

# ASYMPTOTIC ANALYSIS OF A GLR TEST FOR DETECTION WITH LARGE SENSOR ARRAYS: NEW RESULTS

*Sonja Hiltunen, Philippe Loubaton*

LIGM (CNRS, Univ. Paris-Est Marne-la-Vallée), 5 Bd. Descartes 77454 Marne-la-Vallée (France), France

## ABSTRACT

This paper addresses the behaviour of a classical multi-antenna GLR test that allows to detect the presence of a known signal corrupted by a multipath propagation channel when the number of sensors  $M$ , the sample size  $N$ , and the number of paths  $L$  are large and of the same order of magnitude. In the asymptotic regime where  $M, N$  and  $L$  converge towards  $\infty$  at the same rate, the test statistics is shown to be asymptotically Gaussian under each hypothesis. The results of this paper extend previous contributions of the authors that addressed the case where the number of paths  $L$  is much smaller than  $M$  and  $N$ , and used these results in order to propose a pragmatic Gaussian approximation of the test statistics when  $L$  is large. Simulation results indicate that the proposed Gaussian approximation of the test statistics is more reliable than the above mentioned pragmatic approximation.

**Index Terms**— Multichannel detection, asymptotic analysis, large random matrices

## 1. INTRODUCTION

Multivariate signals that are observed at the output of massive MIMO multi-antenna receivers are of course high-dimensional, and in this context, the available number of observations  $N$  is very often of the same order of magnitude as the number of sensors  $M$ . It is now well understood that in this case, standard statistical multivariate signal processing problems such as source localization ([8], [11]), source power estimation ([4]), detection ([7],[9], [2]), or synchronization ([10]) need to be revisited, since standard statistical methodologies provide reliable results only when the ratio  $\frac{M}{N}$  is small enough.

In [10], we addressed the detection problem of a known synchronization sequence transmitted by a single transmitter in an unknown multipath propagation channel when the receiver is equipped with a massive antenna array. We assumed that the observations were corrupted by a temporally white, but spatially correlated (with unknown spatial covariance matrix) additive complex Gaussian noise, and studied the behaviour of the generalized likelihood ratio test (GLRT) when the number of sensors  $M$  and the length  $N$  of the synchronization sequence are large and of the same order of magnitude. Under the assumption that the number of paths  $L$  is much smaller than  $M$  and  $N$ , we proved that the GLR test statistics has a Gaussian behaviour under each hypothesis. However, we observed that even for reasonably small values of  $L$ , this Gaussian approximation does not allow to accurately predict the ROC curves of the GLRT. When  $L$  is of the same order of magnitude as  $M$  and  $N$ , existing results concerning centered large Fisher random matrices ([12], [1]) allowed to conclude that under the null hypothesis  $H_0$ , the GLRT statistics has a Gaussian behaviour whose mean and variance can be

expressed in closed form. The behaviour of the test statistics under hypothesis  $H_1$  is mathematically much harder because it needs to establish a central limit theorem for a certain linear statistics of the eigenvalues of a large random non zero mean Fisher matrix, a question that is not covered by existing results. In [10], we rather proposed a pragmatic Gaussian approximation providing reasonably accurate results.

In the present paper, we study more rigorously the behaviour of the test statistics under hypothesis  $H_1$ , and prove that it has a Gaussian behaviour. Due to space limitations, we do not provide the proof of this technical result. We rather compare the new approximation with the pragmatic proposal of [10], and establish that in a certain SNR range, the two approximations coincide, thus explaining the nice behaviour of the large  $L$  approximation of [10]. However, for large SNR, the two approximations differ, and simulation results show that the new proposed approximation provides more accurate results.

This paper is organized as follows. In section 2, we provide the signal model under hypotheses  $H_0$  and  $H_1$ , recall the expression of the statistics  $\eta_N$  corresponding to the GLRT test, and recall that in order to study the probability distribution of  $\eta_N$  under the two hypotheses, there is no restriction to assume that the additive noise is spatially decorrelated and that the training sequence matrix is orthogonal. In section 3, we review the main results of [10]. In section 4, we present the main result of this paper, and compare the corresponding Gaussian approximation with the pragmatic proposal of [10]. In section 5, we provide numerical results to demonstrate that the new approximation allows to predict in a more reliable way the ROC curves of the GLRT in a reasonable range of probability of false alarm and of probability of detection.

