Multiple Primary User Spectrum Sensing in the Low SNR Regime

Lu Wei, *Member, IEEE*, Prathapasinghe Dharmawansa, *Member, IEEE*, and Olav Tirkkonen, *Member, IEEE*

***Abstract*—We consider multi-antenna cooperative spectrum sensing in cognitive radio networks, when there are multiple primary users and/or multipath channels. A noise-uncertainty-free detector that is optimal in the low signal to noise ratio regime is analyzed. We derive the moments of the test statistics involved, which lead to simple and accurate analytical formulae for the key performance metrics. The approximative false alarm and detection probabilities as well as receiver operating characteristic are given in closed form. From the considered simulation settings, performance gain over several known detection algorithms is observed in scenarios with relatively low signal to noise ratio.**

***Index Terms*—Cognitive radio; locally best invariant test; moment-based approximation; multipath channels; multiple primary users; spectrum sensing.**

I. Introduction

In Cognitive Radio (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 a single active primary user with one antenna, and a single-tap channel. Based on this assumption, several eigenvalue based sensing algorithms have been proposed recently [1–8]. The assumption of a single primary user is made as the investigations in the literature have mainly focussed on CR networks, where the primary users are TV or DVB systems. The single-tap channel assumption, which may fail to reflect the situations in highly frequency selective channels, has been made for simplicity. The single primary user assumption may not be viable 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. Furthermore, in unlicensed bands, several unlicensed systems, such as Wi-Fi, Bluetooth, and DECT, may share the same band without coordination, resulting multiple primary user scenarios [9]. Using existing single primary user detection algorithms in these scenarios will induce performance loss. Despite the need to understand multipath multi-primary user detection with multiple sensors, the results in this direction are rather limited. A spherical test based detection algorithm has been proposed for multi-primary spectrum sensing in [10], and subsequently studied in [11]. Although this detector is the best known multiple primary user detector, it is not the optimal one in the low Signal-to-Noise Ratio (SNR) regime. Numerical evidence suggests that the spherical test based detector does not perform particularly well when the SNR is relatively low [11, 12]. Spectrum sensing in the low SNR regime is both a practical and challenging issue. For example, recent FCC regulations require that the secondary devices must be able to detect signals with SNR as low as -18 dB [13, 14]. However, no optimal low SNR spectrum sensing algorithm is known in the setting of multipath multiple primary users.

To address this issue, in this paper we first map the problem of spectrum sensing in a multipath and/or multi-primary user scenario to a purely multi-primary user sensing problem. Then we continue by analyzing a multiple primary user detector that is optimal in the low SNR regime. We investigate sensing performance by deriving closed-form moment expressions of the test statistics of this low-SNR detector. Using the derived moments, approximations to the false alarm probability, the detection probability and the Receiver Operating Characteristic (ROC) are constructed. The derived approximations are easily computable and reasonably accurate. Numerical results show the effectiveness of the proposed multiple primary user detector over the existing one in relatively low SNR scenarios. To the best of our knowledge the contributions of this paper, regarding the performance of the considered detector, as summarized in the four propositions (including the three lemmas) are new.

The rest of this paper is organized as follows. In Section II we map the problem of multi-sensor spectrum sensing in a multipath multi-primary scenario to a purely multi-primary one. Then we propose the optimal low SNR detector for multi-primary user sensing. Performance analysis of the proposed detection algorithm is addressed in Section III. Section IV presents numerical examples to verify the derived results and to compare the detection performance in diverse scenarios. Finally in Section V we conclude the main results of this paper.

II. Problem Formulation

We consider a spectrum sensing scenario where there may be $P \geq 1$ primary users transmitting within the bandwidth W so that the transmission sample duration is an integer multiple of a minimum duration $T \leq 1/W$. For primaries

L. Wei and O. Tirkkonen are with the Department of Communications and Networking, Aalto University, Finland (E-mail: {lu.wei, olav.tirkkonen}@aalto.fi). P. Dharmawansa is with the Department of Statistics, Stanford University, USA (E-mail: prathapa@stanford.edu).

This paper was presented in part at the 7-th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom), June 2012.

using Time-division Multiplexing (TDM), this sample duration is the symbol interval, for Code-division Multiplexing (CDM) primaries it is the chip duration, and for Orthogonal Frequency Division Multiplexing (OFDM) primaries, it is the inverse Fourier bandwidth. We further assume that there is a minimum bandwidth of the transmissions of the primaries within W, corresponding to a sample duration $s_{\max}T$.

There is a detector with K sensors trying to detect the presence of the primary user(s). The channels between the primary user(s) and the detector may be multipath channels. The transmissions may or may not be asynchronous, and the detector may or may not be synchronized to one or more of the possible primary transmissions. The number of primary users, and the multipath structure of these users, is not known a priori to the detector. We assume that all possible sample durations between T and $s_{\max}T$ that a primary user may use, are known to the detector, and that the sample timing both at the transmitter and receiver are subject to negligible clock drifts.

A. Signal Model

After filtering, the received complex baseband signal at the K sensors in the bandwidth of interest is

$$\mathbf{x}(t) = \sum_m \sum_{p=1}^{P} s_{p,m}\mathbf{h}_p(t - mT_p + d_p) + \mathbf{n}(t) \ , \qquad (1)$$

where $T_p = s_pT$ is the sample duration for transmitter p, $\mathbf{h}_p(t)$ is the convolution of the reception filter, the transmission filter of p, and the vector of channel impulse responses between p and the K sensors. The sum over m is over the train of transmission samples. The differences in symbol timing and propagation delay for the different transmitters are characterized by the variables d_p.

The K-sensor detector selects s_s to be the smallest common multiple of the possible values of s_p, and takes samples every s_sT seconds, resulting in

$$\mathbf{x}_n = \sum_m \sum_{p=1}^{P} s_{p,m}\mathbf{h}_p\big((ns_s - ms_p)T + d_p\big) + \mathbf{n}_n \ . \qquad (2)$$

Denoting $\mathbf{h}_{p,m} = \mathbf{h}_p\big((ms_p)T + d_p\big)$, we have

$$\mathbf{x}_n = \sum_{p=1}^{P} \sum_{m=1}^{M} s_{p,nr_p-m}\mathbf{h}_{p,m} + \mathbf{n}_n \ , \qquad (3)$$

where $r_p = s_s/s_p$ is the number of independent transmission samples transmitted by primary p per receive sample taken. Here we have assumed causality so that there is an earliest time that a sample affects the received signal, as well as a maximum delay spread M. The channels from some primaries may be shorter than this, in which case the corresponding channel vectors vanish. Note that the number of multipath components depends crucially on the pulse shapes used by the transmitters, the receiver filter, and whether the detector is synchronized to one or more of the transmitters. The signal model covers all possibilities.

Stacking the MP transmitted samples s_{p,nr_p-m} to a vector $\mathbf{s}_n$, the received vector can be written as

$$\mathbf{x}_n = \mathbf{H}\mathbf{s}_n + \mathbf{n}_n, \qquad (4)$$

where $\mathbf{x}_n \in \mathbb{C}^K$, and the $K \times PM$ matrix $\mathbf{H} = [\mathbf{h}_{1,1}, \mathbf{h}_{1,2}, \ldots, \mathbf{h}_{1,M}, \mathbf{h}_{2,1}, \ldots, \mathbf{h}_{P,M}]$ represents the M-tap channels between the P primary users and the K sensors. Assuming proper receiver filtering, the $K \times 1$ vector $\mathbf{n}$ represents additive complex Gaussian noise with mean zero and covariance matrix $\sigma^2\mathbf{I}_K$, where σ^2 is the noise power.

We collect N observations from (4) to a $K \times N$ ($K \leq N$) received data matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] = \mathbf{HS}$, where $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_N]$. We assume that the transmitted samples $s_{p,m}$ are independent and follow a standard complex Gaussian distribution. They are by definition independent from the noise.[1] This holds for TDM, pseudorandomly spread CDM, and approximately for OFDM with a large number of subcarriers. We further assume that the channel $\mathbf{H}$ is constant during sensing i.e. deterministic channels, and is subject to negligible inaccuracies arising from clock drifts. Thus a channel model needs not be specified to carry out the analysis in this paper. Our focus is performance analysis for given channel realizations. Analyzing the average performance over fading channels is beyond the scope of this work.

With these assumptions, the sample covariance matrix of the signal part of $\mathbf{X}$ becomes[2]

$$\mathbf{H}\mathbf{S}\mathbf{S}^\dagger\mathbf{H}^\dagger \approx \sum_p \gamma_p \sum_{m=1}^{M} \mathbf{h}_{p,m}\mathbf{h}_{p,m}^\dagger \ , \qquad (5)$$

where $\gamma_p := \mathbb{E}[s_p s_p^\dagger]$ is the signal power of primary transmission p. The approximation is due to the finite number of samples and corresponding deviations from zero mean of sums of symbols. Note that we do not need to assume independence of the elements of the columns of $\mathbf{S}$, i.e. of the symbol vectors $\mathbf{s}_n$ of different samples. To have (5), it is sufficient to have independence between elements within a symbol vector. For this it is sufficient that the detector does not oversample. Accordingly, (5) holds in a multipath situation, when consecutive s_n may partially be cyclic shifts of each other. Also, even if the transmissions apply cyclic prefixes, (5) holds.

We find that the same covariance structure is present for any M and P. To simplify notations, in what follows we assume that $M = 1$, with the understanding that some of the P primary users may equally well be in a multipath situation. The resulting multiple primary user single-tap signal model has been studied in [10, 11]. The considered signal model in this subsection covers the important case where all the primary users occupy the total band with bit rates connected by integer numbers.

