Further Results on the Capacity and Error Probability Analysis of Noncoherent MIMO Systems in the Low SNR Regime

Marko Beko, *Student Member, IEEE*, João Xavier, *Member, IEEE*, and Victor A. N. Barroso, *Senior Member, IEEE*

Abstract—The noncoherent single-user multiple-input–multiple-output (MIMO) channel in the low signal-to-noise ratio (SNR) regime is investigated from two viewpoints: capacity and probability of error analysis. The novelty in both viewpoints is that we allow an arbitrary correlation structure for the Gaussian observation noise. First, we look at the capacity of the spatially correlated Rayleigh fading channel. We investigate the impact of channel and noise correlation on the mutual information for the ON–OFF and Gaussian signaling schemes. Our findings establish that, in the low SNR regime, mutual information is maximized when the transmit antennas are fully correlated (the same holds for the receive array). Then, the deterministic channel setup is considered and a pairwise error probability (PEP) analysis for the generalized likelihood ratio test (GLRT) receiver is performed. This leads to a codebook design criterion on which we base the construction of new space-time constellations. Their performance is assessed by computer simulations and, as a byproduct, we show that our codebooks are also of interest for Bayesian receivers, which decode constellations with nonuniform priors.

Index Terms—Channel capacity, colored noise, correlated Rayleigh fading, Gaussian modulation, generalized likelihood ratio test (GLRT) receiver, low signal-to-noise ratio (SNR), maximum *a posteriori* (MAP) estimation, multiple-input–multiple-output (MIMO) systems, noncoherent communications.

I. INTRODUCTION

IN slowly fading scenarios, channel stability enables the receiver to be trained in order to acquire the channel state information (CSI) necessary for *coherent detection* of the transmitted codeword. The scope of this paper will be fast fading scenarios, where fading coefficients change into new, almost independent values before being learned by the receiver through training signals. Hence, CSI is no more accessible, and the receiver must then operate in a *noncoherent* mode. Further, we focus on the low signal-to-noise ratio (SNR) regime. This stems from the fact that in the third-generation mobile data systems almost 40% of geographical locations experience receiver SNR levels below 0 dB while only less than 10% display levels above 10 dB. High SNR requirement, besides its low-power efficiency, cannot always be satisfied due to the power limitations in the mobile device. Also, recent technological advances have led to the emergence of small, low-power, and possibly mobile devices that, when deployed in large numbers, have the ability to form an intelligent (sensor) network, which can monitor large areas, detect the presence or absence of targets, etc. This motivates the analysis and construction of communication schemes that can cope with the low SNR regime. See [1]–[3] for a more thorough discussion of this topic.

A. Previous Work

Low SNR multiple-input–multiple-output (MIMO) systems when CSI is available at the receiver have been treated in [1]. The interplay of rate, bandwidth, and power is analyzed in the region of energy per bit close to its minimum value. The scenario where no CSI is available at the receiver has been considered in [4]. It has been shown that the optimal signaling at low SNR achieves the same capacity as the known channel case for single transmit antenna systems. Verdu [5] has 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 considered for a discrete-time memoryless channel with peak constrained inputs. Similar results to [6] and [7] have been obtained in [8] under weaker assumptions on the input signals. In the same work, Rao and Hassibi have demonstrated that the ON–OFF signaling presented in [4] generalizes to the multiantenna 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 investigated 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. The symbol error probability point of view

for the analysis of low SNR noncoherent independent and identically distributed (i.i.d.) Rayleigh channel is more recent, although Hochwald *et al.* [10] had reported that in the low SNR and Rayleigh fading channel it seems one should employ only one transmit antenna. Borran *et al.* [2], under the assumption of equally probable codewords, presented a technique that uses Kullback–Liebler (KL) divergence between the probability density functions (pdfs) induced at the receiver by distinct transmitted codewords as a design criterion for codebook design. In low SNR condition, their constellation points occupy multiple level (signal points lie in concentric spheres) with a point usually in the origin. The codes thereby constructed were shown to perform better than some existing noncoherent codebook constructions in low SNR, namely, [10]. Srinivasan *et al.* [11] considered the case of a single transmit antenna in the low SNR regime. Using the information theoretic results over the low SNR noncoherent i.i.d. Rayleigh fading channel under an average power constraint (cf. [8] and [9]), they allow for codewords with unequal priors in a code and optimize over prior probabilities to achieve better performance. This results in constellations that assume a point in the origin with probability 1/2, with the probabilities of the points lying in the sphere being equal. In [12], the correlated Rayleigh fading model was studied and it was shown that at any SNR, any single antenna performs better when used with suitable precoding in a MIMO correlated Rayleigh fading than in a single-input–multiple-output (SIMO) channel. Consequently, code designs that exploit the correlations in the transmit antennas in the MIMO case to provide gains over the corresponding SIMO case in the low SNR regime were presented.

B. Contributions and Paper Organization

We study the noncoherent MIMO channel in the low SNR regime from the capacity and pairwise error probability (PEP) viewpoints. The novel aspect is that we allow the Gaussian observation noise to have an arbitrary correlation structure, though known to the transmitter and the receiver. In Section II, the spatially correlated noncoherent MIMO block Rayleigh fading channel is analyzed. This extends the approach in [8] as we take into account both channel and noise correlation. The impact of channel and noise correlation on the mutual information is obtained for the ON–OFF and Gaussian signaling. The main conclusion is that mutual information is maximized when both the transmit and receive antennas are fully correlated. We also argue that the ON–OFF signaling is optimal for this multiantenna setting. In Section III, contrary to most approaches for the low SNR regime, the channel matrix is assumed deterministic, i.e., no stochastic model is attached to it. A low SNR analysis of the PEP for the generalized likelihood ratio test (GLRT) receiver is introduced, and a codebook design criterion that takes into account the information about noise correlation is obtained. For the special case of single transmit antenna and spatio–temporal white Gaussian noise, it is shown that the problem of finding good codes corresponds to the very well-known packing problem in the complex projective space [22], [25]. New space–time constellations for some particular wireless scenarios are constructed. Computer simulations show that these new codebooks are also of interest for Bayesian receivers

that decode constellations with nonuniform priors. Section IV contains the main conclusions of our paper.

Throughout this paper, the operator ^{T} $(^H)$ denotes transpose (complex conjugate transpose). The multivariate circularly symmetric, complex Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The operator $\mathsf{E}_p[\cdot]$ denotes the expectation with respect to the probability distribution p. When no ambiguity can occur, we simply write $\mathsf{E}[\cdot]$. For any matrix $\boldsymbol{A}$, we write its trace as $\mathrm{tr}(\boldsymbol{A})$. The symbol $\|\boldsymbol{A}\| = \sqrt{\mathrm{tr}(\boldsymbol{A}^H \boldsymbol{A})}$ denotes the Frobenius norm. The N-dimensional identity matrix is denoted by $\boldsymbol{I}_N$ and the $M \times N$ matrix of all zeros by $\boldsymbol{0}_{M \times N}$ (if no ambiguity can occur, we write $\boldsymbol{0}$). The minimum (maximum) eigenvalue of the symmetric matrix $\boldsymbol{A}$ is denoted by $\lambda_{\min}(\boldsymbol{A})$ $(\lambda_{\max}(\boldsymbol{A}))$. The determinant of matrix $\boldsymbol{A}$ is denoted by $\det(\boldsymbol{A})$. The curled inequality symbol $\succeq$ represents matrix inequality between Hermitian matrices.

II. RANDOM FADING CHANNEL: THE LOW SNR MUTUAL INFORMATION ANALYSIS

1) Data Model and Assumptions: We focus on a communication system comprising M transmit and N receive antennas over a narrowband flat Rayleigh fading channel. We assume a block fading channel model that is widely used in the MIMO literature [8], [9], [15], [24], with coherence interval T. In complex baseband 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 popular *separable (Kronecker) spatial correlation* model [1], [14], [15] is used, i.e., $\boldsymbol{H} = \sqrt{\rho/M} \boldsymbol{K}_t^{1/2} \boldsymbol{H}_w \left(\boldsymbol{K}_r^T\right)^{1/2}$, where $\boldsymbol{H}_w$ is an $M \times N$ matrix composed of statistically independent $\mathcal{CN}(0, 1)$ entries, and the coefficient ρ is the model parameter proportional to the SNR. We assume that the matrix $\boldsymbol{H}_w$ remains fixed for the coherence time T after which it changes to a new independent value. The correlation coefficients between the M (N) transmit (receive) antennas are assembled into an $M \times M$ $(N \times N)$ positive-semidefinite Hermitian correlation matrix $\boldsymbol{K}_t$ $(\boldsymbol{K}_r^T$; the operator ^{T} is used for the sake of convenience). The matrix $\boldsymbol{H}_w$ is not known at the receiver or at the transmitter, but its distribution is, in addition to $\boldsymbol{K}_t$ and $\boldsymbol{K}_r$. This model is appropriate for the scenarios where only the objects surrounding the transmitter and the receiver cause the correlation of the local antenna elements, while they have no impact on the correlation at the other end of the link. The model has been found to be satisfactory in certain scenarios [16], [17]. We would also like to point out that there exist other spatial correlation models that take into account coupling between transmit and receive sides; see [18] and references therein. Although these models may characterize realistic channels more accurately for some scenarios (in these cases, the Kronecker model leads

to capacity underestimations), we adopt the Kronecker model since the general case is less tractable for the analytical approach in this work. For a fair comparison of different correlation cases, we assume that $\mathrm{tr}(\boldsymbol{K}_t) = M$ and $\mathrm{tr}(\boldsymbol{K}_r) = N$.

A2) Transmit power constraint: We impose the power constraint $\mathsf{E}[\mathrm{tr}(\boldsymbol{S}^H\boldsymbol{S})] \leq TM$.

A3) Noise distribution: The noise covariance matrix $\boldsymbol{\Upsilon} = \mathsf{E}[\mathrm{vec}(\boldsymbol{E})\,\mathrm{vec}(\boldsymbol{E})^H]$ is known at the transmitter and at the receiver ($\mathrm{vec}(\boldsymbol{E})$ stacks all columns of the matrix $\boldsymbol{E}$ on the top of each other, from left to right). Also, without loss of generality (w.l.o.g.), we assume $\mathrm{tr}(\boldsymbol{\Upsilon}) = NT$. Note that in A3), we let the data model depart from the customary assumption of spatio–temporal white Gaussian observation noise that is clearly an approximation (for which it is known that leads to severe capacity underestimations). In realistic scenarios, the $\boldsymbol{E}$ term often exhibits a very rich correlation structure, e.g., see [1], [19, pp. 10, 159, 171], and [20, pp. 100]. The generalization to arbitrary noise covariance matrices $\boldsymbol{\Upsilon}$ encompasses many scenarios of interest as special cases: spatially colored or not jointly with temporally colored or not observation noise, multiuser environment, etc. A methodology for designing space–time codebooks for noncoherent communications in multiple-antenna wireless systems and an arbitrarily given noise covariance matrix $\boldsymbol{\Upsilon}$ in the high SNR regime has recently been proposed; see [22] for more details.

