Performance of Statistical Tests for Single Source Detection using Random Matrix Theory

Pascal Bianchi, Merouane Debbah, Mylène Maïda, Jamal Najim

Performance of Statistical Tests for Single Source Detection using Random Matrix Theory

P. Bianchi, M. Debbah, M. Maida and J. Najim

April 19, 2010

Abstract

This paper introduces a unified framework for the detection of a single source with a sensor array in the context where the noise variance and the channel between the source and the sensors are unknown at the receiver. The Generalized Maximum Likelihood Test is studied and yields the analysis of the ratio between the maximum eigenvalue of the sampled covariance matrix and its normalized trace. Using recent results from random matrix theory, a practical way to evaluate the threshold and the p-value of the test is provided in the asymptotic regime where the number K of sensors and the number N of observations per sensor are large but have the same order of magnitude. The theoretical performance of the test is then analyzed in terms of Receiver Operating Characteristic (ROC) curve. It is in particular proved that both Type I and Type II error probabilities converge to zero exponentially as the dimensions increase at the same rate, and closed-form expressions are provided for the error exponents. These theoretical results rely on a precise description of the large deviations of the largest eigenvalue of spiked random matrix models, and establish that the presented test asymptotically outperforms the popular test based on the condition number of the sampled covariance matrix.

This work was partially supported by french programs ANR-07-MDCO-012-01 ‘Sesame’, and ANR-08-BLAN-0311-03 ‘GranMa’.

P. Bianchi and J. Najim are with CNRS and Télécom Paristech, France. {bianchi,najim}@telecom-paristech.fr ,

M. Debbah is with SUPELEC and holds Alcatel-Lucent/Supélec Flexible Radio chair, France merouane.debbah@supelec.fr ,

M. Maida is with Université Paris-Sud, UMR CNRS 8628, France. Mylene.Maida@math.u-psud.fr ,

I. INTRODUCTION

The detection of a source by a sensor array is at the heart of many wireless applications. It is of particular interest in the realm of cognitive radio [1], [2] where a multi-sensor cognitive device (or a collaborative network[1]) needs to discover or sense by itself the surrounding environment. This allows the cognitive device to make relevant choices in terms of information to feed back, bandwidth to occupy or transmission power to use. When the cognitive device is switched on, its prior knowledge (on the noise variance for example) is very limited and can rarely be estimated prior to the reception of data. This unfortunately rules out classical techniques based on energy detection [4], [5], [6] and requires new sophisticated techniques exploiting the space or spectrum dimension.

In our setting, the aim of the multi-sensor cognitive detection phase is to construct and analyze tests associated with the following hypothesis testing problem:

$$\boldsymbol{y}(n) = \begin{cases} \boldsymbol{w}(n) & \text{under } H_0 \\ \boldsymbol{h}\, s(n) + \boldsymbol{w}(n) & \text{under } H_1 \end{cases} \quad \text{for} \quad n = 0 : N-1 \ , \tag{1}$$

where $\boldsymbol{y}(n) = [y_1(n), \ldots, y_K(n)]^T$ is the observed $K \times 1$ complex time series, $\boldsymbol{w}(n)$ represents a $K \times 1$ complex circular Gaussian white noise process with unknown variance σ^2, and N represents the number of received samples. Vector $\boldsymbol{h} \in \mathbb{C}^{K\times 1}$ is a deterministic vector and typically represents the propagation channel between the source and the K sensors. Signal $s(n)$ denotes a standard scalar independent and identically distributed (i.i.d.) circular complex Gaussian process with respect to the samples $n = 0 : N-1$ and stands for the source signal to be detected.

The standard case where the propagation channel and the noise variance are known has been thoroughly studied in the literature in the Single Input Single Output case [4], [5], [6] and Multi-Input Multi-Ouput [7] case. In this simple context, the most natural approach to detect the presence of source $s(n)$ is the well-known *Neyman-Pearson* (NP) procedure which consists in rejecting the null hypothesis when the observed likelihood ratio lies above a certain threshold [8]. Traditionally, the value of the threshold is set in such a way that the *Probability of False Alarm* (PFA) is no larger than a predefined *level* $\alpha \in (0, 1)$. Recall that the PFA (resp. the miss

[1]The collaborative network corresponds to multiple base stations connected, in a wireless or wired manner, to form a virtual antenna system[3].

probability) of a test is defined as the probability that the receiver decides hypothesis H_1 (resp. H_0) when the true hypothesis is H_0 (resp. H_1). The NP test is known to be uniformly most powerful *i.e.*, for any level $\alpha \in (0,1)$, the NP test has the minimum achievable miss probability (or equivalently the maximum achievable power) among all tests of level α. In this paper, we assume on the opposite that:

- the noise variance σ^2 is unknown,
- vector $\boldsymbol{h}$ is unknown.

In this context, probability density functions of the observations $\boldsymbol{y}(n)$ under both H_0 and H_1 are unknown, and the classical NP approach can no longer be employed. As a consequence, the construction of relevant tests for (1) together with the analysis fo their perfomances is a crucial issue. The classical approach followed in this paper consists in replacing the unknown parameters by their maximum likelihood estimates. This leads to the so-called *Generalized Likelihood Ratio* (GLR). The *Generalized Likelihood Ratio Test* (GLRT), which rejects the null hypothesis for large values of the GLR, easily reduces to the statistics given by the ratio of the largest eigenvalue of the sampled covariance matrix with its normalized trace, cf. [9], [10], [11]. Nearby statistics [12], [13], [14], [15], with good practical properties, have also been developed, but would not yield a different (asymptotic) error exponent analysis.

In this paper, we analyze the performance of the GLRT in the asymptotic regime where the number K of sensors and the number N of observations per sensor are large but have the same order of magnitude. This assumption is relevant in many applications, among which cognitive radio for instance, and casts the problem into a large random matrix framework.

Large random matrix theory has already been applied to signal detection [16] (see also [17]), and recently to hypothesis testing [15], [18], [19]. In this article, the focus is mainly devoted to the study of the largest eigenvalue of the sampled covariance matrix, whose behaviour changes under H_0 or H_1. The fluctuations of the largest eigenvalue under H_0 have been described by Johnstone [20] by means of the celebrated Tracy-Widom distribution, and are used to study the threshold and the p-value of the GLRT.

In order to characterize the performance of the test, a natural approach would have been to evaluate the *Receiver Operating Characteristic* (ROC) curve of the GLRT, that is to plot the power of the test versus a given level of confidence. Unfortunately, the ROC curve does not admit any simple closed-form expression for a finite number of sensors and snapshots. As the

miss probability of the GLRT goes exponentially fast to zero, the performance of the GLRT is analyzed via the computation of its error exponent, which caracterizes the speed of decrease to zero. Its computation relies on the study of the large deviations of the largest eigenvalue of 'spiked' sampled covariance matrix. By 'spiked' we refer to the case where the eigenvalue converges outside the bulk of the limiting spectral distribution, which precisely happens under hypothesis H_1. We build upon [21] to establish the large deviation principle, and provide a closed-form expression for the rate function.

We also introduce the error exponent curve, and plot the error exponent of the power of the test versus the error exponent for a given level of confidence. The error exponent curve can be interpreted as an asymptotic version of the ROC curve in a $\log$-$\log$ scale and enables us to establish that the GLRT outperforms another test based on the condition number, and proposed by [22], [23], [24] in the context of cognitive radio.

Notice that the results provided here (determination of the threshold of the GLRT test and the computation of the error exponents) would still hold within the setting of real Gaussian random variables instead of complex ones, with minor modifications[2].

The paper is organized as follows.

Section II introduces the GLRT. The value of the threshold, which completes the definition of the GLRT, is established in Section II-B. As the latter threshold has no simple closed-form expression and as its practical evaluation is difficult, we introduce in Section II-C an asymptotic framework where it is assumed that both the number of sensors K and the number N of available snapshots go to infinity at the same rate. This assumption is valid for instance in cognitive radio contexts and yields a very simple evaluation of the threshold, which is important in real-time applications.

In Section III, we recall several results of large random matrix theory, among which the asymptotic fluctuations of the largest eigenvalue of a sample covariance matrix, and the limit of the largest eigenvalue of a spiked model.

These results are used in Section IV where an approximate threshold value is derived, which leads to the same PFA as the optimal one in the asymptotic regime. This analysis yields a relevant practical method to approximate the p-*values* associated with the GLRT.

[2]Details are provided in Remarks 4 and 9.

Section V is devoted to the performance analysis of the GLRT. We compute the error exponent of the GLRT, derive its expression in closed-form by establishing a *Large Deviation Principle* for the test statistic T_N [3], and describe the error exponent curve.

Section VI introduces the test based on the condition number, that is the statistics given by the ratio between the largest eigenvalue and the smallest eigenvalue of the sampled covariance matrix. We provide the error exponent curve associated with this test and prove that the latter is outperformed by the GLRT.

Section VII provides further numerical illustrations and conclusions are drawn in Section VIII.

Mathematical details are provided in the Appendix. In particular, a full rigorous proof of a large deviation principle is provided in Appendix A, while a more informal proof of a nearby large deviation principle, maybe more accessible to the non-specialist, is provided in Appendix B.

Notations

For $i \in \{0,1\}$, $\mathbb{P}_i(\mathcal{E})$ represents the probability of a given event $\mathcal{E}$ under hypothesis H_i. For any real random variable T and any real number γ, notation

$$T \underset{H_0}{\overset{H_1}{\gtrless}} \gamma$$

stands for the test function which rejects the null hypothesis when $T > \gamma$. In this case, the *probability of false alarm (PFA)* of the test is given by $\mathbb{P}_0(T > \gamma)$, while the power of the test is $\mathbb{P}_1(T > \gamma)$. Notation $\xrightarrow[H_i]{a.s.}$ stands for the almost sure (a.s.) convergence under hypothesis H_i. For any one-to-one mapping $T : \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X}$ and $\mathcal{Y}$ are two sets, we denote by T^{-1} the inverse of T w.r.t. composition. For any borel set $A \in \mathbb{R}$, $x \mapsto \mathbf{1}_A(x)$ denotes the indicator function of set A and $\|\boldsymbol{x}\|$ denotes the Euclidian norm of a given vector $\boldsymbol{x}$. If $\boldsymbol{A}$ is a given matrix, denote by $\boldsymbol{A}^H$ its transpose-conjugate. If F is a cumulative distribution function (c.d.f.), we denote by $\bar{F}$ is complementary c.d.f., that is: $\bar{F} = 1 - F$.

[3]Note that in recent papers [25], [14], [15], the fluctuations of the test statistics under H_1, based on large random matrix techniques, have also been used to approximate the power of the test. We believe that the performance analysis based on the error exponent approach, although more involved, has a wider range of validity.

II. Generalized Likelihood Ratio Test

In this section, we derive the Generalized Likelihood Ratio Test (section II-A) and compute the associated threshold and p-value (section II-B). This exact computation raises some computational issues, which are circumvented by the introduction of a relevant asymptotic framework, well-suited for mathematical analysis (Section II-C).

A. Derivation of the Test

Denote by N the number of observed samples and recall that:

$$\boldsymbol{y}(n) = \begin{cases} \boldsymbol{w}(n) & \text{under } H_0 \\ \boldsymbol{h}\, s(n) + \boldsymbol{w}(n) & \text{under } H_1 \end{cases}, \qquad n = 0 : N-1 \ ,$$

where $(\boldsymbol{w}(n), 0 \leq n \leq N-1)$ represents an independent and identically distributed (i.i.d.) process of $K \times 1$ vectors with circular complex Gaussian entries with mean zero and covariance matrix $\sigma^2 \mathbf{I}_K$, vector $\boldsymbol{h} \in \mathbb{C}^{K\times 1}$ is deterministic, signal $(s(n), 0 \leq n \leq N-1)$ denotes a scalar i.i.d. circular complex Gaussian process with zero mean and unit variance. Moreover, $(\boldsymbol{w}(n), 0 \leq n \leq N-1)$ and $(s(n), 0 \leq n \leq N-1)$ are assumed to be independent processes. We stack the observed data into a $K \times N$ matrix $\mathbf{Y} = [\boldsymbol{y}(0), \ldots, \boldsymbol{y}(N-1)]$. Denote by $\hat{\mathbf{R}}$ the sampled covariance matrix:

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H,$$

and respectively, by $p_0(\mathbf{Y}; \sigma^2)$ and $p_1(\mathbf{Y}; \boldsymbol{h}, \sigma^2)$ the likelihood functions of the observation matrix $\mathbf{Y}$ indexed by the unknown parameters $\boldsymbol{h}$ and σ^2 under hypotheses H_0 and H_1.

As $\mathbf{Y}$ is a $K \times N$ matrix whose columns are i.i.d. Gaussian vectors with covariance matrix $\boldsymbol{\Sigma}$ defined by:

$$\boldsymbol{\Sigma} = \begin{cases} \sigma^2 \mathbf{I}_K & \text{under } H_0 \\ \boldsymbol{h}\boldsymbol{h}^H + \sigma^2 \mathbf{I}_K & \text{under } H_1 \end{cases}, \tag{2}$$

the likelihood functions write:

$$p_0(\mathbf{Y}; \sigma^2) = (\pi\sigma^2)^{-NK} \exp\left(-\frac{N}{\sigma^2}\mathrm{tr}\, \hat{\mathbf{R}}\right) \ , \tag{3}$$

$$p_1(\mathbf{Y}; \boldsymbol{h}, \sigma^2) = (\pi^K \det(\boldsymbol{h}\boldsymbol{h}^H + \sigma^2 \mathbf{I}_K))^{-N} \exp\left(-N\mathrm{tr}\, (\hat{\mathbf{R}}(\boldsymbol{h}\boldsymbol{h}^H + \sigma^2 \mathbf{I}_K)^{-1})\right) \ . \tag{4}$$

In the case where parameters $\boldsymbol{h}$ and σ^2 are available, the celebrated Neyman-Pearson procedure yields a uniformly most powerful test, given by the likelihood ratio statistics $\frac{p_1(\mathbf{Y}; \boldsymbol{h}, \sigma^2)}{p_0(\mathbf{Y}; \sigma^2)}$.

However, in the case where $\boldsymbol{h}$ and σ^2 are unknown, which is the problem addressed here, no simple procedure garantees a uniformly most powerful test, and a classical approach consists in computing the GLR:

$$L_N = \frac{\sup_{\boldsymbol{h},\sigma^2} p_1(\mathbf{Y};\boldsymbol{h},\sigma^2)}{\sup_{\sigma^2} p_0(\mathbf{Y};\sigma^2)} \,. \tag{5}$$

In the GLRT procedure, one rejects hypothesis H_0 whenever $L_N > \xi_N$, where ξ_N is a certain threshold which is selected in order that the PFA $\mathbb{P}_0(L_N > \xi_N)$ does not exceed a given level α.

In the following proposition, which follows after straightforward computations from [26] and [9], we derive the closed form expression of the GLR L_N. Denote by $\lambda_1 > \lambda_2 > \cdots > \lambda_K \geq 0$ the ordered eigenvalues of $\hat{\mathbf{R}}$ (all distincts with probability one).

Proposition 1. *Let T_N be defined by:*

$$T_N = \frac{\lambda_1}{\frac{1}{K}\operatorname{tr}\hat{\mathbf{R}}} \,, \tag{6}$$

then, the GLR (cf. Eq. (5)) writes:

$$L_N = \frac{C}{(T_N)^N \left(1 - \frac{T_N}{K}\right)^{(K-1)N}}$$

where $C = \left(1 - \frac{1}{K}\right)^{(1-K)N}$.

