A Bayesian Framework for Collaborative Multi-Source Signal Sensing

Romain Couillet, Merouane Debbah

A Bayesian Framework for Collaborative Multi-Source Signal Sensing

Romain Couillet†, *Student Member, IEEE*, Mérouane Debbah*, *Senior Member, IEEE*

***Abstract*—This paper introduces a Bayesian framework to detect multiple signals embedded in noisy observations, from an array of sensors. For various states of knowledge on the communication channel and the noise at the receiving sensors, a marginalization procedure based on random matrix theory techniques, in conjunction with the maximum entropy principle, is used to compute the Neyman-Pearson hypothesis testing criterion. Quite remarkably, although rather involved, explicit expressions for the Bayesian detector are derived which enable to decide on the presence of signal sources in a noisy wireless environment. Under the hypotheses that the true channel conditions adhere the maximum entropy model, the proposed detector is the optimal Neyman-Pearson detector; if so, the performance of the derived decision criteria can be used as an upper-bound for the performance of alternative detectors. In particular, simulation results are provided that suggest that the classical energy detector is close-to-optimal when the noise power is *a priori* known to the sensor array, especially when many sources simultaneously transmit, while the conditioning number-based detector, used classically when the noise power is unknown, is shown to perform poorly in comparison to the proposed optimal detector.**

I. Introduction

Since a few years, the idea of smart communication devices has made its way through the general framework of cognitive radio [1]. The general idea of an ideal cognitive terminal is that of a device capable of exploring and exploiting the communication environment based on any prior state of knowledge [2]. Such a device should be first able to turn prior information on the transmission channel into a mathematically tractable form. This allows then the terminal to take optimal instantaneous decisions in terms of information to feed back, bandwidth to occupy, transmission power to use etc. It should also be capable of updating its knowledge to continuously adapt to the dynamics of the environment. This vision of a cognitive radio is compliant with Haykin's anticipation of "brain empowered" wireless devices [2].

One of the key features of cognitive receivers is their ability to *sense* free spectrum. When the cognitive device is switched on, its prior knowledge on the environment is very limited but still it is requited to *decide* whether it receives informative data or only thermal noise, on different frequency bands: this will be further referred to as the *signal sensing* procedure. In the single-input single-output (SISO) scenario, the study of Bayesian signal detectors dates back to the work of Urkowitz [3] on additive white Gaussian noise (AWGN) channels. It was later extended to more realistic channel models [4]-[5]. Urkowitz's signal detector is optimal in the sense that his process performs the maximum *correct detection rate*, i.e. the odds for an informative signal to be detected as such, for a given low *false alarm rate*, i.e. the odds for a pure noise input to be wrongly declared an informative signal. To the authors' knowledge, there exists no trivial multiple input multiple output (MIMO) extension of the classical energy detector, apart from that which consists into summing up all powers received at the sensor array, e.g. [6]. In that case, the receiver decides that the incoming signal carries *information* if the total received power exceeds a given threshold; otherwise the received signal is declared pure noise. However, both single sensor or sensor array models often assume perfect *a priori* knowledge of the background noise power and, more generally, does not capitalize on the prior information available at the receiver; e.g. the receiver might be aware that, given the wireless environment, it should expect a strong line-of-sight component from the transmitter or, on the contrary, should expect the fading channel to vary strongly. Note also that the assumed prior knowledge of the noise power, or equivalently of the expected signal-to-noise ratio (SNR), is unrealistic when performing signal sensing. Alternative solutions have been proposed that deal with the issue of *a priori* unknown noise power. In [7]-[8], a method based on the conditioning number of the received sample covariance matrix[1] allows to determine a Neyman-Pearson decision criterion that is independent of the noise power. An extension of this work is found in [9], where the decision criterion is now based on the ratio between the smallest eigenvalue of the received sample covariance matrix and its trace. This method is however not heuristic but is the solution of the generalized likelihood ratio test (GLRT), which is another approach to deal with decision tests with uncertain system parameters. Other contributions propose detection schemes that are well suited for the determination of the exact number of transmit sources, e.g. [10]-[11], which extends the classical signal sensing procedure.

In a cognitive radio context, one would like all aforementioned schemes to be unified into a general framework. This is, for any prior knowledge available at the cognitive receiver, given some received signal, the device should be able to provide "optimal" decisions regarding the presence

†ST-Ericsson - Sophia Antipolis, 635 Route des Lucioles, 06560 Sophia Antipolis, romain.couillet@supelec.fr

*Chair Alcatel Lucent, Supélec - 3 rue Joliot Curie, 91192 Gif sur Yvette, merouane.debbah@supelec.fr. Debbah's work is partially supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++.

[1]the conditioning number of a Hermitian matrix is defined as the ratio between the largest and the smallest of its eigenvalues.

of informative data, the number of transmitting sources etc. This is the main focus of this work. The first barrier to break is the question of optimality under incomplete prior information. Indeed, even when much is unknown about the environment, a transmission model still has to be found. To this end, we shall model all random parameters about the environment using the *maximum entropy principle* [12], whose object is to assign prior probability distributions to unknown parameters along the following rules: (i) the prior distribution assigned to a unknown variable should be consistent with the prior information about this variable, and (ii) among those distributions that satisfy (i) the assigned prior distribution is the one with maximum entropy, i.e. the one which maximizes randomness. The maximum entropy principle, as a maximizer of randomness, is mathematically justified by a Bayesian definition of the uncertainty about a random variable, due to Cox [13], and was largely applied in statistical physics, spurred by the pioneering work from Brillouin on negentropy [14]. In the following, we shall derive the optimal Neyman-Pearson decision criteria for the problems of signal sensing under various prior knowledge on the environment, specifically the knowledge or absence of knowledge on the noise power, and on the number of transmitting sources. Those can be considered both *optimal in the maximum-entropy sense*, when some system variables are unknown, and *truly optimal*, when the true system model matches the derived maximum-entropy model; in the latter case, the derived Neyman-Pearson criteria will provide performance upper-bounds to the classical heuristic detectors. Part of this work is dedicated to evaluate through simulations how close to optimal those classical detectors are.

This paper is structured as follows: In Section II we formulate the signal sensing problem and introduce the channel model. Then in Section III, the Bayesian signal detectors are computed for various prior information on the system model at the sensing device. Simulations are then presented in Section IV. Finally, after a short discussion in Section V on the general framework and its limitations, we provide our conclusions in Section VI.

Notations: In the following, boldface lowercase and uppercase characters are used for vectors and matrices, respectively. We note $(\cdot)^{\mathsf{H}}$ the Hermitian transpose, $\mathrm{tr}(\cdot)$ denotes the matrix trace. $\mathcal{U}(N)$ is the set of unitary square matrices of size N. The notation $P_X(Y)$ denotes the probability density function of the variable X evaluated in the vicinity of Y. The notation $(x)_+$ equals x if $x > 0$ and 0 otherwise.

II. Signal Model

We consider a communication system composed of M transmitter sources, e.g. this can either be an M-antenna single transmitter or M single antenna (not necessarily uncorrelated) information sources, and a receiver composed of N sensors, be they the uncorrelated antennas of a single terminal or a mesh of scattered sensors. To enhance the multiple-antenna (MIMO) analogy, the set of sources and the set of sensors will be referred to as *the transmitter* and *the receiver*, respectively. The communication channel between the transmitter and the receiver is modelled by the matrix $\mathbf{H} \in \mathbb{C}^{N\times M}$, with $(i,j)^{th}$ entries h_{ij}. If, at time l, the transmitter emits data, denoted as an M-dimensional vector $\mathbf{s}^{(l)} = (s_1^{(l)}, \ldots, s_M^{(l)})^{\mathsf{T}} \in \mathbb{C}^M$. The additive white Gaussian noise at the receiver is modelled, at time l, by the vector $\sigma\boldsymbol{\theta}^{(l)} = \sigma(\theta_1^{(l)}, \ldots, \theta_N^{(l)})^{\mathsf{T}} \in \mathbb{C}^N$, where σ^2 denotes the variance of the noise vector entries. Without generality restriction, we shall consider in the following zero mean and unit variance of the entries of both $\boldsymbol{\theta}^{(l)}$ and $\mathbf{s}^{(l)}$, i.e. $\mathrm{E}[|\theta_i^{(l)}|^2] = 1$, $\mathrm{E}[|s_i^{(l)}|^2] = 1$. We then denote $\mathbf{y}^{(l)} = (y_1^{(l)}, \ldots, y_N^{(l)})^{\mathsf{T}}$ the N-dimensional data received at time l. Assuming the channel coherence time is at least as long as L sampling periods, we finally denote $\mathbf{Y} = (\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(L)}) \in \mathbb{C}^{N\times L}$ the matrix of the concatenated receive vectors.

Depending on whether the transmitter emits informative signals, we consider the following hypotheses

- $\mathcal{H}_0$. Only background noise is received.
- $\mathcal{H}_1$. Informative signals plus background noise are received.

