

A Characterization of Maximum Entropy Spatially Correlated Wireless Channel Models

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

Mérouane Debbah
Supélec
Telecommunications Department,
Gif-sur-Yvette, France
merouane.debbah@supelec.fr

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

Abstract— We consider the use of entropy maximization methods as a tool to generate fading models for Multiple-Input Multiple-Output (MIMO) spatially correlated flat-fading channels. We focus in particular on various assumptions about the degree of knowledge on the full covariance matrix of the channel (namely in the power constraint, rank constraint, and average constraint case). We show that this method can provide closed-form probability density functions in all cases. In a second part, we analyze the statistical properties of the singular values of the resulting fading channel. In general, they are more spread out than in the classical Gaussian i.i.d. case, which incurs less optimistic ergodic capacity figures.

I. INTRODUCTION

A large number of models for wireless transmission channels are available in the literature – [1] provides for a good overview. One usually distinguishes between geometric models, built upon the observation and modelling of the electromagnetic propagation phenomenon, and statistical models, which aim to reproduce statistics obtained experimentally. Recently, a new class of channel models was introduced [2], [3], based on Bayesian inference techniques and on entropy maximization. A probabilistic model is obtained by first defining the range of the variable to model, and by choosing the probability distribution function having maximum entropy, given some constraints based on system design (e.g. about finite power) as well as information from the experimental correlation of the channel. The logic behind entropy maximization, popularized by Jaynes [4], is that the Shannon entropy is a good measure for uncertainty, and therefore entropy maximization provides the maximally noncommittal model given the expressed constraints. In other words, among all probability densities complying with the constraints, the entropy-maximizing density is the one that introduces the least unwarranted assumptions. This is in general a departure from the classical methods, where a postulated model is deemed acceptable only if it does not contradict the experimental results.

The use of entropy maximization as a modelling tool was first introduced by Debbah and Müller in [2]. In [5] and later in [3], spatial correlation in Multiple-Input Multiple-Output (MIMO) flat-fading channels was investigated in this context. Spatial correlation is a critical parameter for communications

applications [6], therefore a two-step model was proposed, whereby a probabilistic model for correlation is sought first, and correlation is marginalized out in a second step to obtain the final channel model. Both steps rely on entropy maximization, which enables to obtain a compact analytical description of both the correlation and the channel itself.

The object of the present paper is to extend the spatial correlation model to the case where the mean of the correlation is known (e.g. from measurements, or because it is a consequence of the antenna array design), and to provide a characterization of the properties of the obtained models, through the statistics of the eigenvalues of the obtained channel, and of their consequences on the additive white Gaussian noise (AWGN) fading channel ergodic capacity.

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 $h_{i,j}$. The fading model that we seek to establish consists of the statistics of the multivariate fading process \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 $\mathbf{h} = \text{vec}(\mathbf{H}) = [h_{1,1} \dots h_{n_r,1}, h_{1,2} \dots h_{n_r,n_t}]^T$. 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. MAXIMUM ENTROPY-BASED DISTRIBUTIONS OF CORRELATED MIMO CHANNELS

As already mentioned, the channel spatial covariance is a critical parameter for communications systems, and we will therefore pay a particular attention to it. Therefore, let us consider the full covariance matrix of \mathbf{h} , defined as the $N \times N$ complex positive definite Hermitian matrix $\mathbf{Q} = \int_{\mathbb{C}^N} \mathbf{h} \mathbf{h}^H P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}) d\mathbf{H}$. In the context of geometric

propagation models, \mathbf{Q} can be shown to depend on the number, position and properties of the antennas in the transmit and receive arrays, as well as of the scatterers where the electromagnetic waves can reflect.

Our method can be summarized as follows:

- 1) derive $P_{\mathbf{Q}}(\mathbf{Q})$ through entropy maximization,
- 2) derive $P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}, \mathbf{Q})$ through entropy maximization,
- 3) marginalize over \mathbf{Q} to obtain $P_{\mathbf{H}}(\mathbf{H}) = \int P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}, \mathbf{Q})P_{\mathbf{Q}}(\mathbf{Q})d\mathbf{Q}$.

The logic behind this method, exposed in more detail in [5], is that $P_{\mathbf{H}|\mathbf{Q}}$ describes the channel statistics corresponding to a given correlation situation (e.g. a given location of the transmitter and receiver arrays, and a given set of scatterers), whereas $P_{\mathbf{Q}}$ models the variety of correlation situations that can be encountered. It is therefore expected that in reality, the coherence interval of \mathbf{Q} is longer than that of \mathbf{H} . However, we will not try to model these time-related properties, and will focus on the description of the probability densities of \mathbf{Q} and \mathbf{H} , both being considered ergodic. Note also that due to the introduction of the intermediate variable \mathbf{Q} the obtained $P_{\mathbf{H}}$ is not, in general, the entropy maximizing distribution of \mathbf{H} .