By Proposition 1, $L_N = \phi_{N,K}(T_N)$ where $\phi_{N,K} : x \mapsto Cx^{-N}\left(1 - \frac{x}{K}\right)^{N(1-K)}$. The GLRT rejects the null hypothesis when inequality $L_N > \xi_N$ holds. As $T_N \in (1, K)$ with probability one and as $\phi_{N,K}$ is increasing on this interval, the latter inequality is equivalent to $T_N > \phi_{N,K}^{-1}(\xi_N)$. Otherwise stated, the GLRT reduces to the test which rejects the null hypothesis for large values of T_N:

$$T_N \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_N \tag{7}$$

where $\gamma_N = \phi_{N,K}^{-1}(\xi_N)$ is a certain threshold which is such that the PFA does not exceed a given level α. In the sequel, we will therefore focus on the test statistics T_N.

Remark 1. *There exist several variants of the above statistics [12], [13], [14], [15], which merely consist in replacing the normalized trace with a more involved estimate of the noise*

variance. Although very important from a practical point of view, these variants have no impact on the (asymptotic) error exponent analysis. Therefore, we restrict our analysis to the traditional GLRT for the sake of simplicity.

B. Exact threshold and p-values

In order to complete the construction of the test, we must provide a procedure to set the threshold γ_N. As usual, we propose to define γ_N as the value which maximizes the power $\mathbb{P}_1(T_N > \gamma_N)$ of the test (7) while keeping the PFA $\mathbb{P}_0(T_N > \gamma_N)$ under a desired level $\alpha \in (0,1)$. It is well-known (see for instance [8], [27]) that the latter threshold is obtained by:

$$\gamma_N = p_N^{-1}(\alpha) \tag{8}$$

where $p_N(t)$ represents the complementary c.d.f. of the statistics T_N under the null hypothesis:

$$p_N(t) = \mathbb{P}_0(T_N > t)\,. \tag{9}$$

Note that $p_N(t)$ is continuous and decreasing from 1 to 0 on $t \in [0,\infty)$, so that the threshold $p_N^{-1}(\alpha)$ in (8) is always well defined. When the threshold is fixed to $\gamma_N = p_N^{-1}(\alpha)$, the GLRT rejects the null hypothesis when $T_N > p_N^{-1}(\alpha)$ or equivalently, when $p_N(T_N) < \alpha$. It is usually convenient to rewrite the GLRT under the following form:

$$p_N(T_N) \underset{H_1}{\overset{H_0}{\gtrless}} \alpha\,. \tag{10}$$

The statistics $p_N(T_N)$ represents the *significance probability* or p-value of the test. The null hypothesis is rejected when the p-value $p_N(T_N)$ is below the level α. In practice, the computation of the p-value associated with one experiment is of prime importance. Indeed, the p-value not only allows to accept/reject an hypothesis by (10), but it furthermore reflects how strongly the data contradicts the null hypothesis [8].

In order to evaluate p-values, we derive in the sequel the exact expression of the complementary c.d.f. p_N. The crucial point is that T_N is a function of the eigenvalues $\lambda_1, \ldots, \lambda_K$ of the sampled covariance matrix $\hat{\mathbf{R}}$. We have

$$p_N(t) = \int_{\Delta_t} p^0_{K,N}(x_1, \cdots, x_K)\mathrm{d}x_{1:K} \tag{11}$$

where for each t, the domain of integration Δ_t is defined by:

$$\Delta_t = \left\{ (x_1, \ldots, x_K) \in \mathbb{R}^K,\ \frac{K x_1}{x_1 + \cdots + x_K} > t \right\} ,$$

and $p^0_{K,N}$ is the joint probability density function (p.d.f.) of the ordered eigenvalues of $\hat{\mathbf{R}}$ under H_0 given by:

$$p^0_{K,N}(x_{1:K}) = \frac{\mathbf{1}_{(x_1 \geq \cdots \geq x_K \geq 0)}}{Z^0_{K,N}} \prod_{1 \leq i < j \leq K} (x_j - x_i)^2 \prod_{j=1}^{K} x_j^{N-K} e^{-N x_j} \tag{12}$$

where $\mathbf{1}_{(x_1 \geq \cdots \geq x_K \geq 0)}$ stands for the indicator function of the set $\{(x_1 \ldots x_K) : x_1 \geq \cdots \geq x_K \geq 0\}$ and where $Z^0_{K,N}$ is the normalization constant (see for instance [28], [29, Chapter 4]).

Remark 2. *For each t, the computation of $p_N(t)$ requires the numerical evaluation of a non-trivial integral. Despite the fact that powerful numerical methods, based on representations of such integrals with hypergeometric functions [30], are available (see for instance [31], [32]), an* on line *computation, requested in a number of real-time applications, may be out of reach.*

Instead, tables of the function p_N should be computed off line i.e.*, prior to the experiment. As both the dimensions K and N may be subject to frequent changes*[4]*, all possible tables of the function p_N should be available at the detector's side, for all possible values of the couple (N, K). This both requires substantial computations and considerable memory space. In what follows, we propose a way to overcome this issue.*

In the sequel, we study the asymptotic behaviour of the complementary c.d.f. p_N when both the number of sensors K and the number of snapshots N go to infinity at the same rate. This analysis leads to simpler testing procedure.

C. Asymptotic framework

We propose to analyze the asymptotic behaviour of the complementary c.d.f. p_N as the number of observations goes to infinity. More precisely, we consider the case where both the number K of sensors and the number N of snapshots go to infinity at the same speed, as assumed below

$$N \to \infty,\ K \to \infty,\ \ c_N := \frac{K}{N} \to c,\ \text{with } 0 < c < 1. \tag{13}$$

[4]In cognitive radio applications for instance, the number of users K which are connected to the network is frequently varying.

This asymptotic regime is relevant in cases where the sensing system must be able to perform source detection in a moderate amount of time *i.e.,* the number K of sensors and the number N of samples being of the same order. This is in particular the case in cognitive radio applications (see for instance [33]). Very often, the number of sensors is lower than the number of snapshots, hence the ratio c lower than 1.

In the sequel, we will simply denote $N, K \to \infty$ to refer to the asymptotic regime (13).

Remark 3. *The results related to the GLRT presented in Sections IV and V remain true for $c \geq 1$; in the case of the test based on the condition number and presented in Section VI, extra-work is needed to handle the fact that the lowest eigenvalue converges to zero, which happens if $c \geq 1$.*

III. Large random matrices - Largest eigenvalue - Behaviour of the GLR statistics

In this section, we recall a few facts on large random matrices as the dimensions N, K go to infinity. We focus on the behaviour of the eigenvalues of $\hat{\mathbf{R}}$ which differs whether hypothesis H_0 holds (Section III-A) or H_1 holds (Section III-B).

As the column vectors of $\mathbf{Y}$ are i.i.d. complex Gaussian with covariance matrix $\mathbf{\Sigma}$ given by (2), the probability density of $\hat{\mathbf{R}}$ is given by:

$$\frac{1}{Z(N,K,\mathbf{\Sigma})} e^{-N\mathrm{tr}(\mathbf{\Sigma}^{-1}\hat{\mathbf{R}})} (\det\, \hat{\mathbf{R}})^{N-K},$$

where $Z(N,K,\mathbf{\Sigma})$ is a normalizing constant.

A. Behaviour under hypothesis H_0

As the behaviour of T_N does not depend on σ^2, we assume that $\sigma^2 = 1$; in particular, $\mathbf{\Sigma} = \mathbf{I}_K$. Under H_0, matrix $\hat{\mathbf{R}}$ is a complex Wishart matrix and it is well-known (see for instance [28]) that the Jacobian of the transformation between the entries of the matrix and the eigenvalues/angles is given by the Vandermonde determinant $\prod_{1\leq i<j\leq K}(x_j - x_i)^2$. This yields the joint p.d.f. of the ordered eigenvalues (12) where the normalizing constant $Z(N,K,\mathbf{I}_K)$ is denoted by $Z^0_{K,N}$ for simplicity.

The celebrated result from Marčenko and Pastur [34] states that the limit as $N, K \to \infty$ of the c.d.f. $F_N(x) = \frac{\#\{i,\lambda_i \leq x\}}{K}$ associated to the empirical distribution of the eigenvalues (λ_i) of

$\hat{\mathbf{R}}$ is equal to $\mathbb{P}_{\check{\mathrm{MP}}}\left((-\infty, x]\right)$ where $\mathbb{P}_{\check{\mathrm{MP}}}$ represents the Marčenko-Pastur distribution:

$$\mathbb{P}_{\check{\mathrm{MP}}}(dy) = \mathbf{1}_{(\lambda^-,\lambda^+)}(y)\frac{\sqrt{(\lambda^+ - y)(y - \lambda^-)}}{2\pi c y}\, dy, \tag{14}$$

with $\lambda^+ = (1+\sqrt{c})^2$ and $\lambda^- = (1-\sqrt{c})^2$. This convergence is very fast in the sense that the probability of deviating from $\mathbb{P}_{\check{\mathrm{MP}}}$ decreases as $e^{-N^2 \times \text{const.}}$. More precisely, a simple application of the large deviations results in [35] yields that for any distance d on the set of probability measures on $\mathbb{R}$ compatible with the weak convergence and for any $\delta > 0$,

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}_0\left(d(F_N, \mathbb{P}_{\check{\mathrm{MP}}}) > \delta\right) = -\infty\ . \tag{15}$$

Moreover, the largest eigenvalue λ_1 of $\hat{\mathbf{R}}$ converges a.s. to the right edge of the Marčenko-Pastur distribution, that is $(1+\sqrt{c})^2$. A further result due to Johnstone [20] describes its speed of convergence ($N^{-2/3}$) and its fluctuations (see also [36] for complementary results). Let Λ_1 be defined by:

$$\Lambda_1 = N^{2/3}\left(\frac{\lambda_1 - (1+\sqrt{c_N})^2}{b_N}\right)\ , \tag{16}$$

where b_N is defined by

$$b_N := (1+\sqrt{c_N})\left(\frac{1}{\sqrt{c_N}} + 1\right)^{1/3}\ , \tag{17}$$

then Λ_1 converges in distribution toward a standard Tracy-Widom random variable with c.d.f. F_{TW} defined by:

$$F_{TW}(x) = \exp\left(-\int_x^\infty (u-x)q^2(u)\, du\right) \quad \forall x \in \mathbb{R}\ , \tag{18}$$

where q solves the Painlevé II differential equation:

$$q''(x) = xq(x) + 2q^3(x), \quad q(x) \sim \mathrm{Ai}(x) \quad \text{as} \quad x \to \infty$$

and where $\mathrm{Ai}(x)$ denotes the Airy function. In particular, F_{TW} is continuous. The Tracy-Widom distribution was first introduced in [37], [38] as the asymptotic distribution of the centered and rescaled largest eigenvalue of a matrix from the Gaussian Unitary Ensemble.

Tables of the Tracy-Widom law are available for instance in [39], while a practical algorithm allowing to efficiently evaluate equation (18) can be found in [40].

Remark 4. *In the case where the entries of matrix $\mathbf{Y}$ are real Gaussian random variables, the fluctuations of the largest eigenvalue are still described by a Tracy-Widom distribution whose definition slightly differs from the one given in the complex case (for details, see [20]).*

B. Behaviour under hypothesis H_1

In this case, the covariance matrix writes $\mathbf{\Sigma} = \sigma^2\mathbf{I}_K + \mathbf{h}\mathbf{h}^*$ and matrix $\hat{\mathbf{R}}$ follows a *single spiked* model. Since the behaviour of T_N is not affected if the entries of $\mathbf{Y}$ are multiplied by a given constant, we find it convenient to consider the model where $\mathbf{\Sigma} = \mathbf{I}_K + \frac{\mathbf{h}\mathbf{h}^*}{\sigma^2}$. Denote by

$$\rho_K = \frac{\|\mathbf{h}\|^2}{\sigma^2}$$

the *signal-to-noise* ratio (SNR), then matrix $\mathbf{\Sigma}$ admits the decomposition $\mathbf{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ where $\mathbf{U}$ is a unitary matrix and $\mathbf{D} = \operatorname{diag}(\rho_K, 1, \ldots, 1)$. With the same change of variables from the entries of the matrix to the eigenvalues/angles with Jacobian $\prod_{1\leq i<j\leq K}(x_j - x_i)^2$, the p.d.f. of the ordered eigenvalues writes:

$$p_K^{1,N}(x_{1:K}) = \frac{\mathbf{1}_{(x_1\geq\cdots\geq x_K\geq 0)}}{Z_{K,N}^1} \prod_{1\leq i<j\leq K}(x_j - x_i)^2 \prod_{j=1}^{K} x_j^{N-K} e^{-Nx_j} I_K\left(\frac{N}{K}\mathbf{B}_K, \mathbf{X}_K\right) \tag{19}$$

where the normalizing constant $Z(N, K, \mathbf{I}_K + \mathbf{h}\mathbf{h}^*)$ is denoted by $Z_{K,N}^1$ for simplicity, $\mathbf{X}_K$ is the diagonal matrix with eigenvalues $(x_1, \ldots, x_K)$, $\mathbf{B}_K$ is the $K \times K$ diagonal matrix with eigenvalues $(\frac{\rho_K}{1+\rho_K}, 0, \ldots, 0)$, and for any real diagonal matrices $\mathbf{C}_K, \mathbf{D}_K$, the spherical integral $I_K(\mathbf{C}_K, \mathbf{D}_K)$ is defined as

$$I_K(\mathbf{C}_K, \mathbf{D}_K) = \int e^{K\operatorname{tr}(\mathbf{C}_K\mathbf{Q}\mathbf{D}_K\mathbf{Q}^H)} dm_K(\mathbf{Q}), \tag{20}$$

with m_K the Haar measure on the unitary group of size K (see [30, Chapter 3] for details).

Whereas this rank-one perturbation does not affect the asymptotic behaviour of F_N (the convergence toward $\mathbb{P}_{\check{\text{MP}}}$ and the deviations of the empirical measure given by (15) still hold under $\mathbb{P}_1$), the limiting behaviour of the largest eigenvalue λ_1 can change if the signal-to-noise ratio ρ_K is large enough.

Assumption 1. *The following constant $\rho \in \mathbb{R}$ exists:*

$$\rho = \lim_{K\to\infty} \frac{\|\boldsymbol{h}\|^2}{\sigma^2} \left(= \lim_{K\to\infty} \rho_K\right). \tag{21}$$

We refer to ρ as the limiting SNR. We also introduce

$$\lambda_{\text{spk}}^{\infty} = (1+\rho)\left(1 + \frac{c}{\rho}\right).$$

Under hypothesis H_1, the largest eigenvalue has the following asymptotic behaviour as N, K go to infinity:

$$\lambda_1 \xrightarrow[H_1]{a.s.} \begin{cases} \lambda_{\text{spk}}^{\infty} & \text{if } \rho > \sqrt{c}\,, \\ \lambda^{+} & \text{otherwise,} \end{cases} \tag{22}$$

see for instance [41] for a proof of this result. Note in particular that $\lambda_{\text{spk}}^{\infty}$ is strictly larger than the right edge of the support λ^+ whenever $\rho > \sqrt{c}$. Otherwise stated, if the perturbation is large enough, the largest eigenvalue converges outside the support of Marčenko-Pastur distribution.

C. Limiting behaviour of T_N under H_0 and H_1

Gathering the results recalled in Sections III-A and III-B, we obtain the following:

Proposition 2. *Let Assumption 1 hold true and assume that $\rho > \sqrt{c}$, then:*

$$T_N \xrightarrow[H_0]{a.s.} (1+\sqrt{c})^2 \quad \textit{and} \quad T_N \xrightarrow[H_1]{a.s.} (1+\rho)\left(1+\frac{c}{\rho}\right) \quad \textit{as } N, K \to \infty.$$

IV. Asymptotic threshold and p-values

A. Computation of the asymptotic threshold and p-value

In Theorem 1 below, we take advantage of the convergence results of the largest eigenvalue of $\hat{\mathbf{R}}$ under H_0 in the asymptotic regime $N, K \to \infty$ to express the threshold and the p-value of interest in terms of Tracy-Widom quantiles. Recall that $\bar{F}_{TW} = 1 - F_{TW}$, that $c_N = \frac{K}{N}$, and that b_N is given by (17).