[1] The analytical results in this paper are also applicable in cases with known signal correlation. For the ease of presentation, we assume independent samples here.

[2] $(\cdot)^\dagger$ denotes conjugate-transpose.

B. Sensing Problem

Now the problem of interest is to use the data matrix $\mathbf{X}$ to decide whether there are primary users.[3] By the assumptions in the last subsection, the sample covariance matrix $\mathbf{R} = \mathbf{X}\mathbf{X}^\dagger$ of the received data matrix follows a complex Wishart distribution, denoted by $\mathbf{R} \sim \mathcal{W}_K(N, \mathbf{\Sigma})$. The corresponding population covariance matrix calculated in the absence of primary users, denoted by hypothesis $\mathcal{H}_0$, is

$$\mathcal{H}_0 : \quad \mathbf{\Sigma} := \mathbb{E}[\mathbf{X}\mathbf{X}^\dagger]/N = \sigma^2 \mathbf{I}_K, \tag{6}$$

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

$$\mathcal{H}_1 : \quad \mathbf{\Sigma} = \sigma^2 \mathbf{I}_K + \sum_{p=1}^{P} \gamma_p \mathbf{h}_p \mathbf{h}_p^\dagger. \tag{7}$$

The received SNR of primary user p across the K sensors is $\text{SNR}_p := \gamma_p ||\mathbf{h}_p||^2/\sigma^2$. 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. The differences between the population covariance matrices (6) and (7) can be explored to detect the primary users. Declaring wrongly $\mathcal{H}_0$, or declaring correctly $\mathcal{H}_1$, defines the false alarm probability P_{fa}, and the detection probability P_{d}, respectively. Since the sample covariance matrix $\mathbf{R}$ is a Wishart matrix, it is sufficient statistics for the population covariance matrix $\mathbf{\Sigma}$ [15]. For different assumptions on the number of primary users P, and on the knowledge of the noise power σ^2, different test statistics can be derived as functions of $\mathbf{R}$.

When $P \geq 1$ but not known a priori, the matrix $\sum_{i=1}^{P} \gamma_i \mathbf{h}_i \mathbf{h}_i^\dagger$ in (7) is positive definite. Thus, the notation $\mathbf{A} \succ \mathbf{B}$ is equivalent with $\mathbf{A} - \mathbf{B}$ being positive definite, the hypothesis test is

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

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

where the noise power σ^2 is assumed to be unknown. Essentially, we are testing a null hypothesis $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_K$ against all the other possible alternatives of $\mathbf{\Sigma}$, i.e. the hypothesis test is blind to P. The corresponding GLR-optimal detector under the hypotheses test (8) is based on the so-called Spherical Test (ST)

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

where $|\cdot|$ and $\text{tr}(\cdot)$ denote matrix determinant and trace, respectively, and the ordered eigenvalues of $\mathbf{R}$ are $0 \leq \lambda_K \leq \ldots \leq \lambda_1 < \infty$. In the context of spectrum sensing, ST detection was first proposed in [10] and analyzed in [11].

Although the ST detector achieves good performance in general, it is not the best detector in the low SNR regime. For multiple primary users, a test statistics that is optimal in detecting small deviations from $\mathcal{H}_0$ is based on

$$T_{\text{J}} := \frac{\text{tr}(\mathbf{R}^2)}{\left(\text{tr}(\mathbf{R})\right)^2} = \frac{\sum_{i=1}^{K} \lambda_i^2}{\left(\sum_{i=1}^{K} \lambda_i\right)^2}. \tag{10}$$

This test statistics was first considered by S. John [16]. A more rigorous derivation of (10) can be found in [17, Eq. (1.2–1.7)], where the resulting test procedure is

$$T_{\text{J}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \zeta, \tag{11}$$

ζ being a threshold. It can be verified that the natural support of T_{J} is $[1/K, 1]$. The criterion under which John's detector is derived is known as the Locally Best Invariant (LBI) criterion [17, Eq. (1.1)]. For every σ^2 and for every other test T, there is a neighborhood of $\sigma^2 \mathbf{I}_K$ such that T_{J} achieves no worse performance than T does [17], although the radius of this neighborhood is not known. Considering (6) and (7) it is clear that John's test is optimal when $\sum_{i=1}^{P} \gamma_i \mathbf{h}_i \mathbf{h}_i^\dagger$, measured in a suitable norm, e.g. the sum of its eigenvalues, is small. This effectively requires that the SNRs are low. Note that eigenvalue decomposition is not needed for John's detector as opposed to most other eigenvalue based detectors. Finally, it is worth noting that different detection techniques need to be designed in a more general scenario of an arbitrary but unknown noise covariance matrix. In this case, the detector based on Roy's statistics [18] turns out to be a good choice for single-primary-user detection. For a similar scenario of arbitrary signal covariance matrix, the corresponding test statistics may be derived following the lines of reasoning in [18].

III. Performance Analysis

In this section we derive closed-form expressions for the moments of T_{J} under both hypotheses. Based on the derived results, we construct approximations to the distributions of T_{J}, which lead to analytical formulae for the false alarm probability, the detection probability, as well as the receiver operating characteristic.

A. False Alarm Probability

Firstly, we study the moments of T_{J} under $\mathcal{H}_0$, the first step to which relies on the following lemma.

Lemma 1. *Under $\mathcal{H}_0$ the random variable $\left(\sum_{i=1}^{K} \lambda_i\right)^2$ is independent of the random variable $T_J = \sum_{i=1}^{K} \lambda_i^2 \Big/ \left(\sum_{i=1}^{K} \lambda_i\right)^2$.*

The proof of Lemma 1 is in Appendix A. Note that this independence for the real Wishart case was proven in [21]. However, the method of proof there may not be applied to the complex case, for which we have invoked a more generic approach based on the general polar coordinates transformation. By virtue of Lemma 1, the m-th moment of T_{J} under $\mathcal{H}_0$ now equals

$$\mathbb{E}[T_{\text{J}}^m] = \mathbb{E}\left[\left(\sum_{i=1}^{K} \lambda_i^2\right)^m\right] \Big/ \mathbb{E}\left[\left(\sum_{i=1}^{K} \lambda_i\right)^{2m}\right]. \tag{12}$$

[3] This collaborative sensing scenario and the subsequent formulations are 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 given the limited capability of the individual sensor.

An exact characterization of the m-th non-negative integer moment of T_{J}, denoted by $\mathcal{M}_m$, is summarized in the following result.

Proposition 1. *The m-th non-negative integer moment of random variable T_{J} under $\mathcal{H}_0$ is*

$$\mathcal{M}_m := \frac{C\,\Gamma(KN)}{\Gamma(2m+KN)} \sum_{a_1+\cdots+a_K=m} \frac{m!}{a_1!\cdots a_K!} \times \prod_{1\le i<j\le K} (2a_j - 2a_i + j - i) \prod_{i=1}^{K} \Gamma(2a_i + N - K + i), \quad (13)$$

where the sum is over all the non-negative integer solutions of $a_1+\cdots+a_K = m$. Here $\Gamma(\cdot)$ defines the Gamma function and the constant $C = \left(\prod_{i=1}^{K} \Gamma(N-i+1)\Gamma(K-i+1)\right)^{-1}$.

The proof of Proposition 1 is in Appendix B. The sum over the partition $a_1 + \cdots + a_K = m$ can be implemented by normal sums as $\sum_{a_1=0}^{m} \sum_{a_2=0}^{m-a_1} \cdots \sum_{a_{K-1}=0}^{m-a_1-\cdots-a_{K-2}}$ with a_K replaced by $m - \sum_{i=1}^{K-1} a_i$ in the summand. Note that for real Wishart matrix, up to the second, fourth and sixth moment of T_{J} under $\mathcal{H}_0$ can be found in [21], [25] and [26], respectively. Our derived explicit moment expression of T_{J} (13) for complex Wishart matrix is valid for an arbitrary non-negative moment.

Although in general the exact distribution of T_{J} seems intractable to obtain, one could easily construct an approximative T_{J} distribution by matching its moments to some known distribution with the same support. It is often the case that for the same statistics the functional form of its distributions in both real and complex cases remains the same [19]. For the real Wishart case under $\mathcal{H}_0$, the exact T_{J} distributions for $K = 2, 3$ hold the same polynomial form as the Beta distribution [21]. Moreover, the Beta distribution was shown to accurately model the distribution of T_{J} for arbitrary K [26]. Motivated by these facts, here we choose the Beta distribution to approximate the T_{J} for complex Wishart case as well. Accordingly we have

Proposition 2. *For any sensor size K and sample size N, the Beta approximation to the CDF of T_{J} under $\mathcal{H}_0$, based on the exact two first moments in (13), is*

$$F_J(y) \approx 1 - \frac{B\left(\frac{K(1-y)}{K-1}; \beta_0, \alpha_0\right)}{B(\alpha_0, \beta_0)}, \qquad y \in [1/K, 1], \quad (14)$$

where

$$\alpha_0 = \frac{(K\mathcal{M}_1 - 1)(K\mathcal{M}_1 - K\mathcal{M}_2 + \mathcal{M}_1 - 1)}{(K-1)K(\mathcal{M}_2 - \mathcal{M}_1^2)}, \quad (15a)$$

$$\beta_0 = \frac{(\mathcal{M}_1 - 1)(K\mathcal{M}_1 - K\mathcal{M}_2 + \mathcal{M}_1 - 1)}{(K-1)(\mathcal{M}_1^2 - \mathcal{M}_2)}. \quad (15b)$$

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

The proof of Proposition 2 is in Appendix C. Note that under $\mathcal{H}_0$ several asymptotic T_{J} distributions for real Wishart matrices were established in [17, 27, 28], which may be generalized to the complex Wishart case. However, simulations show that these approximations converge slowly w.r.t. N for a fixed K [17, 27] and w.r.t. both K and N [28]. On the other hand, the proposed approximation (14) for complex Wishart matrices does not involve any asymptotic expansions in K or N, thus its accuracy is not expected to be affected much by the values of K and N. Simulations in Figure 1 of Section IV support this argument.

