LOW SNR ANALYSIS OF THE NON-COHERENT MIMO CHANNEL UNDER ARBITRARY CHANNEL AND NOISE CORRELATION STRUCTURES

*Marko Beko**

Instituto Superior Técnico – Instituto de Sistemas e Robótica
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
Escola Superior de Tecnologia e Gestão – Instituto Politécnico de Leiria
Morro do Lena – Alto do Vieiro, 2411-901 Leiria, Portugal
marko@isr.ist.utl.pt

João Xavier and Victor Barroso

Instituto Superior Técnico – Instituto de Sistemas e Robótica
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
{jxavier,vab}@isr.ist.utl.pt

ABSTRACT

The non-coherent single-user multiple-input multiple-output (MIMO) channel in the low signal-to-noise ratio (SNR) regime is analyzed from the capacity point of view. We investigate the impact of channel and noise correlation on the mutual information for the on-off and Gaussian signaling schemes. The novelty is that we allow an arbitrary structure for both the channel and noise correlation matrices. Our results establish that, in the low SNR regime, mutual information is maximized when the channel correlation matrix is of rank one.

1. INTRODUCTION

In slowly fading scenarios, channel stability allows the receiver to be trained in order to obtain the channel state information (CSI) necessary for *coherent detection* of the transmitted codeword. The scope of this paper will be fast fading scenarios, where the channel coefficients change too quickly to allow reliable channel estimation. Hence, CSI is no more available, and the receiver must operate in a *non-coherent* mode. Furthermore, we focus on the low signal-to-noise ratio (SNR) regime. This is due to the fact that a variety of digital communication systems (more specifically in wireless, sensor and satellite networks) operate in the power-limited region. See [2, 3] for a more thorough discussion of this topic.

Previous work. In [3], low SNR MIMO systems when CSI is available at the receiver have been considered. The interplay of rate, bandwidth, and power is analyzed in the region of energy per bit close to its minimum value. In [4], the scenario where no CSI is available at the receiver has been treated. It has been demonstrated that the optimal signaling at low SNR achieves the same capacity as the known channel case for single transmit antenna systems. In [5], it has been shown that knowledge of the first and second derivatives of capacity at low SNR gives us insight on bandwidth and energy efficiency for signal transmission. More precisely, these quantities tell us how spectral efficiency grows with energy-per-bit. In [6], a formula for the second-order expansion of the input-output mutual information at low SNR is obtained, whereas in [7] the capacity and the reliability function as the peak constraint tends to zero are analyzed for a discrete-time memoryless channel with peak constrained inputs. In [8], Rao and Hassibi have shown that the on-off signaling presented in [4] generalizes to the multi-antenna setting and attains the known channel capacity. The tradeoff between communication rate and average probability of decoding error using a framework of error-exponent theory has been considered in [9]. It is argued that the advantage of having multiple antennas is best realized when the fading is fully correlated, i.e., a performance gain of MN and a peakiness gain of M^2NT can be achieved where M, N and T represent the number of transmit, receive antennas, and the length of the coherence interval, respectively. In [2], the approach in [8] has been extended as both noise and channel correlation have been taken in account.

Contribution and paper organization. We study the non-coherent MIMO channel in the low SNR regime from the capacity viewpoint. The novel aspect is that we allow the channel covariance matrix to have an arbitrary correlation structure. This, together with the fact that the noise covariance matrix is also allowed to have an arbitrary structure, makes this scenario the most comprehensive and challenging one. The data model is introduced in section 2. In section 3, the spatially correlated non-coherent MIMO block Rayleigh fading channel is analyzed and the impact of channel and noise correlation on the mutual information is obtained for the on-off signaling. In section 4, we extend the analysis presented in section 3 to the case of Gaussian signaling. The main conclusion is that for both cases mutual information is maximized when the channel correlation matrix is fully correlated. We also argue that the on-off signaling is optimal for this multi-antenna setting. Some mathematical details are left to an ap-

*This work was partially supported by Fundação para a Ciência e a Tecnologia (ISR/IST plurianual funding) through the POS Conhecimento Program that includes FEDER funds, and by FCT PhD Grant SFRH/BD/12809/2003, and by IST/ISR Plurianual Unidade 101.

pendix.