Theorem 1. *Consider a fixed level $\alpha \in (0,1)$ and let γ_N be the threshold for which the power of test (7) is maximum,* i.e. *$p_N(\gamma_N) = \alpha$ where p_N is defined by* (11)*. Then:*

1) *The following convergence holds true:*

$$\zeta_N \triangleq \frac{N^{2/3}}{b_N}\left(\gamma_N - (1+\sqrt{c_N})^2\right) \xrightarrow[N,K\to\infty]{} \bar{F}_{TW}^{-1}(\alpha)\ .$$

2) *The PFA of the following test*

$$T_N \underset{H_0}{\overset{H_1}{\gtrless}} (1+\sqrt{c_N})^2 + \frac{b_N}{N^{2/3}}\,\bar{F}_{TW}^{-1}(\alpha) \tag{23}$$

converges to α.

3) *The p-value $p_N(T_N)$ associated with the GLRT can be approximated by:*

$$\tilde{p}_N(T_N) = \bar{F}_{TW}\left(\frac{N^{2/3}(T_N - (1+\sqrt{c_N})^2)}{b_N}\right) \tag{24}$$

in the sense that $p_N(T_N) - \tilde{p}_N(T_N) \to 0$.

Remark 5. *Theorem 1 provides a simple approach to compute both the threshold and the p-values of the GLRT as the dimension K of the observed time series and the number N of snapshots are large: The threshold γ_N associated with the level α can be approximated by the righthand side of (23). Similarly, equation (24) provides a convenient approximation for the p-value associated with one experiment. These approaches do not require the tedious computation of the exact complementary c.d.f. (11) and, instead, only rely on tables of the c.d.f. F_{TW}, which can be found for instance in [39] along with more details on the computational aspects (note that function F_{TW} does not depend on any of the problem's characteristic, and in particular not on c). This is of importance in real-time applications, such as cognitive radio for instance, where the users connected to the network must quickly decide for the presence/absence of a source.*

Proof of Theorem 1: Before proving the three points of the theorem, we first describe the fluctuations of T_N under H_0 with the help of the results in Section III-A. Assume without loss of generality that $\sigma^2 = 1$, recall that $T_N = \frac{\lambda_1}{K^{-1}\mathrm{tr}\hat{\mathbf{R}}}$ and denote by:

$$\tilde{T}_N = \frac{N^{2/3}(T_N - (1+\sqrt{c_N})^2)}{b_N} \tag{25}$$

the rescaled and centered version of the statistics T_N. A direct application of Slutsky's lemma (see for instance [42]) together with the fluctuations of λ_1 as reminded in Section III-A yields that $\tilde{T}_N$ converges in distribution to a standard Tracy-Widom random variable with c.d.f. F_{TW} which is continuous over $\mathbb{R}$. Denote by F_N the c.d.f. of $\tilde{T}_N$ under H_0, then a classical result, sometimes called Polya's theorem (see for instance [43]), asserts that the convergence of F_N towards F_{TW} is uniform over $\mathbb{R}$:

$$\sup_{x\in\mathbb{R}} |F_N(x) - F_{TW}(x)| \xrightarrow[N,K\to\infty]{} 0 \ . \tag{26}$$

We are now in position to prove the theorem.

The mere definition of ζ_N implies that $\alpha = p_N(\gamma_N) = \bar{F}_N(\zeta_N)$. Due to (26), $\bar{F}_{TW}(\zeta_N) \to \alpha$. As F_{TW} has a continuous inverse, the first point of the theorem is proved.

The second point is a direct consequence of the convergence of F_N toward the Tracy-Widom distribution: The PFA of test (23) can be written as: $\mathbb{P}_0\left(\tilde{T}_N > \bar{F}_{TW}^{-1}(\alpha)\right)$ which readily converges to α.

The third point is a direct consequence of (26): $p_N(T_N) - \tilde{p}_N(T_N) = \bar{F}_N(\tilde{T}_N) - \bar{F}_{TW}(\tilde{T}_N) \to 0$. This completes the proof of Theorem 1.

■

V. Asymptotic analysis of the power of the test

In this section, we provide an asymptotic analysis of the power of the GLRT as $N, K \to \infty$. As the power of the test goes exponentially to zero, its error exponent is computed with the help of the large deviations associated to the largest eigenvalue of matrix $\hat{\mathbf{R}}$. The error exponent and error exponent curve are computed in Theorem 2, Section V-A; the large deviations of interest are stated in Section V-B. Finally Theorem 2 is proved in Section V-C.

A. Error exponents and error exponent curve

The most natural approach to characterize the performance of a test is to evaluate its power or equivalently its miss probability *i.e.*, the probability under H_1 that the receiver decides hypothesis H_0. For a given level $\alpha \in (0,1)$, the miss probability writes:

$$\beta_{N,T}(\alpha) = \inf_{\gamma} \left\{ \mathbb{P}_1\left(T_N < \gamma\right),\ \gamma \text{ such that } \mathbb{P}_0\left(T_N > \gamma\right) \leq \alpha \right\} \ . \tag{27}$$

Based on Section II-B, the infimum is achieved when the threshold coincides with $\gamma = p_N^{-1}(\alpha)$; otherwise stated, $\beta_{N,T}(\alpha) = \mathbb{P}_1\left(T_N < p_N^{-1}(\alpha)\right)$ (notice that the miss probability depends on the unknown parameters $\boldsymbol{h}$ and σ^2). As $\beta_{N,T}(\alpha)$ has no simple expression in the general case, we again study its asymptotic behaviour in the asymptotic regime of interest (13). It follows from Theorem 1 that $p_N^{-1}(\alpha) \to \lambda^+ = (1+\sqrt{c})^2$ for $\alpha \in (0,1)$. On the other hand, under hypothesis H_1, T_N converges a.s. to $\lambda_{\text{spk}}^{\infty}$ which is strictly greater than λ^+ when the ratio $\frac{\|\mathbf{h}\|^2}{\sigma^2}$ is large enough. In this case, $\mathbb{P}_1\left(T_N < p_N^{-1}(\alpha)\right)$ goes to zero as it expresses the probability that T_N deviates from its limit $\lambda_{\text{spk}}^{\infty}$; moreover, one can prove that the convergence to zero is exponential in N:

$$\mathbb{P}_1\left(T_N < x\right) \propto e^{-N I_\rho^+(x)} \qquad \text{for} \quad x \leq \lambda_{\text{spk}}^{\infty} \ , \tag{28}$$

where I_ρ^+ is the so-called rate function associated to T_N. This observation naturally yields the following definition of the error exponent $\mathcal{E}_T$:

$$\mathcal{E}_T \;=\; \lim_{N,K\to\infty} -\frac{1}{N}\log\beta_{N,T}(\alpha) \tag{29}$$

the existence of which is established in Theorem 2 below (as $N, K \to \infty$). Also proved is the fact that $\mathcal{E}_T$ does not depend on α.

The error exponent $\mathcal{E}_T$ gives crucial information on the performance of the test T_N, provided that the level α is kept fixed when N, K go to infinity. Its existence strongly relies on the study of the large deviations associated to the statistics T_N.

In practice however, one may as well take benefit from the increasing number of data not only to decrease the miss probability, but to decrease the PFA as well. As a consequence, it is of practical interest to analyze the detection performance when both the miss probability and the PFA go to zero at exponential speed. A couple $(a, b) \in (0, \infty) \times (0, \infty)$ is said to be an *achievable* pair of error exponents for the test T_N if there exists a sequence of levels α_N such that, in the asymptotic regime (13),

$$\lim_{N,K\to\infty} -\frac{1}{N}\log\alpha_N = a \quad \text{and} \quad \lim_{N,K\to\infty} -\frac{1}{N}\log\beta_{N,T}(\alpha_N) = b\ . \tag{30}$$

We denote by $\mathcal{S}_T$ the set of achievable pairs of error exponents for test T_N as $N, K \to \infty$. We refer to $\mathcal{S}_T$ as the *error exponent curve* of T_N.

The following notations are needed in order to describe the error exponent $\mathcal{E}_T$ and error exponent curve $\mathcal{S}_T$.

$$\left\{\begin{array}{lcll} \mathbf{f}(x) & = & \int \frac{1}{y-x}\mathbb{P}_{\check{\mathrm{MP}}}(dy) & \text{for } x \in \mathbb{R}\setminus(\lambda^-,\lambda^+) \\ \mathbf{F}^+(x) & = & \int \log(x-y)\mathbb{P}_{\check{\mathrm{MP}}}(dy) & \text{for } x \geq \lambda^+ \end{array}\right. . \tag{31}$$

Remark 6. *Function* $\mathbf{f}$ *is the well-known Stieltjes transform associated to Marčenko-Pastur distribution and admits a closed-form representation formula. So does function* $\mathbf{F}^+$*, although this fact is perhaps less known. These results are gathered in Appendix C.*

Denote by $\Delta(\,\cdot \mid A)$ the convex indicator function *i.e.* the function equal to zero for $x \in A$ and to infinity otherwise. For $\rho > \sqrt{c}$, define the function:

$$I_\rho^+(x) \;=\; \frac{x-\lambda_{\mathrm{spk}}^\infty}{(1+\rho)} - (1-c)\log\left(\frac{x}{\lambda_{\mathrm{spk}}^\infty}\right) - c\left(\mathbf{F}^+(x) - \mathbf{F}^+(\lambda_{\mathrm{spk}}^\infty)\right) + \Delta(x \mid [\lambda^+,\infty)) \tag{32}$$

Also define the function:

$$I_0^+(x) = x - \lambda^+ - (1-c)\log\left(\frac{x}{\lambda^+}\right) - 2c\left(\mathbf{F}^+(x) - \mathbf{F}^+(\lambda^+)\right) + \Delta(x \mid [\lambda^+, \infty)) \ . \tag{33}$$

We are now in position to state the main theorem of the section:

Theorem 2. *Let Assumption 1 hold true, then:*

1) *For any fixed level $\alpha \in (0,1)$, the limit $\mathcal{E}_T$ in (29) exists as $N, K \to \infty$ and satisfies:*

$$\mathcal{E}_T = I_\rho^+(\lambda^+) \tag{34}$$

if $\rho > \sqrt{c}$ and $\mathcal{E}_T = 0$ otherwise.

2) *The error exponent curve of test T_N is given by:*

$$\mathcal{S}_T \ = \ \left\{(I_0^+(x), I_\rho^+(x)) \ : \ x \in (\lambda^+, \lambda_{\rm spk}^\infty)\right\} \tag{35}$$

if $\rho > \sqrt{c}$ and $\mathcal{S}_T = \emptyset$ otherwise.

The proof of Theorem 2 heavily relies on the large deviations of T_N and is postponed to Section V-C. Before providing the proof, it is worth making the following remarks.

Remark 7. *Several variants of the GLRT have been proposed in the literature, and typically consist in replacing the denominator $\frac{1}{K}\mathrm{tr}\,\hat{\mathbf{R}}$ (which converges toward σ^2) by a more involved estimate of σ^2 in order to decrease the bias [12], [13], [14], [15]. However, it can be established that the error exponents of the above variants are as well given by (34) and (35) in the asymptotic regime.*

Remark 8. *The error exponent $\mathcal{E}_T$ yields a simple approximation of the miss probability in the sense that $\beta_{N,T}(\alpha) \simeq e^{-N\,\mathcal{E}_T}$ as $N \to \infty$. It depends on the limiting ratio c and on the value of the SNR ρ through the constant $\lambda_{\rm spk}^\infty$. In the high SNR case, the error exponent turns out to have a simple expression as a function of ρ. If $\rho \to \infty$ then $\lambda_{\rm spk}^\infty$ tends to infinity as well, which simplifies the expression of rate function I_ρ^+. Using $\mathbf{F}^+(\lambda_{\rm spk}^\infty) = \log \lambda_{\rm spk}^\infty + o_\rho(1)$ where $o_\rho(1)$ stands for a term which converges to zero as $\rho \to \infty$, it is straightforward to show that for each $x \geq \lambda^+$, $I_\rho^+(x) = \log\rho - 1 - (1-c)\log x - c\mathbf{F}^+(x) + o_\rho(1)$. After some algebra, we finally obtain:*

$$\mathcal{E}_T = \log\rho - (1+\sqrt{c}) - (1-c)\log(1+\sqrt{c}) - c\log\sqrt{c} + o_\rho(1) \ .$$

At high SNR, this yields the following convenient approximation of the miss probability:

$$\beta_{N,T}(\alpha) \simeq \left(\psi(c)\,\rho\right)^N \ , \tag{36}$$

where $\psi(c) = e^{-(1+\sqrt{c})}(1+\sqrt{c})^{c-1}c^{-\frac{c}{2}}$.

B. Large Deviations associated to T_N

In order to express the error exponents of interest, a rigorous formalization of (28) is needed. Let us recall the definition of a Large Deviation Principle: A sequence of random variables $(X_N)_{N\in\mathbb{N}}$ satisfies a Large Deviation Principle (LDP) under $\mathbb{P}$ in the scale N with good rate function I if the following properties hold true:

- I is a nonnegative function with compact level sets, i.e. $\{x, I(x) \leq t\}$ is compact for $t \in \mathbb{R}$,
- for any closed set $F \subset \mathbb{R}$, the following upper bound holds true:

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(X_N \in F) \leq -\inf_F I \ . \tag{37}$$

- for any open set $G \subset \mathbb{R}$, the following lower bound holds true:

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(X_N \in G) \geq -\inf_G I \ . \tag{38}$$

For instance, if A is a set such that $\inf_{\text{int}(A)} I = \inf_{\text{cl}(A)} I (= \inf_A I)$, (where int$(A)$ and cl(A) respectively denote the interior and the closure of A), then (37) and (38) yield

$$\lim_{N\to\infty} N^{-1} \log \mathbb{P}(X_N \in A) = -\inf_A I \ . \tag{39}$$

Informally stated,

$$\mathbb{P}(X_N \in A) \quad \propto \quad e^{-N \inf_A I} \qquad \text{as } N \to \infty \ .$$

If, moreover $\inf_A I > 0$ (which typically happens if the limit of X_N -if existing- does not belong to A), then probability $\mathbb{P}(X_N \in A)$ goes to zero exponentially fast, hence a *large deviation* (LD); and the event $\{X_N \in A\}$ can be referred to as a *rare* event. We refer the reader to [44] for further details on the subject.

As already mentioned above, all the probabilities of interest are rare events as N, K go to infinity related to large deviations for T_N. More precisely, Theorem 2 is merely a consequence of the following Lemma.

Lemma 1. *Let Assumption 1 hold true and let* $N, K \to \infty$*, then:*

1) *Under H_0, T_N satisfies the LDP in the scale N with good rate function I_0^+, which is increasing from 0 to ∞ on interval $[\lambda^+, \infty)$.*
2) *Under H_1 and if $\rho > \sqrt{c}$, T_N satisfies the LDP in the scale N with good rate function I_ρ^+. Function I_ρ^+ is decreasing from $I_\rho^+(\lambda^+)$ to 0 on $[\lambda^+, \lambda_{\rm spk}^\infty]$ and increasing from 0 to ∞ on $[\lambda_{\rm spk}^\infty, \infty)$.*
3) *For any bounded sequence $(\eta_N)_{N\geq 0}$,*

$$\lim_{N,K\to\infty} -\frac{1}{N}\log \mathbb{P}_1\left(T_N < (1+\sqrt{c_N})^2 + \frac{\eta_N}{N^{2/3}}\right) = \begin{cases} I_\rho^+(\lambda^+) & \text{if } \rho > \sqrt{c} \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

4) *Let $x \in (\lambda^+, \infty)$ and let $(x_N)_{N\geq 0}$ be any real sequence which converges to x. If $\rho \leq \sqrt{c}$, then:*

$$\lim_{N,K\to\infty} -\frac{1}{N}\log \mathbb{P}_1\left(T_N < x_N\right) = 0 \ . \tag{41}$$

The proof of Lemma 1 is provided in Appendix A.