**General notations.** For a complex matrix  $\mathbf{A}$ , we denote by  $\mathbf{A}^T$  and  $\mathbf{A}^*$  its transpose and its conjugate transpose, and by  $\text{Tr}(\mathbf{A})$  and  $\|\mathbf{A}\|$  its trace and spectral norm.  $\mathbf{I}$  represents the identity matrix. The real normal distribution with mean  $m$  and variance  $\sigma^2$  is denoted  $\mathcal{N}_{\mathbb{R}}(m, \sigma^2)$ . A complex random variable  $Z = X + iY$  follows the distribution  $\mathcal{N}_{\mathbb{C}}(\alpha + i\beta, \sigma^2)$  if  $X$  and  $Y$  are independent with respective distributions  $\mathcal{N}_{\mathbb{R}}(\alpha, \frac{\sigma^2}{2})$  and  $\mathcal{N}_{\mathbb{R}}(\beta, \frac{\sigma^2}{2})$ . For a sequence of random variables  $(X_n)_{n \in \mathbb{N}}$  and a random variable  $X$ , we write  $X_n \xrightarrow{\mathcal{D}} X$  when  $X_n$  converges in distribution to  $X$  when  $n \rightarrow +\infty$ .

## 2. PRESENTATION OF THE PROBLEM.

In the following, we assume that a single transmitter sends a known synchronization sequence  $(s_n)_{n=1, \dots, N}$  through a channel with  $L$  paths, and that the corresponding signal is received on a receiver equipped with  $M$  sensors. The received  $M$ -dimensional signal is denoted by  $(\mathbf{y}_n)_{n=1, \dots, N}$ . When the transmitter and the receiver are perfectly synchronized, the  $M \times N$  observed matrix

$\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$  can be written as

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \mathbf{V} \quad (1)$$

where  $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_N)$  is the additive noise matrix,  $\mathbf{H}$  the  $M \times L$  channel matrix and  $\mathbf{S}$  the  $L \times N$  matrix corresponding to the known synchronization sequence. In the following, we assume that  $(\mathbf{v}_n)_{n \in \mathbb{Z}}$  is an additive independent identically distributed complex Gaussian noise verifying

$$\mathbb{E}(\mathbf{v}_n) = 0, \mathbb{E}(\mathbf{v}_n \mathbf{v}_n^T) = 0, \mathbb{E}(\mathbf{v}_n \mathbf{v}_n^*) = \sigma^2 \mathbf{R}$$

where  $\mathbf{R} > 0$  satisfies  $\frac{1}{M} \text{Tr}(\mathbf{R}) = 1$ . We also assume that the size  $N$  of the training sequence satisfies  $N > M + L$ . In this paper, we study the classical problem of testing the hypothesis  $H_1$  characterized by Eq. (1) against the hypothesis  $H_0$  defined by

$$\mathbf{Y} = \mathbf{V} \quad (2)$$

We assume from now on that  $\mathbf{H}$ ,  $\sigma^2$  and  $\mathbf{R}$  are unknown at the receiver side. In this context, it is well established (see e.g. [3]) that the generalized maximum likelihood test consists in comparing the following statistics  $\eta_N$  to a threshold:

$$\eta_N = -\log \det [\mathbf{I}_L - \mathbf{T}_N], \quad (3)$$

where  $\mathbf{T}_N$  is the  $L \times L$  matrix defined by

$$\mathbf{T}_N = \left( \frac{\mathbf{S}\mathbf{S}^*}{N} \right)^{-1/2} \frac{\mathbf{S}\mathbf{Y}^*}{N} \left( \frac{\mathbf{Y}\mathbf{Y}^*}{N} \right)^{-1} \frac{\mathbf{Y}\mathbf{S}^*}{N} \left( \frac{\mathbf{S}\mathbf{S}^*}{N} \right)^{-1/2}. \quad (4)$$