Therefore, under condition $\mathcal{H}_0$, we have the model,

$$\mathbf{Y} = \sigma \begin{bmatrix} \boldsymbol{\theta}^{(1)} & \cdots & \boldsymbol{\theta}^{(L)} \end{bmatrix} \tag{1}$$

and under condition $\mathcal{H}_1$,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{H} & \sigma\mathbf{I}_N \end{bmatrix} \begin{bmatrix} \mathbf{s}^{(1)} & \cdots & \mathbf{s}^{(L)} \\ \boldsymbol{\theta}^{(1)} & \cdots & \boldsymbol{\theta}^{(L)} \end{bmatrix}. \tag{2}$$

Under this hypothesis, we further denote $\boldsymbol{\Sigma}$ the covariance matrix of $\mathbf{y}^{(1)}$,

$$\boldsymbol{\Sigma} = \mathrm{E}[\mathbf{y}^{(1)}\mathbf{y}^{(1)\mathsf{H}}] = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}_N = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathsf{H}}, \tag{3}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\left(\nu_1 + \sigma^2, \ldots, \nu_N + \sigma^2\right)$, with $\nu_1, \ldots, \nu_N$ the eigenvalues of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$ and $\mathbf{U}$ a certain unitary matrix.

The receiver is entitled to decide whether the sensors are transmitting informative signals or not; this is, the receiver makes a decision over hypothesis $\mathcal{H}_0$ or $\mathcal{H}_1$. The receiver is however considered to have very limited information about the transmission channel and is in particular not necessarily aware of the exact number M of sources and of the signal-to-noise ratio. For this reason, the maximum entropy principle requires that all unknown variables be assigned a probability distribution which is both (i) consistent with the prior information (voluntarily discarding information violates the Bayesian requirements) and (ii) has maximal entropy over the set of densities that validate (i). It is known in particular that the entropy maximizing probability distribution of a random vector, whose population covariance matrix is known, is a multivariate Gaussian distribution with zero mean and variance the known population covariance matrix. If the population covariance matrix is unknown but is known to be of unit trace, then the entropy maximizing distribution is now multivariate independent Gaussian with zero mean and normalized identity covariance matrix. Therefore, if $\mathbf{H}$ is only known to satisfy, as is often the case in the short term, $\frac{1}{N}\mathrm{E}[\mathrm{tr}\mathbf{H}\mathbf{H}^{\mathsf{H}}] = 1$, the maximum entropy principle states that the entries h_{ij} are independent and distributed as $h_{ij} \sim \mathcal{CN}(0, 1/M)$, for all (i,j). For the same reason, both noise $\theta_i^{(l)}$ and signal $s_i^{(l)}$

entries are taken independent Gaussian with zero mean and variance $\mathrm{E}[|\theta_i^{(l)}|^2]=1$, $\mathrm{E}[|s_i^{(l)}|^2]=1$.

The Neyman-Pearson criterion for the receiver to establish whether an informative signal was transmitted is based on the ratio C,

$$C(\mathbf{Y})=\frac{P_{\mathcal{H}_1|\mathbf{Y}}(\mathbf{Y})}{P_{\mathcal{H}_0|\mathbf{Y}}(\mathbf{Y})}. \quad (4)$$

For a given receive space-time matrix $\mathbf{Y}$, if $C(\mathbf{Y})>1$, then the odds are that an informative signal was transmitted, while if $C(\mathbf{Y})<1$, it is more likely that no informative signal was transmitted and therefore only background noise was measured. To ensure a low probability of false alarm (or false positive), i.e. the probability to declare a pure noise sample to carry an informative signal, a certain threshold ξ is generally set such that, when $C(\mathbf{Y})>\xi$, the receiver declares an informative signal was sent, while when $C(\mathbf{Y})<\xi$, the receiver declares that no informative signal was sent. The question of what ratio ξ to be set to ensure a given maximally acceptable false alarm rate will not be treated in the following. We will however provide an explicit expression of (4) for the aforementioned model, and shall compare its performance to that achieved by classical detectors.

Thanks to Bayes' rule, (4) becomes

$$C(\mathbf{Y})=\frac{P_{\mathcal{H}_1}\cdot P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y})}{P_{\mathcal{H}_0}\cdot P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y})} \quad (5)$$

with $P_{\mathcal{H}_i}$ the *a priori* probability for hypothesis $\mathcal{H}_i$ to be true. We suppose that no side information allows the receiver to think $\mathcal{H}_1$ is more or less probable than $\mathcal{H}_0$, and therefore set $P_{\mathcal{H}_1}=P_{\mathcal{H}_0}=\frac{1}{2}$, and then $C(\mathbf{Y})$ reduces to a maximum likelihood criterion.

In the next section, we will derive close-form expressions for $C(\mathbf{Y})$ under the hypotheses that the values of M and the SNR are either perfectly or only partially known at the receiver.

III. Signal Detection

A. Known noise variance and number of signal sources

1) Derivation of $P_{\mathbf{Y}|\mathcal{H}_i}$ in SIMO case: We first analyze the situation when the noise power σ^2 and the number M of signal sources are known to the receiver. We also assume in this first scenario that $M=1$. Since it is a common assumption that the number of available samples at the receiver is larger than the number of sensors themselves, we further consider that $L>N$.[2]

a) Pure noise likelihood $P_{\mathbf{Y}|\mathcal{H}_0}$: In this first scenario, the noise entries $\theta_i^{(l)}$ are Gaussian and independent. The distribution for $\mathbf{Y}$, that can be seen as a random vector with NL entries, is then an NL multivariate uncorrelated complex Gaussian with covariance matrix $\sigma^2\mathbf{I}_{NL}$,

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y})=\frac{1}{(\pi\sigma^2)^{NL}}e^{-\frac{1}{\sigma^2}\mathrm{tr}\mathbf{Y}\mathbf{Y}^{\mathsf{H}}}. \quad (6)$$

[2]the more marginal $L<N$ case will not be treated here for lack of space.

By denoting $\mathbf{x}=(x_1,\ldots,x_N)^{\mathsf{T}}$ the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$, (6) only depends on $\sum_{i=1}^N x_i$,

$$P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y})=\frac{1}{(\pi\sigma^2)^{NL}}e^{-\frac{1}{\sigma^2}\sum_{i=1}^N x_i}. \quad (7)$$

b) Informative signal likelihood $P_{\mathbf{Y}|\mathcal{H}_1}$: In the information plus noise scenario $\mathcal{H}_1$, the problem is more involved. The entries of the channel matrix $\mathbf{H}$ were previously modelled as jointly uncorrelated Gaussian, with $\mathrm{E}[|h_{ij}|^2]=1/M$. Therefore, since $M=1$, $\mathbf{H}\in\mathbb{C}^{N\times 1}$ and $\mathbf{\Sigma}=\mathbf{H}\mathbf{H}^{\mathsf{H}}+\sigma^2\mathbf{I}_N$ has $N-1$ eigenvalues equal to σ^2 and another distinct eigenvalue $\lambda_1=\nu_1+\sigma^2=(\sum_{i=1}^N|h_{i1}|^2)+\sigma^2$. The density of $\lambda_1-\sigma^2$ is a complex χ^2_N distribution (which is, up to a scaling factor 2, equivalent to a real χ^2_{2N} distribution). Hence the eigenvalue distribution of $\mathbf{\Sigma}$, defined on $\mathbb{R}^{+N}$,

$$P_{\mathbf{\Lambda}}(\mathbf{\Lambda})=\frac{1}{N}(\lambda_1-\sigma^2)_+^{N-1}\frac{e^{-(\lambda_1-\sigma^2)}}{(N-1)!}\prod_{i=2}^N\delta(\lambda_i-\sigma^2). \quad (8)$$

Given model (2), $\mathbf{Y}$ is distributed as correlated Gaussian,

$$P_{\mathbf{Y}|\mathbf{\Sigma},I_1}(\mathbf{Y},\mathbf{\Sigma})=\frac{1}{\pi^{LN}\det(\mathbf{\Lambda})^L}e^{-\mathrm{tr}(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^{\mathsf{H}})}. \quad (9)$$

where I_k denotes the prior information "$\mathcal{H}_1$ and $M=k$".

Since the channel $\mathbf{H}$ is unknown, we need to integrate out all possible channels of the model (2) over the probability space of $N\times M$ matrices with Gaussian i.i.d. distribution. This is equivalent to integrating out all possible covariance matrices $\mathbf{\Sigma}$ over the space of such positive definite Hermitian matrices

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y})=\int_{\mathbf{\Sigma}}P_{\mathbf{Y}|\mathbf{\Sigma}\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma})P_{\mathbf{\Sigma}}(\mathbf{\Sigma})d\mathbf{\Sigma}. \quad (10)$$

Eventually, the Neyman-Pearson decision ratio (4) for the single-input multiple-output channel takes an explicit expression, given as follows,

Theorem 1: The Neyman-Pearson decision ratio $C_{\mathbf{Y}|I_1}(\mathbf{Y})$ for the presence of an informative signal under prior information I_1, i.e. the receiver knows (i) $M=1$ signal source, (ii) the SNR σ^{-2}, reads

$$C(\mathbf{Y})=\frac{1}{N}\sum_{l=1}^N\frac{\sigma^{2(N+L-1)}e^{\sigma^2+\frac{x_l}{\sigma^2}}}{\prod_{\substack{i=1\\i\neq l}}^N(x_l-x_i)}J_{N-L-1}(\sigma^2,x_l) \quad (11)$$

with $x_1,\ldots,x_N$ the empirical eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and where

$$J_k(x,y)=\int_x^{+\infty}t^k e^{-t-\frac{y}{t}}dt. \quad (12)$$

The main steps of the proof of Theorem 1 are provided in Appendix A. Among the interesting features of (11), note that the Neyman-Pearson test does only depend on the eigenvalues of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$. This suggests that the eigenvectors of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ do not provide any information regarding the presence of an informative signal. The essential reason is that both under $\mathcal{H}_0$ and $\mathcal{H}_1$, the eigenvectors of $\mathbf{Y}$ are isotropically distributed

on the unit N-dimensional complex sphere; as such, a given realization of the eigenvectors of $\mathbf{Y}$ does not carry any relevant information to the hypothesis test. The Gaussian assumption for $\mathbf{H}$ brought by the maximum entropy principle is therefore essential here. Note however that (11) is not reduced to a function of the sum $\sum_i x_i$ of the eigenvalues, as for the classical energy detector.