Remark 9. 1) *The proof of the large deviations for T_N relies on the fact that the denominator $K^{-1}\mathrm{tr}\,\hat{\mathbf{R}}$ of T_N concentrates much faster than λ_1. Therefore, the large deviations of T_N are driven by those of λ_1, a fact that is exploited in the proof.*
2) *In Appendix A, we rather focus on the large deviations of λ_1 under H_1 and skip the proof of Lemma 1-(1), which is simpler and available (to some extent) in [29, Theorem 2.6.6][5]. Indeed, the proof of the LDP relies on the joint density of the eigenvalues. Under H_1, this joint density has an extra-term, the spherical integral, and is thus harder to analyze.*
3) *Lemma 1-(3) is not a mere consequence of Lemma 1-(2) as it describes the deviations of T_N at the vicinity of a point of discontinuity of the rate function. The direct application of the LDP would provide a trivial lower bound ($-\infty$) in this case.*
4) *In the case where the entries of matrix $\mathbf{Y}$ are real Gaussian random variables, the results stated in Lemma 1 will still hold true with minor modifications: The rate functions will be slightly different. Indeed, the computation of the rate functions relies on the joint density of the eigenvalues, which differs whether the entries of $\mathbf{Y}$ are real or complex.*

[5]see also the errata sheet for the sign error in the rate function on the authors webpage.

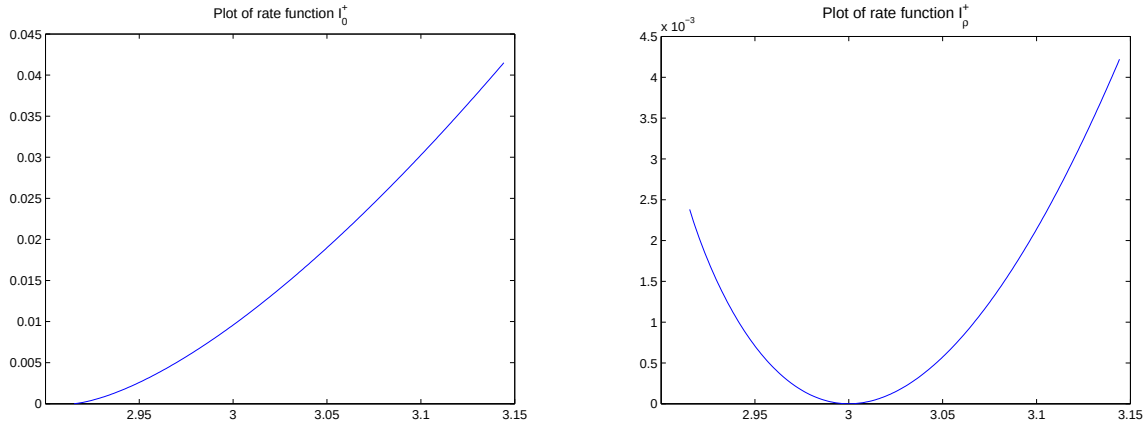

Figure 1. Plots of rate functions I_0^+ and I_ρ^+ in the case where $c = 0.5$ and $\rho = 1\,db$. In this case, $\lambda^+ = 2.9142$, $\lambda_{\rm spk}^\infty = 3$, $I_0^+(\lambda^+) = 0$ and $I_\rho^+(\lambda_{\rm spk}^\infty) = 0$.

C. Proof of Theorem 2

In order to prove (34), we must study the asymptotic behaviour of the miss probability $\beta_{N,T}(\alpha) = \mathbb{P}_1\left(T_N < p_N^{-1}(\alpha)\right)$ as $N, K \to \infty$. Using Theorem 1-(1), we recall that

$$\beta_{N,T}(\alpha) = \mathbb{P}_1\left(T_N < (1+\sqrt{c_N})^2 + \frac{\eta_N}{N^{2/3}}\right) \tag{42}$$

where $c_N = \frac{K}{N}$ converges to c and where η_N is a deterministic sequence such that

$$\lim_{N,K\to\infty} \eta_N = (1+\sqrt{c})\left(\frac{1}{\sqrt{c}}+1\right)^{1/3} \bar{F}_{TW}^{-1}(\alpha)\,.$$

Hence, Lemma 1-(3) yields the first point of Theorem 2. We now prove the second point. Assume that $\rho > \sqrt{c}$. Consider any $x \in (\lambda^+, \lambda_{\rm spk}^\infty)$ and for every N, K, consider the test function which rejects the null hypothesis when $T_N > x$,

$$T_N \underset{H_0}{\overset{H_1}{\gtrless}} x\,. \tag{43}$$

Denote by $\alpha_N = \mathbb{P}_0(T_N > x)$ the PFA associated with this test. By Lemma 1-(1) together with the continuity of the rate function at x, we obtain:

$$\lim_{N,K\to\infty} -\frac{1}{N}\log\alpha_N = \inf_{y\in[x,\infty)} I_0^+(y) = I_0^+(x)\,. \tag{44}$$

The miss probability of this test is given by $\beta_{N,T}(\alpha_N) = \mathbb{P}_1(T_N < x)$. By Lemma 1-(2),

$$\lim_{N,K\to\infty} -\frac{1}{N}\log\beta_{N,T}(\alpha_N) = \inf_{y\in(-\infty,x]} I_\rho^+(y) = I_\rho^+(x)\ . \tag{45}$$

Equations (44) and (45) prove that $(I_0^+(x), I_\rho^+(x))$ is an achievable pair of error exponents. Therefore, the set in the righthand side of (35) is included in $\mathcal{S}_T$. We now prove the converse. Assume that (a,b) is an achievable pair of error exponents and let α_N be a sequence such that (30) holds. Denote by $\gamma_N = p_N^{-1}(\alpha_N)$ the threshold associated with level α_N. As $I_0^+(x)$ is continuous and increasing from 0 to ∞ on interval (λ^+,∞), there exists a (unique) $x \in (\lambda^+,\infty)$ such that $a = I_0^+(x)$. We now prove that γ_N converges to x as N tends to infinity. Consider a subsequence $\gamma_{\varphi(N)}$ which converges to a limit $\gamma \in \mathbb{R}\cup\{\infty\}$. Assume that $\gamma > x$. Then there exists $\epsilon > 0$ such that $\gamma_{\varphi(N)} > x+\epsilon$ for large N. This yields:

$$-\frac{1}{\varphi(N)}\log\mathbb{P}_0\left(T_{\varphi(N)} > \gamma_{\varphi(N)}\right) \geq -\frac{1}{\varphi(N)}\log\mathbb{P}_0\left(T_{\varphi(N)} > x+\epsilon\right)\ . \tag{46}$$

Taking the limit in both terms yields $I_0^+(x) \geq I_0^+(x+\epsilon)$ by Lemma 1, which contradicts the fact that I_0^+ is an increasing function. Now assume that $\gamma < x$. Similarly,

$$-\frac{1}{\varphi(N)}\log\mathbb{P}_0\left(T_{\varphi(N)} > \gamma_{\varphi(N)}\right) \leq -\frac{1}{\varphi(N)}\log\mathbb{P}_0\left(T_{\varphi(N)} > x-\epsilon\right) \tag{47}$$

for a certain ϵ and for N large enough. Taking the limit of both terms, we obtain $I_0^+(x) \leq I_0^+(x-\epsilon)$ which leads to the same contradiction. This proves that $\lim_N \gamma_N = x$. Recall that by definition (30),

$$b = \lim_{N,K\to\infty} -\frac{1}{N}\log\mathbb{P}_1\left(T_N < \gamma_N\right)\ .$$

As γ_N tends to x, Lemma 1 implies that the righthand side of the above equation is equal to $I_\rho^+(x) > 0$ if $x \in (\lambda^+, \lambda_{\text{spk}}^\infty)$ and $\rho > \sqrt{c}$. It is equal to 0 if $x \geq \lambda_{\text{spk}}^\infty$ or $\rho \leq \sqrt{c}$. Now $b > 0$ by definition, therefore both conditions $x \in (\lambda^+, \lambda_{\text{spk}}^\infty)$ and $\rho > \sqrt{c}$ hold. As a conclusion, if (a,b) is an achievable pair of error exponents, then $(a,b) = (I_0^+(x), I_\rho^+(x))$ for a certain $x \in (\lambda^+, \lambda_{\text{spk}}^\infty)$, and furthermore $\rho > \sqrt{c}$. This completes the proof of the second point of Theorem 2.

VI. Comparison with the test based on the condition number

This section is devoted to the study of the asymptotic performances of the test $U_N = \frac{\lambda_1}{\lambda_K}$, which is popular in cognitive radio [22], [23], [24]. The main result of the section is Theorem 3, where it is proved that the test based on T_N asymptotically outperforms the one based on U_N in terms of error exponent curves.

A. Description of the test

A different approach which has been introduced in several papers devoted to cognitive radio contexts consists in rejecting the null hypothesis for large values of the statistics U_N defined by:

$$U_N = \frac{\lambda_1}{\lambda_K}\,, \tag{48}$$

which is the ratio between the largest and the smallest eigenvalues of $\hat{\mathbf{R}}$. Random variable U_N is the so-called *condition number* of the sampled covariance matrix $\hat{\mathbf{R}}$. As for T_N, an important feature of the statistics U_N is that its law does not depend of the unknown parameter σ which is the level of the noise. Under hypothesis H_0, recall that the spectral measure of $\hat{\mathbf{R}}$ weakly converges to the Marčenko-Pastur distribution (14) with support (λ^-, λ^+). In addition to the fact that λ_1 converges toward λ^+ under H_0 and $\lambda_{\rm spk}^\infty$ under H_1, the following result related to the convergence of the lowest eigenvalue is of importance (see for instance [45], [46], [41]):

$$\lambda_K \xrightarrow{a.s.} \lambda^- = \sigma^2(1-\sqrt{c})^2 \tag{49}$$

under both hypotheses H_0 and H_1. Therefore, the statistics U_N admits the following limits:

$$U_N \xrightarrow[H_0]{a.s.} \frac{\lambda^+}{\lambda^-} = \frac{(1+\sqrt{c})^2}{(1-\sqrt{c})^2}, \quad \text{and} \quad U_N \xrightarrow[H_1]{a.s.} \frac{\lambda_{\rm spk}^\infty}{\lambda^-} \quad \text{for } \rho > \sqrt{c}\ . \tag{50}$$

The test is based on the observation that the limit of U_N under the alternative H_1 is strictly larger than the ratio λ^+/λ^-, at least when the SNR ρ is large enough.

B. A few remarks related to the determination of the threshold for the test U_N

The determination of the threshold for the test U_N relies on the asymptotic independence of λ_1 and λ_K under H_0. As we shall prove below that test U_N is asymptotically outperformed by test T_N, such a study, rather involved, seems beyond the scope of this article. For the sake of completeness however, we describe unformally how to set the threshold for U_N. Recall the definition of Λ_1 in (16) and let Λ_K be defined as:

$$\Lambda_K = N^{2/3}\frac{\left(\lambda_K - (1-\sqrt{c_N})^2\right)}{\left(\sqrt{c_N}-1\right)\left(c_N^{-1/2}-1\right)^{1/3}}\ .$$

Then both Λ_1 and Λ_K converge toward Tracy-Widom random variables. Moreover,

$$(\Lambda_1, \Lambda_K) \xrightarrow[N,K\to\infty]{} (X, Y)\ ,$$

where X and Y are independent random variables, both distributed according to F_{TW}[6].

As a corollary of the previous convergence, a direct application of the Delta method [27, Chapter 3] yields the following convergence in distribution:

$$N^{2/3}\left(\frac{\lambda_1}{\lambda_K} - \frac{(1+\sqrt{c_N})^2}{(1-\sqrt{c_N})^2}\right) \to (aX + bY) \ ,$$

where

$$a = \frac{(1+\sqrt{c})}{(1-\sqrt{c})^2}\left(\frac{1}{\sqrt{c}}+1\right)^{1/3} \quad \text{and} \quad b = \frac{(1+\sqrt{c})^2}{(\sqrt{c}-1)^3}\left(\frac{1}{\sqrt{c}}-1\right)^{1/3} \ ,$$

which enables one to set the threshold of the test, based on the quantiles of the random variable $aX + bY$. In particular, following the same arguments as in Theorem 1-1), one can prove that the optimal threshold (for some fixed $\alpha \in (0,1)$), defined by $\mathbb{P}_0(U_N > \gamma_N) = \alpha$, satisfies

$$\xi_N \ \triangleq \ N^{2/3}\left(\gamma_N - \frac{(1+\sqrt{c_N})^2}{(1-\sqrt{c_N})^2}\right) \xrightarrow[N,K\to\infty]{} \bar{F}^{-1}_{aX+bY}(\alpha) \ .$$

In particular, ξ_N is bounded as $N, K \to \infty$.

C. Performance analysis and comparison with the GLRT

We now provide the performance analysis of the above test based on the condition number U_N in terms of error exponents. In accordance with the definitions of section V-A, we define the miss probability associated with test U_N as $\beta_{N,U}(\alpha) = \inf_\gamma \mathbb{P}_1\left(U_N < \gamma\right)$ for any level $\alpha \in (0,1)$, where the infimum is taken w.r.t. all thresholds γ such that $\mathbb{P}_0\left(U_N > \gamma\right) \leq \alpha$. We denote by $\mathcal{E}_U$ the limit of sequence $-\frac{1}{N}\log\beta_{N,U}(\alpha)$ (if it exists) in the asymptotic regime (13). We denote by $\mathcal{S}_U$ the error exponent curve associated with test U_N *i.e.*, the set of couples (a,b) of positive numbers for which $-\frac{1}{N}\log\beta_{N,U}(\alpha_N) \to b$ for a certain sequence α_N which satisfies $-\frac{1}{N}\log\alpha_N \to a$.

Theorem 3 below provides the error exponents associated with test U_N. As for T_N, the performance of the test is expressed in terms of the rate function of the LDPs for U_N under $\mathbb{P}_0$ or $\mathbb{P}_1$. These rate functions combine the rate functions for the largest eigenvalue λ_1, i.e. I^+_ρ and I^+_0 defined in Section V-B, together with the rate function associated to the smallest eigenvalue, I^-, defined below. As we shall see, the positive rank-one perturbation does not affect λ_K whose rate function remains the same under H_0 and H_1.

[6]Such an asymptotic independence is not formally proved yet for $\hat{\mathbf{R}}$ under H_0, but is likely to be true as a similar result has been established in the case of the Gaussian Unitary Ensemble [47],[40].

We first define:

$$\mathbf{F}^{-}(x) = \int \log(y-x) d\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}(y) \quad \text{for } x \leq \lambda^{-} \ . \tag{51}$$

As for $\mathbf{F}^{+}$, function $\mathbf{F}^{-}$ also admits a closed-form expression based on $\mathbf{f}$, the Stieltjes transform of Marčenko-Pastur distribution (see Appendix C for details).