In order to study the behaviour of the test above, we study the limit distribution of  $\eta_N$  under each hypothesis. For this, we recall that it is possible to assume without restriction that the matrix  $\mathbf{S}$  verifies  $\frac{\mathbf{S}\mathbf{S}^*}{N} = \mathbf{I}_L$  and that  $\mathbb{E}(\mathbf{v}_n \mathbf{v}_n^*) = \sigma^2 \mathbf{I}$ , i.e. the matrix  $\mathbf{R}$  is reduced to identity. Moreover, there exist two independent  $M \times (N - L)$  and  $M \times L$  complex Gaussian matrices  $\mathbf{V}_1$  and  $\mathbf{V}_2$  with i.i.d.  $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$  entries such that  $\eta_N$  can be written as

$$\eta_N = \log \det \left( \mathbf{I}_L + \mathbf{V}_2^* / \sqrt{N} (\mathbf{V}_1 \mathbf{V}_1^* / N)^{-1} \mathbf{V}_2 / \sqrt{N} \right) \quad (5)$$

under hypothesis  $H_0$ , and as

$$\eta_N = \log \det (\mathbf{I}_N + \mathbf{G}_N) \quad (6)$$

under hypothesis  $H_1$ . Here, the matrix  $\mathbf{G}_N$  is defined by

$$\mathbf{G}_N = \left( \mathbf{H} + \mathbf{V}_2 / \sqrt{N} \right)^* (\mathbf{V}_1 \mathbf{V}_1^* / N)^{-1} \left( \mathbf{H} + \mathbf{V}_2 / \sqrt{N} \right). \quad (7)$$

For more details, we refer the reader to section II of [10]. In the following sections, we study the asymptotic behaviour of the distribution of  $\eta_N$  in the following asymptotic regime:

**Assumption 1.**  $M, N, L$  converge towards  $+\infty$  in such a way that  $c_N = \frac{M}{N}$  and  $d_N = \frac{L}{N}$  converge towards  $c > 0$  and  $d > 0$  with  $c + d < 1$

In order to simplify the notations, this regime will be denoted as  $N \rightarrow +\infty$ . We note that the condition  $c + d < 1$  is consistent with the assumption that  $M + L < N$ . As the dimensions of the matrix  $\mathbf{H}$  grow with  $M, N, L$ , we also have to specify the behavior of  $\mathbf{H}$ . In the following, we assume that  $\sup_N \|\mathbf{H}\| < +\infty$ , i.e. that the spectral norm of  $\mathbf{H}$  remains finite when its dimensions increase. We finally recall that the behaviour of  $\eta_N$  under hypothesis  $H_0$  follows from existing mathematical results (see below for more details), and therefore concentrate on the behaviour of  $\eta_N$  under hypothesis  $H_1$ .

### 3. REVIEW OF THE RESULTS OF [10].

In this section, we review the main results of [10], concentrating on the pragmatic approximation of  $\eta_N$  under Assumption 1 and hypothesis  $H_1$ . In order to explain the structure of this pragmatic approximation, we first recall the behaviour of  $\eta_N$  when the number of paths  $L$  remains fixed. The following result holds.

**Theorem 1.** *In the asymptotic regime where  $M, N$  converge towards  $+\infty$  in such a way that  $c_N = \frac{M}{N}$  converges towards  $c < 1$  and where  $L$  is a fixed parameter that does not scale with  $M, N$ , under hypothesis  $H_1$ , it holds that*

$$(\delta_{N,1})^{-1/2} (\eta_N - \bar{\eta}_{N,1}) \rightarrow_{\mathcal{D}} \mathcal{N}_{\mathbb{R}}(0, 1) \quad (8)$$

where  $\bar{\eta}_{N,1}$  and  $\delta_{N,1}$  are defined by

$$\begin{aligned} \bar{\eta}_{N,1} &= L \log \frac{1}{1 - c_N} + \log \det (\mathbf{I} + \mathbf{H}^* \mathbf{H} / \sigma^2) \\ \delta_{N,1} &= \frac{L}{N} \frac{c_N}{1 - c_N} + \frac{\kappa}{N} \end{aligned}$$

with  $\kappa = \text{Tr} \left[ \mathbf{I} - \left( \mathbf{I} + \frac{\mathbf{H}^* \mathbf{H}}{\sigma^2} \right)^{-2} \right]$

**Remark 1.** *As  $L$  remains finite, the dimensions of the  $L \times L$  matrix  $\mathbf{G}_N$  defined by (7) do not grow when  $M, N \rightarrow +\infty$ . Therefore, in order to study the behaviour of  $\eta_N = \log \det (\mathbf{I}_L + \mathbf{G}_N)$ , it is sufficient to evaluate the behaviour of each entry of  $\mathbf{G}_N$  and to use a classical linearization argument. We also remark that the asymptotic variance of  $\eta_N$  converges towards 0 at rate  $\frac{1}{N}$ .*

**Remark 2.** *The results related to hypothesis  $H_0$  are obtained from the results concerning  $H_1$  by setting  $\mathbf{H} = 0$ .*

We now recall the behaviour of the distribution of  $\eta_N$  under hypothesis  $H_0$  in the regime defined by Assumption 1.

