A Maximum Entropy Characterization of Spatially Correlated MIMO Wireless Channels

Maxime Guillaud
Forschungszentrum
Telekommunikation Wien (ftw.)
Vienna, Austria
guillaud@ftw.at

Mérouane Debbah
Eurecom Institute
Mobile Communications Department,
Sophia-Antipolis, France
debbah@eurecom.fr

Aris L. Moustakas
Department of Physics
University of Athens
Athens, Greece
arislm@phys.uoa.gr

***Abstract*— We investigate the problem of establishing the joint probability distribution of the entries of a Multiple-Input Multiple-Output (MIMO) spatially correlated flat-fading channel, when little or no information about the channel properties are available. We show that the entropy of a random positive semidefinite matrix is maximized by the Wishart distribution. We subsequently obtain the Maximum Entropy distribution of the MIMO transfer matrix by establishing its distribution conditioned on the covariance, and by later marginalizing over the covariance matrix. The obtained distribution is isotropic, and is described analytically as a function of the Frobenius norm of the channel matrix.**

I. Introduction

While a large number of models for wireless transmission channels can be found in the literature (see e.g. [1] for an overview), most of them rely on some kind of knowledge about the model parameters. In particular, spatial correlation is known to be a critical parameters for MIMO channels, through its influence on the channel capacity [2]. Among the models incorporating spatial correlation as a parameter, let us note the full-correlation model, which specifies the correlation for every pair of scalar variable, the Kronecker model, where the full correlation matrix is assumed to have a Kronecker structure, and the Weichselberger [3] model. This class of models can accurately predict the channel behaviour, on the condition that the correlation properties (or the spatial eigenbases in the case of the Weichselberger model) are known, e.g. through measurements. Essentially, this means that it is easy to study and to replicate the properties of a known channel.

The models proposed in this article have a different goal than those cited previously, since we do not seek a model that matches a particular situation or set of measurements, but rather generally fits a whole class of situations. For instance, the environment in which a mobile device will operate is not known at the time of the design of the channel code, and therefore the code should be adapted to all possible environments. In order to achieve this goal, we propose to use the Maximum Entropy (MaxEnt) principle [4]. The Maximum Entropy principle relies on the fact that, in the absence of any prior information about the process being modeled, the maximally noncommittal distribution (i.e. the one that implies no arbitrary constraint in addition to the available information about the process) is the one with the maximum Shannon entropy.

This paper focuses on the spatial correlation properties of frequency-flat fading channels. In general, in the absence of knowledge about correlation, application of the MaxEnt principle yields a process with independent components (see [5]). However, measurements have shown that this is rarely the case in reality, and that some degree of correlation between the components must be taken into account. Therefore, we first focus on the spatial covariance matrix, and derive the MaxEnt distribution of a general covariance matrix, in both the full-rank and rank-deficient cases. In the full-rank case, the entropy maximizing distribution of the covariance matrix is shown to be a Wishart distribution. In a second step, we construct the analytical model for the MIMO channel itself, by first deriving the MaxEnt distribution of the channel for a known covariance, and later marginalizing over the covariance matrix, using the distribution of the covariance established previously. The obtained distribution is shown to be isotropic, and is described analytically as a function of the Frobenius norm of the channel matrix. In addition to their analytical description, sample generation for simulation purposes according to the channel models proposed in this paper is easily achieved through standard numerical methods.

II. Notations and channel model

We consider a wireless MIMO link with n_t transmit and n_r receive antennas, represented by the $n_r \times n_t$ matrix $\mathbf{H}$. Since we are only concerned with frequency-flat channels, the (i,j)-th coefficient of $\mathbf{H}$ (the attenuation between transmit antenna j and receive antenna i) is a complex scalar that we denote $h_{i,j}$. In this article, we focus on the derivation of the fading characteristics of $\mathbf{H}$ in the form of the probability density function (PDF) $P_{\mathbf{H}}(\mathbf{H})$. We are not concerned with the time-related properties of the channel, i.e. we assume that the process under study is stationary, and refer to the channel realization $\mathbf{H}$ or equivalently to its vectorized notation $\mathbf{h} = \mathrm{vec}(\mathbf{H}) = [h_{1,1} \dots h_{n_r,1}, h_{1,2} \dots h_{n_r,n_t}]^T$. Let us denote $N = n_r n_t$. We will sometimes use the alternative notation where the antenna indices are mapped into $[1 \dots N]$, i.e.

denoting $\mathbf{h} = [h_1 \dots h_N]^T$.