We now briefly recall the results already established. Step 2 is in general not problematic, since it is easy to show that the entropy maximizing distribution $P_{\mathbf{H}|\mathbf{Q}}$ under the covariance constraint, yields

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

i.e. a complex correlated Gaussian distribution of parameter \mathbf{Q} . Special care has to be taken if \mathbf{Q} is rank-deficient, since in this case the distribution is a degenerate Gaussian. However, as shown in [3], this does not change fundamentally our result.

Step 1, on the other hand, admits a richer variety of cases, depending on what is known about \mathbf{Q} . In previous works, we derived the probability density $P_{\mathbf{Q}}(\mathbf{Q})$ for the case where only a finite power constraint is known, and for the case of a known rank of \mathbf{Q} , $L \leq N$. In the present paper, we also introduce the case where the mean of \mathbf{Q} is known.

A. No knowledge about the covariance matrix

Let us define the entropy of \mathbf{Q} as

$$H(P_{\mathbf{Q}}) = \int_{\mathbf{Q}_{\geq 0}} -\log(P_{\mathbf{Q}}(\mathbf{Q}))P_{\mathbf{Q}}(\mathbf{Q})d\mathbf{Q}, \quad (2)$$

where $d\mathbf{Q}$ is a $\mathcal{U}(N)$ -invariant measure. The distribution of \mathbf{Q} is obtained by maximization of $H(P_{\mathbf{Q}})$ under an energy constraint $\text{tr}(\mathbf{Q}) = NE_0$, where E_0 is the average variance per component. A solution to

$$P_{\mathbf{Q}} = \arg \max_{E_{\mathbf{Q}}[\text{tr}(\mathbf{Q})]=NE_0} H(P_{\mathbf{Q}}) \quad (3)$$

can be found by considering the eigendecomposition $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, and yields a complex $N \times N$ Wishart matrix with

N degrees of freedom and covariance $\frac{E_0}{N}\mathbf{I}_N$ (denoted as $\tilde{\mathcal{W}}_N(N, \frac{E_0}{N}\mathbf{I}_N)$). This distribution has isotropic eigenvectors (\mathbf{U} is Haar-distributed), and the distribution of the unordered eigenvalues $\lambda_1 \dots \lambda_N$ on the diagonal of $\mathbf{\Lambda}$ is described by

$$P'_{\mathbf{\Lambda}}(\mathbf{\Lambda}) = \left(\frac{N}{E_0}\right)^{N^2} \prod_{n=1}^N \frac{1}{n!(n-1)!} e^{-\frac{N}{E_0} \sum_{i=1}^N \lambda_i} \prod_{i < j} (\lambda_i - \lambda_j)^2. \quad (4)$$

Note that the isotropic property of the obtained Wishart distribution 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} .

The complete distribution of the correlated channel \mathbf{H} is obtained (step 3 in the method above) by marginalizing over \mathbf{Q} :

$$P_{\mathbf{H}}(\mathbf{H}) = \int_{\mathbf{Q}_{\geq 0}} P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}, \mathbf{Q})P_{\mathbf{Q}}(\mathbf{Q})d\mathbf{Q}. \quad (5)$$

It was shown in [3] that $P_{\mathbf{H}}(\mathbf{H})$ can be described by a scalar function of the Frobenius norm of \mathbf{H} . In particular, $P_{\mathbf{H}}(\mathbf{H}) = \frac{1}{S_N(\|\mathbf{H}\|_F^2)} P_x^{(N)}(\|\mathbf{H}\|_F^2)$, where $S_N(x) = \frac{\pi^N x^{N-1}}{(N-1)!}$, and

$$P_x^{(L)}(x) = \frac{2}{x} \sum_{i=1}^L \left(-L \sqrt{\frac{x}{NE_0}}\right)^{L+i} \frac{K_{i+L-2}\left(2L \sqrt{\frac{x}{NE_0}}\right)}{[(i-1)!]^2 (L-i)!}, \quad (6)$$

where K denotes the Bessel K -function [7, Section 8.432]. This indicates that \mathbf{h} is isotropically distributed (uniformly on the sphere), and that the probability density of its squared Frobenius norm is described by $P_x^{(N)}$.

