MAXIMUM ENTROPY MIMO WIRELESS CHANNEL MODELS WITH LIMITED INFORMATION

M. Guillaud and M. Debbah, Eurecom Institute[1], France

Institut Eurécom, Mobile Communications Department
2229 Route des Crêtes, 06904 Sophia-Antipolis, France
Phone: +33 4 93 00 81 37, Fax: +33 4 93 00 82 00
email: {guillaud,debbah}@eurecom.fr

Abstract

In this contribution, analytical wireless channel models are derived from the maximum entropy principle, when only limited information about the environment is available. These models are useful in situations where analytical models of the fading characteristics of a multiple-antennas wireless channel are needed, and where the classical Rayleigh fading model is too coarse. The issues of the knowledge of the average channel energy, of an energy upper-bound, and of spatial correlation, are studied. First, analytical models are derived for the cases where these parameters are known deterministically. Frequently, these parameters are unknown, but still known to represent meaningful system characteristics (this includes typical scenarios where the received energy or the spatial correlation varies with the user position). In these cases, consistent analytical channel models are derived, based on maximum entropy distributions of the energy or space correlation parameters. In particular, we show that the entropy-maximizing distribution of the covariance matrices is conveniently handled through its eigenvalues, whereas its eigenvectors are uniformly distributed. Using this technique, the modeler can provide consistent models incorporating correlation of the channel antenna gains without the explicit value of these gains. The results are compared in terms of mutual information to the classical i.i.d. Gaussian model in the SISO case.

1 Introduction

The problem of modelling the characteristics of a wireless transmission channel is crucial to the appropriate design of suitable channel codes. The recent shift to the Multiple-Input Multiple-Output (MIMO) paradigm [1] and the corresponding need for MIMO channel models, together with the introduction of codes (such as turbo codes [2]) that can operate very close to the channel capacity, has placed the channel models under scrutiny: initial capacity analyses of MIMO channels assuming i.i.d. Rayleigh fading [3] were touting promising spectral efficiencies, whereas the importance of correlation between channel coefficients [4] and of the channel matrix rank are now understood to be critical parameters. In order to facilitate channel code development, analytical channel models are a desirable asset. Unfortunately, most of the available channel models that capture the complex spatial characteristics of the propagation channel (geometry, reflection coefficients, . . .) are based on ray tracing methods or variations thereof, which model the channel as a superposition of multipath components [5] and therefore do not lend themselves easily to analysis. Conversely, some analytical models were proposed to address the problem of accurate space correlation modeling by assuming a Rayleigh fading with appropriately designed correlation properties [6].

In [7], Debbah *et al.* address the question of channel modeling on the basis of statistical inference. Relying only on the principle of logical consistency, they propose a modeling methodology based only on the knowledge which is available on the environment. In particular, they show that by using the principle of maximum entropy introduced by Jaynes [8], one can translate the information on the environment into a joint distribution of the entries of the MIMO channel matrix. Choosing the distribution with the greatest entropy is justified on the basis of avoiding the arbitrary introduction of information that is not available. This article extends maximum entropy channel modeling to cases where the channel is known to have spatial correlation, but the exact characteristics of this correlation are not known.

[1]This research was supported by France Telecom and by the NEWCOM network of excellence.

2 Notations and channel model

Let us consider the multiple-antenna wireless channel with n_t transmit and n_r receive antennas. Since we are only concerned with non-frequency selective channels, let the complex scalar coefficient $h_{i,j}$ denote the channel attenuation between transmit antenna j and receive antenna i, $j = 1 \dots n_t$, $i = 1 \dots n_r$. Let $\mathbf{H}(t)$ denote the $n_r \times n_t$ channel matrix at time t. We recall the general model for a time-varying flat-fading channel with additive noise

$$\mathbf{y}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{n}(t), \tag{1}$$

where $\mathbf{n}(t)$ is usually modeled as a complex circularly-symmetric Gaussian random variable (r.v.) with independent identically distributed (i.i.d.) coefficients. In this article, we focus on the derivation of the fading characteristics of $\mathbf{H}(t)$. When we are not concerned with the time-related properties of $\mathbf{H}(t)$, we will drop the time index t, and refer to the channel realization $\mathbf{H}$ or equivalently to its vectorized notation $\mathbf{h} \triangleq \mathrm{vec}(\mathbf{H}) = [h_{1,1} \dots h_{n_r,1}, h_{1,2} \dots h_{n_r,n_t}]^T$. Let us also denote $N \triangleq n_r n_t$ and map the antenna indices into $[1 \dots N]$, *i.e.* denoting equivalently $\mathbf{h} = [h_1 \dots h_N]^T$. In the sequel, $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote respectively the real and imaginary parts of a complex number.