**Proposition 1.** *In the asymptotic regime defined by Assumption 1, under  $H_0$ , it holds that*

$$(\delta_{N,2})^{-1/2} (\eta_N - \bar{\eta}_{N,2}) \rightarrow_{\mathcal{D}} \mathcal{N}_{\mathbb{R}}(0, 1) \quad (9)$$

where  $\bar{\eta}_{N,2}$  and  $\delta_{N,2}$  are defined by

$$\begin{aligned} \bar{\eta}_{N,2} &= -N \left( (1 - c_N) \log(1 - c_N) + (1 - d_N) \log(1 - d_N) \right) \\ &\quad + N(1 - c_N - d_N) \log(1 - c_N - d_N) \\ \delta_{N,2} &= \log \left( \frac{(1 - c_N)(1 - d_N)}{1 - c_N - d_N} \right) \end{aligned} \quad (10)$$

**Remark 3.** *We remark that in contrast with the case where  $L$  remains fixed, it is not sufficient to evaluate the behaviour of the individual entries of matrix  $\mathbf{G}_N$  in order to obtain the behaviour of  $\eta_N$ . (9) follows from the observation that*

$$\eta_N = \log \det \left( \mathbf{I}_M + (\mathbf{V}_2 \mathbf{V}_2^* / N) (\mathbf{V}_1 \mathbf{V}_1^* / N)^{-1} \right)$$

and thus coincides with a linear statistics of the eigenvalues of the matrix  $\mathbf{V}_2 \mathbf{V}_2^* / N (\mathbf{V}_1 \mathbf{V}_1^* / N)^{-1}$ . This matrix is a Fisher matrix, and the asymptotic behaviour of  $\eta_N$  follows from the results of [1] and [12].

**Remark 4.** We mention that  $\bar{\eta}_{N,2}$  increases linearly with  $M, N, L$ , and that the asymptotic variance  $\delta_{N,2}$  is a  $\mathcal{O}(1)$  term. This in contrast with the regime  $L$  fixed in which the asymptotic mean is a  $\mathcal{O}(1)$  term and the asymptotic variance is a  $\mathcal{O}(1/N)$  term. We finally mention that it can be shown that the relative error between  $\mathbb{E}(\eta_N)$  and its approximation  $\bar{\eta}_{N,2}$  is  $\mathcal{O}(\frac{1}{N^{2-\epsilon}})$  for each  $\epsilon > 0$ .

As mentioned in [10], the asymptotic mean and the asymptotic variance of  $\eta_N$  in the regime  $M, N$  converge towards  $+\infty$  and  $L$  remains finite have an additive structure:  $\bar{\eta}_{N,1}$  is the sum of two terms: the asymptotic mean of  $\eta_N$  under  $H_0$  in the considered asymptotic regime (i.e.  $L \log \frac{1}{1-c_N}$ ) and the asymptotic mean of  $\eta_N$  in the standard asymptotic regime where  $N \rightarrow +\infty$  and  $M$  and  $L$  are fixed (i.e.  $\log \det(\mathbf{I} + \mathbf{H}^* \mathbf{H} / \sigma^2)$ ). The asymptotic variance is also the sum of the asymptotic variance under  $H_0$  (i.e.  $\frac{L}{N} \frac{c_N}{1-c_N}$ ) with the asymptotic variance under  $H_1$  in the regime  $N \rightarrow +\infty$  and  $M, L$  fixed (i.e.  $\kappa/N$ ). Proposition 1 as well as the additive structure of  $\bar{\eta}_{N,1}$  and  $\delta_{N,1}$  suggested us to propose in [10] to approximate the distribution of  $\eta_N$  under  $H_1$  in regime defined by Assumption 1 by a  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$  where  $\bar{\eta}_{N,3} = \bar{\eta}_{N,2} + \log \det(\mathbf{I} + \mathbf{H}^* \mathbf{H} / \sigma^2)$  (i.e. the sum of the asymptotic mean of  $\eta_N$  under  $H_0$  with the asymptotic mean of  $\eta_N$  when  $N \rightarrow +\infty$  and  $M, L$  remain fixed) and  $\delta_{N,3} = \delta_{N,2} + \frac{\kappa}{N}$  (i.e. the sum of the asymptotic variance of  $\eta_N$  under  $H_0$  with the asymptotic variance of  $\eta_N$  when  $N \rightarrow +\infty$  and  $M, L$  remain fixed). The numerical results of [10] showed that the approximation  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$  is more accurate than the approximation  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,1}, \delta_{N,1})$  even if  $L$  remains small compared to  $M, N$ .