From the test procedure (11) and Proposition 2, the resulting approximation to the false alarm probability, for a given threshold ζ, equals

$$P_{\text{fa}}(\zeta) = 1 - F_{\text{J}}(\zeta) \approx \frac{B\left(\frac{K(1-\zeta)}{K-1}; \beta_0, \alpha_0\right)}{B(\alpha_0, \beta_0)}, \quad (16)$$

where $\zeta \in [1/K, 1]$. Equivalently for any P_{fa} requirement a threshold can be calculated by numerically inverting $F_{\text{J}}(\zeta)$

$$\zeta = F_{\text{J}}^{-1}(1 - P_{\text{fa}}). \quad (17)$$

B. Detection Probability

We first study the moments of T_{J} under $\mathcal{H}_1$. For convenience of the discussion we define the random variables

$$x := \frac{1}{N^2}\text{tr}(\mathbf{R}^2), \qquad y := \frac{1}{N}\text{tr}(\mathbf{R}), \qquad z := \frac{x}{y^2}. \quad (18)$$

Clearly, z is the random variable of interest, T_{J}. Unlike the case of $\mathcal{H}_0$, the equality (12) no longer holds under $\mathcal{H}_1$. In order to estimate the moments of z it is not enough to estimate the moments of random variables x and y separately, estimating their correlation is needed. A standard technique in this situation is the so-called 'Delta method' [29], which relies on the Taylor expansions for the moments of random variable z. Using the 'Delta method', in the following proposition we propose simple and accurate estimates of the mean and variance of z under $\mathcal{H}_1$, which involve the first two exact moments and the covariance of random variables x and y.

Proposition 3. *Denote the means of the random variables x, y by μ_x, μ_y and the variances by ν_x, ν_y. The second order approximation to the mean and the first order approximation to the variance of the random variable z are given by*

$$\mu_z \approx \frac{\mu_x}{\mu_y^2} - \frac{2\mu_{xy}}{\mu_y^3} + \frac{3\mu_x\nu_y}{\mu_y^4}, \quad (19)$$

and

$$\nu_z \approx \frac{\nu_x}{\mu_y^4} - \frac{4\mu_x\mu_{xy}}{\mu_y^5} + \frac{4\mu_x^2\nu_y}{\mu_y^6}, \quad (20)$$

respectively, where μ_{xy} denotes the covariance of x and y. Here the quantities μ_x, μ_y, ν_x, ν_y and μ_{xy} are calculated, in terms of the population covariance matrix under $\mathcal{H}_1$ (7), as

$$\mu_x = \text{tr}(\mathbf{\Sigma}^2) + \frac{1}{N}\big(\text{tr}(\mathbf{\Sigma})\big)^2, \quad (21a)$$

$$\nu_x = \frac{4}{N}\text{tr}(\mathbf{\Sigma}^4) + \frac{2}{N^2}\left(4\text{tr}(\mathbf{\Sigma})\text{tr}(\mathbf{\Sigma}^3) + \big(\text{tr}(\mathbf{\Sigma}^2)\big)^2\right) + \frac{2}{N^3}\left(2\left(\text{tr}(\mathbf{\Sigma})\right)^2\text{tr}(\mathbf{\Sigma}^2) + \text{tr}(\mathbf{\Sigma}^4)\right), \quad (21b)$$

$$\mu_y = \text{tr}(\mathbf{\Sigma}), \quad (21c)$$

$$\nu_y = \frac{1}{N}\text{tr}(\mathbf{\Sigma}^2), \tag{21d}$$

$$\mu_{xy} = \frac{2}{N}\text{tr}(\mathbf{\Sigma}^3) + \frac{2}{N^2}\text{tr}(\mathbf{\Sigma})\text{tr}(\mathbf{\Sigma}^2). \tag{21e}$$

The proof of Proposition 3 is in Appendix D. Note that the higher moments or more accurate moment estimates of z can be similarly obtained by following the proof of Proposition 3. The basic procedure involves enumerating the higher moments of $\mathbf{R}$ as functions of $\mathbf{\Sigma}$, for example the second order approximation to the variance would require up to the 6-th moment of $\mathbf{R}$. This enumeration procedure, although rather tedious, is quite straightforward.

With the estimates of the mean (19) and variance (20), closed-form distributions of T_{J} under $\mathcal{H}_1$ can be constructed. Here we also choose the Beta distribution (41), since it has the same support as T_{J}. An additional motivation comes form the fact that for $K = 2$ the test statistics T_{J} becomes a linear transform of the test statistics T_{ST}, whose exact distribution under $\mathcal{H}_1$ holds the same polynomial form as the Beta distribution [11]. Accordingly we have

Proposition 4. *For any sensor size K and sample size N, the Beta approximation to the CDF of T_{J} under $\mathcal{H}_1$, based on the estimated two first moments (19) and (20), is*

$$G_J(y) \approx 1 - \frac{B\left(\frac{K(1-y)}{K-1}; \beta_1, \alpha_1\right)}{B(\alpha_1, \beta_1)}, \quad y \in [1/K, 1], \tag{22}$$

where

$$\alpha_1 = \frac{(1 - K\mu_z)\left((\mu_z - 1)(K\mu_z - 1) + K\nu_z\right)}{(K-1)K\mu_z}, \tag{23a}$$

$$\beta_1 = \frac{(\mu_z - 1)\left((\mu_z - 1)(K\mu_z - 1) + K\nu_z\right)}{(K-1)\mu_z}. \tag{23b}$$

The proof of Proposition 4 essentially follows that of Proposition 2, and is thus omitted. Note that under $\mathcal{H}_1$, asymptotic T_{J} distributions for real Wishart matrices are available in [17, 32], which may be generalized to the complex Wishart case. However, besides being slowly converging, these asymptotic results are only valid for some specific structures of the population covariance matrix $\mathbf{\Sigma}$. On the other hand, the proposed finite-dimensional approximation (22) for complex Wishart matrices is applicable for arbitrary $\mathbf{\Sigma}$ structures, e.g. arbitrary SNR values, channel realizations and number of primary users.

From the test procedure (11) and Proposition 4, the resulting approximation to the detection probability reads

$$P_{\text{d}}(\zeta) = 1 - G_{\text{J}}(\zeta) \approx \frac{B\left(\frac{K(1-\zeta)}{K-1}; \beta_1, \alpha_1\right)}{B(\alpha_1, \beta_1)}, \tag{24}$$

where $\zeta \in [1/K, 1]$.

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 approximative false alarm probability (16) and detection probability (24), an analytical ROC expression for John's detector is obtained as

$$P_{\text{d}} = 1 - G_{\text{J}}\left(F_{\text{J}}^{-1}(1 - P_{\text{fa}})\right). \tag{25}$$

Note that the parameters α_0, β_0 in (15) and α_1, β_1 in (23) are only elementary functions of the sensor size K, the sample size N and the population covariance matrix $\mathbf{\Sigma}$. Moreover, if we further approximate the values of the parameters in (15) and (23) to their respective nearest integers, both the false alarm probability (16) and detection probability (24) 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 two-moment-based Beta approximations in Propositions 2 and 4 correspond to the simplest form of a general Jacobi polynomial approximation. In the general framework, up to any n-th degree of Jacobi polynomials matching the corresponding first n moments of T_{J} would be used. Since the random variable T_{J} is of a finite support, the Jacobi polynomial expansion for the distribution of T_{J} is exact according to the Weierstrass approximation theorem [33]. Namely, when n goes to infinity the Jacobi polynomial approximation converges to the exact T_{J} distribution. In practise, the choice of n reflects a trade-off between the approximation accuracy and the implementation complexity. In light of the good accuracy as shown in the next section, we consider $n = 2$ in this paper. The general n-moment-based approximation, including the error analysis, can be easily obtained by following the procedures in [26, 34].

The exact false alarm probability and detection probability can be written as a sum of the n-moment-based approximation and an error term. The error term is related to the higher order polynomials left out from the approximation. An explicit expression for the error term $e_{\alpha_i,\beta_i}(\zeta)$ of the two-moment-based Beta approximation can be found, e.g., in [26, Eq. (6)]. Due to the complicated form of this error term, analysis on its behavior seems difficult. However, in the most interesting regions of low false alarm probability $P_{\text{fa}}(\zeta \to 1/K)$ and high detection probability $P_{\text{d}}(\zeta \to 1)$, the behavior of the error can be understood. Consider an infinitesimal ϵ fulfilling $1/K < \epsilon \ll 1$, it follows along the same line of reasoning in [11, Section III-C] that the leading order term in $e_{\alpha_0,\beta_0}(\epsilon)$ for low false alarm probability $P_{\text{fa}}(\epsilon)$ is proportional to ϵ^{β_0} and the leading order error in $e_{\alpha_1,\beta_1}(1-\epsilon)$ for high detection probability $P_{\text{d}}(1-\epsilon)$ is ϵ^{α_1}. Typically, the values β_0 and α_1 are positive and large. For example, (β_0, α_1) equals $(2393.0, 11.9)$, $(1193.0, 10.5)$ and $(593.02, 12.1)$ for the parameters considered in Figure 3, Figure 4 and Figure 5 respectively. Thus, the corresponding errors for low false alarm probability and high detection probability decrease quite fast.