B. Rank constraint on the covariance matrix

Let us now consider an extra constraint on the rank of \mathbf{Q} . This is a commonly used assumption when modelling propagation situations with a limited number of scatterers. Imposing $\text{rank}(\mathbf{Q}) = L \leq N$, it is obvious that the result in (4) directly applies to describe the density of the L non-zero eigenvalues, simply by changing the parameter N to L . In this case, the marginalization of \mathbf{Q} (step 3) simply yields $P_{\mathbf{H}}(\mathbf{H}) = \frac{1}{S_N(\|\mathbf{H}\|_F^2)} P_x^{(L)}(\|\mathbf{H}\|_F^2)$. Note that taking $L = N$, i.e. the full rank case, yields the result of Section III-A.

The models proposed in this section and in the previous one, in addition to having tractable analytical expressions, also admit simple numerical generation according to the PDF $P_{\mathbf{H}}$, since \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. (6) (e.g. by numerical inversion of the corresponding cumulative density function).

C. Knowledge of the covariance mean

Adding the constraint that the mean of the covariance matrix \mathbf{Q} must be an arbitrary positive definite matrix \mathbf{M} , we now consider the entropy-maximizing distribution

$$P_{\mathbf{Q}|\mathbf{M}} = \arg \max_{E_{\mathbf{Q}|\mathbf{M}} = \mathbf{M}} H(P_{\mathbf{Q}}). \quad (7)$$

Let us seek the expression of $P_{\mathbf{Q}|\mathbf{M}}$, by considering the functional

$$L(P_{\mathbf{Q}}) = H(P_{\mathbf{Q}}) + \beta \left[\int_{\mathbf{Q} \geq 0} P_{\mathbf{Q}}(\mathbf{Q}) - 1 \right] + \text{tr} \left(\Omega \left[\mathbf{M} - \int_{\mathbf{Q} \geq 0} \mathbf{Q} P_{\mathbf{Q}}(\mathbf{Q}) \right] \right) \quad (8)$$

where β and the $N \times N$ matrix Ω are Lagrange multipliers. Equating the functional derivative $\frac{\delta L(P_{\mathbf{Q}})}{\delta P_{\mathbf{Q}}}$ to zero yields $P_{\mathbf{Q}}(\mathbf{Q}) = \exp(\beta - 1) \exp(-\text{tr}(\Omega \mathbf{Q}))$. Normalization of the density to 1 imposes that $\exp(\beta - 1) = \frac{\det(\Omega)^N}{\pi^{N(N-1)/2} \prod_{i=1}^N (i-1)!}$, and the constraint on the average lets us identify $\Omega = N\mathbf{M}^{-1}$. Therefore,

$$P_{\mathbf{Q}|\mathbf{M}}(\mathbf{Q}, \mathbf{M}) = \frac{1}{\det\left(\frac{\mathbf{M}}{N}\right)^N \pi^{N(N-1)/2} \prod_{i=1}^N (i-1)!} \cdot \exp\left(-\text{tr}\left\{\left(\frac{\mathbf{M}}{N}\right)^{-1} \mathbf{Q}\right\}\right). \quad (9)$$

We recognize that in this case, \mathbf{Q} is a complex Wishart matrix with N degrees of freedom and covariance $\frac{\mathbf{M}}{N}$, or $\tilde{\mathcal{W}}_N(N, \frac{\mathbf{M}}{N})$.

The distribution of \mathbf{H} is obtained by integration over all positive definite matrices \mathbf{Q} :

$$P_{\mathbf{H}|\mathbf{M}}(\mathbf{H}, \mathbf{M}) = \int_{\mathbf{Q} > 0} P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}, \mathbf{Q}) P_{\mathbf{Q}|\mathbf{M}}(\mathbf{Q}, \mathbf{M}) \quad (10)$$

$$= \int_{\mathbf{Q} > 0} \frac{e^{-\mathbf{h}^H \mathbf{Q}^{-1} \mathbf{h}} e^{-\text{tr}\left\{\left(\frac{\mathbf{M}}{N}\right)^{-1} \mathbf{Q}\right\}}}{\det(\pi \mathbf{Q}) \det\left(\frac{\mathbf{M}}{N}\right)^N \pi^{N(N-1)/2} \prod_{i=1}^N (i-1)!} \quad (11)$$