#### 4. GAUSSIAN APPROXIMATION OF $\eta_N$ UNDER HYPOTHESIS $H_1$ .

In this section, we present the main result of this paper, and point out the connections between the proposed Gaussian approximation of  $\eta_N$  and the pragmatic approximation  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$  proposed in [10]. In order to introduce our theorem, we have to introduce some notations. We first define  $s_N$  as the unique positive solution of the equation

$$s_N = \frac{1}{M} \text{Tr} \left( \sigma^2(1-c_N)\mathbf{I} + \frac{\mathbf{H}\mathbf{H}^*}{1 + \sigma^2 c_N s_N} \right)^{-1} \quad (11)$$

and denote by  $\mathbf{S}_N$  the  $M \times M$  matrix given by

$$\mathbf{S}_N = \left( \sigma^2(1-c_N)\mathbf{I} + \frac{\mathbf{H}\mathbf{H}^*}{1 + \sigma^2 c_N s_N} \right)^{-1} \quad (12)$$

The following result holds.

**Theorem 2.** Under Assumption 1 and hypothesis  $H_1$ , it holds that

$$(\delta_N)^{-1/2}(\eta_N - \bar{\eta}_N) \rightarrow_{\mathcal{D}} \mathcal{N}_{\mathbb{R}}(0, 1) \quad (13)$$

where  $\bar{\eta}_N$  and  $\delta_N$  are defined by

$$\begin{aligned} \bar{\eta}_N &= \log \det \left( (1-c_N)\mathbf{I} + \frac{\mathbf{H}\mathbf{H}^*}{\sigma^2(1 + \sigma^2 c_N s_N)} \right) + \\ &M \sigma^2(1-c_N) \left( \frac{1}{\sigma^2(1-c_N)} - s_N \right) + N \log(1 + \sigma^2 c_N s_N) - \\ &N((1-d_N) \log(1-d_N) - (1-c_N-d_N) \log(1-c_N-d_N)) \end{aligned} \quad (14)$$

and

$$\delta_N = \log \left( \frac{1-d_N}{1-c_N-d_N} \right) - \log \left( 1 + \sigma^4 c_N(1-c_N) \frac{1}{M} \text{Tr}(\mathbf{S}_N^2) \right) \quad (15)$$

The proof of this result is quite intricate, and we will only provide the general strategy. We first mention that using existing results concerning the so-called Information + Noise models (see e.g. [6], [5]), it is easy to establish that  $\mathbb{E}(\eta_N) = \bar{\eta}_N + \mathcal{O}(\frac{1}{N^{1-\epsilon}})$  for each  $\epsilon > 0$ . The difficult part of the proof consists in evaluating the asymptotic behaviour of the characteristic function  $g_N(u)$  of  $\eta_N - \mathbb{E}(\eta_N)$ . The general approach we follow is based on the following expression of  $g_N(u)$ :

$$g_N(u) = \mathbb{E} \left[ e^{iu(\mathbb{E}_{\mathbf{V}_1}(\eta_N) - \mathbb{E}(\eta_N))} \mathbb{E} \left( e^{iu(\eta_N - \mathbb{E}_{\mathbf{V}_1}(\eta_N))} | \mathbf{V}_2 \right) \right]$$

where  $\mathbb{E}_{\mathbf{V}_1}$  represents the mathematical expectation operator w.r.t.  $\mathbf{V}_1$  and where  $\mathbb{E}(\cdot | \mathbf{V}_2)$  is the conditional expectation operator given  $\mathbf{V}_2$ . It is possible to establish that for each  $u$ ,

$$\mathbb{E} \left( e^{iu(\eta_N - \mathbb{E}_{\mathbf{V}_1}(\eta_N))} | \mathbf{V}_2 \right) \simeq e^{-\xi_N u^2/2}$$

where  $\xi_N$  represents a positive deterministic term. From this, we deduce that