3 Previous results: known channel energy constraint

In [7], a probability distribution is derived as the one that maximizes the entropy $\int_{\mathbb{C}^N} -\log(P(\mathbf{H}))P(\mathbf{H})\mathrm{d}\mathbf{H}$, where $\mathrm{d}\mathbf{H} \triangleq \prod_{i=1}^N \mathrm{dRe}(h_i)\mathrm{dIm}(h_i)$ is the Haar measure on $\mathbb{C}^N$, under the only assumption that the channel has a finite average energy NE_0, and the normalization constraint associated to the definition of a probability density, *i.e.*

$$\int_{\mathbb{C}^N} ||\mathbf{H}||_F^2 P(\mathbf{H})\mathrm{d}\mathbf{H} = NE_0, \quad \text{and} \int_{\mathbb{C}^N} P(\mathbf{H})\mathrm{d}\mathbf{H} = 1. \tag{2}$$

This is achieved through the method of Lagrange multipliers, by writing

$$L(P) = \int_{\mathbb{C}^N} -\log(P(\mathbf{H}))P(\mathbf{H})\mathrm{d}\mathbf{H} + \beta \left[1 - \int_{\mathbb{C}^N} P(\mathbf{H})\mathrm{d}\mathbf{H}\right] + \gamma \left[NE_0 - \int_{\mathbb{C}^N} ||\mathbf{H}||_F^2 P(\mathbf{H})\mathrm{d}\mathbf{H}\right] \tag{3}$$

where we introduce the scalar Lagrange coefficients β and γ, and taking the functional derivative [9] w.r.t. P equal to zero:

$$\frac{\delta L(P)}{\delta P} = -\log(P(\mathbf{H})) - 1 - \beta - \gamma||\mathbf{H}||_F^2 = 0. \tag{4}$$

Eq. (4) yields $P(\mathbf{H}) = \exp\left(-(\beta+1) - \gamma||\mathbf{H}||_F^2\right)$, and the normalization of this distribution according to (2) finally yields the coefficients β and γ, and the final distribution is obtained as

$$P_{\mathbf{H}|E_0}(\mathbf{H}) = \frac{1}{(\pi E_0)^N} \exp\left(-\sum_{i=1}^N \frac{|h_i|^2}{E_0}\right) \tag{5}$$

Interestingly, the distribution defined by eq. (5) corresponds to a complex Gaussian r.v. with independently fading coefficients, although neither Gaussianity nor independence were among the initial constraints. These properties 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.

4 Unknown energy constraint

Let us now introduce a new model for situations where the channel energy E can not be assumed constant, for instance when an unknown shadowing must be accounted for. In this case, we propose to take E as a r.v., and marginalize the distribution of $\mathbf{H}$ over E:

$$P(\mathbf{H}) = \int_{\mathbb{R}^+} P_{\mathbf{H},E}(\mathbf{H}, E)\mathrm{d}E = \int_{\mathbb{R}^+} P_{\mathbf{H}|E}(\mathbf{H})P_E(E)\mathrm{d}E. \tag{6}$$

In order to establish the probability distribution P_E, let us find the maximum entropy distribution under the constraints:

- $0 \leq E \leq E_{max}$, where E_{max} represents an absolute constraint on the transmit power, or on the amplitude range of the receiver,
- its average $E_0 \triangleq \int_0^{E_{max}} E P_E(E) \mathrm{d}E$ is known.

Applying the Lagrange multipliers method again, we introduce the scalar unknowns β and γ, and maximize the functional

$$L(P_E) = -\int_0^{E_{max}} log(P_E(E)) P_E(E) \mathrm{d}E + \beta \left[\int_0^{E_{max}} E P_E(E) \mathrm{d}E - E_0 \right] + \gamma \left[\int_0^{E_{max}} P_E(E) \mathrm{d}E - 1 \right]. \tag{7}$$

Taking the derivative equal to zero ($\frac{\delta L(P_E)}{\delta P_E} = 0$) yields $P_E(E) = \exp(\beta E - 1 + \gamma)$, and the Lagrange multipliers are finally eliminated by solving the normalization equations

$$\int_0^{E_{max}} E \exp(\beta E - 1 + \gamma) \mathrm{d}E = E_0, \quad \text{and} \quad \int_0^{E_{max}} \exp(\beta E - 1 + \gamma) \mathrm{d}E = 1. \tag{8}$$