A. 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 multiantenna setting and attains the known channel capacity. Here, we show that this is also the case for the correlated Rayleigh fading channel model with arbitrary noise covariance matrix. Furthermore, we maximize the mutual information with respect to (w.r.t.) the input signal $\boldsymbol{S}_{\mathrm{on}}$, $\boldsymbol{K}_t$, and $\boldsymbol{K}_r$. Hence, we view both $\boldsymbol{K}_t$ and $\boldsymbol{K}_r$ as system parameters that we can introduce and track. This paper completes our previous work in [23]. The ON–OFF signaling is defined as follows: for any $\epsilon > 1$ and assuming $\rho < 1$

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

With an analysis similar to [8], presented in Appendix I, 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

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

where $\otimes$ denotes the Kronecker product. Note that for the special case of spatio–temporal white observation noise and uncorrelated Rayleigh fading channel case, i.e., $\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$, $\boldsymbol{K}_t = \boldsymbol{I}_M$, and $\boldsymbol{K}_r = \boldsymbol{I}_N$, our result in (2) recovers the finding in [8]. In that case, the maximal mutual information (per channel use) is equal to

$$\frac{1}{T}I(\boldsymbol{X};\boldsymbol{S}) = \rho\,N + o(\rho).$$

Now, we address the maximization of the mutual information w.r.t. $\boldsymbol{S}_{\mathrm{on}}$, $\boldsymbol{K}_t$, and $\boldsymbol{K}_r$, i.e.,

$$\max_{\mathrm{tr}(\boldsymbol{S}_{\mathrm{on}}\boldsymbol{S}_{\mathrm{on}}^H) \leq TM, \boldsymbol{K}_t \in \mathcal{P}_M, \boldsymbol{K}_r \in \mathcal{P}_N} \mathrm{tr}\left(\boldsymbol{\Upsilon}^{-1}\left(\boldsymbol{K}_r \otimes \boldsymbol{S}_{\mathrm{on}}\boldsymbol{K}_t\boldsymbol{S}_{\mathrm{on}}^H\right)\right) \quad (3)$$

where

$$\mathcal{P}_n = \{\boldsymbol{X} : n \times n \text{ matrix such that } \boldsymbol{X} \succeq \boldsymbol{0} \text{ and } \mathrm{tr}(\boldsymbol{X}) = n\}. \quad (4)$$

In Appendix II, we show that the maximum in (3) is attained by

$$\widehat{\boldsymbol{S}}_{\mathrm{on}} = \sqrt{TM}[\,\hat{\boldsymbol{s}} \quad \boldsymbol{0}_{T\times(M-1)}\,]$$
$$\widehat{\boldsymbol{K}}_r = N\hat{\boldsymbol{u}}\hat{\boldsymbol{u}}^H \quad \left[\widehat{\boldsymbol{K}}_t\right]_{i,i} = M\delta_{i1} \quad (5)$$

where

$$(\hat{\boldsymbol{u}},\hat{\boldsymbol{s}}) = \underset{\boldsymbol{u}\in\mathbb{C}^N,\|\boldsymbol{u}\|=1,\boldsymbol{s}\in\mathbb{C}^T,\|\boldsymbol{s}\|=1}{\arg\max}\ (\boldsymbol{u}\otimes\boldsymbol{s})^H\boldsymbol{\Upsilon}^{-1}(\boldsymbol{u}\otimes\boldsymbol{s}) \quad (6)$$

with $\delta_{ij} = 1$ for $i = j$ and zero, otherwise. The notation $[\boldsymbol{M}]_{i,i}$ represents the entry of the matrix $\boldsymbol{M}$ on the position (i,i). Note that $\widehat{\boldsymbol{K}}_t$ is a diagonal matrix. The optimization problem in (6) always admits a solution (maximization of a continuous function over a compact set), but, in general, a closed-form solution is not available. The exception is the case when the noise covariance matrix $\boldsymbol{\Upsilon}$ has a Kronecker structure, say $\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}_1 \otimes \boldsymbol{\Upsilon}_2$, for some $N \times N$ matrix $\boldsymbol{\Upsilon}_1$ and $T \times T$ matrix $\boldsymbol{\Upsilon}_2$. In that situation, the optimal $\hat{\boldsymbol{u}}$ (resp., $\hat{\boldsymbol{s}}$) can be taken as any unit-norm eigenvector associated with the minimal eigenvalue of $\boldsymbol{\Upsilon}_1$ (resp., $\boldsymbol{\Upsilon}_2$). For example, in the case of spatially white-temporal colored observation noise, i.e., $\boldsymbol{\Upsilon} = \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}$ for some $T \times T$ positive-definite matrix $\boldsymbol{\Sigma}$, the vector $\hat{\boldsymbol{s}}$ is the unit-norm eigenvector associated with the minimal eigenvalue of $\boldsymbol{\Sigma}$ (in other words, we transmit the codeword in the direction that is least affected by the noise). For the choice in (5),, the maximal mutual information (per channel use) is equal to

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

where $\hat{\lambda} = (\hat{\boldsymbol{u}}\otimes\hat{\boldsymbol{s}})^H\boldsymbol{\Upsilon}^{-1}(\hat{\boldsymbol{u}}\otimes\hat{\boldsymbol{s}})$.

1) Remarks: From (5), it is clear that both the transmit and receive antennas should be made as correlated as possible, as both the optimal $\boldsymbol{K}_t$ and $\boldsymbol{K}_r$ have rank one. Note that in (7) the mutual information is proportional to the number of transmit antennas M. The similar conclusion has been drawn in [9] and [15] where it has been demonstrated that channel correlation can have beneficial impact on the capacity. This is in sharp contrast with the case of uncorrelated Rayleigh fading channel model for which it has been shown that the maximal mutual information is independent of the number of transmit antennas [8]. Also, since $\mathrm{tr}(\boldsymbol{\Upsilon}^{-1}) = \sum_{i=1}^{NT} 1/\lambda_i \geq \sum_{i=1}^{NT}(2 - \lambda_i) = NT$, where λ_i's are the eigenvalues of $\boldsymbol{\Upsilon}$, we can w.l.o.g. assume that, e.g., $[\boldsymbol{\Upsilon}^{-1}]_{1,1} \geq 1$ (otherwise, $\mathrm{tr}(\boldsymbol{\Upsilon}^{-1}) < NT$). Then, by choosing $\boldsymbol{u}_1 = [1 \quad \boldsymbol{0}_{1\times(N-1)}]^T$ and $\boldsymbol{s}_1 = [1 \quad \boldsymbol{0}_{1\times(T-1)}]^T$, we have $\hat{\lambda} \geq (\boldsymbol{u}_1\otimes\boldsymbol{s}_1)^H\boldsymbol{\Upsilon}^{-1}(\boldsymbol{u}_1\otimes\boldsymbol{s}_1) = [\boldsymbol{\Upsilon}^{-1}]_{1,1} \geq 1$. This result confirms the general principle that correlated noise is beneficial from the capacity point of view. See, e.g., [20, p. 100] for more details. In practice, by changing the antenna separation, one can control the eigenvalues of $\boldsymbol{K}_t$ and $\boldsymbol{K}_r$, but not their eigenvectors.

See [15], [9], [26], [27], and [20, Sec. 6.3.3]. As a consequence, the result presented herein has to be intended as the upper bound on the channel capacity. Nevertheless, the previous conclusion holds for the case when we cannot manipulate the eigenvectors of K_r (which are still available to the transmitter through a feedback link): in this case, from (39), we obtain

$$\frac{1}{T}I(\boldsymbol{X};\boldsymbol{S}) = \rho\,N\,M\lambda_{\max}\left(\boldsymbol{F}_{i^*}\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_{i^*}^H\right) + o(\rho)$$

where $\widehat{\boldsymbol{\Upsilon}}$, $\boldsymbol{F}_i$, and i^* are defined in (32), (32), and (36), respectively. Clearly, $\lambda_{\max}(\boldsymbol{F}_{i^*}\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_{i^*}^H) \geq 1$ (otherwise, $\mathrm{tr}(\boldsymbol{\Upsilon}^{-1}) < NT$). Thus, correlated noise is beneficial from the capacity point of view in this case as well. 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 and the noise covariance matrix $\boldsymbol{\Upsilon}$ is arbitrary (when $\boldsymbol{\Upsilon} = \boldsymbol{I}_{NT}$, then $\hat{\lambda} = 1$; from (7), we retrieve the result on [20, p. 94], where the coherent correlated Rayleigh fading channel has been treated). With coherent reception, we know that the mutual information is maximized if the input $\mathrm{vec}(\boldsymbol{S})$ is circularly symmetric, complex Gaussian distribution, i.e., $\mathrm{vec}(\boldsymbol{S}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{P}_{\mathrm{coh}})$ for some covariance matrix $\boldsymbol{P}_{\mathrm{coh}}$ with $\mathrm{tr}(\boldsymbol{P}_{\mathrm{coh}}) \leq TM$ such that the power constraint in the assumption **A2** is satisfied (note that we may assume w.l.o.g. that $\mathrm{vec}(\boldsymbol{S})$ is zero mean). In that case, it can be readily shown (by maximizing the first-order expansion of the very well-known expectation-log-det expression for the capacity w.r.t. $\boldsymbol{P}_{\mathrm{coh}}$, $\boldsymbol{K}_t$, and $\boldsymbol{K}_r$) that the optimal covariance matrix $\widehat{\boldsymbol{P}}_{\mathrm{coh}}$ is given by

$$\widehat{\boldsymbol{P}}_{\mathrm{coh}} = \mathrm{vec}(\widehat{\boldsymbol{S}}_{\mathrm{on}})\mathrm{vec}^H(\widehat{\boldsymbol{S}}_{\mathrm{on}}) \qquad (8)$$

with $\widehat{\boldsymbol{S}}_{\mathrm{on}}$, $\widehat{\boldsymbol{K}}_t$, and $\widehat{\boldsymbol{K}}_r$ defined as in (5).