On the negative side, the integral $J_k(x,y)$ does not take a closed-form expression, but for $x = 0$; see page 561 of [15]. This is rather inconvenient for practical purposes, since $J_k(x,y)$ must be evaluated every time or a rather large lookup table must be kept in memory. It is also difficult to get any insight on the performance of such a detector for different values of σ^2, N and L, let alone marginalizing over σ^2 when the latter is not perfectly known. This last point is discussed in Section III-B1. We now turn to the more general case when $M \geq 1$.

2) Derivation of $P_{\mathbf{Y}|\mathcal{H}_i}$ in MIMO case: In the MIMO configuration, $P_{\mathbf{Y}|\mathcal{H}_0}$ remains unchanged and Equation (6) still holds. For the subsequent derivations, due to space limitation, we only treat the situation when $M \leq N$ but the case $M > N$ is a trivial extension.

In this scenario, $\mathbf{H} \in \mathbb{C}^{N\times M}$ is a random matrix with independent and identically distributed central Gaussian entries. The variance of every row is $\mathrm{E}[\sum_{j=1}^{M} |h_{ij}|^2] = 1$. Therefore $M\mathbf{H}\mathbf{H}^{\mathsf{H}}$ is distributed as a null Wishart matrix. Hence, observing that $\mathbf{\Sigma} - \sigma^2\mathbf{I}_N$ is the diagonal matrix of the eigenvalues of $\mathbf{H}\mathbf{H}^{\mathsf{H}}$,

$$\mathbf{\Sigma} = \mathbf{U}\cdot\mathrm{diag}(\nu_1+\sigma^2,\ldots,\nu_M+\sigma^2,\sigma^2,\ldots,\sigma^2)\cdot\mathbf{U}^{\mathsf{H}} \quad (13)$$

for some unitary matrix $\mathbf{U} \in \mathbb{C}^{N\times N}$, the eigenvalue distribution density of $\mathbf{\Lambda}$ can be derived [16]

$$P_{\mathbf{\Lambda}}(\mathbf{\Lambda}) = \frac{(N-M)!M^{MN}}{N!}\prod_{i=1}^{M}\frac{e^{M\sum_{i=1}^{M}(\lambda_i-\sigma^2)}(\lambda_i-\sigma^2)_+^{N-M}}{(M-i)!(N-i)!} \times \prod_{i<j}^{M}(\lambda_i-\lambda_j)^2\prod_{i>M}^{N}\delta(\lambda_i-\sigma^2). \quad (14)$$

From Equations (13) and (14) above, the MIMO equivalent result to Theorem 1 unfolds as follows,

Theorem 2: The Neyman-Pearson decision ratio $C_{\mathbf{Y}|I_M}(\mathbf{Y})$ for the presence of informative signal under prior information I_M, i.e. when the receiver is aware of (i) $M \leq N$ signal sources, (ii) the noise power σ^2, reads

$$C(\mathbf{Y}) = \frac{\sigma^{2M(N+L-M)}(N-M)!e^{M^2\sigma^2}}{N!M^{(M-1-2L)M/2}\prod_{j=1}^{M-1}j!}\sum_{\mathbf{a}\subset[1,N]}\frac{e^{\frac{\sum_{i=1}^{M}x_{a_i}}{\sigma^2}}}{\prod_{a_i}\prod_{\substack{j\neq a_1\\ \vdots\\ j\neq a_i}}(x_{a_i}-x_j)} \times \sum_{\mathbf{b}\in\mathcal{P}(M)}(-1)^{\mathrm{sgn}(\mathbf{b})+M}\prod_{l=1}^{M}J_{N-L-2+b_l}(M\sigma^2,Mx_{a_l}) \quad (15)$$

with $\mathcal{P}(M)$ the ensemble of permutations of $\{1,\ldots,M\}$, $\mathbf{b} = (b_1,\ldots,b_M)$ and $\mathrm{sgn}(\mathbf{b})$ the signature of the permutation $\mathbf{b}$.

The main steps of the proof of Theorem 2 are provided in Appendix B. Note again that $C_{\mathbf{Y}|I_M}(\mathbf{Y})$ is a function of the empirical eigenvalues $x_1,\ldots,x_N$ of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ only. In the following, we extend the current signal detector to the situations where M and σ^2 are not *a priori* known at the receiver.

B. Number of sources and/or noise variance unknown

1) Unknown noise variance: Efficient signal sensing when the noise level is unknown is highly desirable. Indeed, if the noise level were exactly known, some prior noise detection mechanism would be required. The difficulty here is handily avoided thanks to *ad-hoc* methods that are asymptotically independent of the noise level [7], [8], [9]. Instead, we shall consider here the prior information about the noise power. It might happen though that the receiver has no knowledge whatsoever on the value of the noise power. When such a situation arises, the unknown parameter must be assigned an *uninformative prior* [17]. Assigning uninformative priors of variables defined in a continuum is however, still to this day, a controverted issue of the maximum entropy theory. The classical uninformative priors considered in the literature are (i) the uniform prior, i.e. every two positive values for the noise power are equi-probable, which experiences problems of scaling invariance thoroughly discussed in [12], and (ii) Jeffreys prior [17], i.e. the prior distribution for σ^2 takes the form $\sigma^{-\alpha}$ for any deterministic choice of positive α, which is invariant under scaling but is not fully attractive. Note that Jeffreys uninformative priors, in spite of their inherent bias for some hypotheses, are sometimes used in other areas of signal processing, see e.g. [18] for the estimation of noise covariance matrices.

To alleviate somewhat the problem of uninformative priors, we shall consider that the noise power σ^2 is known at least to be bounded both from under and from above, i.e. $\sigma^2 \in [\sigma_-^2,\ \sigma_+^2]$, and we shall consider the "desirable" assumption of uniform prior for σ^2 over the set $[\sigma_-^2,\ \sigma_+^2]$.

This therefore leads to

$$P_{\mathbf{Y}|I'_M}(\mathbf{Y}) = \frac{1}{\sigma_+^2-\sigma_-^2}\int_{\sigma_-^2}^{\sigma_+^2}P_{\mathbf{Y}|\sigma^2,I'_M}(\mathbf{Y},\sigma^2)d\sigma^2 \quad (16)$$

with I'_k the information "$\mathcal{H}_1$, $M = k$ and $\sigma^2 \in [\sigma_-^2,\sigma_+^2]$". The latter results in the updated decisions of the form,

$$C(\mathbf{Y}) = \frac{\int_{\sigma_-^2}^{\sigma_+^2}P_{\mathbf{Y}|\sigma^2,I'_M}(\mathbf{Y},\sigma^2)d\sigma^2}{\int_{\sigma_-^2}^{\sigma_+^2}P_{\mathbf{Y}|\sigma^2,\mathcal{H}_0}(\mathbf{Y},\sigma^2)d\sigma^2}. \quad (17)$$

The computational difficulty raised by the integrals $J_k(x,y)$ does not allow for any satisfying closed-form formulas for (16) so that only numerical integrations can be performed at this point.

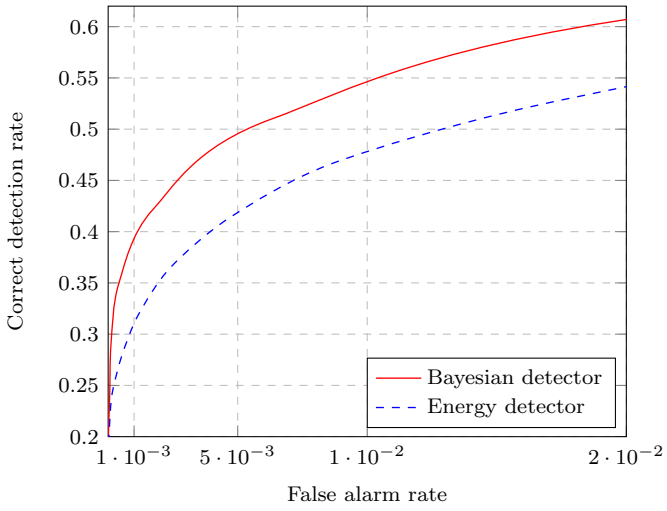

Fig. 1. ROC curve for SIMO transmission, $M = 1$, $N = 4$, $L = 8$, SNR $= -3$ dB, FAR range of practical interest.