Let us denote $\left(\frac{\mathbf{M}}{N}\right)^{-1} = \mathbf{U}\mathbf{D}\mathbf{U}^H$ the eigendecomposition of the mean of the Wishart distribution. Furthermore, let $\mathbf{Q}' = \mathbf{D}^{1/2} \mathbf{U}^H \mathbf{Q} \mathbf{U} \mathbf{D}^{1/2}$ become the integration variable (with the introduction of the corresponding Jacobian $\det(\mathbf{D})^{-N}$). We obtain

$$P_{\mathbf{H}|\mathbf{M}}(\mathbf{H}, \mathbf{M}) = \int_{\mathbf{Q}' > 0} \frac{e^{-\mathbf{h}^H \mathbf{U} \mathbf{D}^{1/2} \mathbf{Q}'^{-1} \mathbf{D}^{1/2} \mathbf{U}^H \mathbf{h}} e^{-\text{tr}\{\mathbf{Q}'\}}}{\pi^{N(N-1)/2} \prod_{i=1}^N (i-1)! \pi^N \det(\mathbf{D})^{-1} \det(\mathbf{Q}')} \quad (12)$$

where we used the fact that \mathbf{D} and $\frac{\mathbf{M}}{N}$ have reciprocal determinants. We note that if we consider the vector $\mathbf{g} = \mathbf{D}^{1/2} \mathbf{U}^H \mathbf{h}$ instead of \mathbf{h} , we obtain its PDF

$$P_{\mathbf{g}|\mathbf{M}}(\mathbf{g}, \mathbf{M}) = \int_{\mathbf{Q}' > 0} \frac{e^{-\mathbf{g}^H \mathbf{Q}'^{-1} \mathbf{g}} e^{-\text{tr}\{\mathbf{Q}'\}}}{\pi^N \det(\mathbf{Q}') \pi^{N(N-1)/2} \prod_{i=1}^N (i-1)!}, \quad (13)$$

where we introduced the Jacobian $\det(\mathbf{D})^{-1}$. Note that eq. (13) is in fact independent of \mathbf{M} . Furthermore, since the term under the integral is the product of the density of

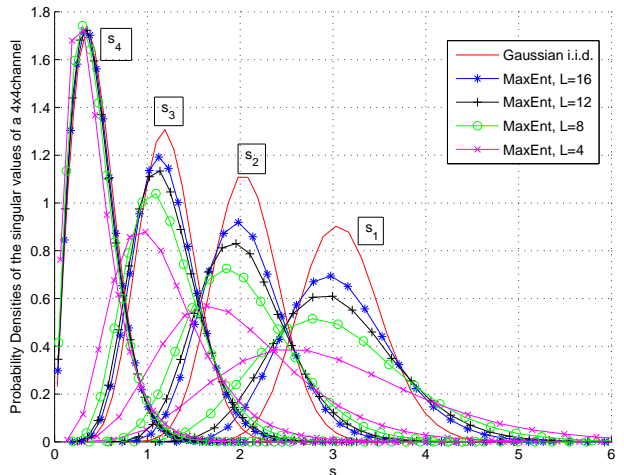


Fig. 1. Probability density functions of the 4 singular values of a 4×4 channel, for the Gaussian i.i.d. channel and for the MaxEnt model with parameter $L = 16, 12, 8$ and 4 .

a complex Gaussian random variable \mathbf{g} with covariance \mathbf{Q}' , and of the density of a Wishart $\tilde{\mathcal{W}}_N(N, \mathbf{I})$ matrix \mathbf{Q}' , we recognize that we are in the case described in Section III-A (with $E_0 = N$). Therefore, we conclude that the PDF of \mathbf{g} is given by $P_{\mathbf{g}}(\mathbf{g}) = \frac{1}{S_N(\|\mathbf{G}\|_F^2)} P_x^{(N)}(\|\mathbf{G}\|_F^2)$.

This indicates that samples of \mathbf{H} can be generated easily for the purpose of simulation, by generating \mathbf{G} using the procedure outlined in the previous section, and by applying the transformation $\mathbf{h} = \mathbf{D}^{-1/2} \mathbf{U} \mathbf{g}$.