$$g_N(u) \simeq e^{-\xi_N u^2/2} \mathbb{E} \left[ e^{iu(\mathbb{E}_{\mathbf{V}_1}(\eta_N) - \mathbb{E}(\eta_N))} \right]$$

and prove that  $\mathbb{E} \left[ e^{iu(\mathbb{E}_{\mathbf{V}_1}(\eta_N) - \mathbb{E}(\eta_N))} \right] \simeq e^{-\chi_N u^2/2}$ . This, in turn, implies that  $g_N(u) \simeq e^{-(\xi_N + \chi_N) u^2/2}$  and that  $(\xi_N + \chi_N)^{-1/2}(\eta_N - \mathbb{E}(\eta_N))$  converges in distribution towards a  $\mathcal{N}(0, 1)$  random variable. We finally comment on Theorem 2.

We first mention that if  $\mathbf{H} = 0$ , then,  $s_N$  defined by Eq. (11) coincides with  $\frac{1}{\sigma^2(1-c_N)}$ , and that  $\bar{\eta}_N$  and  $\delta_N$  coincide with  $\bar{\eta}_{N,2}$  and  $\delta_{N,2}$  respectively.

We now compare the pragmatic approximation  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$  with the approximation provided by Theorem 2 in the case where  $\text{Tr}(\mathbf{H}\mathbf{H}^*)$  remains bounded when  $N \rightarrow +\infty$ . In this case, it is easy to check that  $s_N$  defined by Eq. (11) can be written as

$$s_N = \frac{1}{\sigma^2(1-c_N)} + \mathcal{O}\left(\frac{1}{N}\right)$$

Using standard first order expansions, we obtain immediately that  $\bar{\eta}_{N,3} = \bar{\eta}_N + \mathcal{O}(\frac{1}{N})$  and  $\delta_{N,3} = \delta_N + \mathcal{O}(\frac{1}{N})$ . Therefore, when  $\text{Tr}(\mathbf{H}\mathbf{H}^*)$  remains bounded, the two approximations tend to coincide. We remark that the signal to noise ratio at the receiver side coincides with  $\frac{\text{Tr}(\mathbf{H}\mathbf{H}^*)}{\sigma^2}$ , so assuming that  $\text{Tr}(\mathbf{H}\mathbf{H}^*)$  remains bounded implies in practice that the signal to noise ratio is low. This condition is of course much stronger than the condition that  $\|\mathbf{H}\|$  remains bounded, which, rather, implies that  $\text{Tr}(\mathbf{H}\mathbf{H}^*) = \mathcal{O}(M)$ , and corresponds to high signal to noise ratios. Therefore, at moderate signal to noise ratios, the pragmatic approximation  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$  behaves as the rigorous one, thus explaining the accuracy of the pragmatic Gaussian approximation of [10].

#### 5. NUMERICAL RESULTS

In this section, we compare the accuracies of the approximations  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$  and  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_N, \delta_N)$ . The true mathematical expectation and the true variance of  $\eta_N$ , as well as the ROC curves corresponding to the GLRT, are evaluated by generating  $10^7$  independent realizations of  $\eta_N$ . In each experiment, the matrix  $\mathbf{H}$  is generated as  $\mathbf{H} = \mathbf{G} / (\text{Tr}(\mathbf{G}\mathbf{G}^*))^{-1/2}$  where  $\mathbf{G}$  is a realization of a  $M \times L$  random matrix with i.i.d.  $\mathcal{N}_{\mathbb{C}}(0, 1)$  entries. Therefore,  $\mathbf{H}$  is normalized in such a way that  $\text{Tr}(\mathbf{H}\mathbf{H}^*) = 1$ , and the signal to noise ratio

coincides with  $\frac{1}{\sigma^2}$ . Finally, in all our experiments,  $c_N = M/N$  is fixed to  $1/2$ , while  $d_N = L/N = 1/4$ .

We first represent for various values of  $L$  in figures 1 and 2 the relative errors on  $\mathbb{E}(\eta_N)$  and on  $\text{Var}(\eta_N)$  provided by the two approximations when  $\frac{1}{\sigma^2} = \frac{M}{24}$  (low SNR) and  $\frac{1}{\sigma^2} = \frac{M}{7}$  (high SNR). We recall that for each value of  $L$ ,  $M$  and  $N$  satisfy  $d_N = 1/4$  and  $c_N = 1/2$ . As expected, the proposed approximation  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_N, \delta_N)$  appears much more accurate than  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$ , and the accuracy of the latter tends to decrease when the SNR increases. In other words, for low SNR, both approximations are valid (although the proposed approximation gives a more accurate theoretical expected value), but with high SNR,  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$  is less accurate.