IV. Numerical Results

In this section we first examine the accuracy of the derived approximations to the false alarm and the detection probabilities via Monte-Carlo simulations. Then we compare the performance of John's detector to some optimal noise-uncertainty-free detectors in realistic sensing scenarios. Our focus here is detection in the low SNR regime, which is a practical and challenging issue. The considered values of the

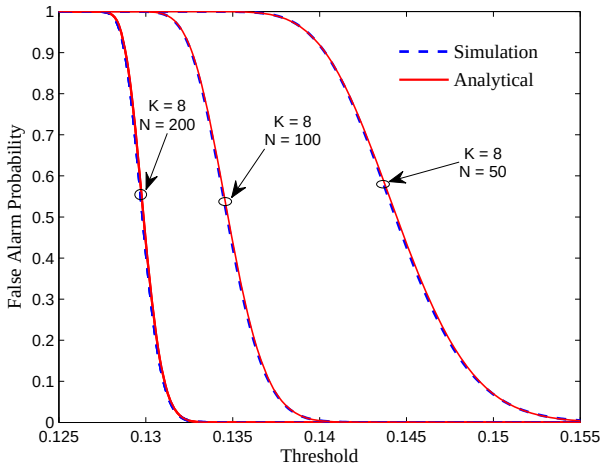

Fig. 1. False alarm probability: analytical (16) versus simulations. For $K = 8$, $N = 50$, 100 and 200, the approximation error is respectively 4.78×10^{-9}, 1.35×10^{-9} and 1.66×10^{-9}.

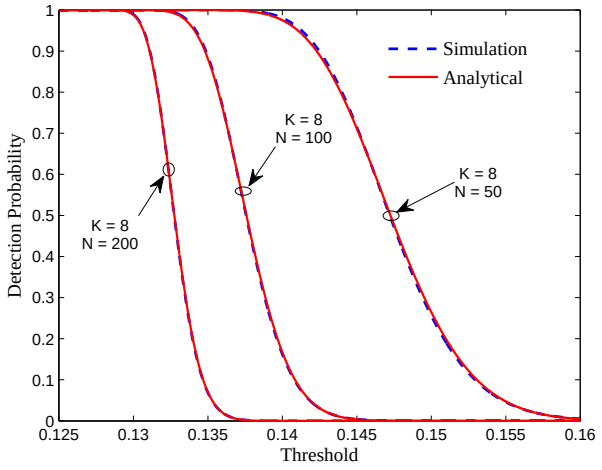

Fig. 2. Detection probability (assuming three primary users with $\text{SNR}_1 = -4$ dB, $\text{SNR}_2 = -5$ dB and $\text{SNR}_3 = -6$ dB): analytical (24) versus simulations. For $K = 8$, $N = 50$, 100 and 200, the approximation error is respectively 1.45×10^{-7}, 4.54×10^{-8} and 1.12×10^{-8}.

parameters K and N in this section reflect practical spectrum sensing situations. The sample size N 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 and Detection Probability

In Figure 1 we plot the approximative false alarm probability (16) and the simulated false alarm probability as a function of the threshold. To quantitatively show the approximation accuracy we calculate the approximation error, measured by the Cramér-von Mises goodness-of-fit criterion, of the proposed false alarm probability (16) with respect to the simulations. The Cramér-von Mises criterion is defined as

$$\int_{\zeta} |F(x) - \hat{F}(x)|^2 \mathrm{d}F(x), \tag{26}$$

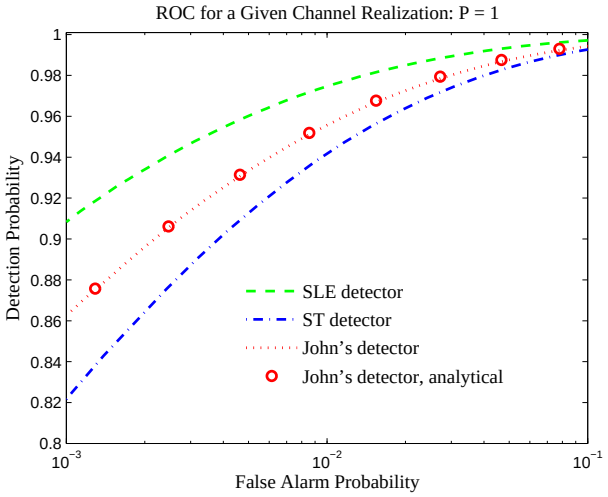

Fig. 3. Instantaneous ROC: assuming one primary user with $\text{SNR}_1 = -4$ dB using $K = 4$ sensors and $N = 400$ samples per sensor. The eigenvalues of $\mathbf{\Sigma}$ are $[1.40, 1, 1, 1]$.

where $F(x)$ and $\hat{F}(x)$ defines a CDF function and its estimate, respectively. In Figures 1 and 2 we assume uniform sampling in $\zeta \in [0.125, 0.2]$ with a sampling size 10^6. The results, summarized in the caption of Figure 1, show that the derived analytical false alarm probability matches the simulations well and the approximation errors are of the same order of magnitude, as expected.

In Figure 2 we plot the derived analytical detection probability (24) versus simulations, assuming three simultaneously transmitting primary users ($P = 3$) with $\text{SNR}_1 = -4$ dB, $\text{SNR}_2 = -5$ dB and $\text{SNR}_3 = -6$ dB. Without loss of generality, we assume unit powers for the zero mean Gaussian signal and noise. The entries of the channel matrix $\mathbf{H}$, which are fixed during sensing, are independently drawn from a standard complex Gaussian distribution. The channel vector for each primary user is normalized as $\mathbf{u}_i = \mathbf{h}_i/||\mathbf{h}_i||$. As a result, the population covariance matrix $\mathbf{\Sigma}$ can be explicitly represented as a function of SNRs, i.e. $\mathbf{\Sigma} = \mathbf{I}_K + \sum_{i=1}^{P} \text{SNR}_i \mathbf{u}_i \mathbf{u}_i^\dagger$. With the same $\mathbf{\Sigma}$, the simulated curve is plotted using 10^6 Monte-Carlo runs with the resulting approximation error as shown in the caption of Figure 2. From Figure 2 it can be observed that the derived detection probability (24) agrees with the simulations well. Unlike the case for the false alarm probability, here we observe that the approximation error decreases noticeably as the number of samples N increases. One intuitive reason is that for a fixed K the density of T_J is more concentrated around its mean as N increases. As a result, the Taylor series expansion (53) about the mean becomes more accurate as represented by the first few terms. This fact leads to refined estimates of the mean (19) and variance (20) under $\mathcal{H}_1$ as N increases.

B. Detection Performance

We compare the detection performance of John's detector to some optimal 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

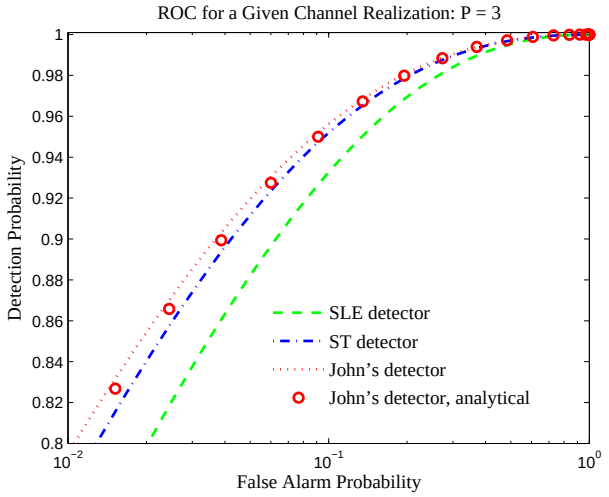

Fig. 4. Instantaneous ROC: assuming three primary users with $\text{SNR}_1 = -4$ dB, $\text{SNR}_2 = -5$ dB and $\text{SNR}_3 = -6$ dB using $K = 4$ sensors and $N = 200$ samples per sensor. The eigenvalues of $\mathbf{\Sigma}$ are $[1.54, 1.25, 1.02, 1]$.

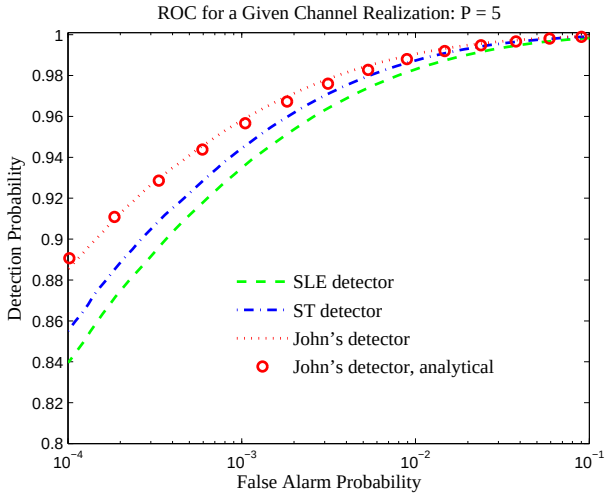

Fig. 5. Instantaneous ROC: assuming five primary users with $\text{SNR}_1 = 0$ dB, $\text{SNR}_2 = -2$ dB, $\text{SNR}_3 = -4$ dB, $\text{SNR}_4 = -6$ dB and $\text{SNR}_5 = -8$ dB using $K = 4$ sensors and $N = 100$ samples per sensor. The eigenvalues of $\mathbf{\Sigma}$ are $[2.55, 1.57, 1.25, 1.06]$

detection performance for a given detector. Our focus here is detection in the presence of multiple primary users, thus we consider for comparison the ST detector [10, 11] which is optimal for multiple primary user detection. In addition the the Scaled Largest Eigenvalue based (SLE) detector [4–8], $T_{\text{SLE}} := \lambda_1 / \sum_{i=1}^{K} \lambda_i$, which is optimal in detecting single primary user, is considered for comparison as well. For the sake of clarity, we neither include the comparisons with detectors that are non-optimal nor detectors that are sensitive to noise power uncertainty [35] such as the energy detector [36]. For performance comparisons in these directions, the readers are referred to [6, 7, 11].