$\beta < 0$ is the solution to the transcendental equation

$$E_{max} \exp(\beta E_{max}) - \left(\frac{1}{\beta} + E_0 \right) (\exp(\beta E_{max}) - 1) = 0, \tag{9}$$

and finally P_E is obtained as the truncated exponential law

$$P_E(E) = \frac{\beta}{\exp(\beta E_{max}) - 1} \exp(\beta E), \quad 0 \leq E \leq E_{max}, \quad 0 \quad \text{elsewhere}. \tag{10}$$

Note that taking $E_{max} = +\infty$ in eq. (9) yields $\beta = -\frac{1}{E_0}$ and the exponential law $P_E(E) = E_0 \exp\left(-\frac{E}{E_0}\right)$.

4.1 Application to the SISO channel

In order to illustrate the difference between the two situations presented so far, let us investigate the Single-Input Single-Output (SISO) case $n_t = n_r = 1$, where the channel is represented by a single complex scalar h. Furthermore, since the distribution is circularly symmetric, it is more convenient to consider the distribution of $r \triangleq |h|$. After the change of variables $h \triangleq r(\cos\theta + i \sin\theta)$, and marginalization over θ, eq. (5) becomes

$$P_r(r) = \frac{2r}{E_0} \exp\left(-\frac{r^2}{E_0} \right), \tag{11}$$

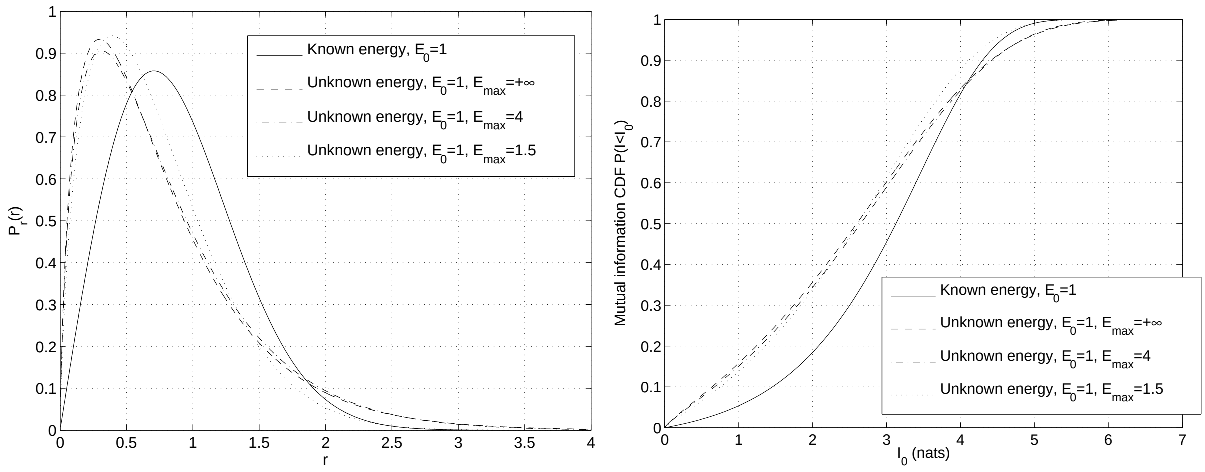
whereas eq. (6) yields

$$P_r(r) = \int_0^{E_{max}} \frac{\beta}{\exp(\beta E_{max}) - 1} \frac{2r}{E} \exp\left(\beta E - \frac{r^2}{E} \right) \mathrm{d}E. \tag{12}$$

Note that the integral always exists since $\beta < 0$. Figure 1(a) depicts the probability density functions (PDFs) of r under the known energy constraint (eq. (11), with $E_0 = 1$), and the known energy distribution constraint (eq. (12) is computed numerically, for $E_{max} = 1.5, 4$ and $+\infty$, taking $E_0 = 1$). Figure 1(b) depicts the cumulative density function (CDF) of the corresponding instantaneous mutual information $I(r) \triangleq \log(1 + \rho r^2)$, for signal-to-noise ratio $\rho = 15$ dB. The lowest range of the CDF is of particular interest for wireless communications since it represents the probability of a channel outage for a given transmission rate. The curves clearly show that the models corresponding to the unknown energy have a lower outage capacity that the Gaussian channel model.