B. Mutual Information: Gaussian Modulation

From a practical point of view, it is unreasonable to allow input signals with large peakiness as the previous ON–OFF signaling. Thus, we compute the low SNR mutual information for the more realistic case of Gaussian modulation. Let $\boldsymbol{s} = \mathrm{vec}(\boldsymbol{S})$ 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 meet the power constraint in the assumption A2), $\mathrm{tr}(\boldsymbol{P}) \leq TM$. Then, at sufficiently low SNR, the mutual information between $\boldsymbol{X}$ and $\boldsymbol{S}$ up to second order in ρ is given by

$$I(\boldsymbol{X};\boldsymbol{S}) = \frac{\rho^2}{2M^2}\,\mathrm{tr}\left(\mathsf{E}[\boldsymbol{Z}^2] - (\mathsf{E}[\boldsymbol{Z}])^2\right) + o(\rho^2) \qquad (9)$$

$$= \frac{\rho^2}{2M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}\sum_{z=1}^{N}\lambda_k\lambda_z$$

$$\times \mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz}\widetilde{\boldsymbol{P}}_{ij})\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{zk}\widetilde{\boldsymbol{P}}_{ji}) + o(\rho^2) \qquad (10)$$

where $\boldsymbol{Z} = \boldsymbol{\Upsilon}^{-1/2}\left(\boldsymbol{K}_r \otimes \boldsymbol{S}\boldsymbol{K}_t\boldsymbol{S}^H\right)\boldsymbol{\Upsilon}^{-1/2}$, $\widetilde{\boldsymbol{P}}_{ij} = \boldsymbol{E}_i\widetilde{\boldsymbol{P}}\boldsymbol{E}_j^H$, $\widetilde{\boldsymbol{P}} = ((\boldsymbol{K}_t^T)^{1/2} \otimes \boldsymbol{I}_T)\boldsymbol{P}((\boldsymbol{K}_t^T)^{1/2} \otimes \boldsymbol{I}_T)$, $\widehat{\boldsymbol{\Upsilon}}_{kz} = \boldsymbol{F}_k\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_z^H$, and λ_k's, for $k = 1, \ldots, N$, are the eigenvalues of $\boldsymbol{K}_r$. The ma-

trices $\widehat{\boldsymbol{\Upsilon}}$ and $\boldsymbol{F}_k$, for $k = 1, \ldots, N$, are defined in (32), whereas the $T \times TM$ matrix $\boldsymbol{E}_i$, for $i = 1, \ldots, M$, is given by

$$\boldsymbol{E}_i = \boldsymbol{e}_i^T \otimes \boldsymbol{I}_T \qquad (11)$$

where $\boldsymbol{e}_i$ represents the ith column of the identity matrix $\boldsymbol{I}_M$. The proof is given in Appendix III. We now address the optimization problem

$$\max_{\boldsymbol{P} \succeq \boldsymbol{0}, \mathrm{tr}(\boldsymbol{P}) \leq TM, \boldsymbol{K}_t \in \mathcal{P}_M, \boldsymbol{K}_r \in \mathcal{P}_N}$$
$$\times \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}\sum_{z=1}^{N}\lambda_k\lambda_z\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz}\widetilde{\boldsymbol{P}}_{ij})\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{zk}\widetilde{\boldsymbol{P}}_{ji}). \quad (12)$$

It can be shown that the maximum of (12) is attained by the following signaling scheme. The optimal correlation matrices $\widehat{\boldsymbol{K}}_r$ and $\widehat{\boldsymbol{K}}_t$ are defined as in (5), and the optimal covariance matrix $\widehat{\boldsymbol{P}}$ is given by

$$\widehat{\boldsymbol{P}} = TM\boldsymbol{K}_{\boldsymbol{P}} \otimes \hat{\boldsymbol{s}}\hat{\boldsymbol{s}}^H \qquad (13)$$

where the vectors $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{s}}$ are, as before, solutions of the optimization problem (6). The $M \times M$ constant matrix $\boldsymbol{K}_{\boldsymbol{P}}$ has all the entries equal to zero except the entry $(1,1)$, which is one. The proof is given in Appendix IV. In this case, the mutual information (per channel use) is given by

$$\frac{1}{T}I(\boldsymbol{X};\boldsymbol{S}) = \frac{\rho^2}{2}N^2TM^2\hat{\lambda}^2 + o(\rho^2) \qquad (14)$$

where $\hat{\lambda} = (\hat{\boldsymbol{u}} \otimes \hat{\boldsymbol{s}})^H\boldsymbol{\Upsilon}^{-1}(\hat{\boldsymbol{u}} \otimes \hat{\boldsymbol{s}})$.

1) Remarks: In [8], it has been proved that for the uncorrelated Rayleigh fading channel only one transmit antenna should be employed. Here, we see from (14) that having more transmit (M) and receive (N) antennas can actually enhance the channel performance in terms of capacity significantly in the correlated setup. We see that the mutual information is proportional to M^2N^2, whereas in [8], the increase is only linear in the number of the receive antennas. Hence, by making the antennas as correlated as possible the total gain is M^2N. Remark that although the mathematics for calculating $\widehat{\boldsymbol{P}}$, $\widehat{\boldsymbol{K}}_t$, and $\widehat{\boldsymbol{K}}_r$ is quite involved, our findings are not surprising, since we see that the optimal values correspond to those of the coherent correlated Rayleigh fading channel case [it is easy to check that $\widehat{\boldsymbol{P}} = \widehat{\boldsymbol{P}}_{\mathrm{coh}}$ defined in (8)]. The conclusions herein presented are in concordance with [9] and [15] and with the results of Section II-A where it has been shown that channel correlation and correlated noise can actually improve the channel performance.

III. DETERMINISTIC FADING CHANNEL: THE LOW SNR PEP ANALYSIS

1) Data Model and Assumptions: We retain the data model (1), but the presumptions under which we work are the following:
- P1) Channel matrix: The channel matrix $\boldsymbol{H}$ is not known at the receiver or at the transmitter, and no stochastic model is assumed for it.

P2) Transmit power constraint: The codeword $\boldsymbol{S}$ is chosen from a finite codebook $\mathcal{S} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_K\}$ known to the receiver, where K is the size of the codebook. We impose the power constraint $\mathrm{tr}(\boldsymbol{S}_k^H \boldsymbol{S}_k) = 1$ for each codeword. We further assume that each codeword is of full rank.

P3) Noise distribution: As in the assumption A3), the noise covariance matrix $\boldsymbol{\Upsilon} = \mathsf{E}[\mathrm{vec}(\boldsymbol{E})\,\mathrm{vec}(\boldsymbol{E})^H]$ is known at the transmitter and at the receiver.

2) Receiver: Under the previous conditions, the conditional pdf of the received vector $\boldsymbol{x} = \mathrm{vec}(\boldsymbol{X})$, given the transmitted matrix $\boldsymbol{S}$ and the unknown realization of the channel $\boldsymbol{g} = \mathrm{vec}(\boldsymbol{H})$, is given by

$$p(\boldsymbol{x}|\boldsymbol{S},\boldsymbol{g}) = k \exp\{-\|\boldsymbol{x} - (\boldsymbol{I}_N \otimes \boldsymbol{S})\boldsymbol{g}\|^2_{\boldsymbol{\Upsilon}^{-1}}\}$$

where $k = 1/\left(\pi^{TN}\det(\boldsymbol{\Upsilon})\right)$ and the notation $\|\boldsymbol{z}\|^2_{\boldsymbol{A}} = \boldsymbol{z}^H \boldsymbol{A} \boldsymbol{z}$ is used. Since no stochastic model is attached to the channel propagation matrix, the receiver faces a multiple hypothesis testing problem with the channel $\boldsymbol{H}$ as a deterministic nuisance parameter. Hence, we will assume a GLRT receiver. The GLRT [32] is composed of a bank of K parallel processors where the kth processor assumes the presence of the kth codeword and computes the likelihood of the observation, after replacing the channel by its maximum-likelihood (ML) estimate. The GLRT detector chooses the codeword associated with the processor exhibiting the largest likelihood of the observation, i.e., $\widehat{k} = \arg\max\{p(\boldsymbol{x}|\boldsymbol{S}_k, \widehat{\boldsymbol{g}}_k) : k = 1, 2, \ldots, K\}$, where $\widehat{\boldsymbol{g}}_k = (\boldsymbol{\mathcal{S}}_k^H \boldsymbol{\mathcal{S}}_k)^{-1} \boldsymbol{\mathcal{S}}_k^H \boldsymbol{\Upsilon}^{-1/2} \boldsymbol{x}$ with $\boldsymbol{\mathcal{S}}_k = \boldsymbol{\Upsilon}^{-1/2}(\boldsymbol{I}_N \otimes \boldsymbol{S}_k)$ represents the ML estimate of the channel. Due to the respective expression for $\widehat{\boldsymbol{g}}_k$, we note that since each codeword of the codebook has full rank (presumption **P2**), the ML channel estimate is well defined.

3) Low SNR Analysis: Recently, in [13], in contrast to other approaches for the low SNR regime, the channel matrix was assumed deterministic and a low SNR analysis of the PEP was introduced, for the special case $M = 1$ and spatio–temporal white Gaussian noise. Here, we generalize the approach in [13] for any number of transmit antennas and arbitrary noise covariance matrix. Let $P_{\boldsymbol{S}_i \to \boldsymbol{S}_j}$ be the probability of the GLRT receiver deciding $\boldsymbol{S}_j$ when $\boldsymbol{S}_i$ is sent. It can be shown that for $T \geq 2M$

$$P_{\boldsymbol{S}_i \to \boldsymbol{S}_j} \approx P\left(Y > \boldsymbol{g}^H \boldsymbol{L}_{ij} \boldsymbol{g}\right) \tag{15}$$

with $\boldsymbol{L}_{ij} = \boldsymbol{\mathcal{S}}_i^H \boldsymbol{\Pi}_j^\perp \boldsymbol{\mathcal{S}}_i$, $\boldsymbol{\Pi}_j^\perp = \boldsymbol{I}_{TN} - \boldsymbol{\mathcal{S}}_j\left(\boldsymbol{\mathcal{S}}_j^H \boldsymbol{\mathcal{S}}_j\right)^{-1}\boldsymbol{\mathcal{S}}_j^H$, and $Y = \sum_{m=1}^{MN} \sin\alpha_m\left(|a_m|^2 - |b_m|^2\right)$, where a_m and b_m are i.i.d. circular complex Gaussian random variables with zero mean and unit variance, i.e., $a_m, b_m \overset{iid}{\sim} \mathcal{CN}(0,1)$ for $m = 1, \ldots, MN$. The angles α_m are the *principal angles* between the subspaces spanned by $\boldsymbol{\mathcal{S}}_i$ and $\boldsymbol{\mathcal{S}}_j$. The proof is given in Appendix V. For the case of spatio–temporal white observation noise, i.e., $\boldsymbol{\Upsilon} = \boldsymbol{I}_{NT}$ and $M = 1$, from (15), we have

$$P_{\boldsymbol{s}_i \to \boldsymbol{s}_j} \approx P\left(\sum_{i=1}^{N}\left(|a_i|^2 - |b_i|^2\right) > \|\boldsymbol{g}\|^2 \sin\alpha_1\right) \tag{16}$$

where we assume $\sin\alpha_1 \neq 0$ (remark that for $\boldsymbol{\Upsilon} = \boldsymbol{I}_{NT}$ there are maximum M different principal angles where each of them is of multiplicity N). In our work [22], we derive the expression for the PEP in the high SNR regime and $T \geq 2M$. For $M = 1$ and $\boldsymbol{\Upsilon} = \boldsymbol{I}_{NT}$, it is given by

$$P_{\boldsymbol{s}_i \to \boldsymbol{s}_j} = \mathcal{Q}\left(\frac{1}{\sqrt{2}}\|\boldsymbol{g}\|\sin\alpha_{ij}\right) \tag{17}$$