IV. CAPACITY ANALYSIS OF THE PROPOSED MODELS

Let us consider a flat-fading AWGN channel, described by

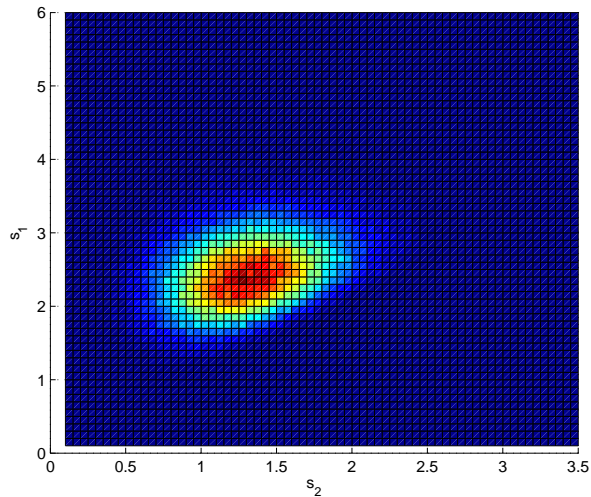
$$\mathbf{r}_n = \mathbf{H} \mathbf{s}_n + \mathbf{n}_n, \quad (14)$$

where matrix \mathbf{H} represents $n_r \times n_t$ matrix of the channel gains between all transmit-receive antenna pairs, and vectors \mathbf{s}_n , \mathbf{r}_n and \mathbf{n}_n represent respectively the transmitted signal (assumed drawn from a Gaussian i.i.d. distribution with identity covariance, denoted $\mathcal{N}(0, \mathbf{I})$), the received signal, and an additive noise sample ($\mathcal{N}(0, \sigma^2 \mathbf{I})$) at instant n .

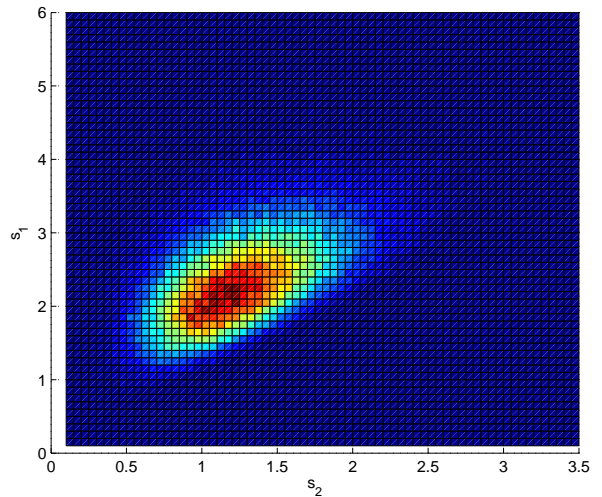
Let us consider the singular value decomposition of the channel matrix $\mathbf{H} = \mathbf{V} \mathbf{S} \mathbf{W}^H$, where \mathbf{V} and \mathbf{W} are unitary matrices, and \mathbf{S} is a diagonal matrices with the (real, positive) singular values s_1, \dots, s_p , where $p = \min(n_t, n_r)$. The capacity of the channel in (14), for a given realization of \mathbf{H} , has been shown by Telatar [8] to be

$$C = \log \det \left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{H} \mathbf{H}^H \right) = \sum_{i=1}^p \log \left(1 + \frac{s_i^2}{\sigma^2} \right). \quad (15)$$

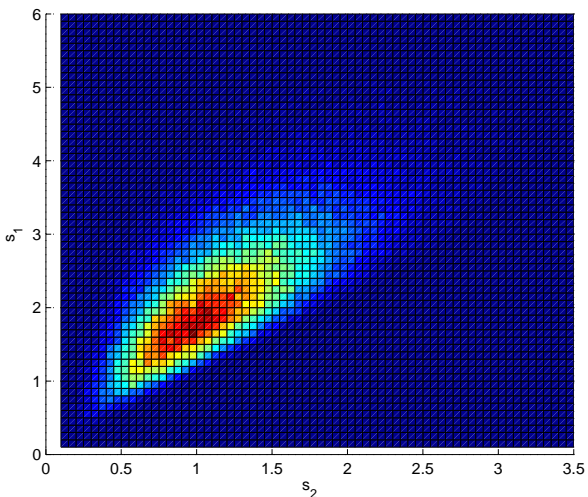
It is obvious from this expression that the various capacity metrics associated to fading channels (ergodic capacity, outage capacity...) depend only on the distribution of the singular