5 Space correlation models

In this section, we shall incorporate several states of knowledge about space correlation characteristics 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) PDF of amplitude r

(b) CDF of instantaneous mutual information $I(r)$

Figure 1: Amplitude and mutual information distributions of the proposed SISO channel models.

5.1 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 $\mathbf{Q} \triangleq \int_{\mathbb{C}^N} \mathbf{h}\mathbf{h}^H P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H}$ is known, where $\mathbf{Q}$ is a $N \times N$ complex Hermitian matrix. Each component of the covariance constraint represents an independent linear constraint of the form

$$\int_{\mathbb{C}^N} h_a h_b^* P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H} = q_{a,b} \tag{13}$$

for $(a,b) \in [1,\ldots,N]^2$. Note that this constraint makes any previous energy constraint redundant since $\int_{\mathbb{C}^N} ||\mathbf{H}||_F^2 P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H} = \mathrm{tr}(\mathbf{Q})$. Proceeding along the lines of the method exposed previously, we introduce N^2 Lagrange coefficients $\alpha_{a,b}$, and maximize

$$\begin{aligned} L(P_{\mathbf{H}|\mathbf{Q}}) = \int_{\mathbb{C}^N} -\log(P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}))P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H} \quad &+ \quad \beta\left[1 - \int_{\mathbb{C}^N} P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H}\right] \\ &+ \sum_{\substack{a\in[1,\ldots,N] \\ b\in[1,\ldots,N]}} \alpha_{a,b}\left[\int_{\mathbb{C}^N} h_a h_b^* P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})\mathrm{d}\mathbf{H} - q_{a,b}\right]. \end{aligned} \tag{14}$$

Denoting $\mathbf{A} = [\alpha_{a,b}]_{(a,b)\in[1,\ldots,N]^2}$ the $N \times N$ matrix of the Lagrange multipliers, the derivative is

$$\frac{\delta L(P_{\mathbf{H}|\mathbf{Q}})}{\delta P_{\mathbf{H}|\mathbf{Q}}} = -\log(P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H})) - 1 - \beta - \mathbf{h}^T\mathbf{A}\mathbf{h}^* = 0. \tag{15}$$

Therefore, $P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}) = \exp\left(-(\beta+1) - \mathbf{h}^T\mathbf{A}\mathbf{h}^*\right)$, or, after elimination of the Lagrange coefficients through proper normalization,

$$P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}) = \frac{1}{\det(\pi\mathbf{Q})}\exp\left(-(\mathbf{h}^H\mathbf{Q}^{-1}\mathbf{h})\right). \tag{16}$$

Again, the maximum entropy principle yields a Gaussian distribution, although of course its components are not independent anymore.

5.2 Knowledge of the existence of a correlation matrix

It was shown in Section 3 that in the absence of information on space correlation, maximum entropy modeling yields i.i.d. coefficients for the channel matrix, and therefore an identity covariance matrix.

We now consider the case where covariance is known to be a parameter of interest, but is not known deterministically. Again, 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 $\mathbf{Q}$.

Let us first establish the distribution of $\mathbf{Q}$, under the energy constraint $\int \mathrm{tr}(\mathbf{Q})P_{\mathbf{Q}}(\mathbf{Q})\mathrm{d}\mathbf{Q} = NE_0$, by maximizing the functional

$$L(P_{\mathbf{Q}}) = \int_{\mathcal{S}} -\log(P_{\mathbf{Q}}(\mathbf{Q}))P_{\mathbf{Q}}(\mathbf{Q})\mathrm{d}\mathbf{Q} + \beta\left[\int_{\mathcal{S}} P_{\mathbf{Q}}(\mathbf{Q})\mathrm{d}\mathbf{Q} - 1\right] + \gamma\left[\int_{\mathcal{S}} \mathrm{tr}(\mathbf{Q})P_{\mathbf{Q}}(\mathbf{Q})\mathrm{d}\mathbf{Q} - NE_0\right]. \quad (17)$$

Due to their structure, covariance matrices are restricted to the space $\mathcal{S}$ of $N \times N$ positive semidefinite complex matrices. Therefore, let us perform the variable change to the eigenvalues/eigenvectors space. Specifically, let us denote $\Lambda \triangleq \mathrm{diag}(\lambda_1 \ldots \lambda_N)$ the diagonal matrix containing the eigenvalues of $\mathbf{Q}$, and let $\mathbf{U}$ be the unitary matrix containing the eigenvectors, such that $\mathbf{Q} = \mathbf{U}\Lambda\mathbf{U}^H$.