In order to see a clear picture of the impact of the number of primary users P on the detection performance, we plot various ROC curves assuming different values of P in realistic scenarios with relatively low SNRs. In Figure 3 we plot the ROC curves for the considered three noise-uncertainty-free detectors in the presence of a single primary user with $\text{SNR}_1 = -4$ dB. In this case, the number of sensors and samples per sensor is $K = 4$ and $N = 400$, respectively. In Figure 4 we assume a scenario of three simultaneously transmitting primary users with $\text{SNR}_1 = -4$ dB, $\text{SNR}_2 = -5$ dB and $\text{SNR}_3 = -6$ dB. The number of sensors is chosen to be $K = 4$, and the number of samples per sensor is $N = 200$. In Figure 5 we consider the case of five primary users with $\text{SNR}_1 = 0$ dB, $\text{SNR}_2 = -2$ dB, $\text{SNR}_3 = -4$ dB, $\text{SNR}_4 = -6$ dB and $\text{SNR}_5 = -8$ dB using $K = 4$ sensors and $N = 100$ samples per sensor. The analytical approximations of the ROC curves are obtained by (25), where we assume the channel matrix is independently drawn from a standard complex Gaussian distribution and is kept constant during sensing. For the specific channel realizations considered in Figure 3, Figure 4 and Figure 5, the eigenvalues[4] of the population covariance matrix $\mathbf{\Sigma}$ are respectively $[1.40, 1, 1, 1]$, $[1.54, 1.25, 1.02, 1]$ and $[2.55, 1.57, 1.25, 1.06]$. With the same $\mathbf{\Sigma}$, the corresponding numerical ROC is plotted. Note that as the considered detectors are not functions of the noise power σ^2, without loss of generality, we set $\sigma^2 = 1$ at the secondary receiver.

In Figure 3 we observe that the SLE detector performs better than the ST and John's detectors in the presence of a single primary user. This is intuitively clear since the SLE detector is optimized for single primary user detection. In the scenarios of more than one primary users, we see from Figure 4 and Figure 5 that both the considered multiple primary user detectors: the ST and John's detectors outperform the SLE detector, as expected. In addition, it is seen from Figure 5 that the advantage of John's detector persists even in the presence of one primary user with a not-so-low SNR ($\text{SNR}_1 = 0$ dB). We also observe that the derived analytical ROC expression (25) matches simulations well. To further examine the relative performance of the competing ST and John's detectors, we compared their detection probabilities as a function of number of primary users P in Table I shown on top of the next page, where we set $P_{\text{fa}} = 10^{-2}$. In order to focus on the impact of P on the relative performance, the SNRs of the primary users are all set at 0 dB with $K = 4$ sensors and $N = 50$ samples per sensor. In Table I the blue color indicates a higher P_{d} value of each column, from which we see that as P increases the ST detector may start to outperform John's detector, though the differences are relatively small. The intuitive reason shall be clear by now: when P is large the probability is high that John's detector may depart from its optimality region, which is in the neighborhood of $\mathcal{H}_0 : \mathbf{\Sigma} = \sigma^2 \mathbf{I}_K$, leading to performance degradation.

Both the ST detector derived from the GLR criterion and John's detector derived from the LBI criterion are optimal detectors in the presence of multiple primary users. In general it is difficult to quantify the conditions, such as SNR values, number of primary users, and channel realizations, under which the ST detector outperforms John's detector and vice

[4]For the considered detectors, the test statistics depend on $\mathbf{\Sigma}$ only through the eigenvalues of $\mathbf{\Sigma}$.

TABLE I
DETECTION PROBABILITY COMPARISONS: $K = 4$, $N = 50$, $P_{\text{FA}} = 10^{-2}$.

number of primary users	$P = 2$	$P = 4$	$P = 6$	$P = 8$	$P = 10$
$P_{\text{d, ST}}$	0.8495	0.9970	0.9871	0.9616	0.9795
$P_{\text{d, J}}$	0.8700	0.9984	0.9913	0.9508	0.9651

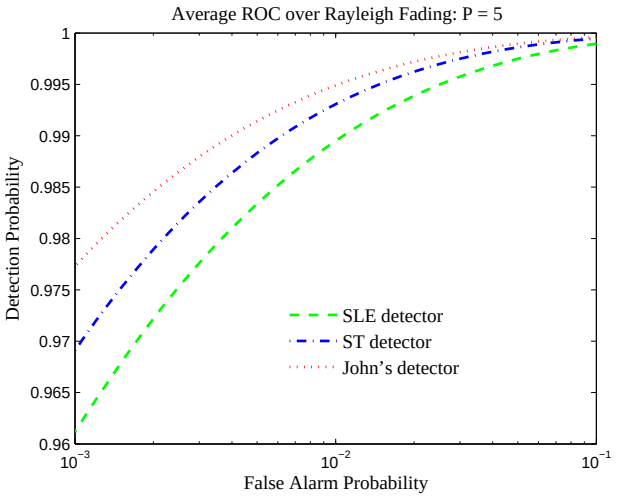

Fig. 6. Average ROC: assuming five primary users with $\text{SNR}_1 = 1$ dB, $\text{SNR}_2 = -3$ dB, $\text{SNR}_3 = -7$ dB, $\text{SNR}_4 = -11$ dB and $\text{SNR}_5 = -15$ dB using $K = 4$ sensors and $N = 100$ samples per sensor.

versa. Despite this, some general understanding can be gained based on the simulations performed and the results in [11] (and references therein):

- The range of the SNR values within which John's detector performs better is quite wide: from low to not-too-large SNRs.
- When the SNR values are very small, John's detector may perform substantially better than the ST detector.
- For high SNR values the ST detector may outperform John's detector, though in this case their performance gap is likely to be very small.

We end this section by studying the average detection performance over fading channel via simulations in Figure 6. Contrary to the previous ROC curves generated for a fixed $\mathbf{\Sigma}$, here we plot average ROC curves over 10^4 realizations of $\mathbf{\Sigma}$. For each $\mathbf{\Sigma} = \mathbf{I}_K + \sum_{i=1}^{P} \gamma_i \mathbf{h}_i \mathbf{h}_i^\dagger$, the channel matrix $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_P]$ is independently drawn from a standard complex Gaussian distribution corresponds to Rayleigh fading. It is seen from Figure 6 that the relative performance of the considered detectors remain the same as these in Figure 4 and Figure 5 for multiple-primary-user detection.

The simulations show excellent performance for John's detector with $N = 100$ samples. The assumption of negligible effect of clock drifts on the signal model (3) can be reassessed in this light. If the clock drift of the transmitters and detectors would be 20 parts-per-million, typical for hand-held devices of today, the maximum timing difference arising within the sample collection time would be 0.4% of the sample duration, which is insignificant for most pulse shapes in use. Accordingly, the signal model considered is viable in such scenarios.

V. CONCLUSION

In this paper, we investigated the performance of John's detector, which is a candidate for multi-sensor spectrum sensing in the presence of multiple primary users and/or multipath channels. John's detector is optimal for detecting small deviations of the covariance matrix from a matrix proportional to identity. Analytical formulae have been derived for its approximative false alarm probability, detection probability as well as receiver operating characteristic. The derived results are simple to calculate and yield an almost-exact fit to simulations. From the simulation setting considered, performance gain over several detection algorithms is observed when SNR is relatively low. Considering the results of this paper, it seems that John's detector is a viable choice for spectrum sensing when there are multiple primary users and/or multipath channels.

ACKNOWLEDGMENT

This work is supported by the Academy of Finland and the Nokia Foundation. We thank the reviewers and the editor for their constructive comments, which significantly improved the presentation and the scope of this paper.

APPENDIX A
PROOF OF LEMMA 1

We state the following definition first.

Definition 1. *Under $\mathcal{H}_0$, the joint density of the unordered eigenvalues $\lambda_i \in [0, \infty)$ for the sample covariance matrix $\mathbf{R}$ reads [19]*

$$\Lambda(\lambda_1, \ldots, \lambda_K) := \frac{C}{K!} \prod_{1 \le i < j \le K} (\lambda_i - \lambda_j)^2 \prod_{i=1}^{K} \lambda_i^{N-K} \mathrm{e}^{-\lambda_i}, \tag{27}$$

where $C = \left(\prod_{i=1}^{K} \Gamma(N - i + 1) \Gamma(K - i + 1) \right)^{-1}$.

We now prove Lemma 1.