III. MAXENT DISTRIBUTIONS OF CORRELATED CHANNELS

In [5], Debbah et al. show that the probability distribution that maximizes the entropy $\int_{\mathbb{C}^N} -\log(P(\mathbf{H}))P(\mathbf{H})\mathrm{d}\mathbf{H}$, where $\mathrm{d}\mathbf{H} = \prod_{i=1}^N \mathrm{dRe}(h_i)\mathrm{dIm}(h_i)$ is the Lebesgue measure on $\mathbb{C}^N$, under the only assumption that the channel has a finite average energy NE_0, is the Gaussian i.i.d. distribution

$$P_{\mathbf{H}|E_0}(\mathbf{H}) = \frac{1}{(\pi E_0)^N} \exp\left(-\sum_{i=1}^N \frac{|h_i|^2}{E_0}\right). \quad (1)$$

Note that the Gaussianity and the independence property of the obtained distribution are the consequence, via the maximum entropy principle, of the ignorance by the modeler of any constraint other than the total average energy NE_0, rather than assumptions. In the following sections, we shall incorporate some knowledge about the spatial correlation characteristics of $\mathbf{H}$ in the framework of maximum entropy channel modeling. We first study the case where the correlation matrix is deterministic, and subsequently extend the result to an unknown covariance matrix.

A. Deterministic knowledge of the correlation matrix

In this section, we establish the maximum entropy distribution of $\mathbf{H}$ under the assumption that the covariance matrix $\mathbf{Q} = \int_{\mathbb{C}^N} \mathbf{h}\mathbf{h}^H P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H}$ is known, where $\mathbf{Q}$ is a $N \times N$ complex positive definite Hermitian matrix. Each component of the covariance constraint represents an independent linear constraint of the form

$$\int_{\mathbb{C}^N} h_a h_b^* P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H} = q_{a,b} \quad (2)$$

for $(a, b) \in [1, \dots, N]^2$. The entropy maximization under these covariance constraints (the energy constraint being implicitly set by $\mathrm{tr}(\mathbf{Q})$) is achieved through the Lagrange multipliers method, by introducing the N^2+1 Lagrange coefficients $\alpha_{a,b}$ and β, and maximizing the functional

$$\begin{aligned} L(P_{\mathbf{H}|\mathbf{Q}}) &= \int_{\mathbb{C}^N} -\log(P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}))P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H} \\ &+ \sum_{\substack{a\in[1,\dots,N]\\ b\in[1,\dots,N]}} \alpha_{a,b}\left[\int_{\mathbb{C}^N} h_a h_b^* P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H} - q_{a,b}\right] \\ &+ \beta\left[1 - \int_{\mathbb{C}^N} P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H}\right]. \end{aligned} \quad (3)$$

A necessary condition for the entropy maximization is obtained by letting $\frac{\delta L(P_{\mathbf{H}|\mathbf{Q}})}{\delta P_{\mathbf{H}|\mathbf{Q}}} = 0$. Letting $\mathbf{A} = [\alpha_{a,b}]_{(a,b)\in[1,\dots,N]^2}$ denote the $N \times N$ matrix of the Lagrange multipliers, and taking the derivative of eq. (3) yields

$$\frac{\delta L(P_{\mathbf{H}|\mathbf{Q}})}{\delta P_{\mathbf{H}|\mathbf{Q}}} = -\log(P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})) - 1 - \beta - \mathbf{h}^T\mathbf{A}\mathbf{h}^* = 0. \quad (4)$$

After elimination of the Lagrange coefficients through proper normalization, this yields

$$P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}, \mathbf{Q}) = \frac{1}{\det(\pi\mathbf{Q})} \exp\left(-(\mathbf{h}^H\mathbf{Q}^{-1}\mathbf{h})\right). \quad (5)$$

Therefore, with the extra constraint of a deterministic correlation matrix, the maximum entropy principle yields a complex Gaussian distribution.

B. Knowledge of the existence of a correlation matrix

Let us consider the case where covariance is known to be a parameter of interest, but is not known deterministically. We will proceed in two steps, first seeking a probability distribution function for the covariance matrix $\mathbf{Q}$, and then marginalizing the channel distribution over $\mathbf{Q}$.

1) Correlation Matrix MaxEnt PDF: Let us first establish the PDF $P_{\mathbf{Q}}$ of $\mathbf{Q}$, with the energy constraint NE_0, by maximizing the functional

$$\begin{aligned} L(P_{\mathbf{Q}}) &= \int_{\mathcal{S}} -\log(P_{\mathbf{Q}}(\mathbf{Q}))P_{\mathbf{Q}}(\mathbf{Q})\mathrm{d}\mathbf{Q} \quad (6) \\ &+\beta\left[\int_{\mathcal{S}} P_{\mathbf{Q}}(\mathbf{Q})\mathrm{d}\mathbf{Q} - 1\right] + \gamma\left[\int_{\mathcal{S}} \mathrm{tr}(\mathbf{Q})P_{\mathbf{Q}}(\mathbf{Q})\mathrm{d}\mathbf{Q} - NE_0\right], \end{aligned}$$

where $\mathcal{S}$ denotes the set of $N \times N$ positive semidefinite complex matrices. Since the trace operator can be expressed simply as a sum of eigenvalues, we perform the variable change to the eigenvalues/eigenvectors space, and denote $\Lambda = \mathrm{diag}(\lambda_1 \dots \lambda_N)$ the diagonal matrix containing the eigenvalues of $\mathbf{Q}$, and $\mathbf{U}$ the unitary matrix containing the corresponding eigenvectors. Therefore, $\mathbf{Q} = \mathbf{U}\Lambda\mathbf{U}^H$. Let $\mathcal{U}(N)$ denote the set of $N \times N$ unitary matrices, endowed with the Haar measure. In order for the variable change to be bijective, Λ is defined over $\mathbb{R}_{\leq}^{+N}$, the space of real N-tuples with non-negative non-decreasing components, and $\mathbf{U}$ is defined over the space of unitary $N \times N$ matrices with real, non-negative first row, which is denoted by $\mathcal{U}(N)/T$ (see [6, Lemma 4.4.6]).