We use the mapping between the space of complex $N \times N$ self-adjoint matrices (of which $\mathcal{S}$ is a subspace), and $\mathcal{U}(N)/T \times \mathbb{R}^N_{\leq}$, where $\mathcal{U}(N)/T$ denotes the space of unitary $N \times N$ matrices with real, non-negative first row, and $\mathbb{R}^N_{\leq}$ is the space of real N-tuples with non-decreasing components (see [10], Lemma 4.4.6). The positive semidefinite property of the covariance matrices further restricts the components of Λ to non-negative values, and therefore $\mathcal{S}$ maps into $\mathcal{U}(N)/T \times \mathbb{R}^{+N}_{\leq}$. Eq. (17) becomes

$$\begin{aligned} L(P_{\mathbf{U},\Lambda}) &= \int_{\mathcal{U}(N)/T\times\mathbb{R}^{+N}_{\leq}} -\log(P_{\mathbf{U},\Lambda}(\mathbf{U},\mathbf{\Lambda}))P_{\mathbf{U},\Lambda}(\mathbf{U},\mathbf{\Lambda})K(\Lambda)\mathrm{d}\mathbf{U}\mathrm{d}\Lambda \\ &+ \beta\left[\int_{\mathcal{U}(N)/T\times\mathbb{R}^{+N}_{\leq}} P_{\mathbf{U},\Lambda}(\mathbf{U},\mathbf{\Lambda})K(\Lambda)\mathrm{d}\mathbf{U}\mathrm{d}\Lambda - 1\right] \\ &+ \gamma\left[\int_{\mathcal{U}(N)/T\times\mathbb{R}^{+N}_{\leq}} \left(\sum_{i=1}^{N}\lambda_i\right) P_{\mathbf{U},\Lambda}(\mathbf{U},\mathbf{\Lambda})K(\Lambda)\mathrm{d}\mathbf{U}\mathrm{d}\Lambda - NE_0\right], \end{aligned} \quad (18)$$

where we introduced the corresponding Jacobian $K(\Lambda) \triangleq \frac{(2\pi)^{N(N-1)/2}}{\prod_{j=1}^{N-1} j!}\prod_{i<j}(\lambda_i - \lambda_j)^2$, and used $\mathrm{tr}(\mathbf{Q}) = \mathrm{tr}(\Lambda) = \sum_{i=1}^{N}\lambda_i$. Maximizing the entropy of the distribution $P_{\mathbf{U},\Lambda}$ by taking $\frac{\delta L(P_{\mathbf{U},\Lambda})}{\delta P_{\mathbf{U},\Lambda}} = 0$ yields

$$-K(\Lambda) - K(\Lambda)\log(P_{\mathbf{U},\Lambda}(\mathbf{U},\mathbf{\Lambda})) + \beta K(\Lambda) + \gamma\left(\sum_{i=1}^{N}\lambda_i\right)K(\Lambda) = 0. \quad (19)$$

Since $K(\Lambda) \neq 0$ except on a set of measure zero, this is equivalent to

$$P_{\mathbf{U},\Lambda}(\mathbf{U},\mathbf{\Lambda}) = \exp\left(\beta - 1 + \gamma\sum_{i=1}^{N}\lambda_i\right). \quad (20)$$

Since this distribution does not explicitly depend on $\mathbf{U}$, it can be factored as $P_{\mathbf{U},\Lambda}(\mathbf{U},\mathbf{\Lambda}) = P_{\mathbf{U}}P_{\mathbf{\Lambda}}(\mathbf{\Lambda})$ where $P_{\mathbf{U}}$ is the (constant) density of $\mathbf{U}$ on $\mathcal{U}(N)/T$, and $P_{\mathbf{\Lambda}}(\mathbf{\Lambda}) \triangleq C\exp\left(\gamma\sum_{i=1\ldots N}\lambda_i\right)$ where C is a constant, is the distribution of the eigenvalues on $\mathbb{R}^{+N}_{\leq}$. The unknowns can be eliminated by solving the normalization equations $\int_{\mathcal{U}(N)/T} P_{\mathbf{U}}\mathrm{d}\mathbf{U} = 1$, $\int_{\mathcal{U}(N)/T\times\mathbb{R}^{+N}_{\leq}} P_{\mathbf{U}}P_{\mathbf{\Lambda}}(\mathbf{\Lambda})K(\Lambda)\mathrm{d}\mathbf{U}\mathrm{d}\Lambda = 1$, $\int_{\mathcal{U}(N)/T\times\mathbb{R}^{+N}_{\leq}} \left(\sum_{i=1}^{N}\lambda_i\right) P_{\mathbf{U}}P_{\mathbf{\Lambda}}(\mathbf{\Lambda})K(\Lambda)\mathrm{d}\mathbf{U}\mathrm{d}\Lambda = NE_0$, and $P_{\mathbf{U}}C = \exp(\beta - 1)$. Note that the uniform distribution on $\mathcal{U}(N)/T$ can be easily generated from Gram-Schmidt orthogonalization (and proper normalization of the first row) of a standard complex Gaussian matrix.