2. DATA MODEL

We focus on a communication system consisting of M transmit and N receive antennas over a narrowband flat Rayleigh fading channel. A block fading channel model which is widely used in the MIMO literature [8, 9]with coherence interval T, is adopted. In complex base band notation we have the system model

$$\boldsymbol{X} = \boldsymbol{S}\boldsymbol{H} + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{S}$ is the $T \times M$ matrix of transmitted symbols, $\boldsymbol{X}$ is the $T \times N$ matrix of received symbols, $\boldsymbol{H}$ is the $M \times N$ channel matrix, and $\boldsymbol{E}$ is the $T \times N$ matrix of zero-mean additive observation noise. We work under the following assumptions: **A1 (Channel matrix)** The stochastic model for the channel matrix $\boldsymbol{H}$ is such that $\text{vec}(\boldsymbol{H}) = \sqrt{\frac{\rho}{M}}\boldsymbol{K}^{\frac{1}{2}}\boldsymbol{h}_w$, where the vector $\boldsymbol{h}_w$ is an $MN \times 1$ vector comprised of statistically independent $\mathcal{CN}(0,1)$ entries and the matrix $\boldsymbol{K}$ is an $MN \times MN$ positive semidefinite Hermitian correlation matrix. The coefficient ρ is the model parameter proportional to the SNR and $\text{vec}(\boldsymbol{H})$ stacks all columns of the matrix $\boldsymbol{H}$ on the top of each other, from left to right. We assume that the vector $\boldsymbol{h}_w$ remains fixed for the coherence time T after which it changes to a new independent value. Also, the vector $\boldsymbol{h}_w$ is not known at the receiver neither at the transmitter, but its distribution is. Further, we view the channel correlation matrix $\boldsymbol{K}$ as a system parameter which we can introduce and track. The generalization to arbitrary channel covariance matrices $\boldsymbol{K}$ comprises many scenarios of interest as special cases: the popular separable (Kronecker) spatial correlation model [2, 3], a recently proposed spatial correlation model that takes into account coupling between transmit and receive sides [10], uncorrelated rayleigh fading channel model [8], etc. For a fair comparison of different correlation cases, we assume that $\text{tr}(\boldsymbol{K}) = MN$; **A2 (Transmit power constraint)** We impose the power constraint $\mathsf{E}[\text{tr}(\boldsymbol{S}^H\boldsymbol{S})] \leq TM$; **A3 (Noise distribution)** The noise covariance matrix $\boldsymbol{\Upsilon} = \mathsf{E}[\text{vec}(\boldsymbol{E})\ \text{vec}(\boldsymbol{E})^H]$ is known at the transmitter and at the receiver. Without loss of generality (w.l.o.g.), we assume $\text{tr}(\boldsymbol{\Upsilon}) = NT$. In **A3**, as in [2], we let the data model depart from the customary assumption of spatio-temporal white Gaussian observation noise. In real scenarios the $\boldsymbol{E}$ term often has a very rich correlation structure, e.g, see [3] and pp. 100 in [11]. In [12], a methodology for designing space-time codebooks for non-coherent communications in multiple-antenna wireless systems and an arbitrary given noise covariance matrix $\boldsymbol{\Upsilon}$ in the high SNR regime has been proposed.

3. MUTUAL INFORMATION: ON-OFF SIGNALING

In [8], it has been demonstrated that the on-off signaling presented in [4], where the single transmit antenna systems were considered, generalizes to the multi-antenna setting and attains the known channel capacity. In [1, 2], the authors have extended the results from [8] to the case of the correlated Rayleigh fading channel model that obeys the Kronecker structure, with arbitrary noise covariance matrix. Here, we argue that this is also the case for the correlated Rayleigh fading channel model with an arbitrary channel covariance matrix (as in [1, 2], the noise term is allowed to have an arbitrary covariance matrix). Moreover, we maximize the mutual information with respect to (w.r.t.) the input signal $\boldsymbol{S}_{on}$ and $\boldsymbol{K}$. The on-off signaling is defined as: for any $\epsilon > 1$ and assuming $\rho < 1$,

$$\boldsymbol{S} = \begin{cases} \boldsymbol{S}_{on}\,\rho^{-\frac{\epsilon}{2}} & \text{with probability (w.p.) } \rho^{\epsilon} \\ \boldsymbol{0}_{T\times M} & \text{w.p. } 1-\rho^{\epsilon} \end{cases}$$

With an analysis similar to [8], not presented here due to lack of space, it can be shown that at sufficiently low SNR the mutual information between $\boldsymbol{X}$ and $\boldsymbol{S}$ up to first order in ρ is given by

$$I(\boldsymbol{X};\boldsymbol{S}) = \frac{\rho}{M}\,\text{tr}\left(\boldsymbol{\Upsilon}^{-1}\left(\boldsymbol{I}_N \otimes \boldsymbol{S}_{on}\right)\boldsymbol{K}\left(\boldsymbol{I}_N \otimes \boldsymbol{S}_{on}^H\right)\right) + o(\rho), \tag{2}$$