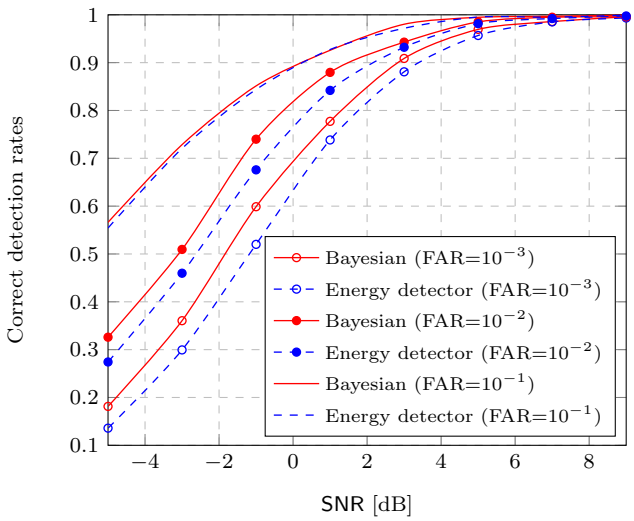

Fig. 2. Correct detection rates under FAR constraints for different SNR levels, $M = 1$, $N = 4$, $L = 8$

C. Unknown number of sources M

In practical cases, the number of transmitting sources is only known to be finite and discrete. If only an upper bound value $M_{\max}$ on M is known, a uniform prior is assigned to positive M (which is here fully compliant with the maximum entropy principle for discrete random processes). The probability distribution of $\mathbf{Y}$ under hypothesis I_0 ="σ^2 known, $M \geq 1$ unknown", reads

$$P_{\mathbf{Y}|I_0}(\mathbf{Y}) = \frac{1}{M_{\max}} \sum_{i=1}^{M_{\max}} P_{\mathbf{Y}|\text{“}M=i'',I_0}(\mathbf{Y}), \qquad (18)$$

which does not meet any computational difficulty.

Assuming again equal probability for hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, this leads to the decision ratio,

$$C(\mathbf{Y}) = \frac{1}{M_{\max}} \frac{\sum_{i=1}^{M_{\max}} P_{\mathbf{Y}|\text{“}M=i'',I_0}(\mathbf{Y})}{P_{\mathbf{Y}|\mathcal{H}_0}(\mathbf{Y})}. \qquad (19)$$

Note now that it is possible to make a decision test on the number of sources itself in a rather straightforward extension of the previous formula. Indeed, given a space-time matrix realization $\mathbf{Y}$, the probability for the number of transmit antennas to be i is, from Bayes' rule,

$$P_{\text{“}M=i''|\mathbf{Y}}(\mathbf{Y}) = \frac{P_{\mathbf{Y}|\text{“}M=i''}(\mathbf{Y}) P_{\text{“}M=i''}}{\sum_{j=0}^{M_{\max}} P_{\mathbf{Y}|\text{“}M=j''}(\mathbf{Y}) P_{\text{“}M=j''}}, \qquad (20)$$

where all the quantities of interest here were derived in the previous sections and appendixes. The multiple hypothesis test on M is then based on a comparison of the *odds* $O(\text{“}M = i'')$ for the events "$M = i''$, for all $i \in \{0, \ldots, M_{\max}\}$. The odds for the event "$M = i''$ is defined as

$$O(\text{“}M = i'') = \frac{P_{\text{“}M=i''|\mathbf{Y}}(\mathbf{Y})}{\sum_{\substack{j=0 \\ j\neq i}}^{M_{\max}} P_{\text{“}M=j''|\mathbf{Y}}(\mathbf{Y})}. \qquad (21)$$

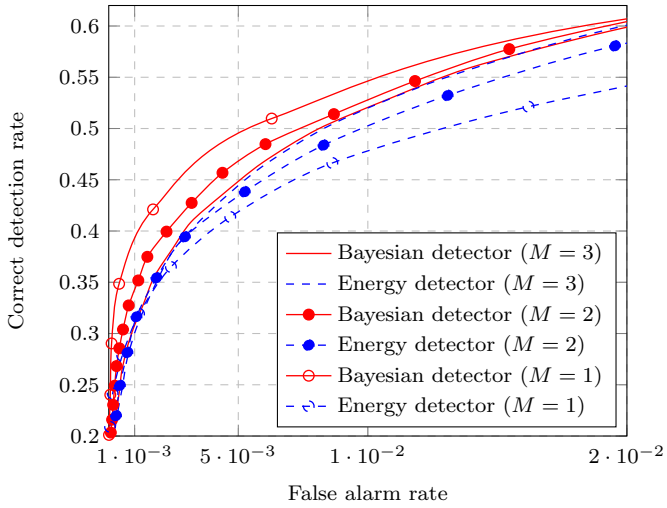

Fig. 3. ROC curve for MIMO transmission, $N = 4$, $L = 8$, SNR $= -3$ dB, varying M. FAR range of practical interest.

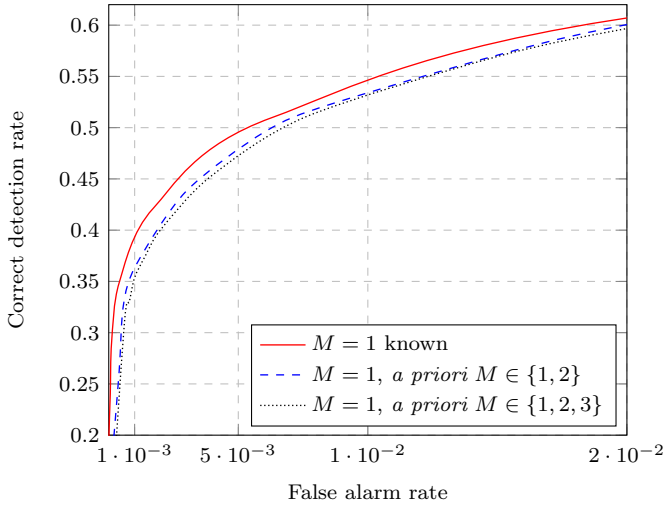

Fig. 4. ROC curve for unknown number of sources, when $M_{\max} = 1, 2, 3$, $M = 1$, $N = 4$, $L = 8$, SNR $= -3$ dB. FAR range of practical interest.

IV. Simulation and Results

In the following, we present results obtained for the aforementioned SIMO and MIMO scenarios, using Theorems 1 and 2 respectively. In the simulations, the hypotheses concerning incoming data, channel aspect and noise figure are those presented in the model of Section II, i.e. the channel, signal and noise matrix entries are i.i.d. Gaussian with zero mean and respective variances $1/M$, 1 and 1. Therefore, under this setting, the Neyman-Pearson decision criteria derived in Theorems 1 and 2 are optimal. If the noise variance σ^2 is *a priori* known at the sensing device, the results are compared against the classical energy detector [3]. If the noise variance is *a priori* unknown at the receiver, our results will be compared against both the conditioning number method of [7]-[8] and the GLRT approach of [9].

In our first example, we consider a SIMO channel with $N = 4$ antennas at the receiver, $L = 8$ sampling instants and a signal to noise ratio $\mathsf{SNR} = -3$ dB. To compare the performances of the different methods, we provide their respective operating characteristic curves (ROC), i.e. for all discussed methods, we compare simulation results in terms of rates of false positives (or false alarm rates (FAR)) against rates of true positives (correct detection rate (CDR)). Figure 1 presents the respective ROC curves for the Bayesian estimator of Theorem 1 against the classical energy detector [3], obtained from $100,000$ Monte Carlo realizations of the model. We remind that the energy detector decision criterion is based on the test

$$C_{\text{energy}}(\mathbf{Y}) = \frac{1}{L\sigma^2}\frac{1}{N}\sum_{i=1}^{N} x_i. \tag{22}$$

Figure 1 suggests that as much as a 10% increase in detection ability is obtained by the Bayesian detector and that this gain increases along with smaller FAR. This tendency is confirmed by Figure 2 in which the performance of the Bayesian signal detector with respect to the energy detector for different FAR is presented against the SNR.

In Figure 3, we take $N = 4$, $L = 8$ and $\mathsf{SNR} = -3$ dB as before but consider now $M = 1$ to $M = 3$ signal sources; we then use Theorem 2 here. In this scenario the energy detector closes in the gap with the Bayesian detector, with a surprising behaviour: the energy detector performs better for $M \in \{2, 3\}$ than for $M = 1$, while the opposite tendency is observed for the Bayesian detector. A possible interpretation comes as follows: for the energy detector, due to MIMO channel hardening effect [19], the variance of the receive signal is smaller with increasing M, reducing then the false positive rates; on the contrary, the performance decay of the Bayesian estimator may be attributed to the increased number of degrees of freedom in the MIMO channel $\mathbf{H}$ for larger M, which makes it more difficult to match $P_{\mathcal{H}_1|\mathbf{Y}}(\mathbf{Y})$ against $P_{\mathcal{H}_0|\mathbf{Y}}(\mathbf{Y})$ from single-shot simulations of $\mathbf{Y}$ matrices. In Figure 4, we assume that the number $M = 1$ of signal sources is *a priori* known only to be in the set $M \in \{1, \ldots, M_{\max}\}$, for $M_{\max} \in \{1, 2, 3\}$. We observe that, while perfect knowledge of M leads to some non-negligible gain with respect to the cases when M is unknown, increasing $M_{\max}$ does not really impact performance. This might be interpreted by the fact that the hypothesis $M = 1$ is naturally selected against hypotheses of larger M. The same remark holds true for the case when the SNR is *a priori* unknown, discussed hereafter.