Now, define for each $x \in \mathbb{R}$:

$$I^{-}(x) = x - \lambda^{-} - (1-c)\log\left(\frac{x}{\lambda^{-}}\right) - 2c\left(\mathbf{F}^{-}(x) - \mathbf{F}^{-}(\lambda^{-})\right) + \Delta(x|(0,\lambda^{-}])^{\cdot} \tag{52}$$

If λ_1 and λ_K were independent random variables, the contraction principle (see e.g. [44]) would imply that the following functions

$$\Gamma_\rho(t) = \inf_{(x,y)} \left\{ I_\rho^{+}(x) + I^{-}(y) \ : \quad \frac{x}{y} = t \right\} \quad \text{and} \quad \Gamma_0(t) = \inf_{(x,y)} \left\{ I_0^{+}(x) + I^{-}(y) \ : \quad \frac{x}{y} = t \right\}$$

defined for each $t \geq 0$, are the rate functions associated with the LDP governing λ_1/λ_K under hypotheses H_1 and H_0 respectively. Of course, λ_1 and λ_K are not independent, and the contraction principle does not apply. However, a careful study of the p.d.f. $p^0_{K,N}$ and $p^1_{K,N}$ shows that λ_1 and λ_K behave as if they were asymptotically independent, from a large deviation perspective:

Lemma 2. *Let Assumption 1 hold true and let $N, K \to \infty$, then:*

1) *Under H_0, U_N satisfies the LDP in the scale N with good rate function Γ_0.*
2) *Under H_1 and if $\rho > \sqrt{c}$, U_N satisfies the LDP in the scale N with good rate function Γ_ρ.*
3) *For any bounded sequence $(\eta_N)_{N\geq 0}$,*

$$\lim_{N,K\to\infty} -\frac{1}{N}\log \mathbb{P}_1\left(U_N < \frac{(1+\sqrt{c_N})^2}{(1-\sqrt{c_N})^2} + \frac{\eta_N}{N^{2/3}}\right) = \begin{cases} \Gamma_\rho(\lambda^{+}) & \text{if } \rho > \sqrt{c} \\ 0 & \text{otherwise.} \end{cases} \tag{53}$$

   Moreover, $\Gamma_\rho(\lambda^{+}) = I_\rho^{+}(\lambda^{+})$.
4) *Let $x \in (\lambda^{+}, \infty)$ and let $(x_N)_{N\geq 0}$ be any real sequence which converges to x. If $\rho \leq \sqrt{c}$, then:*

$$\lim_{N,K\to\infty} -\frac{1}{N}\log \mathbb{P}_1\left(T_N < x_N\right) = 0 \tag{54}$$

Remark 10. *In the context of Lemma 1, both quantities λ_1 and λ_K deviate at the same speed, to the contrary of statistics T_N where the denominator concentrated much faster than the largest eigenvalue λ_1. Nevertheless, proof of Lemma 2 is a slight extension of the proof of Lemma 1,*

based on the study of the joint deviations (λ_1, λ_K), the proof of which can be performed similarly to the proof of the deviations of λ_1. Once the large deviations established for the couple (λ_1, λ_K), it is a matter of routine to get the large deviations for the ratio λ_1/λ_K. A proof is outlined in Appendix B.

We now provide the main result of the section.

Theorem 3. *Let Assumption 1 hold true, then:*

1) *For any fixed level $\alpha \in (0,1)$ and for each ρ, the error exponent $\mathcal{E}_U$ exists and coincides with $\mathcal{E}_T$.*
2) *The error exponent curve of test U_N is given by:*

$$\mathcal{S}_U \;=\; \left\{ (\Gamma_0(t), \Gamma_\rho(t)) \;:\; t \in \left(\frac{\lambda^+}{\lambda^-}, \frac{\lambda^\infty_{\rm spk}}{\lambda^-} \right) \right\} \tag{55}$$

   if $\rho > \sqrt{c}$ and $\mathcal{S}_U = \emptyset$ otherwise.
3) *The error exponent curve $\mathcal{S}_T$ of test T_N uniformly dominates $\mathcal{S}_U$ in the sense that for each $(a,b) \in \mathcal{S}_U$ there exits $b' > b$ such that $(a,b') \in \mathcal{S}_T$.*

Proof: The proof of items (1) and (2) is merely bookkeeping from the proof of Theorem 2 with Lemma 2 at hand.

Let us prove item (3). The key observation lies in the following two facts:

$$\forall x \in (\lambda^+, \lambda^\infty_{\rm spk}), \quad \Gamma_\rho\left(\frac{x}{\lambda^-}\right) \;=\; I^+_\rho(x)\;, \tag{56}$$

$$\forall x \in (\lambda^+, \lambda^\infty_{\rm spk}), \quad \Gamma_0\left(\frac{x}{\lambda^-}\right) \;<\; I^+_0(x)\;. \tag{57}$$

Recall that

$$\begin{aligned} \Gamma_\rho\left(\frac{x}{\lambda^-}\right) \;&=\; \inf_{(u,v)} \left\{ I^+_\rho(u) + I^-(v) \;:\; \frac{u}{v} = \frac{x}{\lambda^-} \right\} \\ &\overset{(a)}{\leq}\; I^+_\rho(x) + I^-(\lambda^-) \;=\; I^+_\rho(x), \end{aligned}$$

where (a) follows from the fact that $I^-(\lambda^-) = 0$ and by taking $u = x, v = \lambda^-$. Assume that inequality (a) is strict. Due to the fact that I^+_ρ is decreasing, the only way to decrease the value of $I^+_\rho(u) + I^-(v)$ under the considered constraint $\frac{u}{v} = \frac{x}{\lambda^-}$ is to find a couple (u,v) with $u > x$, but this cannot happen because this would enforce $v > \lambda^-$ so that the constraint $\frac{u}{v} = \frac{x}{\lambda^-}$ remains

fulfilled, and this would end up with $I^-(v) = \infty$. Necessarily, (a) is an equality and (56) holds true.

Let us now give a sketch of proof for (57). Notice first that $\frac{dI_0^+}{du} \mid_{u=x} > 0$ (which easily follows from the fact that I_0^+ is increasing and differentiable) while $\frac{dI^-}{dv} \mid_{v \nearrow \lambda^-} = 0$. This equality follows from the direct computation:

$$\begin{aligned} \lim_{x \nearrow \lambda^-} \frac{I^-(x)}{x - \lambda^-} &= 1 - \frac{1-c}{\lambda^-} - 2c \left. \frac{d\mathbf{F}^-}{dx} \right|_{x \nearrow \lambda^-} \\ &= 1 - \frac{1+\sqrt{c}}{1-\sqrt{c}} + 2c\mathbf{f}(\lambda^-) \quad = \; 0 \; , \end{aligned}$$

where the last equality follows from the fact that $\frac{d\mathbf{F}^-}{dx} = -\mathbf{f}$ together with the closed-form expression for $\mathbf{f}$ as given in Appendix C. As previously, write:

$$\begin{aligned} \Gamma_0 \left(\frac{x}{\lambda^-} \right) &= \inf_{(u,v)} \left\{ I_0^+(u) + I^-(v) \; : \quad \frac{u}{v} = \frac{x}{\lambda^-} \right\} \\ &\overset{(a)}{\leq} I_0^+(x) + I^-(\lambda^-) \quad = I_0^+(x). \end{aligned}$$

Consider now a small perturbation $u = x - \delta$ and the related perturbation $v = \lambda^- - \delta'$ so that the constraint $\frac{u}{v} = \frac{x}{\lambda^-}$ remains fulfilled. Due to the values of the derivatives of I_0^+ and I^- at respective points x and λ^-, the decrease of $I_0^+(x-\delta)$ will be larger than the increase of $I^-(\lambda^- - \delta')$, and this will result in the fact that

$$\Gamma_0 \left(\frac{x}{\lambda^-} \right) \; \leq \; I_0^+(x-\delta) + I^-(\lambda^- + \delta') \; < \; I_0^+(x) \; ,$$

which is the desired result, which in turn yields (57).

We can now prove Theorem 3-(3). Let $(a,b) \in \mathcal{S}_U$ and $(a,b') \in \mathcal{S}_T$, we shall prove that $b < b'$. Due to the mere definitions of the curves $\mathcal{S}_U$ and $\mathcal{S}_T$, there exist $x \in (\lambda^+, \lambda^\infty_{\rm spk})$ and $t \in (\lambda^+/\lambda^-, \lambda^\infty_{\rm spk}/\lambda^-)$ such that $a = I_0^+(x) = \Gamma_0(t)$. Eq. (57) yields that $\frac{x}{\lambda^-} < t$. As I_ρ^+ is decreasing, we have

$$b' \; = \; I_\rho^+(x) \; > \; I_\rho^+(t\lambda^-) \; = \; \Gamma_\rho(t) \; = \; b \; ,$$

and the proof is completed. ■

Remark 11. *Theorem 3-(1) indicates that when the number of data increases, the powers of tests T_N and U_N both converge to one at the same exponential speed $\mathcal{E}_U = \mathcal{E}_T$, provided that the level α is kept fixed. However, when the level goes to zero exponentially fast as a function of*

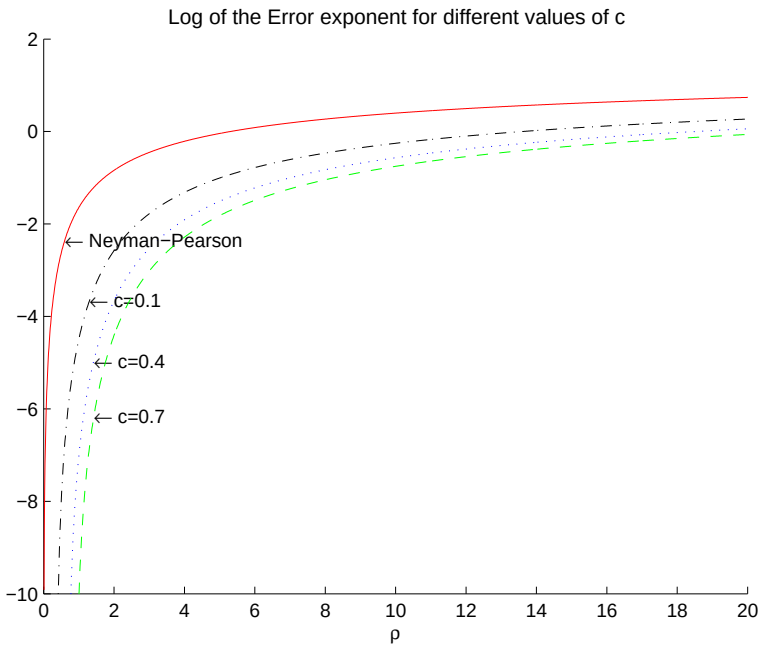

Figure 2. Computation of the logarithm of the error exponent $\mathcal{E}$ associated to the test T_N for different values of c (with $\mathcal{E}_\rho$ defined for $\rho \geq \sqrt{c}$ and $\mathcal{E}_\rho\,|_{\rho=\sqrt{c}} = 0$), and comparison with the optimal result (Neyman-Pearson) obtained in the case where all the parameters are perfectly known.

the number of snapshots, then the test based on T_N outperforms U_N in terms of error exponents: The power of T_N converges to one faster than the power of U_N. Simulation results for N, K fixed sustain this claim (cf. Figure 4). This proves that in the context of interest ($N, K \to \infty$), the GLRT approach should be prefered to the test U_N.

VII. Numerical Results

In the following section, we analyze the performance of the proposed tests in various scenarios.

Figure 2 compares the error exponent of test T_N with the optimal NP test (assuming that all the parameters are known) for various values of c and ρ. The error exponent of the NP test can be easily obtained using Stein's Lemma (see for instance [48]).

In Figure 3, we compare the Error Exponent curves of both tests T_N and U_N. The analytic expressions provided in 2 and 3 for the Error Exponent curves have been used to plot the curves. The asymptotic comparison clearly underlines the gain of using test T_N.

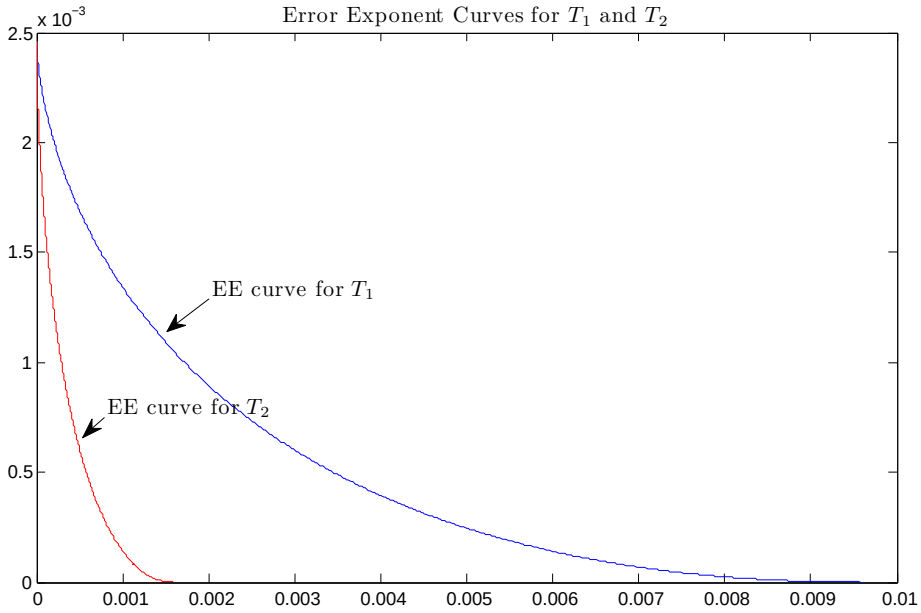

Figure 3. Error Exponent curves associated to the tests T_N (T_1) and U_N (T_2) in the case where $c = \frac{1}{5}$ and $\rho = 10\,dB$. Each point of the curve corresponds to a given error exponent under H_0 (X axis) and its counterpart error exponent under H_1 (Y axis) as described in Theorem 2-(2) for T_N and Theorem 3-(2) for U_N.

Finally, we compare in Figure 4 the powers (computed by Monte-Carlo methods) of tests T_N and U_N for finite values of N and K. We consider the case where $K = 10$, $N = 50$ and $\rho = 1$ and plot the probability of error under H_0 versus the power of the test, that is α versus $\mathbb{P}_1(T_N \geq \gamma_N)$ (resp. $\mathbb{P}_1(U_N \geq \gamma_N)$) where γ_N is fixed by the following condition:

$$\mathbb{P}_0(T_N \geq \gamma_N) = \alpha \quad (\text{resp. } \mathbb{P}_0(U_N \geq \gamma_N) = \alpha) \ .$$

VIII. Conclusion

In this contribution, we have analyzed in detail the GLRT in the case where the noise variance and the channel are unknown. Unlike similar contributions, we have focused our efforts on the analysis of the error exponent by means of large random matrix theory and large deviation techniques. Closed-form expressions were obtained and enabled us to establish that the GLRT asymptotically outperforms the test based on the condition number, a fact that is supported by finite-dimension simulations. We also believe that the large deviations techniques introduced here will be of interest for the engineering community, beyond the problem addressed in this paper.

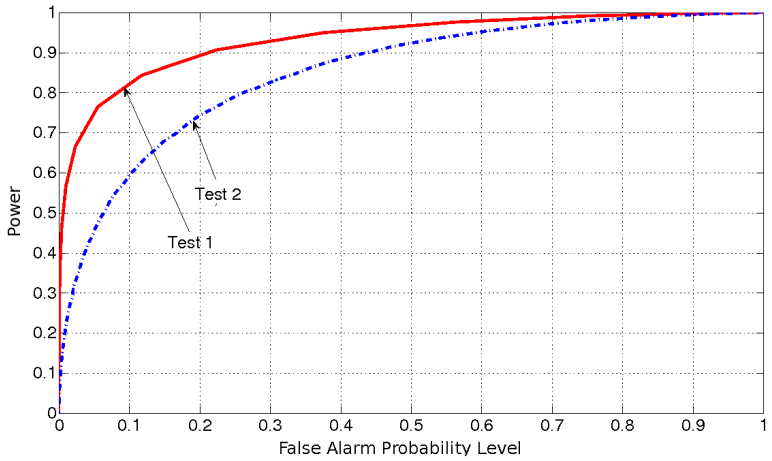

Figure 4. Simulated ROC curves for T_N (test 1) and U_N (test 2) in the case where $K = 10$, $N = 50$ and $\rho = 10\,dB$.

ACKNOWLEGMENT

We thank Olivier Cappé for many fruitful discussions related to the GLRT.

APPENDIX A
PROOF OF LEMMA 1: LARGE DEVIATIONS FOR T_N

The large deviations of the largest eigenvalue of large random matrices have already been investigated in various contexts, Gaussian Orthogonal Ensemble [49] and deformed Gaussian ensembles [21]. As mentionned in [21, Remark 1.2], the proofs of the latter can be extended to complex Wishart matrix models, that is random matrices $\hat{\mathbf{R}}$ under H_0 or H_1.