where $\otimes$ denotes the Kronecker product. Remark that for the special case of the Kronecker correlation model, i.e., the channel covariance matrix $\boldsymbol{K}$ is of the form $\boldsymbol{K} = \boldsymbol{K}_r \otimes \boldsymbol{K}_t$ for some Hermitian positive semidefinite matrices $\boldsymbol{K}_r$ and $\boldsymbol{K}_t$, our result in (2) recovers the finding in [1, 2]. Now, we address the maximization of the mutual information w.r.t. $\boldsymbol{S}_{on}$ and $\boldsymbol{K}$, i.e.,

$$\begin{array}{c} \max \\ \text{tr}\left(\boldsymbol{S}_{on}\boldsymbol{S}_{on}^H\right) \leq TM \\ \boldsymbol{K} \in \mathcal{P}_{MN} \end{array} \quad \text{tr}\left(\boldsymbol{\Upsilon}^{-1}\overline{\boldsymbol{S}}_{on}\boldsymbol{K}\overline{\boldsymbol{S}}_{on}^H\right) \tag{3}$$

where $\mathcal{P}_n = \{\boldsymbol{Y} : n\times n \text{ matrix such that } \boldsymbol{Y} \succeq \boldsymbol{0} \text{ and } \text{tr}\,(\boldsymbol{Y}) = n\}$ and $\overline{\boldsymbol{S}}_{on} = \boldsymbol{I}_N \otimes \boldsymbol{S}_{on}$. Carrying out an analysis similar to the one presented in [2], it can be shown that the maximum of (3) is attained by:

$$\widehat{\boldsymbol{S}}_{on} = \text{ivec}(\widehat{\boldsymbol{s}}_{on}),\ \widehat{\boldsymbol{K}} = MN\widehat{\boldsymbol{u}}\widehat{\boldsymbol{u}}^H, \tag{4}$$

where

$$(\widehat{\boldsymbol{u}}, \widehat{\boldsymbol{s}}_{on}) = \begin{array}{c} \arg\max \\ ||\boldsymbol{u}|| = 1, \\ ||\boldsymbol{s}_{on}|| \leq \sqrt{TM} \end{array} \quad \boldsymbol{s}_{on}^H\boldsymbol{M}_{on}\boldsymbol{s}_{on} \tag{5}$$

with

$$\boldsymbol{M}_{on} = \boldsymbol{K}_{on}^H\left(\left(\boldsymbol{u}\boldsymbol{u}^H\right)^T \otimes \boldsymbol{\Upsilon}^{-1}\right)\boldsymbol{K}_{on}, \tag{6}$$

$\boldsymbol{u} \in \mathbb{C}^{MN}$ and $\boldsymbol{s}_{on} = \text{vec}\,(\boldsymbol{S}_{on}) \in \mathbb{C}^{TM}$. The operation "ivec" operates as an inverse of "vec" (in this case, reshapes the TM-dimensional vector into a $T\times M$ matrix). The matrix $\boldsymbol{K}_{on}$ is the $TMN^2\times TM$ matrix such that $\text{vec}\,(\boldsymbol{I}_N \otimes \boldsymbol{S}_{on}) =$

$\boldsymbol{K}_{on}\text{vec}\left(\boldsymbol{S}_{on}\right)$. The proof of (4) is omitted due to paper length constraints. Essentially, it uses the fact that the maximum of a convex function over a compact, convex set is attained at extreme points. The optimization problem in (5) always admits a solution (maximization of a continuous function over a compact set) but, in general, a closed form solution is not available. For the choice in (4), the maximal mutual information (per channel use (p.c.u.)) is given by

$$\frac{1}{T}I(\boldsymbol{X};\boldsymbol{S}) = \rho\, N\, M\hat{\lambda} + o(\rho) \tag{7}$$

where

$$\hat{\lambda} = \widehat{\underline{\boldsymbol{s}}}^H \boldsymbol{K}_{on}^H \left(\left(\widehat{\boldsymbol{u}}\widehat{\boldsymbol{u}}^H\right)^T \otimes \boldsymbol{\Upsilon}^{-1}\right)\boldsymbol{K}_{on}\widehat{\underline{\boldsymbol{s}}} \tag{8}$$