Letting $F(\mathbf{U}, \mathbf{\Lambda}) = P_{\mathbf{Q}}(\mathbf{U}\Lambda\mathbf{U}^H)$, and introducing the Jacobian $K(\Lambda) = \frac{(2\pi)^{N(N-1)/2}}{\prod_{j=1}^{N-1} j!} \prod_{i<j}(\lambda_i - \lambda_j)^2$, eq. (6) becomes

$$\begin{aligned} L(F) &= \int_{\mathcal{U}(N)/T\times\mathbb{R}_{\leq}^{+N}} -\log(F(\mathbf{U},\mathbf{\Lambda}))F(\mathbf{U},\mathbf{\Lambda})K(\Lambda)\mathrm{d}\mathbf{U}\mathrm{d}\Lambda \\ &+ \gamma\left[\int_{\mathcal{U}(N)/T\times\mathbb{R}_{\leq}^{+N}} \left(\sum_{i=1}^N \lambda_i\right) F(\mathbf{U},\mathbf{\Lambda})K(\Lambda)\mathrm{d}\mathbf{U}\mathrm{d}\Lambda - NE_0\right] \\ &+ \beta\left[\int_{\mathcal{U}(N)/T\times\mathbb{R}_{\leq}^{+N}} F(\mathbf{U},\mathbf{\Lambda})K(\Lambda)\mathrm{d}\mathbf{U}\mathrm{d}\Lambda - 1\right]. \end{aligned} \quad (7)$$

The maximum entropy distribution is obtained by letting $\frac{\delta L(F)}{\delta F} = 0$, which yields

$$\left[-1 - \log(F(\mathbf{U},\mathbf{\Lambda})) + \beta + \gamma\left(\sum_{i=1}^N \lambda_i\right)\right] K(\Lambda) = 0. \quad (8)$$

Since $K(\Lambda) \neq 0$ except on a set of measure zero, this is equivalent to $F(\mathbf{U}, \mathbf{\Lambda}) = \exp\left(\beta - 1 + \gamma \sum_{i=1}^{N} \lambda_i\right)$. Note that the distribution $F(\mathbf{U}, \mathbf{\Lambda})K(\Lambda)$ does not explicitly depend on $\mathbf{U}$. This implies that $\mathbf{U}$ is uniformly distributed, with constant density $P_{\mathbf{U}} = (2\pi)^N$ over $\mathcal{U}(N)/T$. Therefore, the joint density can be factored as $F(\mathbf{U}, \mathbf{\Lambda})K(\Lambda) = P_{\mathbf{U}} P_{\mathbf{\Lambda}}(\mathbf{\Lambda})$, where the eigenvalues are distributed over $\mathbb{R}^{+N}_{\leq}$ according to

$$P_{\Lambda}(\Lambda) = \frac{\mathrm{e}^{\beta-1}}{P_{\mathbf{U}}} \exp\left(\gamma \sum_{i=1}^{N} \lambda_i\right) \frac{(2\pi)^{N(N-1)/2}}{\prod_{j=1}^{N-1} j!} \prod_{i<j} (\lambda_i - \lambda_j)^2. \tag{9}$$

It is worth noting that the form of eq. (9) indicates that the order of the eigenvalues is immaterial, and therefore Λ can be equivalently defined over $\mathbb{R}^{+N}$, with the PDF of the *unordered* eigenvalues becoming

$$P'_{\Lambda}(\Lambda) = \frac{\mathrm{e}^{\beta-1}}{P_{\mathbf{U}}} \exp\left(\gamma \sum_{i=1}^{N} \lambda_i\right) \frac{(2\pi)^{N(N-1)/2}}{N! \prod_{j=1}^{N-1} j!} \prod_{i<j} (\lambda_i - \lambda_j)^2. \tag{10}$$

Finally, the Lagrange coefficients γ and β can be eliminated by solving the normalization equation $\int_{\mathbb{R}^{+N}} P'_{\Lambda}(\Lambda) \mathrm{d}\Lambda = 1$ by way of the Selberg integral (see [7, eq. (17.6.5)]). This yields $\gamma = -\frac{N}{E_0}$, and

$$P'_{\Lambda}(\Lambda) = \left(\frac{N}{E_0}\right)^{N^2} \prod_{n=1}^{N} \frac{1}{n!(n-1)!} \mathrm{e}^{-\frac{N}{E_0} \sum_{i=1}^{N} \lambda_i} \prod_{i<j} (\lambda_i - \lambda_j)^2. \tag{11}$$