where $\mathcal{Q}(x) = \int_x^{+\infty} 1/\sqrt{2\pi}\, e^{-t^2/2}\, dt$. Equations (16) and (17) show that the probability of misdetecting $\boldsymbol{s}_i$ for $\boldsymbol{s}_j$ depends on the channel $\boldsymbol{g} = \boldsymbol{h}$, but more importantly, on the relative geometry of the codewords $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$. Since $P_{\boldsymbol{s}_i \to \boldsymbol{s}_j} = P_{\boldsymbol{s}_j \to \boldsymbol{s}_i}$ (a feature of the scenario $M = 1$ and $\boldsymbol{\Upsilon} = \boldsymbol{I}_{NT}$), the PEPs are symmetric, which gives rise to an intuitive distance measure. Hence, by analyzing the PEP in both extreme cases (low and high SNR), it is clear that one wishes to make the codewords $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ as separate as possible, i.e., the problem of finding good codes corresponds to the very well-known packing problem in the complex projective space [25], [22]. Unfortunately, from (15), it seems difficult to propose a codebook design criteria for $M > 1$ and $\boldsymbol{\Upsilon} \neq \boldsymbol{I}_{TN}$. One of the reasons originates from the fact that PEPs are not symmetric for this general case. Hence, as usually, we resort to an upper bound on the PEP. From (15), an upper bound on the PEP is readily derived

$$P_{\boldsymbol{S}_i \to \boldsymbol{S}_j} \leq P\left(Z > \|\boldsymbol{g}\|^2 \lambda_{\min}(\boldsymbol{L}_{ij})\right) \tag{18}$$

where $Z = \sum_{m=1}^{MN} |a_m|^2$. The bound in (18) is admittedly loose, but allows us to come up with a workable codebook design criterion. The simulation results below will assess its effectiveness. By invoking the second part of the theorem on [33, p. 200], the case when $M \leq T < 2M$, and then by repeating the analysis of the case $T \geq 2M$ presented in Appendix V, it is straightforward to see that the matrix $\boldsymbol{L}_{ij}$ is rank deficient. This can seriously affect the error performance of the system since, by interpreting (15), one wants to maximize $\boldsymbol{g}^H \boldsymbol{L}_{ij} \boldsymbol{g}$. Thus, as in the high SNR regime and GLRT receiver [22], when designing constellation for arbitrary $\boldsymbol{\Upsilon}$ and the low SNR regime, we take $T \geq 2M$. Also, remark that for $M \leq T < 2M$, the bound in (18) is not applicable since $\lambda_{\min}(\bar{\boldsymbol{L}}_{ij}) = 0$.

4) Codebook Construction Methodology: The codebook design criterion in (18) is equivalent to the one for the high SNR regime that has been treated in [21]. Denoting a codebook by $\mathcal{S} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_K\}$, we are led to the following optimization problem:

$$\mathcal{X}^* = \underset{\mathcal{S} \in \mathcal{M}}{\arg\max}\; f(\mathcal{S}) \tag{19}$$

where $f : \mathcal{M} \to \mathbb{R}$, $\mathcal{S} = \{\boldsymbol{S}_1, \ldots, \boldsymbol{S}_K\} \mapsto f(\mathcal{S})$, and $f(\mathcal{S}) = \min\{f_{ij}(\mathcal{S}) : 1 \leq i \neq j \leq K\}$ with $f_{ij}(\mathcal{S}) = \lambda_{\min}(\bar{\boldsymbol{L}}_{ij})$. A codebook $\mathcal{S}$ is a point in the space $\mathcal{M} = \{(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_K) : \mathrm{tr}(\boldsymbol{S}_k^H \boldsymbol{S}_k) = 1\}$. Note that the space $\mathcal{M}$ can be viewed as multi-dimensional torus, i.e., the Cartesian product of K unit spheres: $\mathcal{M} = \mathbb{S}^{2TM-1} \times \cdots \times \mathbb{S}^{2TM-1}$ (K times) and each codeword $\boldsymbol{X}_k$ belongs to $\mathbb{C}^{T \times M}$. The symbol $\mathbb{S}^{n-1}$ denotes the unit sphere
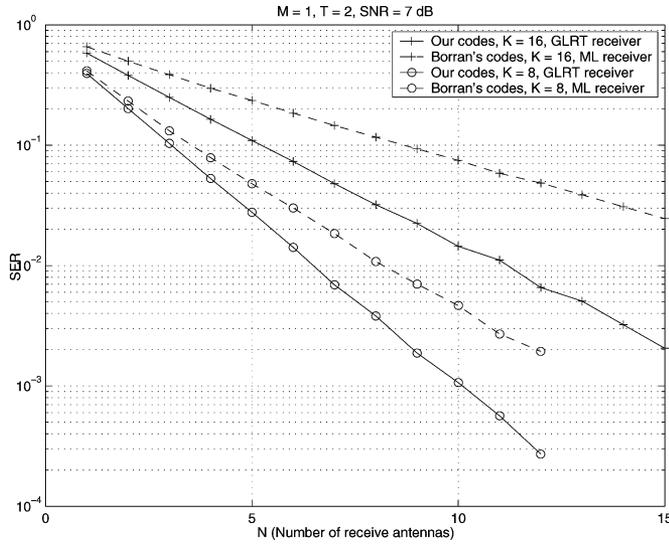
Fig. 1.   Category 1: spatio–temporal white observation noise.
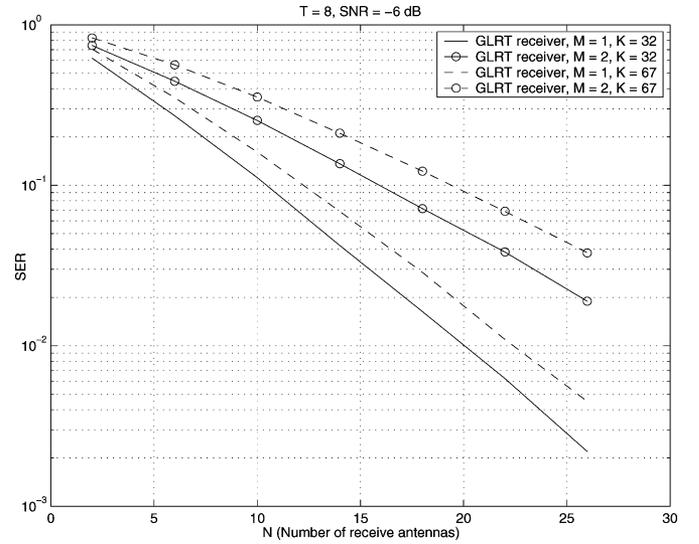
Fig. 2.   Category 1: spatio–temporal white observation noise.

in $\mathbb{R}^n$. From (19), we see that the design of the codebook consists in a high-dimensional, nonlinear, and nonsmooth optimization problem. To solve (19), we employ the algorithm presented in [21] and [22]. The method contains two main steps.

Step 1) It starts by solving a convex semidefinite programming (SDP) relaxation to obtain a rough estimate of the optimal codebook.

Step 2) It refines it through a geodesic descent optimization algorithm, which efficiently exploits the Riemannian geometry of the constraint space $\mathcal{M}$.

Please refer to [22] for more details.

A. Results

Using the codebook construction criterion methodology in [22], we have constructed codes for three special categories of noise covariance matrices $\mathbf{\Upsilon}$. If not stated differently, in all simulations, we assume uncorrelated Rayleigh fading model for the channel matrix, i.e., $h_{ij} \overset{iid}{\sim} \mathcal{CN}\left(0, \sigma^2\right)$. In simulations, we have tried to cover a wide range of low SNR: from very low SNR of -10 to 10 dB.

1) First Category: Spatio–Temporal White Observation Noise: We are not aware of any work concerning the low SNR noncoherent MIMO scenario employing a GLRT receiver. Hence, we will compare the performance of our codes and our GLRT receiver with the codes assuming a Rayleigh fading channel with equally probable codewords [2] and ML receiver. We also show that our codes are of great interest for the constellations with unequal priors [11], [12].

- *Constellations with equal priors for $M = 1$:* In Fig. 1, we compared our codes and our GLRT receiver against the codes found in [2] with the ML receiver proposed therein. We considered the cases where the coherence interval $T = 2$, SNR $= 7$ dB, and codebooks with $K = 8$ and $K = 16$ codewords. The solid and dashed curves represent our codes and Borran codes, respectively. As we can see, for $K = 8$, although the Borran's codes assume the knowledge of actual $\text{SNR} = \mathsf{E}\left[\|\mathbf{s}_k \mathbf{h}\|^2\right] / \mathsf{E}\left[\|\bar{\mathbf{E}}\|^2\right] = 7$ dB,

our codebook constructions can save up to three receive antennas at symbol error rate (SER) of $2 \cdot 10^{-3}$. The same figure plots the results of a similar experiment for $K = 16$. It can be seen that for SER $= 2 \cdot 10^{-2}$, our codes demonstrate a saving of six receive antennas when compared with Borran's codes. The gain of our codes when compared to Borran's can be explained by the following argument. We have seen that for $M = 1$ and spatio–temporal white observation noise the problem of finding good codes corresponds to the packing problem in the complex projective space. This is why for our codes we have used packings presented in [22]. On the other hand, Borran's codes given in [2, Fig. 8] are clearly suboptimal in this example.