and $\widehat{\underline{\boldsymbol{s}}} = 1/\sqrt{TM}\,\widehat{\boldsymbol{s}}_{on}$. Note that $\text{tr}\left(\boldsymbol{\Upsilon}^{-1}\right) = \sum_{i=1}^{NT} 1/\alpha_i \geq \sum_{i=1}^{NT}(2-\alpha_i) = NT$, where α_i's are the eigenvalues of $\boldsymbol{\Upsilon}$. Let $\left[\boldsymbol{\Upsilon}^{-1}\right]_{ij} = \boldsymbol{E}_i\boldsymbol{\Upsilon}^{-1}\boldsymbol{E}_j^T$ where $i,j = 1,...,N$, $\boldsymbol{E}_i = \boldsymbol{e}_i^T \otimes \boldsymbol{I}_T$ and $\boldsymbol{e}_i$ represents the i-th column of $\boldsymbol{I}_N$. Thus, we can w.l.o.g. assume that, e.g., $\lambda_{\mathsf{max}}\left(\left[\boldsymbol{\Upsilon}^{-1}\right]_{11}\right) \geq 1$ (otherwise $\text{tr}\left(\boldsymbol{\Upsilon}^{-1}\right) = \sum_i^N \text{tr}\left(\left[\boldsymbol{\Upsilon}^{-1}\right]_{ii}\right) < \sum_i^N T = NT$). Then, by choosing $\underline{\boldsymbol{u}} = \begin{bmatrix}1 & \boldsymbol{0}_{1\times(MN-1)}\end{bmatrix}^T$ we have

$$\begin{aligned}
\hat{\lambda} &\geq \max_{||\underline{\boldsymbol{s}}|| \leq 1} \underline{\boldsymbol{s}}^H \boldsymbol{K}_{on}^H \left(\left(\underline{\boldsymbol{u}}\,\underline{\boldsymbol{u}}^H\right)^T \otimes \boldsymbol{\Upsilon}^{-1}\right)\boldsymbol{K}_{on}\underline{\boldsymbol{s}} \\
&= \lambda_{\mathsf{max}}\left(\boldsymbol{K}_{on}^H\left(\underline{\boldsymbol{u}}^T \otimes \boldsymbol{I}_{TN}\right)^H \boldsymbol{\Upsilon}^{-1}\left(\underline{\boldsymbol{u}}^T \otimes \boldsymbol{I}_{TN}\right)\boldsymbol{K}_{on}\right) \qquad (9)\\
&= \lambda_{\mathsf{max}}\left(\left(\underline{\boldsymbol{U}}^T \otimes \boldsymbol{I}_T\right)^H \boldsymbol{\Upsilon}^{-1}\left(\underline{\boldsymbol{U}}^T \otimes \boldsymbol{I}_T\right)\right) \qquad (10)\\
&= \lambda_{\mathsf{max}}\left(\left[\boldsymbol{\Upsilon}^{-1}\right]_{11}\right)
\end{aligned}$$

where $\underline{\boldsymbol{U}} = \text{ivec}\left(\underline{\boldsymbol{u}}\right)$ (i.e., we reshape the $MN\times 1$ vector $\underline{\boldsymbol{u}}$ into the $M\times N$ matrix $\underline{\boldsymbol{U}}$). The passage from (9) to (10) is valid due to the following equality: for every $M\times N$ matrix $\boldsymbol{U}$ it holds

$$\boldsymbol{K}_{on}^T\left(\boldsymbol{u}\otimes\boldsymbol{I}_{TN}\right) = \boldsymbol{U}\otimes\boldsymbol{I}_T \tag{11}$$

where $\boldsymbol{u} = \text{vec}\left(\boldsymbol{U}\right)$. The proof of (11) is given in Appendix A. Thus, $\hat{\lambda} \geq 1$. This result confirms the general principle that correlated noise is beneficial from the capacity point of view. See, e.g., pp. 100 in [11] for more details. This is also in agreement with [2] where the case of the Kronecker fading model has been treated.

Fixed eigenvectors of $\boldsymbol{K}$. In practice, by changing the antenna separation one can control the eigenvalues of $\boldsymbol{K}$, but not their eigenvectors. See [9], [11, section 6.3.3]. As a consequence, the result presented in (7) has to be interpreted as the upper bound on the channel capacity. In case where we cannot control the eigenvectors of $\boldsymbol{K}$ the optimization problem defined in (3) becomes

$$\left(\widehat{\boldsymbol{S}}_{on}, \widehat{\boldsymbol{\Lambda}}\right) = \underset{\substack{||\boldsymbol{S}_{on}|| \leq \sqrt{TM}\\ \boldsymbol{\Lambda}\in\mathcal{D}_{MN}}}{\arg\max}\ \text{tr}\left(\boldsymbol{\Upsilon}^{-1}\overline{\boldsymbol{S}}_{on}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H\overline{\boldsymbol{S}}_{on}^H\right)$$