Consider the scenario when the noise variance σ^2 is *a priori* known to belong to the interval $[\sigma^2_-, \sigma^2_+]$. This is presented in Figure 5 which demonstrates the effect of an inaccurate knowledge of the noise power in terms of CDR and FAR. In this simulation, $M = 1$, $N = 4$, $L = 8$ and $\mathsf{SNR} = 0$ dB.[3] Comparison is made between the cases of exact SNR knowledge, short SNR range $[\sigma^2_-, \sigma^2_+] = [-2.5, 2.5]$ dB, large SNR range $[\sigma^2_-, \sigma^2_+] = [-5, 5]$ dB and very large range $[\sigma^2_-, \sigma^2_+] = [-9, 9]$ dB. Observe that the short SNR range provides already a strong performance decay compared to the ideal scenario, which is particularly noticeable in terms of CDR performance at low FAR. Larger SNR ranges are then only slightly worse than the short range scenario and seem to converge to a 'worst-case limit'; this can be interpreted by the fact that the additional hypotheses, i.e. very strong or very little noise power, are automatically discarded as the values of $P_{\mathbf{Y}|\sigma^2, I_1'}(\mathbf{Y}, \sigma^2)$ and $P_{\mathbf{Y}|\sigma^2, \mathcal{H}_0}(\mathbf{Y}, \sigma^2)$ become negligible for unrealistic values of σ^2. Additional simulations for larger SNR ranges were carried out that visually confirm that the FAR and CDR plots are identical here as long as $\sigma^2_- \leq -5$ dB and $\sigma^2_+ \geq 5$ dB. Therefore, simulations suggest that the proposed Bayesian signal detector is able to cope even with totally unknown SNR, which is obviously not the case of the classical energy detector that relies on an SNR-dependent decision threshold.

We now consider totally unknown SNR value, which we demonstrated above to be equivalent as taking the *a priori* information $\sigma^2_- \leq \sigma^2 \leq \sigma^2_+$, with $\sigma^2_- \leq -5$ dB and $\sigma^2_+ \geq 5$ dB. The performance, already provided in Figure 5, is now compared in Figure 6 against the conditioning number method, the GLRT method and the Bayesian approach with Jeffreys prior $P_{\sigma^2}(\sigma^2) = \sigma^{-2\alpha}$ and $\alpha = 1$. We recall briefly those methods. The conditioning number method is an heuristic method, independent of the SNR, which considers the decision ratio

$$C_{\text{cond}}(\mathbf{Y}) = \frac{\max_i(x_i)}{\min_i(x_i)} \tag{23}$$

between the largest and the smallest eigenvalue of the empirical $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ matrix. The GLRT method considers instead the ratio [9]

$$C_{\text{GLRT}}(\mathbf{Y}) = \frac{\sup_{\mathbf{H},\sigma^2} P_{\mathcal{H}_1|\mathbf{Y},\mathbf{H},\sigma^2}(\mathbf{Y})}{\sup_{\sigma^2} P_{\mathcal{H}_0|\mathbf{Y},\sigma^2}(\mathbf{Y})} \tag{24}$$

$$= \left(\left(1 - \frac{1}{N}\right)^{N-1} \frac{\max_i(x_i)}{\frac{1}{N}\sum_{i=1}^N x_i} \left(1 - \frac{\max_i(x_i)}{\sum_{i=1}^N x_i}\right)^{N-1} \right)^{-L}. \tag{25}$$

We observe first that the intuitive, though naive, conditioning number approach is greatly outperformed by the GLRT method, by more than 30% of correction detection performance for low FAR. In turn the Bayesian detector appears to outperform the GLRT method, but only for large FAR.

[3] the specific choices of signal-to-noise ratios were made such that the ROC curves do not lead to trivial all zeros or all ones CDR at low FAR.

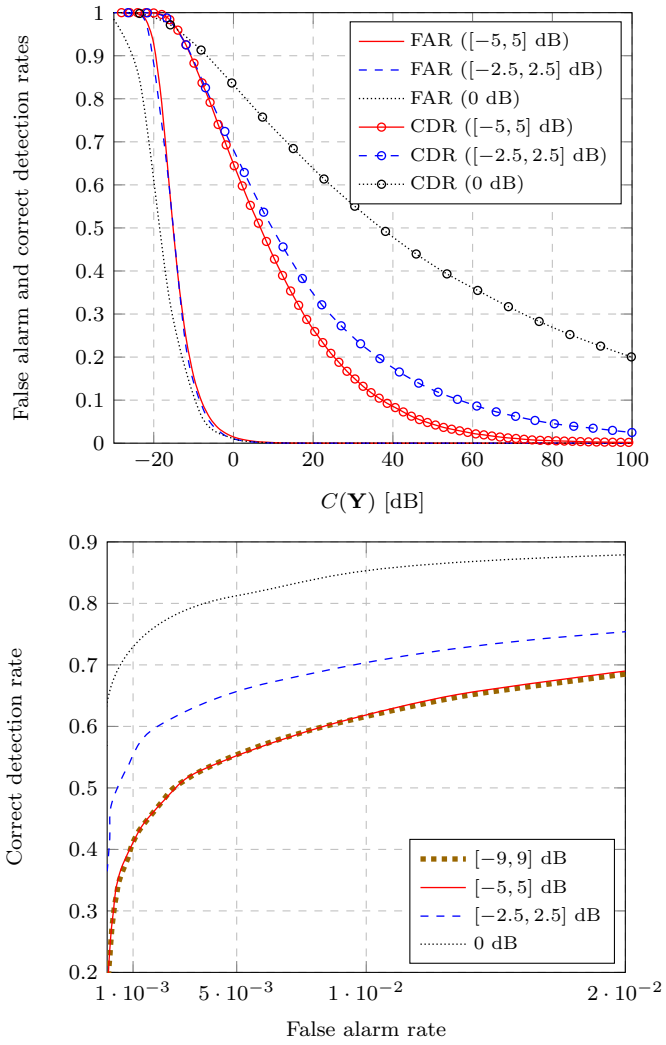

Fig. 5. FAR, CDR and associated ROC curve for different *a priori* information on the SNR: exact SNR knowledge (0 dB), short range SNR ($[-2.5, 2.5]$ dB) and large range SNR ($[-5, 5]$ dB) SNR, $M = 1$, $N = 4$, $L = 8$, true $\mathsf{SNR} = 0$ dB

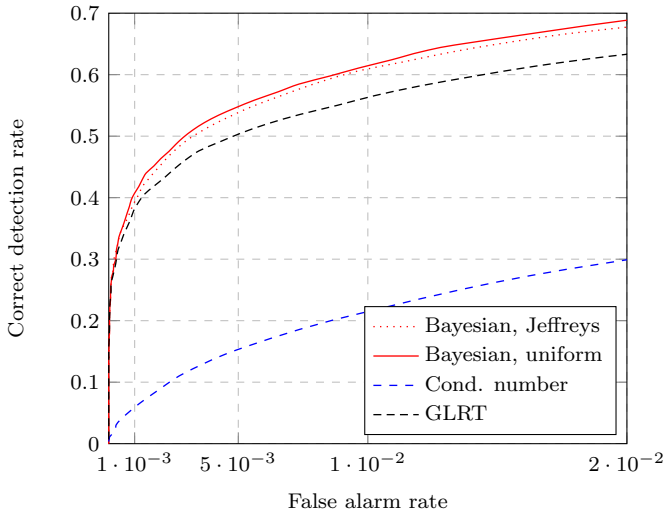

Fig. 6. ROC curve for *a priori* unknown σ^2 of the Bayesian method, conditioning number method and GLRT method, $M = 1$, $N = 4$, $L = 8$, $\mathsf{SNR} = 0$ dB. For the Bayesian method, both uniform and Jeffreys prior, with exponent $\alpha = 1$, are provided.

This seems to suggest that the computationally simple GLRT approach seems to be extremely adequate and close-to-optimal for low FAR constraints.

V. Discussion

In the previous framework, we relied on the maximum entropy principle in order to derive unique *a priori* distributions for the various unknown system parameters. The provided Bayesian solutions, derived from the channel state of knowledge available at the receiver, were claimed consistent in the proposed probability framework. This framework is in particular extensible to whatever prior knowledge the receiver might have on the transmission environment. However, some limitations can be raised. First, as stated in III-B1, uninformative priors modeling is still an incomplete and controversial theory, for which no definite answer is available to this day. When such a prior information is to be treated, the proposed signal sensing framework is not capable of singling out a proper maximum entropy model; this constitutes a major coherency issue of our source detection framework. Also, the mathematical tools to derive maximum entropy distributions, e.g. Lagrangian multipliers, only cope with statistical prior knowledge, such as the moments of the underlying density functions, and are rarely able to treat deterministic knowledge.