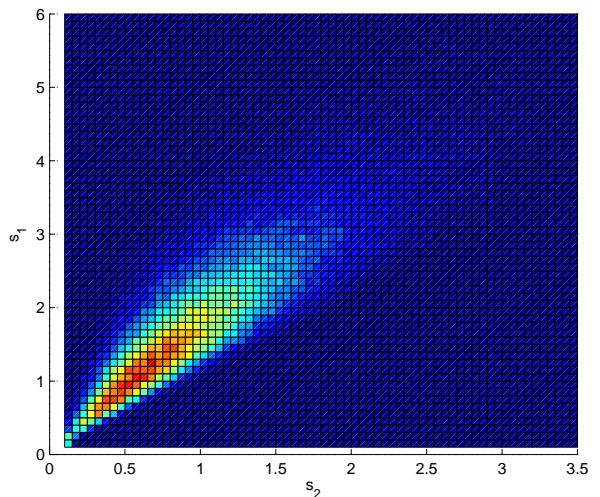
(a) Gaussian i.i.d. channel



(b) MaxEnt channel, L=8



(c) MaxEnt channel, L=4



(d) MaxEnt channel, L=2

Fig. 2. Joint PDF of (s_1, s_2) for a 4×2 channel, for various channel models (blue = low values, red = high values).

values of the channel matrix. In this section, we characterize experimentally the singular value distributions corresponding to the channel statistics presented in Section III, and we compare them to the singular value distributions associated to the classical Gaussian model with i.i.d. components.

Figure 1 depicts the probability density functions of the 4 ordered singular values ($s_1 \geq s_2 \geq s_3 \geq s_4$) of \mathbf{H} , for various fading models, namely the Gaussian i.i.d. model, and the maximum entropy model (denoted by MaxEnt) of section III-B, for several values of L . The energy parameter E_0 was fixed at 1 for all cases. It is noticeable that the distributions corresponding to the MaxEnt model are more spread out than for the Gaussian i.i.d. model. Interestingly, it appears that the distribution of the energy among the singular

value is the same for all considered models (Gaussian and MaxEnt), i.e. the expectation of the i -th singular value $E[s_i^2]$ is independent from the model, for all i . Of course, we also have $\sum_{i=1}^p E[s_i^2] = NE_0$. Interestingly, together with the concavity of the $\log(1 + \cdot)$ function, this indicates that more spread out the distributions will yield lower ergodic capacities. Therefore, the MaxEnt models show in general a lower ergodic capacity than the Gaussian i.i.d. model. Furthermore this phenomenon becomes more sensible as L decreases.

Figure 2 depicts the joint PDF of the two ordered eigenvalues (s_1, s_2) of a 4×2 MIMO channel, corresponding to the Gaussian channel model and to the MaxEnt channel model with parameter $L = 8, 4$ and 2 . Again, it is noticeable

that the MaxEnt channel model has singular values more spread out than the Gaussian i.i.d. channel (Fig. 2(a)), even for the case where the covariance matrix \mathbf{Q} is assumed to be full rank ($L=8$ in this case, Fig. 2(b)). Comparison of the PDF corresponding to the MaxEnt channel model with various values of L (Figs. 2(b), 2(c) and 2(d)) shows that a decreasing rank L of the covariance has the effect, already noted before, of increasing the spread of the distribution, as well as increasing the correlation between the singular values.

V. CONCLUSION

In this paper, we presented several analytical models for wireless MIMO flat-fading channels, based on various degrees of knowledge about the environment, using a modeling approach based on the maximum entropy principle. Our approach consists in first modeling the spatial covariance properties, considering the case where the covariance matrix has only a power constraint, a fixed rank constraint, and the case where the mean of the covariance matrix is known. We presented closed-form formulas for the probability densities of the channel matrix, as well as results of the numerical evaluation of the statistics of the singular values of the corresponding channels, and their implication on the channel capacity.

ACKNOWLEDGMENT

This work was funded by the I0 project of the Forschungszentrum Telekommunikation Wien (ftw.). ftw. is supported by the Kplus competence center program of the Austrian government.

REFERENCES

- [1] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, et al., "Survey of channel and radio propagation models for wireless MIMO systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, 2007.
- [2] 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.
- [3] M. Guillaud, M. Debbah, and A. L. Moustakas, "A maximum entropy characterization of spatially correlated MIMO wireless channels," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Hong Kong, Mar. 2007.
- [4] E. T. Jaynes, *Probability Theory*. Cambridge University Press, 2003.
- [5] M. Guillaud and M. Debbah, "Maximum entropy MIMO wireless channel models with limited information," in *Proc. MATHMOD Conference on Mathematical Modeling*, Wien, Austria, Feb. 2006.
- [6] 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.
- [7] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, 6th edition, 2000.
- [8] İ. Emre Telatar, "Capacity of multi-antenna Gaussian channels," Technical memorandum, Bell Laboratories, Lucent Technologies, October 1995, available at <http://mars.bell-labs.com/papers/proof/>.