- *Constellations with equal priors and $M > 1$:* We present some results to study the impact of employing $M > 1$ transmit antennas in the low SNR regime. Fig. 2 plots the result of the experiment for $T = 8$, SNR $= -6$ dB, $K = 32$, and $M = 1, 2$. We see that at SER $= 2 \cdot 10^{-2}$, our codes for $M = 1$ demonstrate a saving of eight receive antennas when compared with our codes constructed for $M = 2$. The same plot presents the result of the experiment for $T = 8$, SNR $= -6$ dB, $K = 67$, and $M = 1, 2$. For SER $= 4 \cdot 10^{-2}$, our codes for $M = 1$ can spare nine receive antennas compared with our codes constructed for $M = 2$. Then, we compare our codebook constructions for $M = 1$ against Borran's codes with $M = 2$. This is done because one could wonder whether our codebook designs for $M = 2$ are good enough. Fig. 3 plots the result of the experiment. The solid signed and the solid circled curve show the performance of our codes for $K = 32$, $T = 4$, $M = 1$, and $K = 16$, $T = 3$, $M = 1$, respectively. The dashed signed and the dashed circled curve represent the performances of the Borran's codes for $K = 32$, $T = 4$, $M = 2$ and $K = 16$, $T = 3$, $M = 2$, respectively. For 32-point constellations, we see that our codes can save seven receive antennas at SER $= 4 \cdot 10^{-2}$. For 16-point constellations, we witness the gain of more than ten receive antennas at SER $= 10^{-1}$. We think that the results
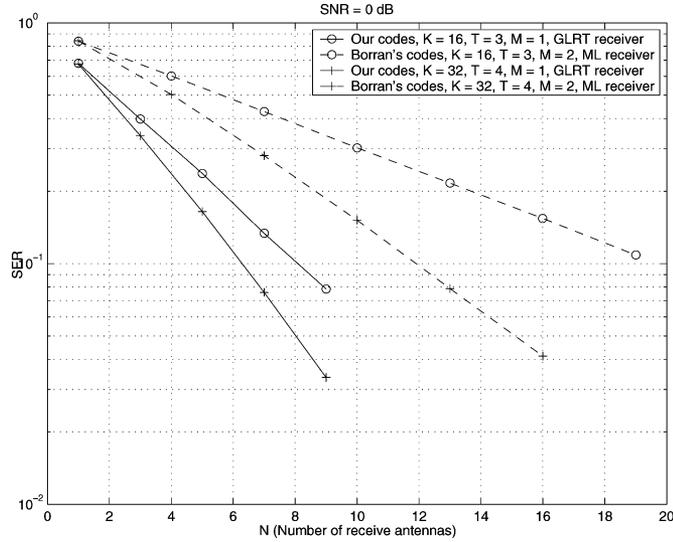
Fig. 3. Category 1: spatio–temporal white observation noise.
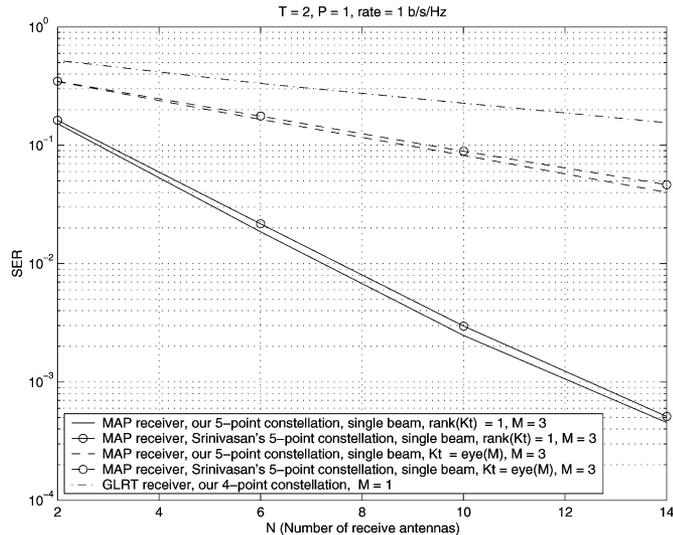
Fig. 4. Category 1: spatio–temporal white observation noise: $\boldsymbol{K}_r = \boldsymbol{I}_N$. Correlated Rayleigh fading: solid curves. Uncorrelated Rayleigh fading: dashed curves. Our and Srinivasan's five-point constellations use MAP receiver; our four-point constellation uses GLRT receiver.

presented in the Figs. 2 and 3 further strengthen the motivation of using a single transmit antenna codebooks in the low SNR regime when GLRT receiver is employed.

- *Constellations with unequal priors:* Now, we depart from our GLRT receiver and show that our codebook designs for $M = 1$ are nevertheless of interest for schemes that allow for nonuniform priors, e.g., the Bayesian receiver in [11] and [12]. In Fig. 4, we show the results of the simulations. We considered the case where the coherence interval $T =$

2, $P = 1$, and rate = 1 bps/Hz, with $P = \mathsf{E}[\text{tr}\left(\boldsymbol{S}_k^H \boldsymbol{S}_k\right)]$. For simplicity, we assume $\boldsymbol{K}_r = \boldsymbol{I}_N$. We consider codes with codewords of the form $\boldsymbol{S}_k = [\,\boldsymbol{s}_k \quad \boldsymbol{0}_{T\times(M-1)}\,]$ since this form of the code resembles the capacity achieving distribution at sufficiently low SNR presented in Section II. We call them single-beam constellations. First, we treat the case of uncorrelated Rayleigh fading channel, i.e., $\boldsymbol{K}_t = \boldsymbol{I}_M$ where $M = 3$. Note that in this case only one transmit antenna is effectively used. The dashed and dashed–circled curves represent our codes, and Srinivasan's five-point constellations with unequal priors [11] (the constellations assume a point in the origin with probability 1/2, with the probabilities of the points lying in the sphere being equal). Next, we assume the correlated Rayleigh fading case with $\boldsymbol{K}_t = \widehat{\boldsymbol{K}}_t$ with $\widehat{\boldsymbol{K}}_t$ defined in Section II (when referring to $\boldsymbol{K}_t = \widehat{\boldsymbol{K}}_t$ case, we simply write rank$(\boldsymbol{K}_t) = 1$). The solid and solid–circled curves represent our codes, and Srinivasan's five-point constellations with unequal priors. As expected, high improvements are possible when codes are used in correlated MIMO scenarios, which is in concordance with the information-theoretic result presented herein. The gain of our five-point constellations with unequal priors compared with Srinivasan's codes is because we use optimal packings in complex projective space (in the outer sphere), whereas Srinivasan uses optimal packings in the real projective space (one can expect larger gains as K increases, where K represents the number of the codewords on the sphere). The improvement obtained can be explained by the optimality of our designed packings. Rankin bound is an upper bound on the packing radius of K subspaces in the Grassmanian space $G(M, \mathcal{C}^T)$. When $M = 1$, the bound applies to packings in the complex projective space, and in this case, it holds

$$\min\{\sin^2\alpha_{ij} : 1 \leq i \neq j \leq K\} \leq \frac{T-1}{T}\frac{K}{K-1}$$

where α_{ij} is the acute angle between codewords $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$. Please refer to [25] for more details. One can easily check that our designed codebook indeed meets the Rankin bound which is 2/3 for $T = 2$ and $K = 4$. Our codebook is represented in the following matrix $\boldsymbol{B} = [\boldsymbol{b}_1 \, \boldsymbol{b}_2]$ where the equations at the bottom of the page hold. The dashed–dotted curve represents our four-point constellation with equal priors and $M = 1$, and it is plotted only to confirm that if the receiver knows the channel statistics, then constellations with nonuniform priors are the best option.

2) Second Category: Spatially White-Temporal Colored Observation Noise: The second category corresponds to spatially white-temporally colored observation noise, i.e., $\boldsymbol{\Upsilon} = \boldsymbol{I}_N \otimes \Sigma(\boldsymbol{\rho})$, where the vector $\boldsymbol{\rho} : T \times 1$ is the first column of the Toeplitz matrix $\Sigma(\boldsymbol{\rho})$. To the best of our knowledge, we are not

$$\boldsymbol{b}_1 = [\,0.4946 - 0.6268i; \quad -0.8183 - 0.4446i; \quad 0.4908 - 0.4101i; \quad -0.0955 - 0.2776i\,]^T$$
$$\boldsymbol{b}_2 = [\,-0.2375 + 0.5533i; \quad -0.3392 + 0.1328i; \quad 0.7326 + 0.2329i; \quad -0.8817 + 0.3693i\,]^T.$$

Fig. 5. Category 2: spatially white–temporal colored observation noise.
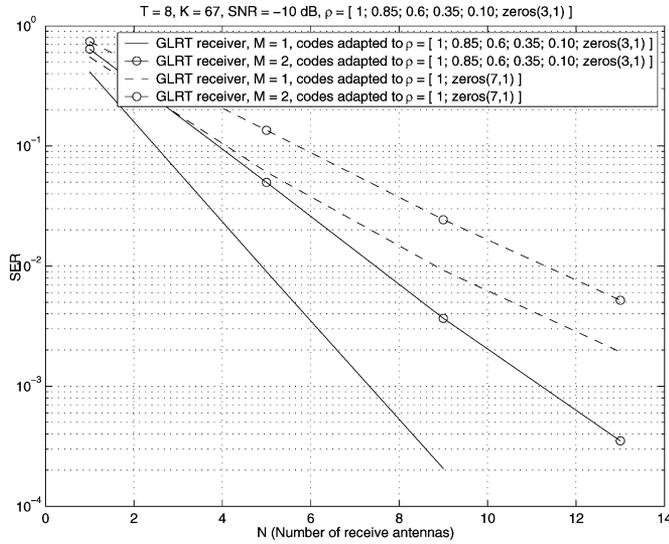
Fig. 6. Category 2: spatially white–temporal colored observation noise.

aware of any work that treats the problem of codebook constructions in the presence of spatially white–temporal colored observation noise for the low SNR regime. Hence, we compare our codes designed (adapted) to this specific scenario with codes designed when the presence of spatio–temporal white observation noise is assumed. The goal here is to demonstrate the increase of performance obtained by matching the codebook construction to the noise statistics.

- *Constellations with equal priors:* Fig. 5 plots the result of the experiment for $T = 8$, SNR $= -10$ dB, $K = 67$, and $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.10; 0; 0; 0]$. The solid and solid–circled curve represent the performance of our codes that match the noise statistics for $M = 1$ and $M = 2$, respectively. The dashed and dashed–circled curve show the performance of our codes adapted to $\boldsymbol{\rho} = [1; \text{zeros}(7,1)]$. For SER $= 2 \cdot 10^{-3}$, our $M = 1$ codes adapted to the noise statistics demonstrate the gain of three receive antennas over our $M = 2$ adapted codes, and more than six receive antennas over $M = 1$ noise mismatched codes. We conclude that for sufficiently low SNR one should construct codebook constellations with just one transmit antenna that match the noise statistics.

- *Constellations with unequal priors:* Although our primal goal in this work is to address the deterministic channel case, Fig. 6 further shows that our codebook designs for $M = 1$ are also of interest for maximum *a posteriori* (MAP) receivers that assume knowledge of the channel statistics. Fig. 6 plots the result of the experiment for $T = 6$, SNR $= -6$ dB, and $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.1; 0]$. The solid, dashed–dotted, and dashed lines represent the performances of our eight-point constellations that match the noise statistics, when the GLRT receiver is implemented for $M = 1$, $M = 2$, and $M = 3$, respectively. The plus-signed dotted line represents the performance of our eight-point constellation that is constructed for the spatio–temporal white noise case ($\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$), when GLRT receiver is implemented and $M = 1$. The plus-signed solid curve represent our 17-point constellation that match the noise
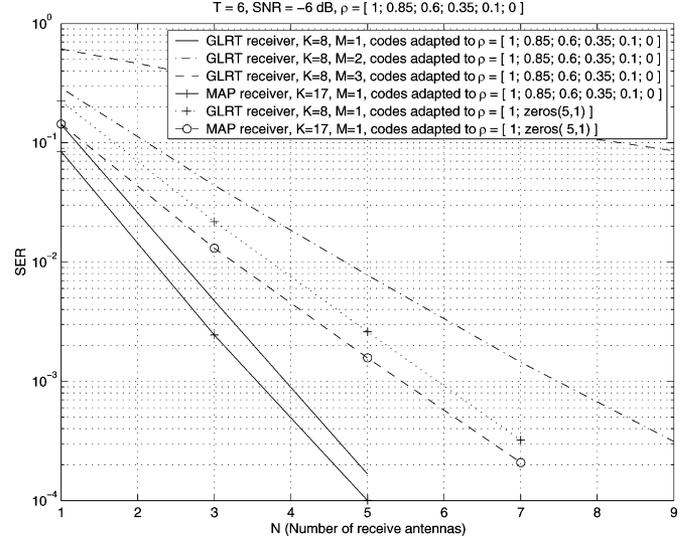
statistics and $M = 1$. The dashed–circled curve shows the performance of our 17-point constellation that is constructed for $\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$ and $M = 1$. Both 17-point constellations are with unequal priors [11] (there is a point in the origin with probability 1/2, with the probabilities of the points lying in the sphere being equal), and they use MAP receiver. The gain that 17-point constellation with unequal priors demonstrate over eight-point constellation with equal priors can be explained by the fact that the signaling scheme proposed in [11] only resembles optimal, capacity achieving distribution. The information theoretic results presented here, over the low SNR noncoherent Rayleigh fading channel with arbitrary noise correlation structure under an average power constraint, suggest that the capacity achieving distribution becomes peaky. We see that for SER of $2 \cdot 10^{-4}$, we can save two receive antennas when we compare our 17-point constellation matched to the noise statistics with the mismatched constellation constructed for $\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$. Also, as expected, for SER of $2 \cdot 10^{-4}$ and $M = 1$, two receive antennas can be spared when we compare our eight-point constellation matched to the noise statistics with the mismatched constellation constructed for $\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$.

