrika*, vol. 55, no. 2, pp. 335-342, 1968.
- [41] T. Sakata, "Likelihood ratio test for one-sided hypothesis of covariance matrices of two normal populations," *Comm. Statist. Theory Methods*, 16(11), pp. 3157-3168, 1987.