where $\mathcal{D}_n = \{\boldsymbol{Y} : n\times n \text{ diagonal matrix such that } \boldsymbol{Y} \succeq \boldsymbol{0} \text{ and } \text{tr}\left(\boldsymbol{Y}\right) = n\}$ and $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$ represents the eigenvalue decomposition (EVD) of the matrix $\boldsymbol{K}$. Let $\boldsymbol{u}_i$'s (λ_i's) be the eigenvectors (eigenvalues) of $\boldsymbol{K}$ and let's define $\boldsymbol{M}_{on}^i = \boldsymbol{K}_{on}^H\left(\left(\boldsymbol{u}_i\boldsymbol{u}_i^H\right)^T \otimes \boldsymbol{\Upsilon}^{-1}\right)\boldsymbol{K}_{on}$. Then, we can write

$$\begin{aligned}
&\max_{\substack{||\boldsymbol{S}_{on}|| \leq \sqrt{TM}\\ \boldsymbol{\Lambda}\in\mathcal{D}_{MN}}} \text{tr}\left(\boldsymbol{\Upsilon}^{-1}\overline{\boldsymbol{S}}_{on}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H\overline{\boldsymbol{S}}_{on}^H\right)\\
=\ &\max_{\substack{||\boldsymbol{S}_{on}|| \leq \sqrt{TM}\\ \boldsymbol{\Lambda}\in\mathcal{D}_{MN}}} \text{vec}^H\left(\overline{\boldsymbol{S}}_{on}\right)\text{vec}\left(\boldsymbol{\Upsilon}^{-1}\overline{\boldsymbol{S}}_{on}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H\right)\\
=\ &\max_{||\boldsymbol{s}_{on}|| \leq \sqrt{TM}}\ \max_{\boldsymbol{\Lambda}\in\mathcal{D}_{MN}}\ \sum_i \lambda_i \boldsymbol{s}_{on}^H\boldsymbol{M}_{on}^i\boldsymbol{s}_{on} \qquad (12)\\
=\ &\max_{||\boldsymbol{s}_{on}|| \leq \sqrt{TM}}\left[MN\ \max_i\ \boldsymbol{s}_{on}^H\boldsymbol{M}_{on}^i\boldsymbol{s}_{on}\right] \qquad (13)\\
=\ &M^2NT\ \max_i\ \lambda_{\mathsf{max}}\left(\boldsymbol{M}_{on}^i\right). \qquad (14)
\end{aligned}$$

The justification behind (14) is the following. It is easy to see that for a given unit-norm vector $\boldsymbol{u}$, the maximum of (5) is attained at an extreme point of the convex, compact set

$$\Delta = \{\boldsymbol{s} : TM\times 1 \text{ vectors such that } ||\boldsymbol{s}|| \leq \sqrt{TM}\}$$

(remark that the objective function in (5) is a convex function since $\boldsymbol{M}_{on}^i = \left(\boldsymbol{M}_{on}^i\right)^H \succeq \boldsymbol{0}$). The same line of reasoning can be used to establish the equivalence between (12) and (13). In this case, the objective function in (12) is a linear combination of λ_i's and defined on the convex, compact set $\Omega = \{(\lambda_1,...,\lambda_{MN}) : MN\times 1 \text{ vectors such that } \sum_{i=1}^{MN}\lambda_i = MN \text{ and } \lambda_i \geq 0\}$. It is readily seen that the extreme points of Ω are the vectors whose all entries are zero except one which is equal to MN. Thus, we have

$$\widehat{\boldsymbol{S}}_{on} = \sqrt{TM}\,\text{ivec}(\boldsymbol{s}_{\mathsf{max}}),\ \widehat{\boldsymbol{K}} = MN\boldsymbol{u}_{i^*}\boldsymbol{u}_{i^*}^H, \tag{15}$$

where $\boldsymbol{s}_{\mathsf{max}}$ is an unit-norm eigenvector associated to the maximal eigenvalue of

$$\boldsymbol{M}_{on}^{i^*} = \boldsymbol{K}_{on}^H\left(\left(\boldsymbol{u}_{i^*}\boldsymbol{u}_{i^*}^H\right)^T \otimes \boldsymbol{\Upsilon}^{-1}\right)\boldsymbol{K}_{on} \tag{16}$$

and

$$i^* = \underset{i=1,2,...,MN}{\arg\max}\ \lambda_{\mathsf{max}}\left(\boldsymbol{M}_{on}^i\right). \tag{17}$$

For the choice in (15), the maximal mutual information (p.c.u.) is given by

$$\frac{1}{T}I(\boldsymbol{X};\boldsymbol{S}) = \rho\, N\, M\lambda_{\mathsf{max}}\left(\boldsymbol{M}_{on}^{i^*}\right) + o(\rho).$$