So far, the maximum entropy distribution of $\mathbf{Q} = \mathbf{U}\Lambda\mathbf{U}^H$ has been established to be a product distribution, with Λ distributed according to the exponential distribution $P_{\mathbf{\Lambda}}(\mathbf{\Lambda})$ on $\mathbb{R}^{+N}_{\leq}$, and $\mathbf{U}$ uniformly distributed. The distribution of $\mathbf{H}$ is finally obtained through

$$P_{\mathbf{H}}(\mathbf{H}) = \int_{\mathcal{S}} P_{\mathbf{H}|\mathbf{Q}}(\mathbf{H}) P_{\mathbf{Q}}(\mathbf{Q}) \mathrm{d}\mathbf{Q} = \int_{\mathcal{U}(N)/T \times \mathbb{R}_{\leq}^{+N}} P_{\mathbf{H}|\mathbf{U},\Lambda}(\mathbf{H}) P_{\mathbf{U}} P_{\mathbf{\Lambda}}(\mathbf{\Lambda}) K(\Lambda) \mathrm{d}\mathbf{U} \mathrm{d}\Lambda \quad (21)$$

$$= \int_{\mathcal{U}(N)/T \times \mathbb{R}_{\leq}^{+N}} C \prod_{i=1}^{N} \left(\lambda_i^{-1} \exp\left(\gamma \lambda_i - \frac{|\mathbf{u}_i^H \mathbf{h}|^2}{\lambda_i} \right) \right) \frac{(2\pi)^{N(N-1)/2}}{\pi^N \prod_{j=1}^{N-1} j!} \prod_{i<j} (\lambda_i - \lambda_j)^2 P_{\mathbf{U}} \mathrm{d}\mathbf{U} \mathrm{d}\Lambda, \quad (22)$$

where $\mathbf{u}_i$ denotes the i-th column of $\mathbf{U}$ and $P_{\mathbf{H}|\mathbf{U},\mathbf{\Lambda}}(\mathbf{h}) = \frac{1}{\pi^N} \prod_{i=1}^{N} \lambda_i^{-1} \exp\left(-\frac{|\mathbf{u}_i^H \mathbf{h}|^2}{\lambda_i} \right)$ comes from eq. (16).

6 Conclusion

We proposed analytical models for wireless channels using the maximum entropy method, when the constraints are expressed in terms of channel energy and spatial correlation. When the channel energy is unknown, numerical simulations show that the proposed model correctly exhibits a reduced mutual information w.r.t. the known energy case. When the spatial correlation matrix is unknown, but is known to be a preponderant parameter, we have shown that the maximum entropy method leads to convenient analytical expressions in the eigenspace of the covariance matrix.

Natural extensions of this method to account for other peculiarities of the wireless channel include the modeling of time correlation, by jointly modeling a number of successive channel realizations to take into account the time-varying properties of the channel, or frequency correlation in the case of a frequency-selective channel.

References

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, March 1998.

[2] C. Berrou, C. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo codes," in *Proc. IEEE International Conference on Communications (ICC)*, 1993, pp. 1064–1070.

[3] 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/`.

[4] P. Kyritsi, D. C. Cox, R. A. Valenzuela, and P. W. Wolniansky, "Correlation analysis based on MIMO channel measurements in an indoor environment," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 713–720, June 2003.

[5] Alister G. Burr, "Capacity bounds and estimates for the finite scatterers MIMO wireless channel," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 812–818, June 2003.

[6] 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*, to appear.

[7] 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.

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

[9] S. V. Fomin and I. M. Gelfand, *Calculus of Variations*, Dover Publications, 2000.

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