Note however that the advances in the field of random matrix theory provide new answers to problems of high dimensionality, even for finite N, L values. Those problems, such as the present maximum-likelihood multi-sensor signal sensing, are often considered intractable and suffer in practice from the so-called *curse of dimensionality*. The current study relies nonetheless on the important property that the transmission channel $\mathbf{H}$ is modelled as i.i.d. Gaussian; if $\mathbf{H}$ were more structured, it would have been more difficult to obtain an integral expression similar to (45) and the final results, be they derivable, would implicate not only the eigenvalues but also the eigenvectors of $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$. Recent mathematical advances [20], [21] provide some hope though that integration over Gaussian matrices with non trivial correlation models may actually be feasible.

More importantly, the proposed Bayesian framework allows one to answer a wider scope of problems than just the present multi-source signal sensing. In particular, we discussed in the introduction the somewhat different problem of counting the number of transmitting sources. Lately, the problem of identifying the transmit power of the individual sources has also received a lot of interest in the random matrix community, see e.g. [22], [23] and [24].

VI. Conclusion

In this work, we introduced a general Bayesian framework for multi-source detection from an array of sensors. This framework is based on a consistent treatment of the information about the transmission channel available at the sensing device. The resulting Bayesian estimators were proven optimal

with respect to the Neyman-Pearson detection criterion. The performance of these novel multi-source detectors is compared in simulations against (i) the classical energy detector when the signal-to-noise ratio is perfectly known, (ii) recent techniques from the field of random matrix theory when the signal-to-noise ratio is unknown. ROC curve comparisons suggested that the energy detector performs close-to-optimally, especially when the number of transmit sources increases. We also observed that the recently proposed GLRT detector when the noise power is *a priori* unknown performs also close-to-optimally, especially when the tolerated rates of false alarms is low. We finally showed that this general framework can be easily extended to problems such as number of sources identification or source power inference.

APPENDIX A
PROOF OF THEOREM 1

We start by noticing that $\mathbf{H}$ is Gaussian and therefore the joint density of its entries is invariant by left and right unitary products. As a consequence, the distribution of the matrix $\mathbf{\Sigma} = \mathbf{H}\mathbf{H}^{\mathsf{H}} + \sigma^2\mathbf{I}$ is unitarily invariant, i.e. for any unitary matrix $\mathbf{V}$, $\mathbf{V}\mathbf{\Sigma}\mathbf{V}^{\mathsf{H}}$ has the same joint density as $\mathbf{\Sigma}$. The latter density does not as a consequence depend on $\mathbf{U}$ in its singular value decomposition (3). This allows us to write, similarly as in [27],

$$P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y}) = \int_{\mathbf{\Sigma}} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma}) P_{\mathbf{\Sigma}}(\mathbf{\Sigma}) d\mathbf{\Sigma} \tag{26}$$

$$= \int_{\mathcal{U}(N)\times\mathbb{R}^+} P_{\mathbf{Y}|\mathbf{\Sigma},\mathcal{H}_1}(\mathbf{Y},\mathbf{\Sigma}) P_{\lambda_1}(\lambda_1) d\mathbf{U} d\lambda_1 . \tag{27}$$

Equation (27) leads then to

$$P_{\mathbf{Y}|I_1}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times\mathbb{R}^{+N}} \frac{e^{-\mathrm{tr}\left(\mathbf{Y}\mathbf{Y}^{\mathsf{H}}\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^{\mathsf{H}}\right)}}{\pi^{NL}\det(\mathbf{\Lambda})^L} (\lambda_1 - \sigma^2)_+^{N-1} \times \frac{e^{-(\lambda_1-\sigma^2)}}{N!} \prod_{i=2}^{N} \delta(\lambda_i - \sigma^2) d\mathbf{U} d\lambda_1 \ldots d\lambda_N . \tag{28}$$

To go further, we utilize the Harish-Chandra identity [26]

$$\int_{\mathcal{U}(N)} e^{\kappa \mathrm{tr}(\mathbf{A}\mathbf{U}\mathbf{B}\mathbf{U}^{\mathsf{H}})} d\mathbf{U} = \left(\prod_{n=1}^{N-1} n!\right) \kappa^{N(N-1)/2} \frac{\det\left(\left\{e^{-A_i B_j}\right\}_{\substack{1\le i\le N \\ 1\le j\le N}}\right)}{\Delta(\mathbf{A})\Delta(\mathbf{B})} \tag{29}$$

in which, for a matrix $\mathbf{X}$ with eigenvalues $x_1, \ldots, x_N$, $\Delta(\mathbf{X}) = \prod_{i>j}(x_i - x_j)$ is the Vandermonde determinant.

In order to avoid divisions by zero when applying (29), we move $\lambda_2, \ldots, \lambda_N$ away from σ^2, in such a way that $P_{\mathbf{Y}|I_1}(\mathbf{Y})$ writes

$$P_{\mathbf{Y}|I_1}(\mathbf{Y}) = \lim_{\lambda_2,\ldots,\lambda_N\to\sigma^2} \frac{e^{\sigma^2}(-1)^{\frac{N(N-1)}{2}} \prod_{j=1}^{N-1} j!}{\pi^{LN}\sigma^{2L(N-1)}N!} \int_{\sigma^2}^{+\infty} \frac{1}{\lambda_1^L}(\lambda_1 - \sigma^2)^{N-1} e^{-\lambda_1} \frac{\det\left(\left\{e^{-\frac{x_i}{\lambda_j}}\right\}_{i,j}\right)}{\Delta(\mathbf{X})\Delta(\mathbf{\Lambda}^{-1})} d\lambda_1 \tag{30}$$

$$= \lim_{\lambda_2,\ldots,\lambda_N\to\sigma^2} \frac{e^{\sigma^2}\prod_{j=1}^{N-1} j!}{\pi^{LN}\sigma^{2L(N-1)}N!} \int_{\sigma^2}^{+\infty} \frac{(\lambda_1 - \sigma^2)^{N-1}}{\lambda_1^L} e^{-\lambda_1} \det\left(\mathbf{\Lambda}^{N-1}\right) \frac{\det\left(\left\{e^{-\frac{x_i}{\lambda_j}}\right\}_{i,j}\right)}{\Delta(\mathbf{X})\Delta(\mathbf{\Lambda})} d\lambda_1 \tag{31}$$

$$= \lim_{\lambda_2,\ldots,\lambda_N\to\sigma^2} \frac{e^{\sigma^2}\sigma^{2(N-1)(N-L-1)}\prod_{j=1}^{N-1} j!}{\pi^{LN}N!} \int_{\sigma^2}^{+\infty} {\lambda_1}^{N-L-1}(\lambda_1 - \sigma^2)^{N-1} e^{-\lambda_1} \frac{\det\left(\left\{e^{-\frac{x_i}{\lambda_j}}\right\}_{i,j}\right)}{\Delta(\mathbf{X})\Delta(\mathbf{\Lambda})} d\lambda_1 \tag{32}$$

in which $\mathbf{X}$ and $x_1, \ldots, x_N$ respectively correspond to $\mathbf{Y}\mathbf{Y}^{\mathsf{H}}$ and its eigenvalues. The equality (31) comes from the fact that $\Delta(\mathbf{\Lambda}^{-1}) = (-1)^{N(N+3)/2}\frac{\Delta(\mathbf{\Lambda})}{\det(\mathbf{\Lambda})^{N-1}}$.

By denoting $\mathbf{y} = (y_1, \ldots, y_{N-1}, y_N) = (\lambda_2, \ldots, \lambda_N, \lambda_1)$ and the functions $f(x_i, y_j) = e^{-\frac{x_i}{y_j}}$ and $f_i(y_j) = f(x_i, y_j)$, we can perform a similar derivation as in [25] to obtain

$$\lim_{\lambda_2,\ldots,\lambda_N\to\sigma^2} \frac{\det\left(\left\{e^{-\frac{x_i}{\lambda_j}}\right\}_{i,j}\right)}{\Delta(\mathbf{X})\Delta(\mathbf{\Lambda})} \tag{33}$$

$$= \lim_{\substack{y_1,\ldots,y_{N-1}\to\sigma^2 \\ y_N\to\lambda_1}} (-1)^{N-1} \frac{\det\left(\{f_i(x_j)\}_{i,j}\right)}{\Delta(\mathbf{X})\Delta(\mathbf{\Lambda})} \tag{34}$$

$$= (-1)^{N-1} \frac{\det\left[f_i(\sigma^2),\ f_i'(\sigma^2), \ldots,\ f^{(N-2)}(\sigma^2),\ f_i(\lambda_1)\right]}{\prod_{i<j}(x_i - x_j)(\lambda_1 - \sigma^2)^{N-1}\prod_{j=1}^{N-2} j!}. \tag{35}$$