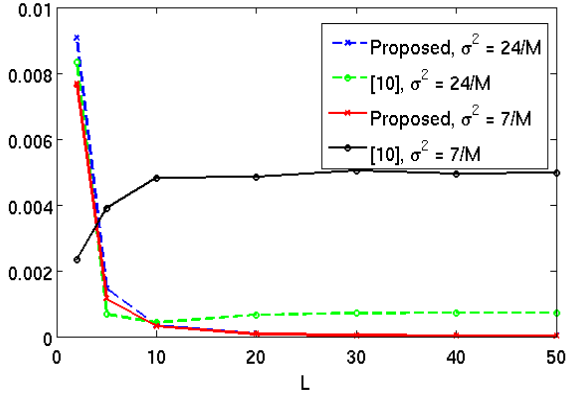


Fig. 1. Relative errors in expected value under hypothesis  $H_1$

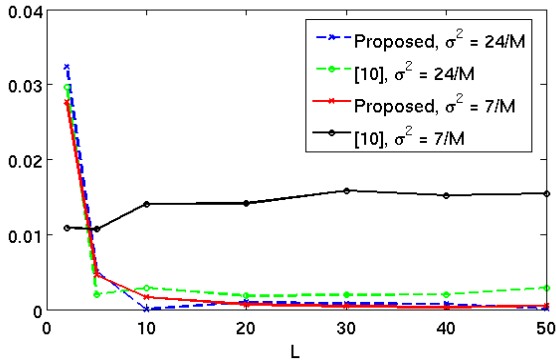


Fig. 2. Relative errors in variance under hypothesis  $H_1$

In order to validate the asymptotic Gaussianity of  $\eta_N$  under  $H_1$ , we show in figure 3 the histogram of the realizations of  $\eta_N$ , together with the probability densities of  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$  and  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_N, \delta_N)$ . In figure 3,  $L = 20$  and  $\frac{1}{\sigma^2} = M/7$  (high SNR). It is clear that the proposed asymptotic distribution offers a better fit with the empirical distribution of  $\eta_N$ .

We finally evaluate the accuracy of the two approximations by comparing their theoretical ROC curves to the empirical ROC curve evaluated from the  $10^7$  realizations of  $\eta_N$ . For each approximation, the threshold corresponding to each false alarm probability is evaluated using the Gaussian approximation  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,2}, \delta_{N,2})$  of  $\eta_N$  under  $H_0$ . Figures 4 and 5 represents the three ROC curves for two different SNR values and  $L = 20$  and  $L = 50$  respectively. The ROC curve corresponding to the proposed approximation appears more accurate than the ROC curve associated to  $\mathcal{N}_{\mathbb{R}}(\bar{\eta}_{N,3}, \delta_{N,3})$

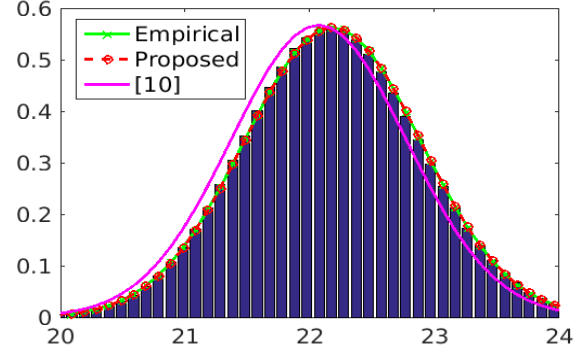


Fig. 3. Histogram of  $\eta_N$  with Gaussian approximations for  $L = 20$  and  $\sigma^2 = 7/M$