In both cases, the large deviations of λ_1 rely on a close study of the density of the eigenvalues, either given by (12) (under H_0) or by (19) for the spiked model (under H_1). The study of the spiked model, as it involves the study of the asymptotics of the spherical integral (see Lemma 3 below), is more difficult. We therefore focus on the proof of the LDP under H_1 (Lemma 1-(2)) and omit the proof of Lemma 1-(1). Once Lemma 1-(2) is proved, proving Lemma 1-(1) is a matter of bookkeeping, with the spherical integral removed at each step.

Recall that $\lambda_1 \geq \cdots \geq \lambda_K$ are the ordered eigenvalues of $\hat{\mathbf{R}}$ and that T_N is the statistics defined in (6).

In the sequel, we shall prove the upper bound of the LDP in Lemma 1-(2) (which gives also the upper bound in Lemma 1-(3)). The proof of the lower bound in Lemma 1-(3) requires more

precise arguments than the lower bound of the LDP. One has indeed to study what happens at the vicinity of λ^+, which is a point of discontinuity of the rate function I_ρ^+. Thus, we skip the proof of the lower bound of the LDP in Lemma 1-(2) to avoid repetition. Note that the proof of Lemma 1-(4) is a mere consequence of the fact that T_N converges a.s. to λ^+ if $\rho \leq \sqrt{c}$, thus $\mathbb{P}_1(T_N < x_N)$ converges to 1 whenever x_N converges to $x > \lambda^+$.

For sake of simplicity and with no loss of generality as the law of T_N does not depend on σ, we assume all along this appendix that $\sigma^2 = 1$. We first recall important asymptotic results for spherical integrals.

A. Useful facts about spherical integrals

Recall that the joint distributions of the ordered eigenvalues under hypothesis H_0 and H_1 are respectively given by (12) and (19). In the latter, the so-called spherical integral (20) is introduced. We recall here results from [21] related to the asymptotic behaviour of the spherical integral in the case where one diagonal matrix is of rank one and the other has the limiting distribution $\mathbb{P}_{\check{\text{MP}}}$. We first introduce the function defined for $x \geq \lambda^+$ by:

$$J_\rho(x) = \begin{cases} \frac{\rho}{c} - \log\left(\frac{\rho}{c(1+\rho)}\right) - \mathbf{F}^+(\lambda_{\text{spk}}^\infty), & \text{if } \rho \leq \sqrt{c} \text{ and } \lambda^+ \leq x \leq \lambda_{\text{spk}}^\infty, \\ \frac{\rho x}{c(1+\rho)} - 1 - \log\left(\frac{\rho}{c(1+\rho)}\right) - \mathbf{F}^+(x), & \text{otherwise.} \end{cases} \tag{58}$$

Consider a K-tuple $(x_1, \cdots, x_K)$ and denote by $\hat{\pi}_{K,\mathbf{x}} = \frac{1}{K-1}\sum_{i=2}^N \delta_{x_2}$ the empirical distribution associated to $(x_2, \cdots, x_K)$; let d be a metric compatible with the topology of weak convergence of measures (for example the Dudley distance - see for instance [50]). A strong version of the convergence of the spherical integral in the exponential scale with speed N, established in [21] can be summarized in the following Lemma:

Lemma 3. *Assume that $N, K \to \infty$ and $\frac{K}{N} \to c \in (0,1)$ and let Assumption 1 hold true. Let $x_1 \geq x_2 \geq \cdots \geq x_K \geq 0$ and $\delta > 0$. If, for N large enough, $|x_1 - x| \leq \delta$ and $d(\hat{\pi}_{K,\mathbf{x}}, \mathbb{P}_{\check{\text{MP}}}) \leq N^{-1/4}$ then:*

$$\left| \frac{1}{N} \log I_K\left(\frac{N}{K}\mathbf{B}_K, \mathbf{X}_K\right) - cJ_\rho(x) \right| \leq \delta,$$

where J_ρ is given by (58)*, $\mathbf{B}_K = \text{diag}\left(\frac{\rho_K}{1+\rho_K}, 0, \ldots, 0\right)$ and $\mathbf{X}_K = \text{diag}(x_1, \cdots, x_K)$.*

Recall that the spherical integral I_K, defined in (20), appears in the joint density (19) of the eigenvalues under H_1. Lemma 3 provides a simple asymptotic equivalent $cJ_\rho(x)$ of the

normalized integral $N^{-1}\log I_K$. Roughly speaking, this will enable us to replace I_K by the quantity $e^{-N\times cJ_\rho(x)}$ when establishing the large deviations of λ_1, which rely on a careful study of density (19).

B. Proof of Lemma 1-(2)

In order to establish the LDP under hypothesis H_1 and condition $\rho > \sqrt{c}$, (that is the bounds (37) and (38)), we first notice that intervals $(x, x+\delta)$ for $x, \delta \in \mathbb{R}^+$ form a basis of the topology of $\mathbb{R}^+$. The LDP will be therefore a consequence of the following bounds:

- (Exponential tightness) there exists a function $f : \mathbb{R}^+ \to \mathbb{R}^+$ going to infinity at infinity such that for all N,
$$\mathbb{P}_1\left(\lambda_1 \geq M\right) \;\leq\; e^{-Nf(M)} \;. \tag{59}$$
Condition (59) is technical (see for instance [44, Lemma 1.2.18]): Instead of proving the large deviation upper bound for every closed set, the exponential tightness (59), if established, enables one to restrict to the compact sets.
- (Upper bound) For any x, for any M such that $0 < x < M$,
$$\lim_{\delta\downarrow 0}\limsup_{N,K\to\infty}\frac{1}{N}\log\mathbb{P}_1\left(x \leq T_N \leq x+\delta,\, \lambda_1 \leq M\right) \;\leq\; -I_\rho^+(x)\;, \tag{60}$$
Due to the exponential tightness, it is sufficient to establish the upper bound for compact sets. As each compact can be covered by a finite number of balls, it is therefore sufficient to establish upper estimate (60) in order to establish the LD upper bound.
- (Lower bound) For any x,
$$\lim_{\delta\downarrow 0}\liminf_{N,K\to\infty}\frac{1}{N}\log\mathbb{P}_1\left(x \leq T_N \leq x+\delta\right) \;\geq\; -I_\rho^+(x)\;. \tag{61}$$
The fact that (61) implies the LD lower bound (38) is standard in LD and can be found in [44, Chapter 1] for instance.

As the arguments are very similar to the ones developed in [21], we only prove in detail the upper bound (60). Proofs of (59) and (61) are left to the reader.

The idea is that the empirical measure $\hat{\pi}_{K,\boldsymbol{\lambda}} := \frac{1}{K-1}\sum_{j=2}^{K}\delta_{\lambda_j}$ (of all but the largest eigenvalues) and the trace concentrate faster than the largest eigenvalue. In the exponential scale with speed N, $\hat{\pi}_{K,\boldsymbol{\lambda}}$ and the trace can be considered as equal to their limit, respectively $P_{\check{\mathrm{MP}}}$ and 1. In particular, the deviations of T_N arise from those of the largest eigenvalue and they both satisfy

the same LDP with the same rate function I_ρ^+. We therefore isolate the terms depending on λ_1 and gather the others through their empirical measure $\hat{\pi}_{K,\boldsymbol{\lambda}}$.

Recall the notations introduced in (12) and (19) and let $x > \lambda^+$, $\delta > 0$. Consider the following domain:

$$\mathcal{D} = \left\{ (x_1, \cdots, x_K) \in [0, M]^K, \ \frac{K x_1}{x_1 + \cdots + x_K} \in (x, x+\delta) \right\}$$

For N large enough:

$$\begin{aligned}
\mathbb{P}_1(x \leq T_N \leq x+\delta, \ \lambda_1 \leq M) &= \int_{\mathcal{D}} dp^1_{K,N}(x_{1:K}) \\
&= \frac{1}{Z^1_{K,N}} \int_{\mathcal{D}} dx_1 \, e^{-N x_1} e^{(N-K)\log x_1} e^{2(K-1)\int \log(x_1-u) d\hat{\pi}_{K,\mathbf{x}}(u)} \\
&\quad \times I_K\left(\frac{N}{K}\mathbf{B}_K, \mathbf{X}_K\right) \prod_{1<i<j} |x_i - x_j|^2 e^{-N\sum_{j=2}^K x_j} \prod_{j=2}^K x_j^{N-K} d\, x_{2:K} \times \mathbf{1}_{(x_1 \geq \cdots \geq x_K \geq 0)} \\
&= \frac{\left(1-\frac{1}{N}\right)^{(K-1)(N-1)} Z^0_{K-1,N-1}}{Z^1_{K,N}} \int_{\mathcal{D}} dx_1 e^{-N x_1} e^{(N-K)\log x_1} e^{2(K-1)\int \log(x_1-u) d\hat{\pi}_{K,\mathbf{y}}(u)} \\
&\quad \times I_K\left(\frac{N}{K}\mathbf{B}_K, \mathbf{X}_K\right) dp^0_{K-1,N-1}(y_{2:K}),
\end{aligned}$$

where we performed the change of variables $y_i := \frac{N}{N-1} x_i$ for $i = 2 : K$, and the related modifications $\hat{\pi}_{K,\mathbf{x}} \leftrightarrow \hat{\pi}_{K,\mathbf{y}}$ and $\mathbf{X}_K = \mathrm{diag}\left(x_1, \frac{N-1}{N} y_2, \cdots, \frac{N-1}{N} y_2\right)$. Note also that strictly speaking, the domain of integration $\mathcal{D}$ would express differently with the y_i's and in particular, we should have changed constant M which majorizes the x_i's into a larger constant as the y_i's can theoretically be slightly above M - we keep the same notation for the sake of simplicity.

To proceed, one has to study the asymptotic behaviour of the normalizing constant:

$$\frac{\left(1-\frac{1}{N}\right)^{(K-1)(N-1)} Z^0_{K-1,N-1}}{Z^1_{K,N}} \ ,$$

which turns out to be difficult. Instead of establishing directly the bounds (59)-(61), we proceed as in [21] and establish similar bounds replacing the probability measures $\mathbb{P}_1$ by the measures $\mathbb{Q}_1$ defined as:

$$\mathbb{Q}_1 := \frac{Z^1_{K,N}}{Z^0_{K-1,N-1}\left(1-\frac{1}{N}\right)^{(K-1)(N-1)}} \mathbb{P}_1$$

and the rate function I_ρ^+ by the function G_ρ defined by:

$$G_\rho(x) = \frac{x}{1+\rho} - (1-c)\log x - c\mathbf{F}^+(x) + c + c\log\left(\frac{\rho}{c(1+\rho)}\right)$$

for $x > \lambda^+$. Notice that these positive measures $\mathbb{Q}_1$ are not probability measures any more, and as a consequence, the function G_ρ is not necessarily positive and its infimum might not be equal to zero, as it is the case for a rate function.

Writing the upper bound for $\mathbb{Q}_1$, we obtain:

$$\begin{aligned}&\mathbb{Q}_1(x \leq T_N \leq x+\delta,\ \lambda_1 \leq M)\\ &\quad\leq \int_{\mathcal{D}} dx_1 e^{-N\Phi(x_1,c_N,\hat{\pi}_{K,\mathbf{y}})} I_K\left(\frac{N}{K}\mathbf{B}_K, \mathbf{X}_K\right) dp^0_{K-1,N-1}(y_{2:K}),\end{aligned}$$

where, for any compactly supported probability measure μ and any real number y greater than the right edge of the support of μ,

$$\Phi(y, c_N, \mu) = -y + (1-c_N)\log y + 2c_N \int \log(y-\lambda) d\mu(\lambda).$$

Let us now localise the empirical measure $\hat{\pi}_{K,\mathbf{y}}$ around $\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}$[7] and the trace around 1. The continuity and convergence properties of the spherical integral recalled in Lemma 3 yield, for K large enough:

$$\begin{aligned}\mathbb{Q}_1(x \leq T_N \leq x+\delta\ ,\ \lambda_1 \leq M) \ \leq\ & \int_x^{x+\delta} dx_1 \int_{\mathcal{E}} e^{-N\Phi(x_1,c_N,\hat{\pi}_{K,\mathbf{y}})} e^{Nc(J_\rho(x_1)+\delta)} dp^0_{K-1,N-1}(y_{2:K})\\ &+4^K M^{N+K} e^{NM\frac{\rho_K}{1+\rho_K}} \int_{\mathcal{E}^C} dp^0_{K-1,N-1}(y_{2:K}),\end{aligned} \tag{62}$$

with

$$\mathcal{E} := \left\{(y_2,\cdots,y_K) \in [0,M]^{K-1},\ d(\hat{\pi}_{K,\mathbf{y}}, \mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}) \leq \frac{1}{N^{1/4}} \quad \text{and} \quad \frac{1}{K}\sum_{j=2}^K y_j \in \left[1-\delta^2, 1+\delta^2\right]\right\}.$$

The second term in (62) is easily obtained considering the fact that all the eigenvalues are less than M so that for $1 \leq j \leq K$, $|x_1 - x_j| \leq 2M$, $x_j^{N-K} \leq M^{N-K}$ and $(UX_KU^*)_{11} \leq M$. Now, standard concentration results under H_0 yield that:

$$\limsup_{N,K\to\infty} \frac{1}{N}\log \mathbb{P}_0\left(\hat{\pi}_{K,\lambda} \notin B(\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}, N^{-1/4}) \text{ or } \frac{1}{K}\sum_{j=2}^K \lambda_j \notin \left[1-\delta^2, 1+\delta^2\right]\right) = -\infty.$$

More precisely, one knows using [51] that the empirical measure $\frac{1}{K}\sum_{j=2}^K \lambda_j$ is close enough to its expectation and then using [52] one knows that the expectation is close enough to its limit $\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}$. The arguments are detailed in the Wigner case in [21] and we do not give more details here.

[7]Notice that if $\hat{\pi}_{K,\mathbf{x}}$ is close to $\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}$, so is $\hat{\pi}_{K,\mathbf{y}}$ due to the change of variable $y_i = \frac{N}{N-1}x_i$.

As $c_N \to c$ for $N, K \to \infty$, $c \mapsto \Phi(y, c, \mu)$ is continuous and $\mu \mapsto \Phi(y, c, \mu)$ is lower semi-continuous, we obtain:

$$\limsup_{N,K\to\infty} \frac{1}{N} \log \mathbb{Q}_1(x \le \lambda_1 \le x+\delta \ , \ \lambda_1 \le M) \ \le \sup_{u\in[x,x+\delta]} (\Phi(u, c, P_{\check{\mathrm{MP}}}) + cJ_\rho(u)) + 2\delta.$$

By continuity in u of the two involved functions, we finally get:

$$\lim_{\delta\downarrow 0} \limsup_{N,K\to\infty} \frac{1}{N} \log \mathbb{Q}_1(x \le \lambda_1 \le x+\delta \ , \ \lambda_1 \le M) \ \le \ \Phi(x, c, \mathbb{P}_{\check{\mathrm{MP}}}) + cJ_\rho(x) = G_\rho(x) \ ,$$

and the counterpart of Eq. (60) is proved for $\mathbb{Q}_1$ and function G_ρ. The proof of the lower bound is quite similar and left to the reader. It remains now to recover (60). As $\mathbb{P}_1$ is a probability measure and the whole space $\mathbb{R}^+$ is both open and closed, an application of the upper and lower bounds for $\mathbb{Q}_1$ immediately yields:

$$\begin{aligned}
&\liminf_{N,K\to\infty} \frac{1}{N} \log \frac{Z^1_{K,N}}{Z^0_{K-1,N-1}\left(1-\frac{1}{N}\right)^{(K-1)(N-1)}} \mathbb{P}_1(T_N \in \mathbb{R}^+) \\
&= \limsup_{N,K\to\infty} \frac{1}{N} \log \frac{Z^1_{K,N}}{Z^0_{K-1,N-1}\left(1-\frac{1}{N}\right)^{(K-1)(N-1)}} \mathbb{P}_1(T_N \in \mathbb{R}^+) \\
&= \lim_{N,K\to\infty} \frac{1}{N} \log \frac{Z^1_{K,N}}{Z^0_{K-1,N-1}\left(1-\frac{1}{N}\right)^{(K-1)(N-1)}} \\
&= -\inf_{\mathbb{R}^+} G_\rho \ . \qquad (63)
\end{aligned}$$

This implies that the LDP holds for $\mathbb{P}_1$ with rate function $G_\rho - \inf_{\mathbb{R}^+} G_\rho$.