The change of variables led to a switch of one column and explains the $(-1)^{N-1}$ factor when computing the resulting determinant. The partial derivatives of f along the second variable is

$$\frac{\partial f}{\partial y^k}(a,b) = \sum_{m=1}^{k} \frac{(-1)^{k+m}}{b^{m+k}} C_k^m \frac{(k-1)!}{(m-1)!} a^m e^{-\frac{a}{b}} \tag{36}$$

$$\triangleq \kappa_k(a,b) e^{-\frac{a}{b}}. \tag{37}$$

Back to the full expression of $P_{\mathbf{Y}|\mathcal{H}_1}(\mathbf{Y})$, we then have

$$
\begin{aligned}
&P_{\mathbf{Y}|I_1}(\mathbf{Y}) \\
&= \frac{e^{\sigma^2}\sigma^{2(N-1)(N-L-1)}}{N\pi^{LN}} \int_{\sigma^2}^{+\infty} (-1)^{N-1}\lambda_1^{N-L-1}e^{-\lambda_1} \\
&\times \frac{\det\left[f_i(\sigma^2),\ f_i'(\sigma^2),\ldots,\ f^{(N-2)}(\sigma^2),\ f_i(\lambda_1)\right]}{\prod_{i<j}(x_i-x_j)} d\lambda_1 \quad (38) \\
&= \frac{e^{\sigma^2}\sigma^{2(N-1)(N-L-1)}}{N\pi^{LN}\prod_{i<j}(x_i-x_j)} \int_{\sigma^2}^{+\infty} d\lambda_1 (-1)^{N-1}\lambda_1^{N-L-1}e^{-\lambda_1} \\
&\times \det\left[\begin{array}{c:c:c} e^{-\frac{x_1}{\sigma^2}} & \left(\kappa_j(x_i,\sigma^2)e^{-\frac{x_i}{\sigma^2}}\right)_{\substack{1\le i\le N\\ 1\le j\le N-2}} & e^{-\frac{x_1}{\lambda_1}} \\ \vdots & & \vdots \\ e^{-\frac{x_N}{\sigma^2}} & & e^{-\frac{x_N}{\lambda_1}} \end{array}\right]. \quad (39)
\end{aligned}
$$

Before going further, we need the following result,

Lemma 1: Given a family $\{a_1,\ldots,a_N\} \in \mathbb{R}^N$, $N \ge 2$, and $b \in \mathbb{R}^*$, we have

$$
\det\left[\begin{array}{c:c} 1 & \\ \vdots & (\kappa_j(a_i,b))_{\substack{1\le i\le N\\ 1\le j\le N-1}} \\ 1 & \end{array}\right] = \frac{1}{b^{N(N-1)}}\prod_{i<j}(a_j-a_i). \quad (40)
$$

This identity follows from the observation that column k of the matrix above is a polynomial of order k. Since summations of linear combinations of the columns do not affect the determinant, each polynomial can be replaced by the monomial of higher order, i.e. $b^{-2(k-1)}a_i^k$. Extracting the product $1\cdot b^{-2}\cdots b^{-2(N-1)} = b^{-(N-1)N}$ from the determinant, what remains is the determinant of a Vandermonde matrix based on the vector $a_1,\ldots,a_N$.

By factorizing every row of the matrix by $e^{-\frac{x_i}{\sigma^2}}$ and developing the determinant on the last column, one obtains

$$
\begin{aligned}
&P_{\mathbf{Y}|I_1}(\mathbf{Y}) \\
&= \frac{e^{\sigma^2}\sigma^{2(N-1)(N-L-1)}}{N\pi^{LN}\prod_{i<j}(x_i-x_j)} \int_{\sigma^2}^{+\infty} \lambda_1^{N-L-1}e^{-\lambda_1}e^{-\frac{\sum_{i=1}^N x_i}{\sigma^2}} \\
&\times (-1)^{N-1}\sum_{l=1}^{N}(-1)^{N+l}\frac{e^{-x_l\left(\frac{1}{\lambda_1}-\frac{1}{\sigma^2}\right)}}{\sigma^{2(N-1)(N-2)}}\prod_{\substack{i<j\\ i\ne l\\ j\ne l}}(x_i-x_j)d\lambda_1 \quad (41) \\
&= \frac{e^{\sigma^2-\frac{1}{\sigma^2}\sum_{i=1}^N x_i}}{N\pi^{LN}\sigma^{2(N-1)(L-1)}}\sum_{l=1}^{N}(-1)^{l-1}\int_{\sigma^2}^{+\infty}\lambda_1^{N-L-1}e^{-\lambda_1} \\
&\times \frac{e^{-x_l\left(\frac{1}{\lambda_1}-\frac{1}{\sigma^2}\right)}}{\prod_{i<l}(x_i-x_l)\prod_{i>l}(x_l-x_i)}d\lambda_1 \quad (42) \\
&= \frac{e^{\sigma^2-\frac{1}{\sigma^2}\sum_{i=1}^N x_i}}{N\pi^{LN}\sigma^{2(N-1)(L-1)}}\sum_{l=1}^{N}\frac{e^{\frac{x_l}{\sigma^2}}}{\prod_{\substack{i=1\\ i\ne l}}^N(x_l-x_i)} \\
&\times \int_{\sigma^2}^{+\infty}\lambda_1^{N-L-1}e^{-\left(\lambda_1+\frac{x_l}{\lambda_1}\right)}d\lambda_1, \quad (43)
\end{aligned}
$$

which finally gives

$$
P_{\mathbf{Y}|I_1}(\mathbf{Y}) = \frac{e^{\sigma^2-\frac{1}{\sigma^2}\sum_{i=1}^N x_i}}{N\pi^{LN}\sigma^{2(N-1)(L-1)}}\sum_{l=1}^{N}\frac{e^{\frac{x_l}{\sigma^2}}J_{N-L-1}(\sigma^2,x_l)}{\prod_{\substack{i=1\\ i\ne l}}^N(x_l-x_i)}. \quad (44)
$$

with $J_k(x,y) = \int_x^{+\infty} t^k e^{-t-\frac{y}{t}}dt$, and we finally have the desired decision criterion.

Appendix B
Proof of Theorem 2

Since $\mathbf{H}$ is still unitarily invariant in the case $M > 1$,

$$
P_{\mathbf{Y}|I_M}(\mathbf{Y}) = \int_{\mathcal{U}(N)\times\mathbb{R}^{+M}} P_{\mathbf{Y}|\mathbf{\Sigma},I_M}(\mathbf{Y},\mathbf{\Sigma})P_{\bar{\mathbf{\Lambda}}}(\bar{\mathbf{\Lambda}})d\mathbf{U}d\bar{\mathbf{\Lambda}}, \quad (45)
$$

which, using the same technique as previously, further develops into

$$
\begin{aligned}
&P_{\mathbf{Y}|I_M}(\mathbf{Y}) \\
&= \frac{(N-M)!M^{MN}e^{M^2\sigma^2}\sigma^{2(N-M)(N-L-1)}(-1)^{MN-\frac{M(M+1)}{2}}}{N!\pi^{NL}\prod_{j=1}^{M-1}j!} \\
&\times \int_{\sigma^2}^{+\infty}\cdots\int_{\sigma^2}^{+\infty}\prod_{i=1}^{M}\lambda_i{}^{N-L-1}\frac{\prod_{i<j}^{M}(\lambda_i-\lambda_j)}{\prod_{i<j}^{N}(x_i-x_j)}e^{-M\sum_{i=1}^M\lambda_i} \\
&\times \det\left[\begin{array}{c:c:ccc} e^{-\frac{x_1}{\sigma^2}} & \kappa_j(x_i,\sigma^2)e^{-\frac{x_i}{\sigma^2}} & e^{-\frac{x_1}{\lambda_M}} & \cdots & e^{-\frac{x_1}{\lambda_1}} \\ \vdots & & \vdots & \cdots & \vdots \\ e^{-\frac{x_N}{\sigma^2}} & & e^{-\frac{x_N}{\lambda_M}} & \cdots & e^{-\frac{x_N}{\lambda_1}} \end{array}\right] \quad (46)
\end{aligned}
$$

in which the term $(-1)^{MN-\frac{M(M+1)}{2}}$ originates from the M exchanges between the k^{th} column and the $(N-k+1)^{th}$ column, $k \in [1,M]$. By factorizing the determinant by $e^{-\frac{1}{\sigma^2}\sum_{i=1}^N x_i}$, developing along the M last columns, we have from Lemma 1,

$$
\begin{aligned}
&\det\left[\begin{array}{c:c:ccc} e^{-\frac{x_1}{\sigma^2}} & \kappa_j(x_i,\sigma^2)e^{-\frac{x_i}{\sigma^2}} & e^{-\frac{x_1}{\lambda_M}} & \cdots & e^{-\frac{x_1}{\lambda_1}} \\ \vdots & & \vdots & \cdots & \vdots \\ e^{-\frac{x_N}{\sigma^2}} & & e^{-\frac{x_N}{\lambda_M}} & \cdots & e^{-\frac{x_N}{\lambda_1}} \end{array}\right] \quad (47) \\
&= \sum_{\mathbf{a}\subset[1,N]}\frac{e^{-\sum_{i=1}^M x_{a_i}\left(\frac{1}{\lambda_i}-\frac{1}{\sigma^2}\right)}}{\sigma^{2(N-M-1)(N-M)}}\frac{e^{-\frac{\sum_{i=1}^N x_i}{\sigma^2}}\prod_{i<j}^N(x_i-x_j)}{\prod_{a_i}\prod_{j\notin[a_1,\ldots,a_i]}(x_{a_i}-x_j)}. \quad (48)
\end{aligned}
$$