Note [8], [9] that eq. (11) describes the unordered eigenvalue density of a complex $N \times N$ Wishart matrix with N degrees of freedom and covariance $\frac{E_0}{N}\mathbf{I}_N$ (denoted as $\tilde{\mathcal{W}}_N(N, \frac{E_0}{N}\mathbf{I}_N)$). Since the eigenvectors of $\mathbf{Q}$ are isotropically distributed, we can conclude that $\mathbf{Q}$ is itself a $\tilde{\mathcal{W}}_N(N, \frac{E_0}{N}\mathbf{I}_N)$ matrix. This constitutes a fairly general result, since it shows that the entropy-maximizing distribution for a $N \times N$ positive semidefinite matrix under an average trace constraint is a Wishart distribution with N degrees of freedom. A similar result, with a slightly different constraint, was obtained by Adhikari in [10], where it is shown that the entropy-maximizing distribution of a positive definite matrix with known mean $\mathbf{G}$ follows a Wishart distribution with $N+1$ degrees of freedom, more precisely the $\tilde{\mathcal{W}}_N(N+1, \frac{\mathbf{G}}{N+1})$ distribution.

The isotropic property of the obtained Wishart distribution (since $\mathbf{U}$ is Haar distributed, there is not privileged direction for the eigenvalues of the covariance matrix $\mathbf{Q}$), is a consequence of the fact that no spatial constraints were imposed on the correlation. The energy constraint (imposed through the trace) only affects the distribution of the eigenvalues of $\mathbf{Q}$. Note also that the generation for simulation purposes of $\mathbf{Q}$ according to the Wishart distribution obtained above is easy, since it can be obtained as $\mathbf{Q} = \frac{E_0}{N}\mathbf{B}\mathbf{B}^H$, where $\mathbf{B}$ is a $N \times N$ matrix with i.i.d. complex circularly-symmetric Gaussian coefficients of unit variance.

2) Marginalization over $\mathbf{Q}$*:* The complete distribution of the correlated channel is obtained by marginalizing out $\mathbf{Q}$, using its distribution as established in the previous section. The distribution of $\mathbf{H}$ is obtained through

$$P_{\mathbf{H}}(\mathbf{H}) = \int_{\mathcal{S}} P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}, \mathbf{Q}) P_{\mathbf{Q}}(\mathbf{Q}) \mathrm{d}\mathbf{Q} \tag{12}$$

$$= \int_{\mathcal{U}(N) \times \mathbb{R}^{+N}} P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}, \mathbf{U}, \Lambda) P'_{\mathbf{\Lambda}}(\mathbf{\Lambda}) \mathrm{d}\mathbf{U} \mathrm{d}\Lambda. \tag{13}$$

Let us rewrite the conditional probability density of eq. (5) as

$$P_{\mathbf{H}|\mathbf{Q}}(\mathbf{h}, \mathbf{U}, \Lambda) = \frac{1}{\pi^N \det(\Lambda)} \mathrm{e}^{-\mathrm{tr}(\mathbf{h}\mathbf{h}^H \mathbf{U} \Lambda^{-1} \mathbf{U}^H)}. \tag{14}$$

Using this expression in (13), we obtain $P_{\mathbf{H}}(\mathbf{H})$ as

$$\frac{1}{\pi^N} \int_{\mathbb{R}^{+N}} \int_{\mathcal{U}(N)} \mathrm{e}^{-\mathrm{tr}(\mathbf{h}\mathbf{h}^H \mathbf{U} \Lambda^{-1} \mathbf{U}^H)} \mathrm{d}\mathbf{U} \det(\Lambda)^{-1} P'_{\mathbf{\Lambda}}(\mathbf{\Lambda}) \mathrm{d}\Lambda. \tag{15}$$

Following the notations of [11], let $\det(f(i,j))$ denote the determinant of a matrix with the (i,j)-th element given by an arbitrary function $f(i,j)$. Let also $\Delta(\mathbf{X})$ denote the Vandermonde determinant of the eigenvalues x_i of matrix $\mathbf{X}$,

$$\Delta(\mathbf{X}) = \det(x_i^{j-1}) = \prod_{i>j} (x_i - x_j). \tag{16}$$

Using these notations, let us recall the Harish-Chandra-Itzykson-Zuber (HCIZ) integral [12]

$$\int_{\mathcal{U}(N)} \mathrm{e}^{\kappa \mathrm{tr}(\mathbf{A}\mathbf{U}\mathbf{B}\mathbf{U}^H)} \mathrm{d}\mathbf{U} = \left(\prod_{n=1}^{N-1} n!\right) \kappa^{N(N-1)/2} \frac{\det\left(e^{-A_i B_j}\right)}{\Delta(\mathbf{A})\Delta(\mathbf{B})}, \tag{17}$$