Proof: Consider the transform $\lambda_i = \xi_i^2$, $i = 1, \ldots, K$ on the density (27), the resulting joint density of ξ_i, up to a constant, equals

$$\prod_{1 \le i < j \le K} \left(\xi_i^2 - \xi_j^2 \right)^2 \prod_{i=1}^{K} \xi_i^{2(N-K)+1} \mathrm{e}^{-\xi_i^2}, \tag{28}$$

and the random variable T_{J} now becomes

$$T_{\text{J}} = \frac{\sum_{i=1}^{K} \xi_i^4}{\left(\sum_{i=1}^{K} \xi_i^2 \right)^2}. \tag{29}$$

For reasons that will become clear later, we consider the general polar coordinates transformation

$$\xi_1 = \rho\cos\phi_1 \tag{30a}$$

$$\xi_i = \rho\left(\prod_{j=1}^{i-1}\sin\phi_j\right)\cos\phi_i, \quad i=2,\ldots,K-1 \tag{30b}$$

$$\xi_K = \rho\left(\prod_{j=1}^{K-1}\sin\phi_j\right), \tag{30c}$$

where $\rho > 0$, $0 < \phi_i \leq \pi$, $i = 1,\ldots,K-2$ and $0 < \phi_{K-1} \leq 2\pi$. The Jacobian of this transformation is given by [20, Theorem 1.25]

$$\rho^{K-1}\left(\prod_{j=1}^{K-1}|\sin\phi_j|^{K-j-1}\right). \tag{31}$$

Consequently the joint distribution of ρ and $\vec{\phi} = (\phi_1,\ldots,\phi_{K-1})$, up to a constant, is calculated to be

$$\rho^{2KN-1}\mathrm{e}^{-\rho^2}\times g(\vec{\phi}), \tag{32}$$

where the function $g(\vec{\phi})$ does not depend on ρ. Clearly, ρ and $\vec{\phi}$ are independent. Moreover, inserting (30) into (29) it is seen that T_{J} does not involve ρ. Finally, by using the fact that $\rho^2 = \sum_{i=1}^K \xi_i^2 = \sum_{i=1}^K \lambda_i$ the proof of Lemma 1 is completed. ∎

APPENDIX B
PROOF OF PROPOSITION 1

Before we prove Proposition 1 we need the following two lemmas, the proofs of which are omitted and can be found in [12].

Lemma 2. *The expected value of the function $\prod_{i=1}^K \lambda_i^{2a_i}$, where a_i are non-negative integers, equals*

$$\int_{[0,\infty)^K}\left(\prod_{i=1}^K\lambda_i^{2a_i}\right)\Lambda(\lambda_1,\ldots,\lambda_K)\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_K = \frac{C}{K!}\sum_{\pi}\left|\Gamma(2a_{\pi(i)}+N-K+i+j-1)\right|_{i,j=1,\ldots,K}, \tag{33}$$

where $\pi = \pi(1),\ldots,\pi(K)$ defines a permutation of the integers $1,\ldots,K$ and the sum is over all the $K!$ permutations.

Alternatively, Lemma 2 may be established by simplifying the tensor notation in [22, Theorem 2]. Note that Lemma 2 is an extension to the Selberg type integrals considered in [23, Eq. (17.6.5) and (17.8.1)].

Lemma 3. *For positive integers b_i the following equality holds*

$$\left|\Gamma(b_i+j-1)\right|_{i,j=1,\ldots,K} = \prod_{1\leq i<j\leq K}(b_j-b_i)\prod_{i=1}^K\Gamma(b_i). \tag{34}$$

Now we are in a position to prove Proposition 1.

Proof: By virtue of the independence (12), we can consider the moments of the random variables $\sum_{i=1}^K\lambda_i^2$ and $\left(\sum_{i=1}^K\lambda_i\right)^2$ separately under $\mathcal{H}_0$. It is well known that the linear statistics $2\,\mathrm{tr}(\mathbf{R}) = 2\sum_{i=1}^K\lambda_i$ of (27) follows a chi-squared distribution with $2KN$ degrees of freedom. Using the moment expression for chi-squared distribution [24, Eq. (2.35)], the $2m$-th moment of $\sum_{i=1}^K\lambda_i$ is obtained as

$$\mathbb{E}\left[\left(\sum_{i=1}^K\lambda_i\right)^{2m}\right] = \frac{\Gamma(2m+KN)}{\Gamma(KN)}. \tag{35}$$

The next step is to calculate the moments of $\sum_{i=1}^K\lambda_i^2$. Using the multinomial expansion

$$\left(\sum_{i=1}^K\lambda_i^2\right)^m = \sum_{a_1+\cdots+a_K=m}\frac{m!}{a_1!\cdots a_K!}\prod_{i=1}^K\lambda_i^{2a_i} \tag{36}$$

and Lemma 2, we immediately have

$$\mathbb{E}\left[\left(\sum_{i=1}^K\lambda_i^2\right)^m\right] = \frac{C}{K!}\sum_{a_1+\cdots+a_K=m}\frac{m!}{a_1!\cdots a_K!}\times \sum_{\pi}\left|\Gamma(2a_{\pi(i)}+N-K+i+j-1)\right|_{i,j=1,\ldots,K}. \tag{37}$$

For any given permutation π, it is observed that the sum over $a_1+\cdots+a_K=m$ in (37) is invariant under the permutation π and since the number of possible permutations is $K!$, (37) now equals

$$\sum_{a_1+\cdots+a_K=m}\frac{m!C\left|\Gamma(2a_i+N-K+i+j-1)\right|_{i,j=1,\ldots,K}}{a_1!\cdots a_K!}. \tag{38}$$

Simplifying the above determinant using Lemma 3 with $b_i = 2a_i+N-K+i$, we have

$$\mathbb{E}\left[\left(\sum_{i=1}^K\lambda_i^2\right)^m\right] = \sum_{a_1+\cdots+a_K=m}\frac{m!C}{a_1!\cdots a_K!}\times \prod_{1\leq i<j\leq K}(2a_j-2a_i+j-i)\prod_{i=1}^K\Gamma(2a_i+N-K+i). \tag{39}$$

Inserting (39) and (35) into (12) completes the proof of Proposition 1. ∎

APPENDIX C
PROOF OF PROPOSITION 2

Proof: The transform $x = \frac{(K-1)z+1}{K}$ on a standard Beta density $z^{\alpha-1}(1-z)^{\beta-1}/B(\alpha,\beta)$, $z\in[0,1]$ leads to a generalized Beta density

$$f_{\mathrm{GB}}(x) := \frac{K^{\alpha+\beta-1}}{B(\alpha,\beta)(K-1)^{\alpha+\beta-1}}\left(x-\frac{1}{K}\right)^{\alpha-1}(1-x)^{\beta-1}, \tag{40}$$

having the same support $x\in[1/K,1]$ as that of T_{J}. By definition, the CDF of this generalized Beta is calculated as

$$F_{\mathrm{GB}}(y) := \int_{1/K}^y f_{\mathrm{GB}}(x)\,\mathrm{d}x = 1-\frac{B\left(\frac{K(1-y)}{K-1};\beta,\alpha\right)}{B(\alpha,\beta)}. \tag{41}$$

The next step is to obtain the parameters α and β of (41) as functions of the moments of T_{J} (13) by moment matching. Specifically, since the m-th moment of a standard Beta random variable equals $\mathbb{E}[z^m] = (\alpha)_m/(\alpha+\beta)_m$, where $(\alpha)_m =$

$\Gamma(\alpha+m)/\Gamma(\alpha)$ defines the Pochhammer symbol, the m-th moment of the generalized Beta random variable is obtained by binomial expansion as

$$\mathbb{E}[x^m] = \frac{1}{K^m}\sum_{i=0}^{m}\binom{m}{i}\frac{(K-1)^i(\alpha)_i}{(\alpha+\beta)_i}, \tag{42}$$

where $\binom{m}{i}$ denotes the binomial coefficient. In particular, by matching the first two moments in (42) to the moments of T_{J} (13) we have

$$\mathcal{M}_1 = \frac{\alpha K+\beta}{(\alpha+\beta)K}, \quad \mathcal{M}_2 = \frac{(\alpha K+\beta)^2+\alpha K^2+\beta}{(\alpha+\beta)(\alpha+\beta+1)K^2}. \tag{43}$$

From the above equations the parameters α, β, denoted by α_0 and β_0, are solved as in (15). This completes the proof. ■

APPENDIX D
PROOF OF PROPOSITION 3

Proof: The proof is divided into two parts. First, we prove the equalities in (21), with which we then establish the estimates of (19) and (20) via the 'Delta method'. For the first part we need up to the 4-th moment of the sample covariance matrix $\mathbf{R}$ under $\mathcal{H}_1$. By definition, the n-th moment of $\mathbf{R}$ can be represented as

$$\mathbb{E}\left[\prod_{k=1}^{n}\mathbf{R}_{i_k,j_k}\right] = \left.\frac{\partial\Phi(\boldsymbol{\Theta})}{\partial\theta_{j_1,i_1}\partial\theta_{j_2,i_2}\ldots\partial\theta_{j_n,i_n}}\right|_{\boldsymbol{\Theta}=\mathbf{0}}, \tag{44}$$

where $\Phi(\boldsymbol{\Theta}) := \mathbb{E}\left[\mathrm{e}^{\mathrm{tr}(\mathbf{R}\boldsymbol{\Theta})}\right] = |\boldsymbol{\Sigma}|^{-N}|\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Theta}|^{-N}$ denotes the moment generating function of $\mathbf{R}$ under $\mathcal{H}_1$ with the i,j-th entry of the Hermitian parameter matrix $\boldsymbol{\Theta}$ being $\theta_{i,j}$. It was proven in [30] using properties of permutation group representations that (44) admits the following combinatorial structure

$$\left.\frac{\partial\Phi(\boldsymbol{\Theta})}{\partial\theta_{j_1,i_1}\partial\theta_{j_2,i_2}\ldots\partial\theta_{j_n,i_n}}\right|_{\boldsymbol{\Theta}=\mathbf{0}} = \sum_{i=1}^{n}N^{n-i+1}A_{i-1}, \tag{45}$$

where the scalar A_i is a sum of *distinct* terms of the form $\prod_{k=1}^{n}\boldsymbol{\Sigma}_{i_k,j_{\pi(k)}}$ with an index distance[5] i, where π defines a permutation of integers $1,\ldots,n$. This combinatorial structure (45) was first observed in [31], where up to the 4-th moment of $\mathbf{R}$ were calculated. Utilizing these moment expressions and the property of matrix trace operation[6]

$$\mathrm{tr}\left(\mathbf{D}^m\right) = \sum_{i_1,\ldots,i_m\in\{1,\ldots,K\}} d_{i_1,i_2}d_{i_2,i_3}\cdots d_{i_m,i_1}, \tag{46}$$

the equalities in (21) can now be proven. Here we prove only the relatively involved cases of (21b) and (21e) as shown on top of next page, other equalities in (21) can be easily obtained in a similar manner.