Next, we consider correlated Rayleigh fading channel case. We treat single-beam constellations, i.e., codes with codewords of the form $\boldsymbol{S}_k = \begin{bmatrix} \boldsymbol{s}_k & \boldsymbol{0}_{T \times (M-1)} \end{bmatrix}$, and we show that significant improvements are possible over uncorrelated Rayleigh fading. Fig. 7 plots the result of the experiment. We considered the case where the coherence interval $T = 6$, $P = 0.1$, $\boldsymbol{K}_r = \boldsymbol{I}_N$, $\boldsymbol{\rho} = [1; 0.85; 0.6; 0.35; 0.1; 0]$, and rate $= 0.5$ bps/Hz, with $P = \mathsf{E}[\text{tr}(\boldsymbol{S}_k^H \boldsymbol{S}_k)]$. The correlated Rayleigh fading case where $\boldsymbol{K}_t = \widehat{\boldsymbol{K}}_t$ with $\widehat{\boldsymbol{K}}_t$ defined in Section II is assumed (again, the notation $\text{rank}(\boldsymbol{K}_t) = 1$ implies that the case when $\boldsymbol{K}_t = \widehat{\boldsymbol{K}}_t$ is treated). The dotted plus-signed and dashed–dotted curves represent our 17-point single-beam constellation with unequal priors [11], and our eight-point single-beam constellation with equal priors, respectively.
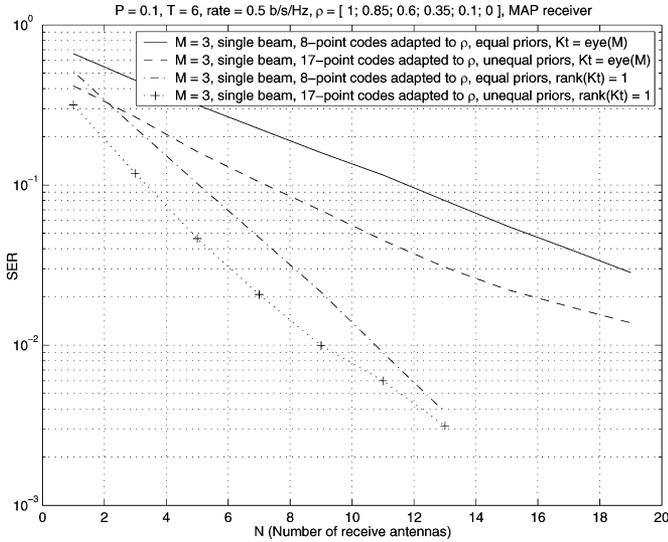
Fig. 7. Category 2: spatially white–temporal colored observation noise: $\boldsymbol{K}_r = \boldsymbol{I}_N$. All codes use MAP receiver.

Fig. 8. Category 3: solid curve—our codes for $M = 2$ adapted to colored noise, dashed curve—our codes for $M = 1$ adapted to colored noise, dashed–dotted curve—our codes for $M = 1$ adapted to white noise.

Then, we investigate the case of uncorrelated Rayleigh fading channel, i.e., $\boldsymbol{K}_t = \boldsymbol{I}_M$ where $M = 3$. The dashed curve represents our 17-point single-beam constellation with unequal priors and the solid curve represents our eight-point single-beam constellation with equal priors. For SER $= 10^{-2}$, our eight- and 17-point codes in the correlated regime perform substantially better than in uncorrelated regime (more than ten receive antennas can be spared if one decides to make the transmit antennas fully correlated). This is in concordance with the results presented in Section II where it has been proved that channel correlation can actually improve the performance of the channel in arbitrary noise environment. We witness the effect of our 17-point codes losing their superiority over eight-point codes as the number of receive antennas increases. This can be justified by the fact that the signaling scheme proposed in [11] only resembles the optimal, capacity achieving distribution.

3) Third Category: $\boldsymbol{E} = \boldsymbol{s}\boldsymbol{\alpha}^T + \boldsymbol{E}_{\text{temp}}$. We considered the case where the noise matrix is of the form $\boldsymbol{E} = \boldsymbol{s}\boldsymbol{\alpha}^T + \boldsymbol{E}_{\text{temp}}$. This models an interfering source $\boldsymbol{s}$ (with known covariance matrix $\boldsymbol{\Upsilon}_s$) where the complex vector $\boldsymbol{\alpha}$ is the known channel attenuation between each receive antenna and the interfering source. The matrix $\boldsymbol{E}_{\text{temp}}$ has a noise covariance matrix belonging to the second category. Thus, the noise covariance matrix is given by $\boldsymbol{\Upsilon} = \boldsymbol{\alpha}\boldsymbol{\alpha}^H \otimes \boldsymbol{\Upsilon}_s + \boldsymbol{I}_N \otimes \Sigma(\boldsymbol{\rho})$. As in the second category, we will compare our codes adapted to this specific scenario with codes designed for spatio–temporal white observation noise. We demonstrate the increase of performance obtained by matching the codebook construction to the noise statistics.

- *Constellations with equal priors:* Fig. 8 shows the result of the experiment for $T = 4$, $N = 2$, $K = 16$, $\boldsymbol{s} = [1; 0.7; 0.4; 0]$, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0]$, and $\boldsymbol{\alpha} = [-0.433 + 0.125\text{i}; -1.665 + 0.288\text{i}]$. The solid (dashed) curve represents our codes for $M = 2$ $(M = 1)$ adapted to colored noise, respectively, and the dashed–dotted curve represents our codes for $M = 1$ adapted to spatio–temporal white observation noise. For SER $= 10^{-2}$, we witness the
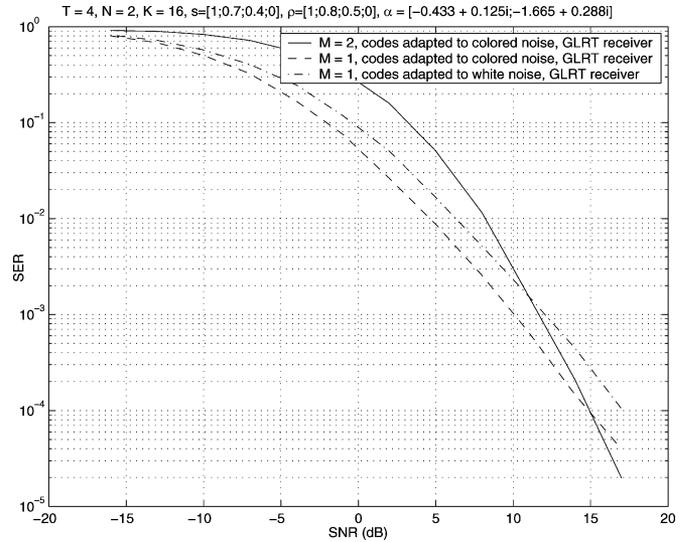
gain of 1.5 dB when we compare the one transmit antenna constellation constructed taking into account the noise statistics, with the one transmit constellation constructed for $\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$. As SNR increases, the $M = 2$ codebook construction, adapted to the noise statistics, outperforms the one antenna constellation.

The foregoing results for the cases when the noise matrix is of the form of the second and third category give rise to the following conclusion:

For a GLRT receiver, at sufficiently low SNR, one should construct codebook constellations with just one transmit antenna, but which are adapted to the noise statistics.

- *Constellations with unequal priors:* As for the case when the noise covariance matrix belongs to the second category, we demonstrate that our codebook designs for $M = 1$ are of interest for MAP receivers that assume knowledge of the channel statistics. Fig. 9 plots the result of the experiment for $T = 4$, $M = 1$, $N = 2$, $\boldsymbol{s} = [1; 0.7; 0.4; 0]$, $\boldsymbol{\rho} = [1; 0.8; 0.5; 0]$, and $\boldsymbol{\alpha} = [-0.4326 + 0.1253\text{i}; -1.6656 + 0.2877\text{i}]$. Solid–circled curve represents our 17-point codes with unequal priors [11] adapted to colored noise, which use MAP receiver. The plus-signed solid curve represents our eight-point codes with equal priors adapted to colored noise, which use ML receiver. The solid curve represents our eight-point codes with equal priors adapted to colored noise, which use GLRT receiver. The dashed–circled curve represents our 17-point codes with unequal priors adapted to white noise, which use MAP receiver. The plus-signed dashed curve represents our 8-point codes with equal priors adapted to white noise, and use ML receiver. The dashed curve represents our 8-point codes with equal priors adapted to white noise, which use GLRT receiver. For SNR from -5 to 5 dB, our one transmit antenna constellations adapted to the noise statistics demonstrate
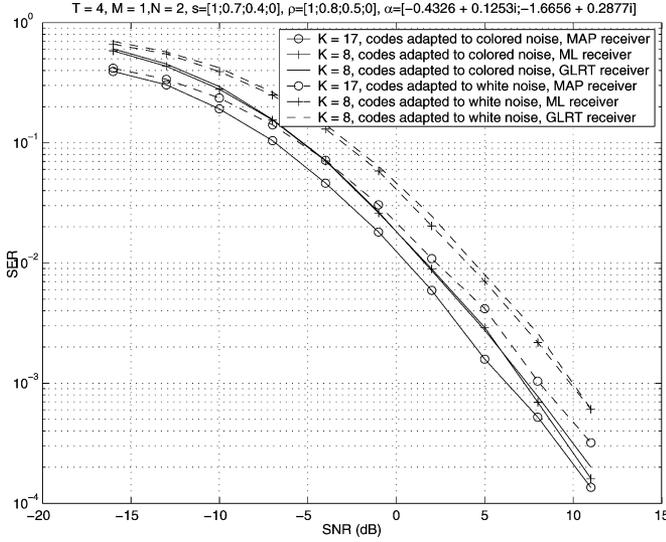
Fig. 9. Category 3.

the gain of 2 dB when compared to the codes designed for the white noise case. For SER = 10^{-2}, our 17-point codes demonstrate the gain of 1 dB when compared to the eight-point codes, which is in concordance with the theoretic-information results presented in this paper. Also, we see that our GLRT receiver performs suboptimally w.r.t. the ML receiver.