where $\mathbf{A}$ and $\mathbf{B}$ are any hermitian matrices with respective eigenvalues $A_1, \ldots, A_N$ and $B_1, \ldots, B_N$. We will now explicit the Haar integral in (15) using the Harish-Chandra-Itzykson-Zuber result by identifying $\mathbf{A} = \mathbf{h}\mathbf{h}^H$ and $\mathbf{B} = \Lambda^{-1}$. Note however that we can not directly apply (17) since $\mathbf{A}$ is rank one, and therefore $\Delta(\mathbf{A}) = 0$. This can be resolved by taking the limit of all other eigenvalues to zero one by one, and applying the l'Hospital rule. Therefore, let $\mathbf{A}$ be an Hermitian matrix which has its Nth eigenvalue A_N equal to $\mathbf{h}^H\mathbf{h}$, and the others $A_1, \ldots, A_{N-1}$ are arbitrary, positive values that will eventually be set to 0. Letting $I(\mathbf{H}, A_1, \ldots, A_{N-1}) =$

$$\frac{1}{\pi^N} \int_{\mathbb{R}^{+N}} \int_{\mathcal{U}(N)} \mathrm{e}^{-\mathrm{tr}(\mathbf{A}\mathbf{U}\Lambda^{-1}\mathbf{U}^H)} P_{\mathbf{U}} \mathrm{d}\mathbf{U} \det(\Lambda)^{-1} P'_{\mathbf{\Lambda}}(\mathbf{\Lambda}) \mathrm{d}\Lambda, \tag{18}$$

$P_{\mathbf{H}}(\mathbf{H})$ can be determined as the limit distribution when the first $N-1$ eigenvalues of $\mathbf{A}$ go to zero:

$$P_{\mathbf{H}}(\mathbf{H}) = \lim_{A_1, \ldots, A_{N-1} \to 0} I(\mathbf{H}, A_1, \ldots, A_{N-1}). \tag{19}$$

Applying the HCIZ to integrate over $\mathbf{U}$ yields after some transformations $I(\mathbf{H}, A_1, \ldots, A_{N-1}) =$

$$C \int_{\mathbb{R}^{+N}} \frac{\det(e^{-\frac{A_i}{\lambda_j}}) \det(\Lambda)^{N-2} \Delta(\Lambda)}{\Delta(\mathbf{A})} \mathrm{e}^{-\frac{N}{E_0}\mathrm{tr}(\Lambda)} \mathrm{d}\Lambda. \tag{20}$$

where we let the constant $C = \pi^{-N} \left(\frac{N}{E_0}\right)^{N^2} \left[\prod_{n=1}^{N} n!\right]^{-1}$.

Then, let us decompose the determinant product using the expansion formula: for an arbitrary $N\times N$ matrix $\mathbf{X}=(X_{i,j})$,

$$\det(\mathbf{X})=\sum_{\mathbf{a}\in\mathcal{P}_N}(-1)^{\mathbf{a}}\prod_{n=1}^{N}X_{n,a_n}=\frac{1}{N!}\sum_{\mathbf{a},\mathbf{b}\in\mathcal{P}_N}(-1)^{\mathbf{a}+\mathbf{b}}\prod_{n=1}^{N}X_{a_n,b_n}, \tag{21}$$

where $\mathbf{a}=[a_1,\ldots,a_N]$, $\mathcal{P}_N$ denotes the set of all permutations of $[1,\ldots,N]$, and $(-1)^{\mathbf{a}}$ is the sign of the permutation. Using the first form of the expansion twice, we obtain

$$\Delta(\Lambda)\det\left(e^{-A_i/\lambda_j}\right)=\sum_{\mathbf{a},\mathbf{b}\in\mathcal{P}_N^2}(-1)^{\mathbf{a}+\mathbf{b}}\prod_{n=1}^{N}\lambda_n^{a_n-1}\mathrm{e}^{-A_{b_n}/\lambda_n}. \tag{22}$$

Therefore,

$$\begin{aligned} I(\mathbf{H},A_1,\ldots,A_{N-1}) &= \frac{C}{\Delta(\mathbf{A})}\cdot \\ &\sum_{\mathbf{a},\mathbf{b}\in\mathcal{P}_N}(-1)^{\mathbf{a}+\mathbf{b}}\prod_{n=1}^{N}\int_{\mathbb{R}^+}\lambda_n^{N+a_n-3}\mathrm{e}^{-\frac{A_{b_n}}{\lambda_n}}\mathrm{e}^{-\frac{N}{E_0}\lambda_n}\mathrm{d}\lambda_n \quad (23)\\ &= CN!\frac{\det[f_i(A_j)]}{\Delta(\mathbf{A})}, \quad (24)\end{aligned}$$

where we let $f_i(x)=\int_{\mathbb{R}^+}t^{N+i-3}\mathrm{e}^{-x/t}\mathrm{e}^{-\frac{N}{E_0}t}\mathrm{d}t$, and obtain (24) by identification of the second form of the determinant expansion in eq. (23). The limit of I as $A_1,\ldots A_{N-1}$ go to zero, is obtained by using a result from [11, Appendix III], about the limit of the ratio $\frac{\det(f_i(x_j))}{\Delta(\mathbf{X})}$ as several eigenvalues converge to the same x_0. In this particular case, this yields

$$\begin{aligned} P_{\mathbf{H}}(\mathbf{H}) &= \frac{(-\gamma)^{N^2}}{\pi^N x_N^{N-1}}\prod_{n=1}^{N-1}[n!(n-1)!]^{-1}\cdot \quad (25)\\ &\det\left[f_i(0);f_i'(0);\ldots;f_i^{(N-2)}(0);f_i(x_N)\right].\end{aligned}$$