With the derived moments and covariance of x and y, some estimates of the moments of z can now be established via the 'Delta method' [29]. Specifically, consider the bi-variate Taylor series expansion (53) of the function $z(x,y) = x/y^2$ in the neighborhood of the means of x and y. The partial derivatives of non-zero value at (μ_x,μ_y) in (53) are calculated as

$$\left.\frac{\partial z}{\partial x}\right|_{x=\mu_x,y=\mu_y} = \frac{1}{\mu_y^2}, \quad \left.\frac{\partial z}{\partial y}\right|_{x=\mu_x,y=\mu_y} = -\frac{2\mu_x}{\mu_y^3}, \tag{54a}$$

$$\left.\frac{\partial^2 z}{\partial x\partial y}\right|_{x=\mu_x,y=\mu_y} = -\frac{2}{\mu_y^3}, \quad \left.\frac{\partial^2 z}{\partial y^2}\right|_{x=\mu_x,y=\mu_y} = \frac{6\mu_x}{\mu_y^4}. \tag{54b}$$

Inserting (54) into (53) and taking the expectation of (53), the second order approximation to the mean of z is obtained as

$$\begin{aligned}\mu_z &\approx z(\mu_x,\mu_y) + \frac{1}{2!}\left(\frac{6\mu_x}{\mu_y^4}\mathbb{E}\left[(y-\mu_y)^2\right] + \right.\\ &\qquad \left.\left(-\frac{2}{\mu_y^3}\right)2\mathbb{E}\left[(x-\mu_x)(y-\mu_y)\right]\right) \qquad (55)\\ &= \frac{\mu_x}{\mu_y^2} - \frac{2\mu_{xy}}{\mu_y^3} + \frac{3\mu_x\nu_y}{\mu_y^4}, \qquad (56)\end{aligned}$$

and similarly, calculating the variance of (53) considering up to the linear terms, the first order approximation to the variance of z reads

$$\begin{aligned}\nu_z &\approx \left(\frac{1}{\mu_y^2}\right)^2\mathbb{V}\left[x-\mu_x\right] + \left(-\frac{2\mu_x}{\mu_y^3}\right)^2\mathbb{V}\left[y-\mu_y\right] + \\ &\qquad \frac{2}{\mu_y^2}\left(-\frac{2\mu_x}{\mu_y^3}\right)\mu_{xy} \qquad (57)\\ &= \frac{\nu_x}{\mu_y^4} - \frac{4\mu_x\mu_{xy}}{\mu_y^5} + \frac{4\mu_x^2\nu_y}{\mu_y^6}, \qquad (58)\end{aligned}$$

where $\mathbb{V}[\cdot]$ denotes the variance operation. This completes the proof. ■

[5]The index distance is defined as the minimum index permutations (restricted to row-to-row or column-to-column permutations) required such that a term $\prod_{k=1}^{n}\boldsymbol{\Sigma}_{i_k,j_{\pi(k)}}$ is permutated to the canonical form $\prod_{k=1}^{n}\boldsymbol{\Sigma}_{i_k,j_k}$. For example, the index distance of the term $\boldsymbol{\Sigma}_{i_1,j_3}\boldsymbol{\Sigma}_{i_2,j_4}\boldsymbol{\Sigma}_{i_3,j_2}\boldsymbol{\Sigma}_{i_4,j_1}$ is 3.

[6]Here $\mathbf{D}$ a $K\times K$ matrix with i,j-th entry $d_{i,j}$.

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] L. Wei and O. Tirkkonen, "Cooperative spectrum sensing of OFDM signals using largest eigenvalue distributions," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.,* Sep. 2009.

[4] 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.

[5] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *IEEE Trans. Veh. Technol.,* vol. 59, no. 4, pp. 1791-1800, May 2010.

[6] 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.

[7] 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.

[8] L. Wei and O. Tirkkonen, "Analysis of scaled largest eigenvalue based detection for spectrum sensing," in *Proc. IEEE Int. Conf. Commun.,* June 2011.

[9] B. Zhao, Y. Chen, C. He, and L. Jiang, "Performance analysis of spectrum sensing with multiple primary users," *IEEE Trans. Veh. Technol.,* vol. 61, no. 2, pp. 914-919, Feb. 2012.

[10] 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.

[11] 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.

$$\nu_x = \frac{1}{N^4}\mathbb{E}\left[\left(\mathrm{tr}(\mathbf{R}^2)\right)^2\right] - \mu_x^2 \tag{47}$$

$$= \frac{1}{N^4}\sum_{a,b,c,d\in\{1,\ldots,K\}} \mathbb{E}\left[\mathbf{R}_{a,b}\mathbf{R}_{b,a}\mathbf{R}_{c,d}\mathbf{R}_{d,c}\right] - \mu_x^2 \tag{48}$$

$$= \frac{1}{N^4}\sum_{a,b,c,d\in\{1,\ldots,K\}} \Big(N^4\boldsymbol{\Sigma}_{a,b}\boldsymbol{\Sigma}_{b,a}\boldsymbol{\Sigma}_{c,d}\boldsymbol{\Sigma}_{d,c} + N^3\big(\boldsymbol{\Sigma}_{a,a}\boldsymbol{\Sigma}_{b,b}\boldsymbol{\Sigma}_{c,d}\boldsymbol{\Sigma}_{d,c} + \boldsymbol{\Sigma}_{a,d}\boldsymbol{\Sigma}_{b,a}\boldsymbol{\Sigma}_{c,b}\boldsymbol{\Sigma}_{d,c} + \boldsymbol{\Sigma}_{a,c}\boldsymbol{\Sigma}_{b,a}\boldsymbol{\Sigma}_{c,d}\boldsymbol{\Sigma}_{d,b} + \boldsymbol{\Sigma}_{a,b}\boldsymbol{\Sigma}_{b,d}\boldsymbol{\Sigma}_{c,a}\boldsymbol{\Sigma}_{d,c} + \boldsymbol{\Sigma}_{a,b}\boldsymbol{\Sigma}_{b,c}\boldsymbol{\Sigma}_{c,d}\boldsymbol{\Sigma}_{d,a} + \boldsymbol{\Sigma}_{a,b}\boldsymbol{\Sigma}_{b,a}\boldsymbol{\Sigma}_{c,c}\boldsymbol{\Sigma}_{d,d}\big) + N^2\big(\boldsymbol{\Sigma}_{a,a}\boldsymbol{\Sigma}_{b,d}\boldsymbol{\Sigma}_{c,b}\boldsymbol{\Sigma}_{d,c} + \boldsymbol{\Sigma}_{a,d}\boldsymbol{\Sigma}_{b,b}\boldsymbol{\Sigma}_{c,a}\boldsymbol{\Sigma}_{d,c} + \boldsymbol{\Sigma}_{a,a}\boldsymbol{\Sigma}_{b,c}\boldsymbol{\Sigma}_{c,d}\boldsymbol{\Sigma}_{d,b} + \boldsymbol{\Sigma}_{a,c}\boldsymbol{\Sigma}_{b,b}\boldsymbol{\Sigma}_{c,d}\boldsymbol{\Sigma}_{d,a} + \boldsymbol{\Sigma}_{a,d}\boldsymbol{\Sigma}_{b,a}\boldsymbol{\Sigma}_{c,c}\boldsymbol{\Sigma}_{d,b} + \boldsymbol{\Sigma}_{a,c}\boldsymbol{\Sigma}_{b,a}\boldsymbol{\Sigma}_{c,b}\boldsymbol{\Sigma}_{d,d} + \boldsymbol{\Sigma}_{a,b}\boldsymbol{\Sigma}_{b,d}\boldsymbol{\Sigma}_{c,c}\boldsymbol{\Sigma}_{d,a} + \boldsymbol{\Sigma}_{a,b}\boldsymbol{\Sigma}_{b,c}\boldsymbol{\Sigma}_{c,a}\boldsymbol{\Sigma}_{d,d} + \boldsymbol{\Sigma}_{a,a}\boldsymbol{\Sigma}_{b,b}\boldsymbol{\Sigma}_{c,c}\boldsymbol{\Sigma}_{d,d} + \boldsymbol{\Sigma}_{a,d}\boldsymbol{\Sigma}_{b,c}\boldsymbol{\Sigma}_{c,b}\boldsymbol{\Sigma}_{d,a} + \boldsymbol{\Sigma}_{a,c}\boldsymbol{\Sigma}_{b,d}\boldsymbol{\Sigma}_{c,a}\boldsymbol{\Sigma}_{d,b}\big) + N\big(\boldsymbol{\Sigma}_{a,a}\boldsymbol{\Sigma}_{b,d}\boldsymbol{\Sigma}_{c,c}\boldsymbol{\Sigma}_{d,b} + \boldsymbol{\Sigma}_{a,a}\boldsymbol{\Sigma}_{b,c}\boldsymbol{\Sigma}_{c,b}\boldsymbol{\Sigma}_{d,d} + \boldsymbol{\Sigma}_{a,d}\boldsymbol{\Sigma}_{b,b}\boldsymbol{\Sigma}_{c,c}\boldsymbol{\Sigma}_{d,a} + \boldsymbol{\Sigma}_{a,d}\boldsymbol{\Sigma}_{b,c}\boldsymbol{\Sigma}_{c,a}\boldsymbol{\Sigma}_{d,b} + \boldsymbol{\Sigma}_{a,c}\boldsymbol{\Sigma}_{b,b}\boldsymbol{\Sigma}_{c,a}\boldsymbol{\Sigma}_{d,d} + \boldsymbol{\Sigma}_{a,c}\boldsymbol{\Sigma}_{b,d}\boldsymbol{\Sigma}_{c,b}\boldsymbol{\Sigma}_{d,a}\big)\Big) - \mu_x^2, \tag{49}$$