IV. CONCLUSION

We have studied the MIMO channel in the low SNR regime from two perspectives: capacity and PEP analysis. The novel aspect is that we allow the Gaussian observation noise to have an arbitrary correlation structure. From the capacity analysis perspective for correlated Rayleigh fading channel, we have shown that, by maximizing the mutual information for the ON–OFF and Gaussian signalings over the system's parameters (antenna correlation), the transmit (receive) antennas should be made as correlated as possible. Further, we have presented the PEP analysis for the low SNR deterministic channel setup and have shown

how the noise statistics could be taken into account when constructing codebook constellations. We argued that one should construct codebooks for just one transmit antenna that match the noise statistics.

APPENDIX I
MUTUAL INFORMATION FOR ON–OFF SIGNALING IN THE LOW SNR REGIME

In this section, we show that the expression for the mutual information between $\boldsymbol{X}$ and $\boldsymbol{S}$ up to first order in ρ, for ON–OFF signaling and sufficiently low SNR, is given by (2). Since

$$I(\boldsymbol{X}; \boldsymbol{S}) = h(\boldsymbol{X}) - h(\boldsymbol{X}|\boldsymbol{S})$$

we will find the mutual information by computing the conditional entropy of $\boldsymbol{X}$ given $\boldsymbol{S}$, i.e., $h(\boldsymbol{X}|\boldsymbol{S})$, and in sequel, the entropy of $\boldsymbol{X}$, i.e., $h(\boldsymbol{X})$.

- We start by computing $h(\boldsymbol{X}|\boldsymbol{S})$. Given $\boldsymbol{S}$, $\boldsymbol{x} = \text{vec}(\boldsymbol{X})$ is a zero-mean complex Gaussian with covariance $\mathsf{E}[\boldsymbol{x}\boldsymbol{x}^H|\boldsymbol{S}] = \boldsymbol{K}_r \otimes (\rho/M)\boldsymbol{S}\boldsymbol{K}_t\boldsymbol{S}^H + \boldsymbol{\Upsilon}$ and the conditional pdf of $\boldsymbol{X}$ given $\boldsymbol{S}$ is described by

$$p(\boldsymbol{X}|\boldsymbol{S}) = \frac{\exp\left\{-\boldsymbol{x}^H\left(\boldsymbol{K}_r \otimes \frac{\rho}{M}\boldsymbol{S}\boldsymbol{K}_t\boldsymbol{S}^H + \boldsymbol{\Upsilon}\right)^{-1}\boldsymbol{x}\right\}}{\pi^{TN}\det\left(\boldsymbol{K}_r \otimes \frac{\rho}{M}\boldsymbol{S}\boldsymbol{K}_t\boldsymbol{S}^H + \boldsymbol{\Upsilon}\right)}. \tag{20}$$

From (20), it is possible to compute the conditional entropy $h(\boldsymbol{X}|\boldsymbol{S})$. By definition (we use logarithms to base e), we have (21) and (22) shown at the bottom of the page.

- Next, we compute $h(\boldsymbol{X})$. We have

$$p(\boldsymbol{X}) = \rho^\epsilon \frac{\exp\left\{-\boldsymbol{x}^H\left(\boldsymbol{K}_r \otimes \frac{\rho^{1-\epsilon}}{M}\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H + \boldsymbol{\Upsilon}\right)^{-1}\boldsymbol{x}\right\}}{\pi^{TN}\det\left(\boldsymbol{K}_r \otimes \frac{\rho^{1-\epsilon}}{M}\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H + \boldsymbol{\Upsilon}\right)}$$
$$+ (1 - \rho^\epsilon)\frac{\exp\{-\boldsymbol{x}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{x}\}}{\pi^{TN}\det(\boldsymbol{\Upsilon})}. \tag{23}$$

It is not difficult to see that the second term of (23) is much larger than the first (since ρ is small and $\epsilon > 1$, ρ^ϵ is small,

$$h(\boldsymbol{X}|\boldsymbol{S}) = -\mathsf{E}_{p(\boldsymbol{X},\boldsymbol{S})}[\log p(\boldsymbol{X}|\boldsymbol{S})]$$
$$= NT\log\pi + \mathsf{E}_{p(\boldsymbol{X},\boldsymbol{S})}\left[\boldsymbol{x}^H\left(\boldsymbol{K}_r \otimes \frac{\rho}{M}\boldsymbol{S}\boldsymbol{K}_t\boldsymbol{S}^H + \boldsymbol{\Upsilon}\right)^{-1}\boldsymbol{x}\right] + \log\det(\boldsymbol{\Upsilon}) + \mathsf{E}_{p(\boldsymbol{X},\boldsymbol{S})}\left[\log\det\left(\boldsymbol{I}_{TN} + \frac{\rho}{M}\boldsymbol{Z}\right)\right],$$
$$\text{where } \boldsymbol{Z} = \boldsymbol{\Upsilon}^{-1/2}\left(\boldsymbol{K}_r \otimes \boldsymbol{S}\boldsymbol{K}_t\boldsymbol{S}^H\right)\boldsymbol{\Upsilon}^{-1/2} \quad (21)$$
$$= NT\log\pi + \log\det(\boldsymbol{\Upsilon}) + \rho^\epsilon\log\det\left(\boldsymbol{I}_{TN} + \frac{\rho^{1-\epsilon}}{M}\boldsymbol{Z}_{\text{on}}\right) + \mathsf{E}_{p(\boldsymbol{S})}\left[\mathsf{E}_{p(\boldsymbol{X}|\boldsymbol{S})}\left[\text{tr}\left(\boldsymbol{x}\boldsymbol{x}^H\left(\boldsymbol{K}_r \otimes \frac{\rho}{M}\boldsymbol{S}\boldsymbol{K}_t\boldsymbol{S}^H + \boldsymbol{\Upsilon}\right)^{-1}\right)\right]\right],$$
$$\text{where } \boldsymbol{Z}_{\text{on}} = \boldsymbol{\Upsilon}^{-1/2}\left(\boldsymbol{K}_r \otimes \boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\right)\boldsymbol{\Upsilon}^{-1/2}$$
$$\approx NT\log\pi + \log\det(\boldsymbol{\Upsilon}) + \mathsf{E}_{p(\boldsymbol{S})}[\text{tr}(\boldsymbol{I}_{NT})] + \rho^\epsilon\log\prod_{i=1}^{r}\frac{\rho^{1-\epsilon}}{M}\lambda_i,$$
$$\text{as } \rho^{1-\epsilon} \text{ is large for } \epsilon > 1 \text{ and } \rho \text{ small, } \lambda_i's \text{ are the positive eigenvalues of } \boldsymbol{Z}_{\text{on}}, \text{ and } r \text{ is its rank}$$
$$= NT\log\pi e + \log\det(\boldsymbol{\Upsilon}) + o(\rho), \qquad \text{since } \lim_{\rho \to 0^+}\rho^\epsilon\log\rho = 0, \qquad \text{for } \epsilon > 1. \tag{22}$$

the determinant in the denominator of the first term is large, whereas the numerator is bounded above by one), therefore

$$
\begin{aligned}
h(\boldsymbol{X}) \\
&= -\mathsf{E}_{p(\boldsymbol{X})}[\log p(\boldsymbol{X})] \\
&\approx -\log\left(1-\rho^\epsilon\right)+NT\log\pi+\log\det(\boldsymbol{\Upsilon})+\mathsf{E}_{p(\boldsymbol{X})}[\boldsymbol{x}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{x}] \\
&\approx \rho^\epsilon + NT\log\pi + \log\det(\boldsymbol{\Upsilon}) + \mathsf{E}_{p(\boldsymbol{X})}\left[\operatorname{tr}\left(\boldsymbol{x}\boldsymbol{x}^H\boldsymbol{\Upsilon}^{-1}\right)\right]
\end{aligned}
$$
(24)

$$
\begin{aligned}
&= NT\log\pi e + \log\det(\boldsymbol{\Upsilon}) \\
&\quad + \frac{\rho}{M}\operatorname{tr}\left(\boldsymbol{\Upsilon}^{-1}\left(\boldsymbol{K}_r\otimes\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\right)\right) + o(\rho).
\end{aligned}
$$
(25)

Equation (25) is valid because

$$
\begin{aligned}
\mathsf{E}_{p(\boldsymbol{X})}&\left[\operatorname{tr}\left(\boldsymbol{x}\boldsymbol{x}^H\boldsymbol{\Upsilon}^{-1}\right)\right] \\
&= \mathsf{E}_{p(\boldsymbol{S},\boldsymbol{h}_w,\boldsymbol{e})}\left[\operatorname{tr}\left(f_{(\boldsymbol{S},\boldsymbol{h}_w,\boldsymbol{e})}f_{(\boldsymbol{S},\boldsymbol{h}_w,\boldsymbol{e})}^H\boldsymbol{\Upsilon}^{-1}\right)\right]
\end{aligned}
$$
(26)

where

$$
f_{(\boldsymbol{S},\boldsymbol{h}_w,\boldsymbol{e})} = \left(\boldsymbol{K}_r^{1/2}\otimes\sqrt{\frac{\rho}{M}}\boldsymbol{S}\boldsymbol{K}_t^{1/2}\right)\boldsymbol{h}_w + \boldsymbol{e},\ \boldsymbol{h}_w = \operatorname{vec}(\boldsymbol{H}_w)
$$

and $\boldsymbol{e} = \operatorname{vec}(\boldsymbol{E})$. Since $\boldsymbol{S}$, $\boldsymbol{h}_w$, and $\boldsymbol{e}$ are independent random variables, from (26), we easily get

$$
\begin{aligned}
\mathsf{E}_{p(\boldsymbol{X})}&\left[\operatorname{tr}\left(\boldsymbol{x}\boldsymbol{x}^H\boldsymbol{\Upsilon}^{-1}\right)\right] \\
&= NT + \frac{\rho}{M}\operatorname{tr}\left(\boldsymbol{\Upsilon}^{-1}\left(\boldsymbol{K}_r\otimes\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\right)\right).
\end{aligned}
$$
(27)

Substituting (27) into (24) results in (25). Then, using (22) and (25), we have

$$
\begin{aligned}
I(\boldsymbol{X};\boldsymbol{S}) &= h(\boldsymbol{X}) - h(\boldsymbol{X}|\boldsymbol{S}) \\
&= \frac{\rho}{M}\operatorname{tr}\left(\boldsymbol{\Upsilon}^{-1}\left(\boldsymbol{K}_r\otimes\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\right)\right) + o(\rho)
\end{aligned}
$$

as required.