Remarks. From (4) and (15) it is clear that the channel covariance matrix should be made as correlated as possible, as

the optimal $\boldsymbol{K}$ has rank one. A short exercise would show that the first order term in (7) corresponds to that of the capacity when the channel is known to the receiver. For the coherent case, the mutual information is maximized if the input vec($\boldsymbol{S}$) is circularly symmetric, complex Gaussian distribution, i.e., vec($\boldsymbol{S}$)$\sim\mathcal{CN}(\boldsymbol{0},\boldsymbol{P}_{coh})$ for some covariance matrix $\boldsymbol{P}_{coh}$ with tr($\boldsymbol{P}_{coh}$) $\leq TM$ such that the power constraint in the assumption **A2** is satisfied. In that case, by maximizing the first-order expansion of the very well known expectation-log-det expression for the capacity w.r.t. $\boldsymbol{P}_{coh}$ and $\boldsymbol{K}$ it can be shown that the optimal correlation matrix $\widehat{\boldsymbol{K}}$ is defined as in (4) and the optimal covariance matrix $\widehat{\boldsymbol{P}}_{coh}$ is given by

$$\widehat{\boldsymbol{P}}_{coh} = \widehat{\boldsymbol{s}}_{on}\widehat{\boldsymbol{s}}_{on}^{H}, \tag{18}$$

with $\widehat{\boldsymbol{s}}_{on}$ defined as in (5).

4. MUTUAL INFORMATION: GAUSSIAN MODULATION

In this section, we compute the low SNR mutual information for the more realistic and practical case of Gaussian modulation. Let $\boldsymbol{s}$ = vec($\boldsymbol{S}$) be a zero-mean random variable with covariance matrix $\boldsymbol{P}$ that follows a circularly symmetric, complex Gaussian distribution, i.e., $\boldsymbol{s}\sim\mathcal{CN}(\boldsymbol{0},\boldsymbol{P})$. Clearly, in order to satisfy the power constraint in the assumption **A2**, tr($\boldsymbol{P}$) $\leq TM$. Then, at sufficiently low SNR, the mutual information between $\boldsymbol{X}$ and $\boldsymbol{S}$ up to second order in ρ is given by

$$I(\boldsymbol{X};\boldsymbol{S}) \;=\; \frac{\rho^2}{2M^2}\,\mathrm{tr}\left(\mathsf{E}[\boldsymbol{Z}^2]-(\mathsf{E}[\boldsymbol{Z}])^2\right)+o(\rho^2), \tag{19}$$

where $\boldsymbol{Z}=\boldsymbol{\Upsilon}^{-\frac{1}{2}}\overline{\boldsymbol{S}}_{on}\boldsymbol{K}\overline{\boldsymbol{S}}_{on}^{H}\boldsymbol{\Upsilon}^{-\frac{1}{2}}$. The proof of (19) is omitted due to paper length constraints. We now address the optimization problem

$$\max_{\boldsymbol{P}\,\in\,\mathcal{H}_{TM},\,\boldsymbol{K}\,\in\,\mathcal{P}_{MN}} \quad \mathrm{tr}\left(\mathsf{E}[\boldsymbol{Z}^2]-(\mathsf{E}[\boldsymbol{Z}])^2\right) \tag{20}$$

where $\mathcal{H}_n=\{\boldsymbol{P}: n\times n$ matrices such that $\boldsymbol{P}=\boldsymbol{P}^H\succeq \boldsymbol{0}$ and tr$(\boldsymbol{P})\leq n\}$. It can be shown (proof omitted) that the maximum of (20) is attained by the following signaling scheme: the optimal correlation matrix $\widehat{\boldsymbol{K}}$ is defined as in (4), and the optimal covariance matrix $\widehat{\boldsymbol{P}}$ is given by

$$\widehat{\boldsymbol{P}} = \widehat{\boldsymbol{s}}_{on}\widehat{\boldsymbol{s}}_{on}^{H} \tag{21}$$

where the vectors $\widehat{\boldsymbol{u}}$ and $\widehat{\boldsymbol{s}}_{on}$ are, as before, solutions of the optimization problem (5). In this case, the mutual information (p.c.u.) is given by

$$\frac{1}{T}I(\boldsymbol{X};\boldsymbol{S}) = \frac{\rho^2}{2}\,N^2\,T\,M^2\,\hat{\lambda}^2+o(\rho^2) \tag{22}$$

with $\hat{\lambda}$ defined as in (8).