where by repeated use of the property (46) and after some manipulations the equality (21b) is established.

$$\mu_{xy} = \frac{1}{N^3}\mathbb{E}\left[\mathrm{tr}(\mathbf{R}^2)\mathrm{tr}(\mathbf{R})\right] - \mu_x\mu_y \tag{50}$$

$$= \frac{1}{N^3}\sum_{a,b,c\in\{1,\ldots,K\}} \mathbb{E}\left[\mathbf{R}_{a,b}\mathbf{R}_{b,a}\mathbf{R}_{c,c}\right] - \mu_x\mu_y \tag{51}$$

$$= \frac{1}{N^3}\sum_{a,b,c\in\{1,\ldots,K\}} \Big(N^3\boldsymbol{\Sigma}_{a,b}\boldsymbol{\Sigma}_{b,a}\boldsymbol{\Sigma}_{c,c} + N^2\left(\boldsymbol{\Sigma}_{a,c}\boldsymbol{\Sigma}_{b,a}\boldsymbol{\Sigma}_{c,b} + \boldsymbol{\Sigma}_{a,b}\boldsymbol{\Sigma}_{b,c}\boldsymbol{\Sigma}_{c,a} + \boldsymbol{\Sigma}_{a,a}\boldsymbol{\Sigma}_{b,b}\boldsymbol{\Sigma}_{c,c}\right) + N\left(\boldsymbol{\Sigma}_{a,c}\boldsymbol{\Sigma}_{b,b}\boldsymbol{\Sigma}_{c,a} + \boldsymbol{\Sigma}_{a,a}\boldsymbol{\Sigma}_{b,c}\boldsymbol{\Sigma}_{c,b}\right)\Big) - \mu_x\mu_y, \tag{52}$$

where by repeated use of the property (46) and after some manipulations the equality (21e) is established.

$$z(x,y) = z(\mu_x,\mu_y) + \left(\frac{\partial z}{\partial x}\Big|_{x=\mu_x,y=\mu_y}\right)(x-\mu_x) + \left(\frac{\partial z}{\partial y}\Big|_{x=\mu_x,y=\mu_y}\right)(y-\mu_y) + \frac{1}{2!}\Bigg(\left(\frac{\partial^2 z}{\partial x^2}\Big|_{x=\mu_x,y=\mu_y}\right)(x-\mu_x)^2 + \left(\frac{\partial^2 z}{\partial y^2}\Big|_{x=\mu_x,y=\mu_y}\right)(y-\mu_y)^2 + \left(\frac{\partial^2 z}{\partial x\partial y}\Big|_{x=\mu_x,y=\mu_y}\right)2(x-\mu_x)(y-\mu_y)\Bigg) + \cdots, \tag{53}$$

[12] L. Wei, P. Dharmawansa, and O. Tirkkonen, "Locally best invariant test for multiple primary user spectrum sensing," in *Proc. Int. Conf. Cognitive Radio Oriented Wireless Net. Commun.,* June 2012.

[13] FCC, "In the matter of unlicensed operation in the TV broadcast bands: second memorandum opinion and order, Federal Communications Commission," FCC 10-174, Sep. 2010.

[14] 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.

[15] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis.* Wiley, 2003.

[16] S. John, "Some optimal multivariate tests," *Biometrika*, vol. 58, no. 1, pp. 123-127, Apr. 1971.

[17] N. Sugiura, "Locally best invariant test for sphericity and the limiting distributions," *The Annals of Mathematical Statistics,* vol. 43, no. 4, pp. 1312-1316, Aug. 1972.

[18] 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.

[19] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *The Annals of Mathematical Statistics,* vol. 35, no. 2, pp. 475-501, 1964.

[20] A. M. Mathai, *Jacobians of Matrix Transformations and Functions of Matrix Arguments.* World Scientific, 1997.

[21] S. John, "The distribution of a statistic used for testing sphericity of normal distributions," *Biometrika,* vol. 59, no. 1, pp. 169-173, Apr. 1972.

[22] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. Inf. Theory,* vol. 49, no. 10, pp. 2363-2371, Oct. 2003.

[23] M. L. Mehta, *Random Matrices.* 3rd Edition, Elsevier, 2004.

[24] M. K. Simon, *Distributions Involving Gaussian Random Variables.* New York: Springer, 2002.

[25] S. John, "Fitting sampling distribution agreeing in support and moments and tables of critical values of sphericity criterion," *Journal of Multivariate Analysis,* vol. 6, pp. 601-607, 1976.

[26] R. J. Boik, "Algorithm AS 284: Null distribution of a statistics for testing sphericity and additivity: a Jacobi polynomial expansion," *Journal of the Royal Statistical Society, series C (Applied Statistics),* vol. 42, no. 3, pp. 567-576, 1993.

[27] H. Nagao, "On some test criteria for covariance matrix," *The Annals of Statistics,* vol. 1, no. 4, pp. 700-709, 1973.

[28] O. Ledoit and M. Wolf, "Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size," *The Annals of Statistics,* vol. 30, no. 4, pp. 1081-1102, 2002.

[29] H. Cramér, *Mathematical Methods of Statistics.* Princeton University Press, 1999.
[30] P. Graczyk, G. Letac, and H. Massam, "The complex Wishart distribution and the symmetric group," *The Annals of Statistics,* vol. 31, no. 1, pp. 287-309, 2003.
[31] 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.
[32] H. Nagao and M. S. Srivastava, "On the distribution of some test criteria for a covariance matrix under local alternatives and bootstrap approximations," *Journal of Multivariate Analysis,* vol. 43, pp. 331-350, 1992.
[33] H. Hochstadt, *Special Functions of Mathematical Physics.* Holt, Rinehart and Winston, New York, 1961.
[34] H. T. Ha, *Advances in Moment-Based Density Approximation Methods.* Ph.D thesis, University of Western Ontario, 2006.
[35] R. Tandra and A. Sahai, "SNR walls for signal detetion," *IEEE J. Sel. Topics in Signal Process.,* vol. 2, no. 1, Feb. 2008.
[36] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," in *Proc. IEEE Int. Conf. Commun.,* May 2003.

Lu Wei (S'08–M'13) received his B.Eng. degree from Xi'an Jiaotong University, China, in 2006, and M.Sc. degree (with distinction) from Helsinki University of Technology (now Aalto university), Finland, in 2008. Currently, he is pursuing a Ph.D degree in Aalto University, Finland. His research interests are in multi-antenna techniques and cognitive radio.

Prathapasinghe Dharmawansa (S'05–M'09) received the B.Sc. and M.Sc. degrees in Electronic and Telecommunication Engineering from the University of Moratuwa, Moratuwa, Sri Lanka in 2003 and 2004 respectively, and the D.Eng. degree in Information and Communications Technology from the Asian Institute of Technology, Thailand, in 2007. He subsequently joined the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST) as a Research Associate. From October 2011 to October 2012, he was with the Department of Communications and Networking, Aalto University (formerly Helsinki University of Technology) School of Electrical Engineering, Finland as a Postdoctoral Researcher. Currently he is a Research Associate with the Department of Statistics, Stanford University, CA. His research interests are in random matrix theory, multivariate statistics, communication theory and signal processing. He received a Best Paper Award in Communication Theory at the IEEE International Conference on Communications (ICC 2011) held in Kyoto, Japan.

Olav Tirkkonen (M'01) Olav Tirkkonen is associate professor in communication theory at the Department of Communications and Networking in Aalto University, Finland, where he has held a faculty position since August 2006. He received his M.Sc. and Ph.D. degrees in theoretical physics from Helsinki University of Technology in 1990 and 1994, respectively. Between 1994 and 1999 he held post-doctoral positions at the University of British Columbia, Vancouver, Canada, and the Nordic Institute for Theoretical Physics, Copenhagen, Denmark. From 1999 to 2010 he was with Nokia Research Center (NRC), Helsinki, Finland, most recently acting as Research Fellow. He has published some 150 papers, and is coauthor of the book "Multiantenna transceiver techniques for 3G and beyond". His current research interests are in multiantenna techniques, cognitive cellular systems, and self-organization of wireless networks.