Eq. (25) shows that the probability of $\mathbf{H}$ depends only on $x_N=\mathbf{h}^H\mathbf{h}$, the Frobenius norm of $\mathbf{H}$. The distribution of $\mathbf{h}$ is isotropic, and is completely determined by the probability density of $x=\mathbf{h}^H\mathbf{h}$. The PDF $P_x(x)$ of x can be obtained by integration of $P_{\mathbf{H}}(\mathbf{h})$ over the zero-centered complex hypersphere of radius x, and of surface $S_N(x)=\frac{\pi^N x^{N-1}}{(N-1)!}$:

$$\begin{aligned} P_x(x) &= \frac{(-\gamma)^{N^2}}{(N-1)!}\prod_{n=1}^{N-1}[n!(n-1)!]^{-1}\cdot \\ &\det\left[f_i(0);f_i'(0);\ldots;f_i^{(N-2)}(0);f_i(x)\right]. \quad (26)\end{aligned}$$

In order to simplify the expression of the determinant, it is useful to identify the Bessel K-function [13, Section 8.432] in f_i:

$$f_i(x) = 2\left(\sqrt{\frac{x}{-\gamma}}\right)^{i+N-2}K_{i+N-2}(2\sqrt{-\gamma x}). \tag{27}$$

Noting that the pth derivative (for $0\le p\le N-2$) of f_i at 0 is simply $f_i^{(p)}(0)=(-1)^{-i-N}\gamma^{p-i-N+2}(i+N-3-p)!$, we expand the determinant along the column containing the $f_i(x)$. After some calculus (omitted due to space constraints), we obtain

$$P_x(x)=-\sum_{n=1}^{N}f_n(x)\frac{(-N/E_0)^{N+n-1}}{[(n-1)!]^2(N-n)!}. \tag{28}$$

C. Limited-rank covariance matrix

In this section, we address the situation where the modeler takes into account the existence of a covariance matrix of rank $L<N$ (we assume that L is known). As in the full-rank case, we will use the eigendecomposition $\mathbf{Q}=\mathbf{U}\Lambda\mathbf{U}^H$ of the covariance matrix, with $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_L,0,\ldots,0)$. Let us denote $\Lambda_L=\mathrm{diag}(\lambda_1,\ldots,\lambda_L)$. The maximum entropy probability density of $\mathbf{Q}$ with the extra rank constraint is unsurprisingly similar to the one derived in Section III-B.1, with the difference that all the energy is carried by the first L eigenvalues, i.e. $\mathbf{U}$ is uniformly distributed over $\mathcal{U}(N)$, while the joint probability of the non-zero eigenvalues is $P_{\Lambda_L}(\Lambda_L)=$

$$\left(\frac{L^2}{NE_0}\right)^{L^2}\prod_{n=1}^{L}\frac{1}{n!(n-1)!}\mathrm{e}^{-\frac{L^2}{NE_0}\sum_{i=1}^{L}\lambda_i}\prod_{i<j\le L}(\lambda_i-\lambda_j)^2. \tag{29}$$

However, when $\mathbf{Q}$ is not full rank, the conditional probability distribution of $\mathbf{H}|\mathbf{Q}$ is a degenerate Gaussian, and eq. (5) does not hold anymore. Only the projection of $\mathbf{h}$ in the subspace associated to the L non-zero eigenvalues (and the L eigenvectors forming the $N\times L$ projector matrix which we denote $\mathbf{U}_{[L]}$) is Gaussian, and therefore eq. (14) must be rewritten as

$$P_{\mathbf{H}|\mathbf{Q}}(\mathbf{h},\mathbf{U},\Lambda_L)=\begin{cases}\frac{\mathrm{e}^{-\mathbf{h}^H\mathbf{U}_{[L]}\Lambda_L^{-1}\mathbf{U}_{[L]}{}^H\mathbf{h}}}{\pi^L\prod_{i=1}^{L}\lambda_i} & \text{if } \mathbf{h}\in\mathrm{Span}(\mathbf{U}_{[L]}),\\ 0 & \text{elsewhere.}\end{cases} \tag{30}$$