Fixed eigenvectors of $\boldsymbol{K}$. In the case where we cannot control the eigenvalues of $\boldsymbol{K}$ it can be shown that the optimal correlation matrix $\widehat{\boldsymbol{K}}$ is defined as in (15) and the optimal covariance matrix $\widehat{\boldsymbol{P}}$ is given by

$$\widehat{\boldsymbol{P}} = TM\boldsymbol{s}_{\mathsf{max}}\boldsymbol{s}_{\mathsf{max}}^{H} \tag{23}$$

where $\boldsymbol{s}_{\mathsf{max}}$ is, as before, an unit-norm eigenvector associated to the maximal eigenvalue of the matrix $\boldsymbol{M}_{on}^{i^*}$ defined in (16). In this case, the mutual information (p.c.u.) is given by

$$\frac{1}{T}I(\boldsymbol{X};\boldsymbol{S}) = \frac{\rho^2}{2}\,N^2\,T\,M^2\,\lambda_{\mathsf{max}}^2\left(\boldsymbol{M}_{on}^{i^*}\right)+o(\rho^2). \tag{24}$$

The proof of (23) is omitted due to lack of space. However, in what follows, we outline its main steps here for the completeness of the presentation. First, we note that the objective in (20) is a quadratic (and always positive) function of the eigenvalues of $\boldsymbol{K}$. Hence, the maximum is attained at an extreme point of the set Ω and, consequently, the optimal correlation matrix is rank one, i.e., it can be written in the form $MN\boldsymbol{u}\boldsymbol{u}^H$ for some $\boldsymbol{u}$. Then, for a given MN dimensional vector $\boldsymbol{u}$ and $\boldsymbol{Z}_u=\boldsymbol{\Upsilon}^{-\frac{1}{2}}\overline{\boldsymbol{S}}_{on}\boldsymbol{u}\boldsymbol{u}^H\overline{\boldsymbol{S}}_{on}^{H}\boldsymbol{\Upsilon}^{-\frac{1}{2}}$ one can show (proof omitted) that the following hold:

$$\mathrm{tr}\left(\mathsf{E}\left[\boldsymbol{Z}_u^2\right]\right)=\mathrm{tr}^2\left(\boldsymbol{P}\boldsymbol{M}_{on}\right)+\mathrm{tr}\left(\left(\boldsymbol{P}\boldsymbol{M}_{on}\right)^2\right) \tag{25}$$

and

$$\mathrm{tr}\left(\left(\mathsf{E}\left[\boldsymbol{Z}_u\right]\right)^2\right)=\mathrm{tr}\left(\left(\boldsymbol{P}\boldsymbol{M}_{on}\right)^2\right) \tag{26}$$

where $\boldsymbol{M}_{on}$ is defined as in (6). By combining (25) and (26) with (20) and by invoking Fan's theorem (see pp. 17 in [13]) we get (23).

Remarks. We see that the mutual information is proportional to M^2N^2. Hence, the results herein presented are in concordance with [2, 9] and with the results of the previous subsection where it has been shown that channel correlation and correlated noise can improve the channel performance. Also, note that the optimal values correspond to those of the coherent correlated Rayleigh fading channel case (remark that $\widehat{\boldsymbol{P}}=\widehat{\boldsymbol{P}}_{coh}$ defined in (18)).

5. CONCLUSIONS

We have studied the MIMO channel in the low SNR regime from the capacity analysis perspective. The novel aspect is that we allow the channel correlation matrix to have an arbitrary correlation structure. This paper completes our previous work in [1, 2] where the Kronecker correlation model has been adopted for the channel covariance matrix. We have shown that, by maximizing the mutual information for the on-off and Gaussian signallings over the system's parameters, the channel correlation matrix should be as correlated as possible.

6. REFERENCES

[1] M. Beko, J. Xavier, and V. Barroso, "Capacity and error probability analysis of non-coherent MIMO systems in the low SNR regime," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, HI USA, 2007.

[2] M. Beko, J. Xavier, and V. Barroso, "Further results on the capacity and error probability analysis of non-coherent MIMO systems in the low SNR regime," accepted for publication in *IEEE Transactions on Signal Processing*.

[3] A. Lozano, A. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2527-2544, Oct. 2003.

[4] I. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290-1301, May 2001.

[5] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319-1343, June 2002.

[6] V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1567-1580, Aug. 2004.

[7] B. Hajek and V. Subramaniam, "Capacity and reliability function for small peak signal constraints," *IEEE Trans. Inform. Theory*, vol. 48, pp. 828-839, Apr. 2002.

[8] C. Rao and B. Hassibi, "Analysis of multiple-antenna wireless links at low SNR," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2123-2130, Sep. 2004.

[9] X. Wu and R. Srikant, "MIMO channels in the low SNR regime: communication rate, error exponent and signal peakiness," in *Proc. of IEEE Information Theory Workshop* , 2004.

[10] W. Weichselberger, M. Herdin, H. Ozcelik, E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90-100, Jan. 2006.