It remains to check that $I^+_\rho = G_\rho - \inf_{\mathbb{R}^+} G_\rho$, which easily follows from the fact to be proved that:

$$\inf_{x\in[\lambda^+,\infty)} G_\rho(x) = G_\rho(\lambda^\infty_{\mathrm{spk}}) \ . \qquad (64)$$

We therefore study the variations of G_ρ over $[\lambda^+, \infty)$. Note that $(\mathbf{F}^+)' = -\mathbf{f}$, and thus that $G'_\rho(x) = (1+\rho)^{-1} - (1-c)x^{-1} + c\mathbf{f}(x)$. Function $\mathbf{f}$ being a Stieltjes transform is increasing for $x > \lambda^+$, and so is G'_ρ, whose limit at infinity is $(1+\rho)^{-1}$. Straightforward but involved computations using the explicit representation (67) for $\mathbf{f}$ yield that $G'_\rho(\lambda^\infty_{\mathrm{spk}}) = 0$. Therefore, G_ρ is decreasing on $[\lambda^+, \lambda^\infty_{\mathrm{spk}}]$ and increasing on $[\lambda^\infty_{\mathrm{spk}}, \infty)$, and (64) is proved.

This concludes the proof of the upper bound in Lemma 1-(2). The proof of Lemma 1-(1) is very similar and left to the reader.

C. Proof of Lemma 1-(3)

The proof of this point requires an extra argument as we study the large deviations of T_N near the point $(1+\sqrt{c})^2$ where the rate function is not continuous. In particular, the limit (53) does not follow from the LDP already established. As we shall see when considering $\mathbb{P}_1\left(T_N < (1+\sqrt{c_N})^2 + \eta_N N^{-2/3}\right)$, the fact that the scale $(N^{-2/3})$ is the same as the one of the fluctuations of the largest eigenvalue of the complex Wishart model is crucial.

We detail the proof in the case when $\rho > \sqrt{c}$ and, as above, consider the positive measures $\mathbb{Q}_1$. We need to prove that:

$$\liminf_{N,K\to\infty} \frac{1}{N} \log \mathbb{Q}_1\left(T_N < (1+\sqrt{c_N})^2 + \frac{\eta}{N^{2/3}}\right) \geq -G_\rho(\lambda^+), \quad \eta \in \mathbb{R}, \tag{65}$$

the other bound being a direct consequence of the LDP. As previously, we will carefully localize the various quantities of interest. Denote by $g_N(\eta) = (1+\sqrt{c_N})^2 + \eta N^{-2/3}$ for $\eta \in \mathbb{R}$ and by $h_N(r) = 1 - rN^{-2/3}$ for $r > 0$. Notice also that $\lambda_1 \leq g_N(\eta)h_N(r)$ together with $\frac{1}{K-1}\sum_{j=2}^K \lambda_j > h_N(r)$ imply that $T_N < g_N(\eta)$. We shall also consider the further constraints:

$$g_N(\eta-1)h_N(r) \leq \lambda_1 \quad \text{and} \quad \lambda_2 < g_N(\eta-2)h_N(r)$$

which enable us to properly separate λ_1 from the support of $\hat{\pi}_{K,\boldsymbol{\lambda}}$. Now, with the localisation indicated above, we have for N large enough,

$$\begin{aligned}\mathbb{Q}_1\left(T_N < g_N(\eta)\right) \;\geq\; \mathbb{Q}_1\Bigg(& g_N(\eta-1)h_N(r) \leq \lambda_1 \leq g_N(\eta)h_N(r), \\ & \frac{1}{K-1}\sum_{j=2}^K \lambda_j > h_N(r),\ \lambda_2 < g_N(\eta-2)h_N(r),\ \hat{\pi}_{K,\boldsymbol{\lambda}} \in B(\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}, N^{-1/4})\Bigg).\end{aligned}$$

As previously, we consider the variables $y_j = \frac{N}{N-1}x_j$ for $2 \leq j \leq K$ and obtain, with the help of Lemma 3:

$$\mathbb{Q}_1\left(T_N < g_N(\eta)\right) \quad \geq \quad \int_{g_N(\eta-1)h_N(r)}^{g_N(\eta)h_N(r)} dx_1 \int_{\mathcal{F}} e^{-N\Phi(x_1,c_N,\hat{\pi}_{K,\mathbf{y}})} e^{Nc(J_\rho(x_1)-\delta)} dp^0_{K-1,N-1}(y_{2:K})$$

with

$$\mathcal{F} := \Bigg\{ (y_2,\cdots,y_K) \in \left[0, \frac{N\, g_N(\eta-2)\, h_N(r)}{N-1}\right]^{K-1}, \\ \frac{1}{K-1}\sum_{j=2}^K y_j > \frac{N}{N-1}h_N(r), \hat{\pi}_{K,\mathbf{y}} \in B(\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}, N^{-1/4}) \Bigg\}.$$

Therefore:

$$\mathbb{Q}_1\left(T_N < g_N(\eta)\right) \;\geq\; h_N(r)\;\left(g_N(\eta) - g_N(\eta-1)\right)\, e^{N(G_\rho(\lambda^+)-2\delta)}\mathbb{P}_0\left((\lambda_2,\cdots,\lambda_K)\in\mathcal{F}\right)$$

(recall that $G_\rho(x) = \Phi(x, c, \mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}) + cJ_\rho(x)$). Now, as $h_N(r)\;\left(g_N(\eta) - g_N(\eta-1)\right) = (1 - rN^{-2/3})N^{-2/3}$, its contribution vanishes at the LD scale:

$$\lim_{N\to\infty}\frac{1}{N}\log\left(h_N(r)\;\left(g_N(\eta) - g_N(\eta-1)\right)\right) = 0\;.$$

It remains to check that $\mathbb{P}_0\left((\lambda_2,\cdots,\lambda_K)\in\mathcal{F}\right)$ is bounded below uniformly in N. This will yield the convergence of $\frac{1}{N}\log\mathbb{P}_0\left((\lambda_2,\cdots,\lambda_K)\in\mathcal{F}\right)$ towards zero, hence (65). Consider:

$$\begin{aligned}\mathbb{P}_0\left((\lambda_2,\cdots,\lambda_K)\in\mathcal{F}^c\right) &\leq \mathbb{P}_0\left(\hat{\pi}_{K,\boldsymbol{\lambda}}\notin B(\mathbb{P}_{\check{\mathrm{M}}\mathrm{P}}, N^{-1/4})\right)\\ &\quad+\mathbb{P}_0\left(\frac{1}{K-1}\sum_{j=2}^{K}\lambda_j < \frac{N}{N-1}h_N(r)\right)+\mathbb{P}_0\left(\lambda_2 > \frac{N}{N-1}g_N(\eta-2)h_N(r)\right)\;.\end{aligned}$$

We have already used the fact that the first term goes to zero when N grows to infinity. Recall that the fluctuations of $\frac{1}{K-1}\sum_{j=2}^{K}\lambda_j$ are of order $\frac{1}{N}$, therefore the second term also goes to zero as we consider deviations of order $N^{-2/3}$. Now, $N^{2/3}(\lambda_2 - (1+\sqrt{c_N})^2)$ converges in distribution to the Tracy-Widom law, therefore the last term converges to $F_{\mathrm{TW}}\left(\eta - 2 + r(1+\sqrt{c})^2\right) < 1$. This concludes the proof.

APPENDIX B
SKETCH OF PROOF FOR LEMMA 2: LARGE DEVIATIONS FOR U_N

As stated in Remark 10, we shall first study the LDP for the joint quantity (λ_1, λ_K). The purpose here is to outline the following convergence:

$$\frac{1}{N}\log\mathbb{P}\left(\lambda_1\in A, \lambda_K\in B\right)\xrightarrow[N,K\to\infty]{} -\inf_{x\in A} I_\rho^+(x) - \inf_{y\in B} I^-(x)\;,$$

which is an illustrative way, although informal[8], to state the LDP for (λ_1, λ_K) (see (39)).

Consider the quantity $\mathbb{P}\left(\lambda_1\in(\alpha_1,\beta_1), \lambda_K\in(\alpha_K,\beta_K)\right)$. As we are interested in the deviations of λ_1 and λ_K, the interesting scenario is $\lambda^+\notin(\alpha_1,\beta_1)$ and $\lambda^-\notin(\alpha_K,\beta_K)$ (recall that $\lambda^\pm$ are the edgepoints of the support of Marčenko-Pastur distribution). More precisely, the interesting case is when the deviations of the extreme eigenvalue occur outside of the bulk: $\alpha_1 > \lambda^+$ and

[8]All the statements, computations and approximations below can be made precise as in the proof of Lemma 1.

$\beta_K < \lambda^-$; such deviations happen at the rate $e^{-N\times const.}$. The case where the deviations would occur within the bulk is unlikely to happen because it would enforce the whole eigenvalues to deviate from the limiting support of Marčenko-Pastur distribution, which happens at the rate $e^{-N^2\times const.}$. Denote by $A = (\alpha_1, \beta_1)$ and $B = (\alpha_K, \beta_K)$.

$$
\begin{aligned}
&\mathbb{P}\left(\lambda_1 \in A, \lambda_K \in B\right) \\
&= \frac{1}{Z^1_{K,N}} \int_{A\times\mathbb{R}^{(K-2)}\times B} 1_{(\lambda_1\geq\cdots\geq\lambda_K\geq 0)} \prod_{1\leq i<j\leq K} (x_i - x_j)^2 \\
&\qquad\qquad \times \prod_{j=1}^{K} x_j^{N-K} e^{-Nx_j} I_K\left(\frac{N}{K}B_K, X_K\right) d\,x_{1:K} \\
&= \int_A d\,x_1\; e^{2\sum_{j=2}^{K-1}\log(x_1-x_j)} e^{(N-K)\log x_1 - Nx_1} I_K\left(\frac{N}{K}B_K, X_K\right) \\
&\quad \times \int_B d\,x_K\; e^{2\sum_{i=2}^{K-1}\log(x_i-x_K)} e^{(N-K)\log x_K - Nx_K} e^{2\log(x_1-x_K)} \\
&\quad\quad \times \frac{Z^0_{K-2,N-2}}{Z^1_{K,N}} \int_{x_1\geq x_2\geq\cdots\geq x_K} \prod_{j=2}^{K-1} e^{-2x_j} \prod_{j=2}^{K-1} \frac{x_j^{N-K} e^{-(N-2)x_j}}{Z^0_{K-2,N-2}} \prod_{2\leq i<j\leq K-1} (x_i - x_j)^2 d\,x_{2:K-1}
\end{aligned}
$$

We shall now perform the following approximations:

$$
\begin{aligned}
\sum_{j=2}^{K-1} \log(x_1 - x_j) &\approx (K-2)\int \log(x_1 - x)\mathbb{P}_{\check{\mathrm{MP}}}(\,dx) = (K-2)\mathbf{F}^+(x_1)\;, \\
\sum_{j=2}^{K-1} \log(x_j - x_K) &\approx (K-2)\int \log(x - x_K)\mathbb{P}_{\check{\mathrm{MP}}}(\,dx) = (K-2)\mathbf{F}^-(x_K)\;, \\
\sum_{j=2}^{K-1} x_j &\approx (K-2)\int x\mathbb{P}_{\check{\mathrm{MP}}}(\,dx) = (K-2)\;, \\
I_K\left(\frac{N}{K}B_K, X_K\right) &\approx e^{NcJ_\rho(x_1)}\;.
\end{aligned}
$$

The three first approximations follow from the fact that $\frac{1}{K-2}\sum_2^{K-1}\delta_{x_i} \approx \mathbb{P}_{\check{\mathrm{MP}}}$, the last one from

Lemma 3. Plugging these approximations into the expression of $\mathbb{P}\left(\lambda_1 \in A, \lambda_K \in B\right)$ yields:

$$\begin{aligned}
&\mathbb{P}\left(\lambda_1 \in A, \lambda_K \in B\right) \\
&\quad\approx \int_A d\,x_1\; e^{2(K-2)\mathbf{F}^+(x_1)} e^{(N-K)\log x_1 - N x_1} e^{N c J_\rho(x_1)} \\
&\qquad \times \int_B d\,x_K\; e^{2(K-2)\mathbf{F}^-(x_K)} e^{(N-K)\log x_K - N x_K} e^{2\log(x_1 - x_K)} \\
&\qquad\quad \times \frac{Z^0_{K-2,N-2}}{Z^1_{K,N}} e^{-2(K-2)} \int_{x_1 \geq x_2 \geq \cdots \geq x_K} \prod_{j=2}^{K-1} \frac{x_j^{N-K} e^{-(N-2)x_j}}{Z^0_{K-2,N-2}} \prod_{2 \leq i < j \leq K-1} (x_i - x_j)^2 d\,x_{2:K-1}\;.
\end{aligned}$$

As $x_1 \geq \alpha_1 \geq \lambda^+$ and $x_K \leq \beta_K \leq \lambda^-$, the last integral goes to one as $K, N \to \infty$ and:

$$\begin{aligned}
&\mathbb{P}\left(\lambda_1 \in A, \lambda_K \in B\right) \\
&\quad\approx \int_A d\,x_1\; e^{-N\left(\frac{2(K-2)}{N}\mathbf{F}^+(x_1) - \left(1-\frac{K}{N}\right)\log x_1 + x_1 - c J_\rho(x_1)\right)} \\
&\qquad \times \int_B d\,x_K\; e^{-N\left(\frac{2(K-2)}{N}\mathbf{F}^-(x_K) - \left(1-\frac{K}{N}\right)\log x_K + x_K + \frac{2\log(x_1 - x_K)}{N}\right)} \\
&\qquad\quad \times \frac{Z^0_{K-2,N-2}}{Z^1_{K,N}} e^{-2(K-2)}\;.
\end{aligned}$$

Recall that we are interested in the limit $N^{-1}\log\mathbb{P}\left(\lambda_1 \in A, \lambda_K \in B\right)$. The last term will account for a constant Υ (see for instance (63)):

$$\frac{1}{n}\log\left(\frac{Z^0_{K-2,N-2}}{Z^1_{K,N}} e^{-2(K-2)}\right) \quad \xrightarrow[N,K\to\infty]{} \quad \Upsilon\;.$$

The term $\frac{2\log(x_1 - x_K)}{N}$ within the exponential in the integral accounts for the interraction between λ_1 and λ_K and its contribution vanishes at the desired rate. In order to evaluate the two remaining integrals, one has to rely on Laplace's method (see for instance [53]) to express the leading term of the integrals (replacing KN^{-1} by c below):

$$\begin{aligned}
\int_A d\,x_1\; e^{-N\left(2c\mathbf{F}^+(x_1) - (1-c)\log x_1 + x_1 - c J_\rho(x_1)\right)} &\approx e^{-N\inf_{x\in A}\left(2c\mathbf{F}^+(x) - (1-c)\log x + x - c J_\rho(x)\right)}\;, \\
\int_B d\,x_K\; e^{-N\left(2c\mathbf{F}^-(x_K) - (1-c)\log x_K + x_K - c J_\rho(x_K)\right)} &\approx e^{-N\inf_{y\in B}\left(2c\mathbf{F}^-(y) - (1-c)\log y + y\right)}\;.
\end{aligned}$$

Finally, we get the desired limit:

$$\frac{1}{N}\log\mathbb{P}\left\{\lambda_1 \in A, \lambda_K \in B\right\} \xrightarrow[N,K\to\infty]{} -\inf_{x\in A}\Phi^+(x) - \inf_{y\in B}\Phi^-(y) + \Upsilon\;,$$

where

$$\begin{aligned}\Phi^+(x) &= 2c\mathbf{F}^+(x) - (1-c)\log x + x - cJ_\rho(x)\ ,\\ \Phi^-(y) &= 2c\mathbf{F}^-(y) - (1-c)\log y + y\ .\end{aligned}$$

It remains to replace J_ρ by its expression (58) and to spread the constant Υ over Φ^+ and Φ^-, which are not a priori rate functions (recall that a rate function is nonnegative). If $\lambda^- \in B$, then the event $\{\lambda_K \in B\}$ is "typical" and no deviation occurs, otherwise stated, the rate function I^- should satisfy $I^-(\lambda^-) = 0$. Similarly, $I_0^+(\lambda^+) = 0$ under H_0 and $I_\rho^+(\lambda_{\rm spk}^\infty) = 0$ under H_1. Necessarily, Υ should write $\Upsilon = \Phi(\lambda^-) + \Phi(\lambda^+)$ under H_0 (resp. $\Upsilon = \Phi(\lambda^-) + \Phi(\lambda_{\rm spk}^\infty)$ under H_1) and the rate functions should be given by: $I^- = \Phi^- - \Phi(\lambda^-)$, $I_0^+ = \Phi^+ - \Phi(\lambda^+)$ under H_0 (resp. $I_\rho^+ = \Phi^+ - \Phi(\lambda_{\rm spk}^\infty)$ under H_1), which are the desired results.