This expression of $P_{\mathbf{H}|\mathbf{Q}}(\mathbf{h},\mathbf{U},\Lambda_L)$ does not lend itself directly to the marginalization described in Section III-B.2, since the zero eigenvalues of $\mathbf{Q}$ complicate the analysis. We solve this by performing the marginalization of the covariance in an L-dimensional subspace: consider an $L\times L$ unitary matrix $\mathbf{B}_L$, and note that the $N\times N$ block-matrix $\mathbf{B}=\begin{pmatrix}\mathbf{B}_L & 0\\ 0 & \mathbf{I}_{N-L}\end{pmatrix}$ is unitary as well. Since the uniform distribution over $\mathcal{U}(N)$ is unitarily invariant, $\mathbf{UB}$ is uniformly distributed over $\mathcal{U}(N)$. Furthermore, since $\int_{\mathcal{U}(L)}\mathrm{d}\mathbf{B}_L=1$, we have $P_{\mathbf{H}}(\mathbf{h})=$

$$\begin{aligned} &\int_{\mathcal{U}(L)}\int_{\mathcal{U}(N)\times\mathbb{R}^{+L}}P_{\mathbf{H}|\mathbf{Q}}(\mathbf{h},\mathbf{UB},\Lambda_L)P_{\Lambda_L}(\Lambda_L)\mathrm{d}\mathbf{U}\mathrm{d}\Lambda_L\mathrm{d}\mathbf{B}_L \quad (31)\\ &= \int_{\mathbf{U}\in\mathcal{U}(N)}\mathbb{1}_{\{\mathbf{h}\in\mathrm{Span}(\mathbf{U}_{[L]})\}}P_{\mathbf{k}}(\mathbf{U}_{[L]}{}^H\mathbf{h})\mathrm{d}\mathbf{U}, \quad (32)\end{aligned}$$

where eq. (32) was obtained by letting $\mathbf{k}=\mathbf{U}_{[L]}{}^H\mathbf{h}$ and

$$P_{\mathbf{k}}(\mathbf{k})=\int_{\mathcal{U}(L)\times\mathbb{R}^{+L}}\frac{\mathrm{e}^{-\mathbf{k}^H\mathbf{B}_L\Lambda_L^{-1}\mathbf{B}_L^H\mathbf{k}}}{\pi^L\prod_{i=1}^{L}\lambda_i}P_{\Lambda_L}(\Lambda_L)\mathrm{d}\mathbf{B}_L\mathrm{d}\Lambda_L \tag{33}$$

Exploiting the similarity of eqs. (33) and (15), we conclude, using the same arguments as in Section III-B.2, that $\mathbf{k}$ is isotropically distributed in $\mathcal{U}(L)$, and that its PDF depends only on its Frobenius norm, following

$$P_{\mathbf{k}}(\mathbf{k}) = \frac{1}{S_L(\mathbf{k}^H\mathbf{k})} P_x^{(L)}(\mathbf{k}^H\mathbf{k}), \tag{34}$$

where

$$P_x^{(L)}(x) = \frac{2}{x}\sum_{i=1}^{L}\left(-L\sqrt{\frac{x}{NE_0}}\right)^{L+i} \frac{K_{i+L-2}\left(2L\sqrt{\frac{x}{NE_0}}\right)}{[(i-1)!]^2 (L-i)!}. \tag{35}$$

Finally, note that $\mathbf{h}^H\mathbf{h} = \mathbf{k}^H\mathbf{k}$, and that the marginalization over the random rotation that transforms $\mathbf{k}$ into $\mathbf{h}$ in eq. (32) preserves the isotropic property of the distribution. Therefore,

$$P_{\mathbf{h}}(\mathbf{h}) = \frac{1}{S_N(\mathbf{h}^H\mathbf{h})} P_x^{(L)}(\mathbf{h}^H\mathbf{h}), \tag{36}$$

or in other words, the Frobenius norm of $\mathbf{H}$ is distributed according to $P_x^{(L)}$ (note that $P_x^{(N)} = P_x$).

D. Further remarks on the proposed model

In addition to the fully analytical description given above, the proposed class of channel model can be easily simulated, since each realization $\mathbf{h}$ can be obtained by generating separately a normalized vector process uniformly distributed over the sphere of radius 1, and a scalar process representing the norm according to eq. (35) (e.g. by numerical inversion of the corresponding cumulative density function).

Note also that the proposed model for H was derived under the assumption that the structure of the covariance matrix was unconstrained. However, the result of Section III-B.1 and applies to various other situations, since it was shown generally that the Wishart distribution with N degrees of freedom maximizes the entropy of a semi-definite positive $N \times N$ matrix. Therefore, this result can be applied to any situation where covariance matrices are used as parameters. For instance, the Kronecker channel model, which is based on the assumption of the separability of the transmit and receive covariance, constraints the structure of the MIMO channel matrix according to given (usually, experimentally estimated) covariance matrices. However, in the absence of experimental data, no specific model for those covariance matrices was available. Our result dictates that, in the absence of measurements, the Wishart model is the least committal choice for the covariance model.

IV. Simulation Results

Examples of the channel Frobenius norm PDFs (eq. (35)) for $L = 1, 2, 4, 8, 12$ and 16 are represented on Fig. 1 for a 4×4 channel ($N = 16$), together with the PDF of the instantaneous power of a Gaussian i.i.d. channel of the same size and mean power. As expected, the energy distribution of the proposed MaxEnt model is more spread out than the energy of a Gaussian i.i.d. channel.