[11] H. Bolcskei, D. Gesbert, C. B. Papadias, A. -J. van der Veen (eds.), *Space-Time Wireless Systems: From Array Processing to MIMO Communications.*, Cambridge University Press, 2006.

[12] M. Beko, J. Xavier, and V. Barroso, "Non-coherent communication in multiple-antenna systems: Receiver design and codebook construction," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5703 - 5715, Dec. 2007.

[13] J. M. Borwein, A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples.*, Springer-Verlag, 2000.

[14] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics.*, Revised Edition, John Wiley & Sons, 1999.

A. APPENDIX

In this appendix, we establish expression (11). We start by using the following result: $\boldsymbol{K}_{on} = \mathcal{H} \otimes \boldsymbol{I}_T$ where

$$\mathcal{H} = (\boldsymbol{I}_N \otimes \boldsymbol{K}_{MN}) (\text{vec}(\boldsymbol{I}_N) \otimes \boldsymbol{I}_M) \tag{27}$$

and $\boldsymbol{K}_{MN}$ is the $MN \times MN$ commutation matrix. See pp. 46 and 48 in [14]. Hence, it holds

$$\boldsymbol{K}_{on}^T (\boldsymbol{u} \otimes \boldsymbol{I}_{TN}) = \left(\mathcal{H}^T (\boldsymbol{u} \otimes \boldsymbol{I}_N)\right) \otimes \boldsymbol{I}_T. \tag{28}$$

In (28) the following property is used: if $\boldsymbol{AC}$ and $\boldsymbol{BD}$ exist then $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}$ and $\boldsymbol{I}_{TN} = \boldsymbol{I}_N \otimes \boldsymbol{I}_T$. See pp. 28 in [14]. This implies that if one manages to prove that for every $M \times N$ matrix $\boldsymbol{U}$

$$\mathcal{H}^T (\boldsymbol{u} \otimes \boldsymbol{I}_N) = \boldsymbol{U}, \tag{29}$$

where $\boldsymbol{u} = \text{vec}(\boldsymbol{U})$, the equality in (11) will be automatically proved. Thus, in the sequel, we prove (29). To this end, note that it is sufficient to show that for every $i = 1, .., M$ and $j = 1, ..., N$

$$\boldsymbol{f}_i^T \left(\mathcal{H}^T (\boldsymbol{u} \otimes \boldsymbol{I}_N)\right) \boldsymbol{e}_j = \boldsymbol{U}_{ij} \tag{30}$$

where $\boldsymbol{U}_{ij}$ is the ij-th element of $\boldsymbol{U}$ and $\boldsymbol{e}_j$, $\boldsymbol{f}_i$ represent the j-th and i-th column of $\boldsymbol{I}_N$ and $\boldsymbol{I}_M$, respectively. Using (27) and (30) we can write

$$\begin{aligned} \boldsymbol{f}_i^T \mathcal{H}^T &= (\text{vec}(\boldsymbol{I}_N) \otimes \boldsymbol{f}_i)^T \left(\boldsymbol{I}_N \otimes \boldsymbol{K}_{MN}^T\right) \\ &= \left[(\boldsymbol{e}_1 \otimes \boldsymbol{f}_i); \quad ... \quad ; (\boldsymbol{e}_N \otimes \boldsymbol{f}_i)\right]^T \left(\boldsymbol{I}_N \otimes \boldsymbol{K}_{MN}^T\right) \end{aligned} \tag{31}$$

and

$$(\boldsymbol{u} \otimes \boldsymbol{I}_N) \boldsymbol{e}_j = \left[(\boldsymbol{u}_1 \otimes \boldsymbol{e}_j)^T \quad ... \quad (\boldsymbol{u}_N \otimes \boldsymbol{e}_j)^T\right]^T \tag{32}$$

where $\boldsymbol{u}_k$, for $k = 1, ..., N$, is the k-th column of the matrix $\boldsymbol{U}$. By combining (31) and (32) with (30) we have

$$\begin{aligned} \boldsymbol{f}_i^T \left(\mathcal{H}^T (\boldsymbol{u} \otimes \boldsymbol{I}_N)\right) \boldsymbol{e}_j &= \sum_{k=1}^{N} (\boldsymbol{e}_k \otimes \boldsymbol{f}_i)^T \boldsymbol{K}_{MN}^T (\boldsymbol{u}_k \otimes \boldsymbol{e}_j) \\ &= \sum_{k=1}^{N} (\boldsymbol{f}_i \otimes \boldsymbol{e}_k)^T (\boldsymbol{u}_k \otimes \boldsymbol{e}_j) \\ &= \boldsymbol{U}_{ij}. \end{aligned} \tag{33}$$

To get (33) Theorem 9 on pp. 47 in [14] is invoked. This completes the proof.