We have proved (informally) that the LDP holds true for (λ_1, Λ_K) with rate function $I_{0/\rho}^+(x) + I^-(y)$. The contraction principle [44, Chap. 4] immediatly yields the LDP for the ratio $\frac{\lambda_1}{\lambda_K}$ with rate function:

$$\Gamma_{0/\rho}(t) = \inf_{(x,y),\frac{x}{y}=t} \{I_{0/\rho}^+(x) + I^-(y)\}\ , \tag{66}$$

which is the desired result. We provide here intuitive arguments to understand this fact.

For this, interpret the value of the rate function $I_\rho^+(x)$ as the cost associated to a deviation of λ_1 (under H_1) around x: $\mathbb{P}\{\lambda_1 \in (x, x+dx)\} \approx e^{-NI_\rho^+(x)}$. If a deviation occurs for the ratio $\frac{\lambda_1}{\lambda_K}$, say $\frac{\lambda_1}{\lambda_K} \in (t, t+dt)$ where $t > \frac{\lambda_{\rm spk}^\infty}{\lambda^-}$ (which is the typical behaviour of U_N under H_1), then necessarily λ_1 must deviate around some value ty, so does λ_K around some value y, so that the ratio is around t. In terms of rate functions, the cost of the joint deviation $(\lambda_1 \approx ty, \lambda_K \approx y)$ is $I_\rho^+(ty) + I^-(y)$. The true cost associated to the deviation of the ratio will be the minimum cost among all these possible joint deviations of λ_1 and λ_K, hence the rate function (66).

APPENDIX C
CLOSED-FORM EXPRESSIONS FOR FUNCTIONS $\mathbf{f}$, $\mathbf{F}^+$ AND $\mathbf{F}^-$

Consider the Stieltjes transform $\mathbf{f}$ of Marčenko-Pastur distribution:

$$\mathbf{f}(z) = \int \frac{\mathbb{P}_{\check{\rm MP}}(d\lambda)}{\lambda - z}\ .$$

We gather without proofs a few facts related to $\mathbf{f}$, which are part of the folklore.

Lemma 4 (Representation of $\mathbf{f}$)**.** *The following hold true:*

1) *Function* $\mathbf{f}$ *is analytic in* $\mathbb{C} - [\lambda^-, \lambda^+]$.
2) *If* $z \in \mathbb{C} - [\lambda^-, \lambda^+]$ *with* $\Re(z) \geq \frac{\lambda^+ + \lambda^-}{2}$*, then*

$$\mathbf{f}(z) = \frac{(1-z-c) + \sqrt{(1-z-c)^2 - 4cz}}{2cz} \ ,$$

where $\sqrt{z}$ *stands for the principal branch of the square-root.*

3) *If* $z \in \mathbb{C} - [\lambda^-, \lambda^+]$ *with* $\Re(z) < \frac{\lambda^+ + \lambda^-}{2}$*, then*

$$\mathbf{f}(z) = \frac{(1-z-c) - \sqrt{(1-z-c)^2 - 4cz}}{2cz} \ ,$$

where $-\sqrt{z}$ *stands for the branch of the square-root whose image is* $\{z \in \mathbb{C}, \ \Re(z) \leq 0\}$.

4) *As a consequence, the following hold true:*

$$\mathbf{f}(x) = \frac{(1-x-c) + \sqrt{(1-x-c)^2 - 4cx}}{2cx} \quad \textit{if} \quad x \geq \lambda^+ \ , \tag{67}$$

$$\mathbf{f}(x) = \frac{(1-x-c) - \sqrt{(1-x-c)^2 - 4cx}}{2cx} \quad \textit{if} \quad 0 \leq x \leq \lambda^- \ . \tag{68}$$

5) *Consider the following function* $\tilde{\mathbf{f}}(z) = c\mathbf{f}(z) - \frac{1-c}{z}$*. Functions* $\mathbf{f}$ *and* $\tilde{\mathbf{f}}$ *satisfy the following system of equations:*

$$\left\{ \begin{array}{rcl} \mathbf{f}(z) & = & -\frac{1}{z(1+\tilde{\mathbf{f}}(z))} \\ \tilde{\mathbf{f}}(z) & = & -\frac{1}{z(1+c\mathbf{f}(z))} \end{array} \right. , \tag{69}$$

Recall the definition (31) and (51) of function $\mathbf{F}^+$ and $\mathbf{F}^-$. In the following lemma, we provide closed-form formulas of interest.

Lemma 5. *The following identities hold true:*

1) *Let* $x \geq \lambda^+$*, then*

$$\mathbf{F}^+(x) = \log(x) + \frac{1}{c}\log(1 + c\mathbf{f}(x)) + \log(1 + \tilde{\mathbf{f}}(x)) + x\mathbf{f}(x)\tilde{\mathbf{f}}(x) \ .$$

2) *Let* $0 \leq x \leq \lambda^-$*, then*

$$\mathbf{F}^-(x) = \log(x) + \frac{1}{c}\log(1 + c\mathbf{f}(x)) + \log(-(1 + \tilde{\mathbf{f}}(x))) + x\mathbf{f}(x)\tilde{\mathbf{f}}(x) \ .$$

Proof: Consider the case where $x \geq \lambda^+$. First write

$$\log(x-y) = \log(x) + \int_x^\infty \left(\frac{1}{u} + \frac{1}{y-u} \right) du \ .$$

Integrating with respect with $\mathbb{P}_{\check{\mathrm{MP}}}$ and applying Funini's theorem yields:

$$\int \log(x-y)\mathbb{P}_{\check{\mathrm{MP}}}(\,dy) = \log(x) + \int_x^\infty \left(\frac{1}{u} + \mathbf{f}(u)\right)\,du$$

in the case where $x > \lambda^+$. Recall that $\mathbf{f}$ and $\tilde{\mathbf{f}}$ are holomorphic functions over $\mathbb{C} - (\{0\} \cup [\lambda^-, \lambda^+])$ and satisfy system (69) (notice in particular that $1 + c\mathbf{f}$ and $1 + \tilde{\mathbf{f}}$ never vanish). Using the first equation of (69) implies that:

$$\int \log(x-y)\mathbb{P}_{\check{\mathrm{MP}}}(\,dy) = \log(x) - \int_x^\infty \mathbf{f}(u)\tilde{\mathbf{f}}(u)\,du\ . \tag{70}$$

Consider $\Gamma(u, \mathbf{f}, \tilde{\mathbf{f}}) = \frac{1}{c}\log(1+c\mathbf{f}) + \log(1+\tilde{\mathbf{f}}) + u\mathbf{f}\tilde{\mathbf{f}}$. By a direct computation of the derivative, we get:

$$\begin{aligned}\frac{d}{du}\Gamma(u, \mathbf{f}(u), \tilde{\mathbf{f}}(u)) &= \mathbf{f}'\left(\frac{1}{1+c\mathbf{f}} + u\tilde{\mathbf{f}}\right) + \tilde{\mathbf{f}}'\left(\frac{1}{1+\tilde{\mathbf{f}}} + u\mathbf{f}\right) + \mathbf{f}\tilde{\mathbf{f}} \\ &= \mathbf{f}(u)\tilde{\mathbf{f}}(u)\ .\end{aligned}$$

Hence

$$\begin{aligned}\int_x^\infty \mathbf{f}(u)\tilde{\mathbf{f}}(u)\,du &= \left[\frac{1}{c}\log(1+c\mathbf{f}) + \log(1+\tilde{\mathbf{f}}) + u\mathbf{f}\tilde{\mathbf{f}}\right]_x^\infty \\ &= -\left(\frac{1}{c}\log(1+c\mathbf{f}(x)) + \log(1+\tilde{\mathbf{f}}(x)) + x\mathbf{f}(x)\tilde{\mathbf{f}}(x)\right).\end{aligned}$$

It remains to plug this identity into (70) to conclude. The representation of $\mathbf{F}^-$ can be established similarly.

■

REFERENCES

[1] J. Mitola III and GQ Maguire Jr. Cognitive radio: making software radios more personal. *IEEE Wireless Communications*, 6:13–18, 1999.

[2] S. Haykin. Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected areas in Comm.*, 23, 2005.

[3] M. Dohler, E. Lefranc, and A.H. Aghvami. Virtual antenna arrays for future wireless mobile communication systems. *ICT, Beijing, China*, 2002.

[4] H. Urkowitz. Energy detection of unknown deterministic signals. *Proc. of the IEEE*, 55:523–531, 1967.

[5] V. I. Kostylev. Energy detection of a signal with random amplitude. *Proc IEEE Int. Conf. on Communications, New York City*.

[6] M. K. Simon, F. F. Digham, and M.-S. Alouini. On the energy detection of unknown signals over fading channels. *ICC 2003 Conference Record, Anchorage, Alaska*, 2003.

[7] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor. Spatial-spectral joint detection for wideband spectrum sensing in cognitive radio networks. *Proc. ICASSP, Las Vegas*, 2008.

[8] E.L. Lehman and J.P. Romano. *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, 2006.

[9] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. on Signal, Speech, and Signal Processing*, 33(2):387–392, April 1985.

[10] P-J Chung, J. F. Böhme, C. F. Mecklenbräuker, and A. Hero. Detection of the number of signals using the benjamin-hochberg procedure. *IEEE Trans. on Signal Processing*, 55, 2007.

[11] A. Taherpour, M. Nasiri-Kenari, and S. Gazor. Multiple antenna spectrum sensing in cognitive radios. *IEEE Transactions on Wireless Communications*, 9(2):814–823, 2010.

[12] X. Mestre. Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans on Inform. Theory*, 54(11):5113–5129, Nov. 2008.

[13] X. Mestre. On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices. *IEEE Trans. on Signal Processing*, 56(11):5353–5368, Nov. 2008.

[14] S. Kritchman and B. Nadler. Determining the number of components in a factor model from limited noisy data. *Chemometrics and Intelligent Laboratory Systems, 2008, 1932*, 94.

[15] S. Kritchman and B. Nadler. Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *in press in IEEE Transactions Signal Processing, 2009*.

[16] J. Silverstein and P. Combettes. Signal detection via spectral theory of large dimensional random matrices. *IEEE Transactions on Signal Processing*, 40(8):2100–2105, 1992.

[17] R. Couillet and M. Debbah. A Bayesian Framework for Collaborative Multi-Source Signal Detection. *IEEE Transactions on Signal Processing (under review)*, 2010. *arxiv:0811.0764*.

[18] N. Raj Rao, J. A. Mingo, R. Speicher, and A. Edelman. Statistical eigen-inference from large Wishart matrices. *Ann. Statist.*, 36(6):2850–2885, 2008.

[19] N. Raj Rao and J. Silverstein. Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. *arXiv:0902.4250*, 2009.

[20] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.

[21] M. Maida. Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.*, 12:1131–1150 (electronic), 2007.

[22] Y. H. Zeng and Y. C. Liang. Eigenvalue based spectrum sensing algorithms for cognitive radio. *to appear in IEEE Trans on Communications volume=arXiv:0804.2960v1*.

[23] L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim. Cooperative Spectrum Sensing Using Random Matrix Theory. *Proceedings of ISPWC*, 2008.

[24] F. Penna, R. Garello, and M. A. Spirito. Cooperative Spectrum Sensing based on the Limiting Eigenvalue Ratio Distribution in Wishart Matrices. *IEEE Communication Letters*, submitted, 2009.

[25] T. Abbas, N-K Masoumeh, and G. Saeed. Multiple antenna spectrum sensing in cognitive radios. *submitted to IEEE Transactions on Wireless Communications, 2009*.

[26] T. W. Anderson. Asymptotic theory for principal component analysis. *J. Math. Stat.*, 34:122–148, 1963.

[27] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[28] M. L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[29] G. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[30] R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York, 1982. Wiley Series in Probability and Mathematical Statistics.

[31] P. Koev. *Random Matrix statistics toolbox*. http://math.mit.edu/ plamen/software/rmsref.html.

[32] P. Koev and A. Edelman. The efficient evaluation of the hypergeometric function of a matrix argument. *Math. Comp.*, 75(254):833–846 (electronic), 2006.

[33] J. Mitola. *Cognitive Radio An Integrated Agent Architecture for Software Defined Radio*. PhD thesis, Royal Institute of Technology (KTH), May 2000.

[34] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.

[35] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields*, 108(4):517–542, 1997.

[36] B. Nadler. On the distribution of the ratio of the largest eigenvalue to the trace of a wishart matrix. *submitted to the Annals of Statistics*, 2010. available at *http://www.wisdom.weizmann.ac.il/∼nadler/*.

[37] C. A. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.

[38] C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177(3):727–754, 1996.

[39] A. Bejan. Largest eigenvalues and sample covariance matrices. tracy-widom and painleve ii: computational aspects and realization in s-plus with applications. *http://www.vitrum.md/andrew/TWinSplus.pdf*, 2005.

[40] F. Bornemann. Asymptotic independence of the extreme eigenvalues of GUE. *arXiv:0902.3870*, 2009.

[41] J. Baik and J. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, 97(6):1382–1408, 2006.

[42] A. W. Van der Vaart. *Asymptotic Statistics*, chapter 14. Cambridge Univ. Press., 1998.

[43] P. J. Bickel and P. W. Millar. Uniform convergence of probability measures on classes of functions. *Statist. Sinica*, 2(1):1–15, 1992.

[44] A. Dembo and O. Zeitouni. *Large Deviations Techniques And Applications*. Springer Verlag, New York, second edition, 1998.

[45] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields*, 78(4):509–521, 1988.

[46] Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.*, 21(3):1275–1294, 1993.

[47] P. Bianchi, M. Debbah, and J. Najim. Asymptotic independence in the spectrum of the gaussian unitary ensemble. *arXiv:0811.0979*.

[48] P.-N. Chen. General Formulas For The Neyman-Pearson Type-II Error Exponent Subject To Fixed And Exponential Type-I Error Bounds. *IEEE Transactions on Information Theory*, 42(1):316–323, 1996.

[49] G. Ben Arous, A. Dembo, and A. Guionnet. Aging of spherical spin glasses. *Probab. Theory Related Fields*, 120(1):1–67, 2001.

[50] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[51] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136 (electronic), 2000.

[52] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.*, 21(2):649–672, 1993.

[53] J. Dieudonné. *Infinitesimal calculus*. Hermann, Paris, 1971. Translated from the French.