The CDF of the mutual information (computed as $\log\det(\mathbf{I} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H)$, where ρ is the SNR) achieved over the limited-rank ($L < 16$) and full rank ($L = 16$) covariance MaxEnt channel at a SNR of 15 dB is pictured on Figure 2 for various ranks L, together with the CDF of the mutual information achieved over the Gaussian i.i.d. channel. The proposed model differs in particular in the tails of the distribution. In particular, the outage capacity for low outage probability is greatly reduced w.r.t. the Gaussian i.i.d. channel model.

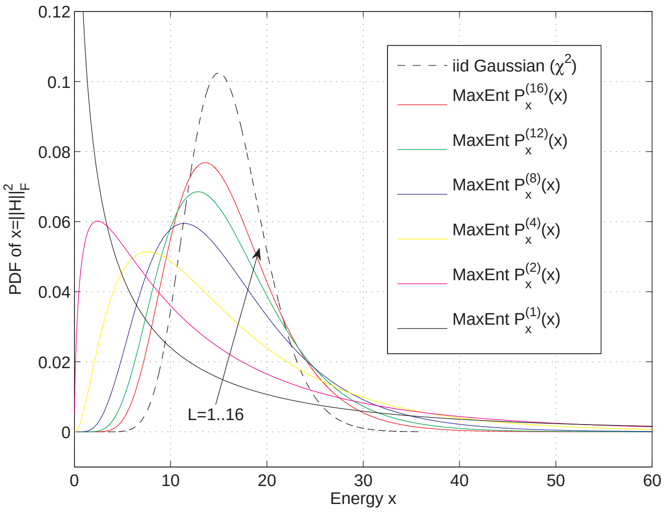

Fig. 1. Limited-rank covariance distribution $P_{x_N}^{(L)}(x)$ for $L = 1, 2, 4, 8, 12$ and 16, and χ^2 with 16 degrees of freedom, for $NE_0 = 16$.

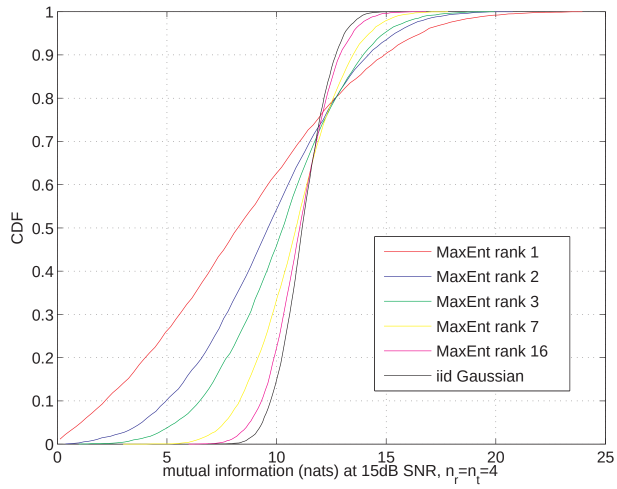

Fig. 2. CDF of the instantaneous mutual information of a 4×4 flat-fading channel for the MaxEnt model with various covariance ranks, at 15dB SNR.

V. Conclusion

We proposed analytical models for MIMO spatially correlated flat-fading wireless channels, based on the maximum entropy method. We first demonstrated that the entropy-maximizing probability distribution of a semi-definite positive matrix is a particular case of the Wishart distribution. We then incorporated this result by conditioning the channel matrix on the full covariance matrix, and by later marginalizing out the covariance parameter. Both the full-rank covariance matrix and the rank-deficient cases were treated. The obtained channel distribution was shown to be isotropic, and was described analytically as a function of the Frobenius norm of the channel matrix.

ACKNOWLEDGMENTS

The authors would like to thank an anonymous reviewer for pointing out reference [10]. This work has been carried out while Maxime Guillaud was with Eurecom Institute. This research was supported by France Telecom and by the NEWCOM network of excellence.

REFERENCES

[1] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Espoti, et al., "Survey of channel and radio propagation models for wireless MIMO systems," *EURASIP Journal on Wireless Communications and Networking*, accepted for publication.

[2] D.-S. Shiu, G.J. Foschini, M.J. Gans, and J.M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[3] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Transactions on Wireless Communications*, vol. 5, no. 1, pp. 90– 100, Jan. 2006.

[4] E. T. Jaynes, *Probability Theory*, Cambridge University Press, 2003.

[5] Merouane Debbah and Ralf R. Müller, "MIMO channel modelling and the principle of maximum entropy," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1667–1690, May 2005.

[6] Fumio Hiai and Dénes Petz, *The Semicircle Law, Free Random Variables and Entropy*, vol. 77 of *Mathematical Surveys and Monographs*, American Mathematical Society, 2000.

[7] M.L. Mehta, *Random Matrices*, Academic Press, 2nd edition, 1991.

[8] Antonia M. Tulino and Sergio Verdú, *Random Matrix Theory and Wireless Communications*, now Publishers, Delft, The Netherlands, 2004.

[9] Alan Edelman, *Eigenvalues and Condition Numbers of Random Matrices*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1989.

[10] S. Adhikari, "A non-parametric approach for uncertainty quantification in elastodynamics," in *Proc. 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Newport, RI, USA, May 2006.

[11] S.H. Simon, A.L. Moustakas, and 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.

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

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