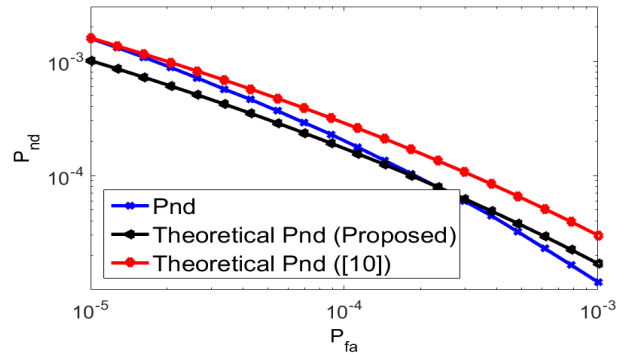


Fig. 4. Empirical and theoretical ROC curves for  $L=20$ ,  $\sigma^2 = 7/M$

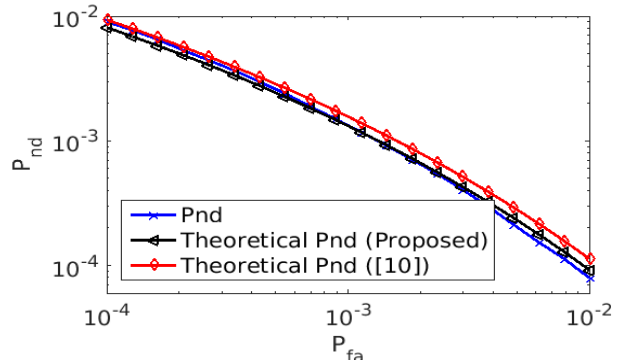


Fig. 5. Empirical and theoretical ROC curves for  $L=50$ ,  $\sigma^2 = \frac{24}{M}$

## 6. CONCLUSION

In this paper, we have established a central limit theorem for the GLRT statistics  $\eta_N$  under hypothesis  $H_1$  in the asymptotic regime where the dimensions  $N, M, L$  all go to infinity at the same rate. The corresponding asymptotic expected value and variance are expressed in closed form. It is established that the pragmatic Gaussian approximation proposed in [10] tends to have the same accuracy than the new approximation at low SNR, but that the performance of the new proposal is much better at high SNR. Simulation results confirm these theoretical claims, and demonstrate that the proposed Gaussian approximation allows to predict in a reliable way the ROC curves of the GLRT.

## REFERENCES

- [1] Z. Bai, J.W. Silverstein, "Spectral analysis of large dimensional random matrices", Springer Series in Statistics, 2nd ed., 2010.
- [2] P. Bianchi, M. Debbah, M. Maëda and J. Najim, "Performance of Statistical Tests for Single Source Detection using Random Matrix Theory", *IEEE Trans. Inf. Theory*, vol. 57, no.4, pp. 2400–2419, April 2011
- [3] D. W. Bliss, P. A. Parker "Temporal synchronization of MIMO wireless communication in the presence of interference", *IEEE Trans. Signal Process.*, vol. 58, no.3, pp.1794-1806, march 2010
- [4] R. Couillet, J.W. Silverstein, Z.D. Bai, M. Debbah, "Eigen-Inference for Energy Estimation of Multiple Sources" *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2420-2439, April 2011.
- [5] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, J. Najim, "On the capacity achieving covariance matrix for Rician MIMO channels: an asymptotic approach", *IEEE Trans. Inf. Theory*, Vol. 56, no. 3, pp. 1048-1069, March 2010.
- [6] W. Hachem, M. Kharouf, J. Najim and J. W. Silverstein. "A CLT for information-theoretic statistics of non-centered Gram random matrices", *Random Matrices: Theory and Applications*, Vol. 1 (2), April 2012.
- [7] S. Kritchman, B. Nadler, "Non-parametric Detection of the Number of Signals, Hypothesis Testing and Random Matrix Theory, *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, 2009.
- [8] X. Mestre, L.A Lagunas, "Modified subspace algorithms for DoA estimation with large arrays", *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 598-614, February 2008.
- [9] R.R. Nadakuditi, J.W. Silverstein, "Fundamental Limit of Sample Generalized Eigenvalue Based Detection of Signals in Noise Using Relatively Few Signal-Bearing and Noise-Only Samples", *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 3, pp. 468 -480, June 2010.
- [10] S. Hiltunen, P. Loubaton, P. Chevalier, "Large system analysis of a GLRT for detection with large sensor arrays in temporally white noise", *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5409 - 5423, October 15 2015
- [11] P. Vallet, X. Mestre, P. Loubaton, "Performance analysis of an improved MUSIC DoA estimator", *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6407-6422, December 1 2015
- [12] S. Zheng, "Central limit theorems for linear spectral statistics of large dimensional  $F$ -matrices", *Ann. Inst. Henri Poincaré*, vol. 48, no. 2, pp. 444-476, 2012.