Together, this becomes,

$$P_{\mathbf{Y}|I_M}(\mathbf{Y})$$
$$= \frac{(N-M)! e^{M^2\sigma^2 - \frac{\sum_{i=1}^N x_i}{\sigma^2}}}{N! M^{(M-2L-1)M/2} \pi^{NL} \sigma^{2(N-M)(L-M)} \prod_{j=1}^{M-1} j!}$$
$$\times \sum_{\mathbf{a}\subset[1,N]} \frac{e^{\frac{\sum_{i=1}^M x_{a_i}}{\sigma^2}}}{\prod_{a_i} \prod_{j\notin\{a_1,\ldots,a_i\}} (x_{a_i} - x_j)} \int_{M\sigma^2}^{+\infty} \cdots \int_{M\sigma^2}^{+\infty}$$
$$e^{-\sum_{i=1}^M \left(\lambda_i + \frac{M x_{a_i}}{\lambda_i}\right)} \prod_{i=1}^M \lambda_i^{N-L-1} \prod_{i<j}^M (\lambda_i - \lambda_j) d\lambda_1 \ldots d\lambda_M. \quad (49)$$

Remind now the Vandermonde determinant identity

$$\prod_{i<j}^M (X_j - X_i) = \sum_{\mathbf{b}\in\mathcal{P}(M)} \mathrm{sgn}(\mathbf{b}) \prod_{i=1}^M X_i^{b_i - 1}, \quad (50)$$

where $\mathcal{P}(k)$ is the ensemble of permutations of k and $\mathrm{sgn}(\mathbf{b})$ designs the signature of the permutation $\mathbf{b}$. We finally obtain

$$P_{\mathbf{Y}|I_M}(\mathbf{Y}) = \frac{(N-M)! M^{(2L-M+1)M/2} e^{M^2\sigma^2 - \frac{\sum_{i=1}^N x_i}{\sigma^2}}}{N! \pi^{NL} \sigma^{2(N-M)(L-M)} \prod_{j=1}^{M-1} j!}$$
$$\times \sum_{\mathbf{a}\subset[1,N]} \frac{e^{\frac{\sum_{i=1}^M x_{a_i}}{\sigma^2}}}{\prod_{a_i} \prod_{\substack{j\neq a_1 \\ \cdots \\ j\neq a_i}} (x_{a_i} - x_j)} \sum_{\mathbf{b}\in\mathcal{P}(M)} (-1)^{\mathrm{sgn}(\mathbf{b})+M}$$
$$\times \prod_{l=1}^M J_{N-L-2+b_l}(M\sigma^2, M x_{a_i}), \quad (51)$$

which completes the proof.

Romain Couillet received his Msc. in Mobile Communications at the Eurecom Institute (France) in 2007. He received his Msc. in Communication Systems in Telecom ParisTech, France in 2007. In September 2007, he joined ST-Ericsson, where he works as an Algorithm Development Engineer on the Long Term Evolution Advanced (LTE-A) project. In parallel to his position at ST-Ericsson, he is currently a PhD student at Supélec (France). His research topics include mobile communications, multi-users multi-antenna detection, cognitive radio, Bayesian probability and random matrix theory.

Mérouane Debbah entered the Ecole Normale Supérieure de Cachan (France) in 1996 where he received the M.Sc and the Ph.D. degrees respectively in 1999 and 2002. From 1999 to 2002, he worked for Motorola Labs on Wireless Local Area Networks and prospective fourth generation systems (OFDM and MC-CDMA). From 2002 until 2003, he was appointed Senior Researcher at the Vienna Research Center for Telecommunications (Austria) working on MIMO wireless channel modeling issues. From 2003 until 2007, he joined the Mobile Communications department of the Institute Eurecom (France) as an Assistant Professor. He is presently a Professor at Supélec (France), holder of the Alcatel-Lucent Chair on flexible radio. His research interests are in information theory, signal processing and wireless communications.

REFERENCES

[1] Mitola III, J., and Maguire Jr, GQ, "Cognitive radio: making software radios more personal", Personal Communications, IEEE [see also IEEE Wireless Communications] 6(4), 13-18, 1999.

[2] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications", IEEE Selected areas in Comm. , vol 23., no. 2, 2005.

[3] H. Urkowitz, "Energy detection of unknown deterministic signals," Proc. of the IEEE, vol. 55, no. 4, pp. 523-531, 1967.

[4] V. I. Kostylev, "Energy detection of a signal with Random Amplitude", Proc. IEEE Int. Conf. on Communications (ICC'02). New York City, pp. 1606-1610, May 2002.

[5] Simon, M.K., F. F. Digham, M.-S. Alouini, "On the Energy Detection of Unknown Signals over Fading Channels," ICC 2003 Conference Record, Anchorage, Alaska, May 11-15, 2003.

[6] 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, April 2008.

[7] L. S. Cardoso, M. Debbah, P. Bianchi, J. Najim, "Cooperative Spectrum Sensing Using Random Matrix Theory", International Symposium on Wireless Pervasive Computing 2008, Santorini, Greece.

[8] Y. Zeng, Y.-Ch. Liang, "Eigenvalue based Spectrum Sensing Algorithms for Cognitive Radio", arXiv:0804.2960.

[9] P. Bianchi, J. Najim, M. Maida, M. Debbah, "Performance of Some Eigen-based Hypothesis Tests for Collaborative Sensing," Proceedings of IEEE Statistical Signal Processing Workshop, 2009.

[10] M. Wax, T. Kailath, "Detection of Signals by Information Theoretic Criteria," IEEE Trans. on Acoustic, Speech and Signal Processing, vol. ASSP-33, no. 2, 1985.

[11] P-J Chung, J. F. Böhme, C. F. Mecklenbräuker, A. Hero, "Detection of the Number of Signals Using the Benjamin-Hochberg Procedure," IEEE Trans. on Signal Processing, vol. 55, no. 6, 2007.

[12] E. T. Jaynes, "Probability Theory: The Logic of Science", Cambridge University Press, June 2003.

[13] R. T. Cox, "Probability, frequency and reasonable expectation", American journal of physics, vol. 14, pp. 1-13, 1946.

[14] L. Brillouin, "Science and Information Theory," Academic Press Inc, 1962.

[15] I. S. Gradshteyn, I. M. Ryzhik, "Table of Integrals, Series and Products", Academic Press, 6th edition, 2000.

[16] A. M. Tulino, S. Verdú, "Random Matrix Theory and Wireless Communications", Now Publishers, vol. 1, Issue 1, 2004.

[17] H. Jeffreys, "An Invariant Form for the Prior Probability in Estimation Problems", Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, no. 186, pp. 453-461, 1946.

[18] S. Bausson, F. Pascal, P. Forster, J-P. Ovarlez and P. Larzabal, "First and Second Order Moments of the Normalized Sample Covariance Matrix of Spherically Invariant Random Vectors," IEEE Signal Processing Letters, vol. 14, pp. 425-428, 2007.

[19] B. Hochwald, T. Marzetta, V. Tarokh, "Multiple-Antenna Channel Hardening and Its Implications for Rate Feedback and Scheduling", IEEE Trans. on Information Theory, vol. 50, no. 9, pp. 1893-1909, 2004.

[20] A. Y. Orlov, T. Shiota, "Schur function expansion for normal matrix model and associated discrete matrix models," arXiv Preprint 0501017, 2005.

[21] T. Ratnarajah, R. Vaillancourt, M. Alvo, "Complex random matrices and Ricean channel capacity," Problems of Information Transmission, vol. 41, no. 1, pp. 1-22, 2005.

[22] N. R. Rao, J. A. Mingo, R. Speicher, A. Edelman, "Statistical eigen-inference from large Wishart matrices," Annals of Statistics, vol. 36, no. 6, pp. 2850-2885, 2008.

[23] Ø. Ryan, A. Masucci, M. Debbah, "Finite Dimensional Statistical Inference," IEEE Trans. on Information Theory, 2009, *submitted.*

[24] R. Couillet, J. W. Silverstein, M. Debbah, "Eigen-inference for multi-source power estimation," IEEE International Symposium on Information Theory, 2010.

[25] S.H. Simon, A.L. Moustakas, L. Marinelli, "Capacity and character expansions: Moment generating function and other exact results for MIMO correlated channels", IEEE Transactions on Information Theory, vol. 52, no. 12, pp. 5336-5351, Dec. 2006.

[26] A. B. Balantekin, "Character expansions, Itzykson-Zuber integrals, and the QCD partition function", *Phys. Rav. D*, vol. 62, no. 8, Oct. 2000.

[27] M. Guillaud, M. Debbah, A. L. Moustakas, "Maximum Entropy MIMO Wireless Channel Models", Submitted to IEEE Trans. Information Theory Dec. 2006, http://arxiv.org/abs/cs.IT/0612101