APPENDIX II
OPTIMIZATION PROBLEM FOR ON–OFF SIGNALLING

In this appendix, we prove that the maximum of the optimization problem defined in (3) is attained by the signaling scheme presented in (5) and (6).

Approach: It is known that we can always maximize (minimize) a function by first maximizing (minimizing) over some of the variables, and then maximizing (minimizing) over the remaining ones, if the constraints are independent (in the sense that each constraint function depends only on one of the variables). See, e.g., [34, p. 133] for more details. This simple and general procedure will help us to transform the problem in (3) into equivalent forms.

Let p^* represent the maximum of (3), i.e.,

$$
p^* = \max_{\substack{\operatorname{tr}\left(\boldsymbol{S}_{\text{on}}\boldsymbol{S}_{\text{on}}^H\right)\leq TM \\ \boldsymbol{\Lambda}_t\in\mathcal{D}_M,\boldsymbol{\Lambda}_r\in\mathcal{D}_N,\boldsymbol{U}_t\in\mathcal{U}_M,\boldsymbol{U}_r\in\mathcal{U}_N \\ \boldsymbol{K}_t=\boldsymbol{U}_t\boldsymbol{\Lambda}_t\boldsymbol{U}_t^H,\boldsymbol{K}_r=\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{U}_r^H}} \operatorname{tr}\left(\boldsymbol{\Upsilon}^{-1}\left(\boldsymbol{K}_r\otimes\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\right)\right)
$$
(28)

where

$$
\begin{aligned}
\mathcal{D}_n &= \{\boldsymbol{X} : n\times n \text{ diagonal matrix such that } \boldsymbol{X}\succeq\boldsymbol{0} \text{ and} \\
&\qquad \operatorname{tr}(\boldsymbol{X}) = n\}
\end{aligned}
$$
(29)
$$
\mathcal{U}_n = \{\boldsymbol{X} : n\times n \text{ unitary matrix, i.e., } \boldsymbol{X}^H\boldsymbol{X}=\boldsymbol{X}\boldsymbol{X}^H=\boldsymbol{I}_n\}.
$$
(30)

$\boldsymbol{U}_t$ ($\boldsymbol{U}_r$) is the $M\times M$ ($N\times N$) unitary matrix of the eigenvectors of $\boldsymbol{K}_t$ ($\boldsymbol{K}_r$) and $\boldsymbol{\Lambda}_t$ ($\boldsymbol{\Lambda}_r$) is the diagonal matrix of the corresponding eigenvalues (for which we assume that they are arranged in the decreasing order). That is, $\boldsymbol{U}_t\boldsymbol{\Lambda}_t\boldsymbol{U}_t^H$ ($\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{U}_r^H$) represents the singular value decomposition (SVD) of the matrix $\boldsymbol{K}_t$ ($\boldsymbol{K}_r$). Let $\widehat{\boldsymbol{S}}_{\text{on}}, \widehat{\boldsymbol{U}}_r, \widehat{\boldsymbol{\Lambda}}_r, \widehat{\boldsymbol{U}}_t$, and $\widehat{\boldsymbol{\Lambda}}_t$ denote optimal values for the matrices $\boldsymbol{S}_{\text{on}}, \boldsymbol{U}_r, \boldsymbol{\Lambda}_r, \boldsymbol{U}_t$, and $\boldsymbol{\Lambda}_t$, respectively. There are two main steps in the proof. In the first step, $\widehat{\boldsymbol{\Lambda}}_r, \widehat{\boldsymbol{U}}_t$, and $\widehat{\boldsymbol{\Lambda}}_t$ will be determined for given $\boldsymbol{S}_{\text{on}}$ and $\boldsymbol{U}_r$, whereas the second step furnishes $\widehat{\boldsymbol{S}}_{\text{on}}$ and $\widehat{\boldsymbol{U}}_r$.

Step 1: We start by rewriting the problem in (28) in the equivalent form

$$
p^* = \max_{\substack{\operatorname{tr}\left(\boldsymbol{S}_{\text{on}}\boldsymbol{S}_{\text{on}}^H\right)\leq TM \\ \boldsymbol{\Lambda}_t\in\mathcal{D}_M,\boldsymbol{\Lambda}_r\in\mathcal{D}_N,\boldsymbol{U}_t\in\mathcal{U}_M,\boldsymbol{U}_r\in\mathcal{U}_N}} \sum_{i=1}^{N}\lambda_i\operatorname{tr}\left(\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\boldsymbol{F}_i\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_i^H\right)
$$
(31)

where λ_i is the diagonal element of $\boldsymbol{\Lambda}_r$ on the position (i,i)

$$
\widehat{\boldsymbol{\Upsilon}} = \left(\boldsymbol{U}_r^H\otimes\boldsymbol{I}_T\right)\boldsymbol{\Upsilon}\left(\boldsymbol{U}_r\otimes\boldsymbol{I}_T\right) \quad \boldsymbol{F}_i = \boldsymbol{f}_i^T\otimes\boldsymbol{I}_T
$$
(32)

and $\boldsymbol{f}_i$ represents the ith column of $\boldsymbol{I}_N$. The equivalence of (28) with (31) can be established by a series of simple manipulations of the objective function of (28)

$$
\begin{aligned}
&\operatorname{tr}\left(\boldsymbol{\Upsilon}^{-1}\left(\boldsymbol{K}_r\otimes\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\right)\right) \\
&= \operatorname{tr}\left(\boldsymbol{\Upsilon}^{-1}\left(\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{U}_r^H\otimes\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\right)\right) \\
&= \operatorname{tr}\left(\boldsymbol{\Upsilon}^{-1}\left(\boldsymbol{U}_r\otimes\boldsymbol{I}_T\right)\left(\boldsymbol{\Lambda}_r\otimes\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\right)\left(\boldsymbol{U}_r^H\otimes\boldsymbol{I}_T\right)\right) \\
&= \operatorname{tr}\left(\widehat{\boldsymbol{\Upsilon}}^{-1}\left(\boldsymbol{\Lambda}_r\otimes\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\right)\right) \\
&= \sum_{i=1}^{N}\lambda_i\operatorname{tr}\left(\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H\boldsymbol{F}_i\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_i^H\right).
\end{aligned}
$$
(33)

In (33), we used the property of Kronecker product: if $\boldsymbol{AC}$ and $\boldsymbol{BD}$ exist, then $(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D}) = \boldsymbol{AC}\otimes\boldsymbol{BD}$; see [36, p. 28].

Idea: The idea that we will use in order to determine the optimal $\boldsymbol{\Lambda}_r, \boldsymbol{U}_t$, and $\boldsymbol{\Lambda}_t$ is to relax the problem in (31) [hence, by doing this, the optimal value of the new, relaxed problem will be equal or larger than of the original one in (31)], and then to show that the maximum of the relaxed problem is also attainable for the original problem. The relaxed problem is defined as

$$
q^* = \max_{\substack{\boldsymbol{A}=\boldsymbol{A}^H\succeq\boldsymbol{0},\operatorname{tr}(\boldsymbol{A})\leq TM^2 \\ \boldsymbol{\Lambda}_r\in\mathcal{D}_N,\boldsymbol{U}_r\in\mathcal{U}_N}} \sum_{i=1}^{N}\lambda_i\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{F}_i\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_i^H\right).
$$
(34)

The motivation behind the definition of the new variable $\boldsymbol{A}$, a $T\times T$ Hermitian semidefinite-positive matrix that obeys the

constraint $\text{tr}(\boldsymbol{A}) \leq TM^2$, is the fact that $\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H$ is also a Hermitian semidefinite-positive matrix with the same upper bound on the trace constraint, i.e., $\text{tr}(\boldsymbol{S}_{\text{on}}\boldsymbol{K}_t\boldsymbol{S}_{\text{on}}^H) \leq TM^2$ (which is easily verified since $\text{tr}(\boldsymbol{S}_{\text{on}}\boldsymbol{S}_{\text{on}}^H) \leq TM$). Clearly, $p^* \leq q^*$. The problem in (34) can be readily maximized by maximizing it first w.r.t. $\boldsymbol{\Lambda}_r$, and then w.r.t. $\boldsymbol{A}$. We will maximize (34) w.r.t. $\boldsymbol{\Lambda}_r$ by invoking the following fact: if a convex function f is defined on the bounded, closed convex set Ω, then, if f has a maximum over Ω it is achieved at an extreme point of Ω (see, e.g., [28, p. 119]). In our case, the function f is a linear combination of λ_i's and the set

$$\Omega = \left\{ (\lambda_1, \ldots, \lambda_N) : N \times 1 \text{ vectors} \right.$$
$$\left. \text{such that} \sum_{i=1}^{N} \lambda_i = N \text{ and } \lambda_i \geq 0 \right\}.$$

It is readily seen that the extreme points of Ω are the vectors whose all entries are zero except one which is equal to N. Regarding maximization of (34) w.r.t. $\boldsymbol{A}$, the Fan's theorem is used, see [35, p. 17]. That is, beamforming in the direction of $\lambda_{\max}(\boldsymbol{F}_i\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_i^H)$ is performed such that an upper bound on $\text{tr}(\boldsymbol{A}\boldsymbol{F}_i\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_i^H)$ is attained $(\text{tr}(\boldsymbol{A}\boldsymbol{F}_i\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_i^H) \leq \text{tr}(\boldsymbol{A})\lambda_{\max}(\boldsymbol{F}_i\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_i^H))$. Hence

$$\left[\widehat{\boldsymbol{\Lambda}}_r\right]_{i,i} = N\delta_{ii^*} \quad \widehat{\boldsymbol{A}} = TM^2\boldsymbol{U}_{i^*}\boldsymbol{K}_A\boldsymbol{U}_{i^*}^H \quad (35)$$

where the matrix $\boldsymbol{U}_{i^*}$ is the $T \times T$ unitary matrix of the eigenvectors of $\boldsymbol{F}_{i^*}\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_{i^*}^H$ (that is, we assume that $\boldsymbol{U}_{i^*}\boldsymbol{D}_{i^*}\boldsymbol{U}_{i^*}^H$ represents the SVD of the matrix $\boldsymbol{F}_{i^*}\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_{i^*}^H$ where the matrix $\boldsymbol{D}_{i^*}$ is the diagonal matrix of the corresponding eigenvalues sorted in the decreasing order); the $T \times T$ constant matrix $\boldsymbol{K}_A$ has all the entries equal to zero except the $(1, 1)$ entry which is one, $\delta_{ij} = 1$ for $i = j$, and zero otherwise; and the matrix $\bar{\boldsymbol{F}}_{i^*}$ is defined as in (32), and

$$i^* = \underset{i=1,\ldots,N}{\arg\max} \lambda_{\max}\left(\boldsymbol{F}_i\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_i^H\right). \quad (36)$$

For the choice in (35) and from (34), we get

$$q^* = NTM^2 \max_{\boldsymbol{U}_r \in \mathcal{U}_N} \lambda_{\max}\left(\boldsymbol{F}_{i^*}\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_{i^*}^H\right). \quad (37)$$

Next, we prove that the optimal value of the relaxed problem in (37) is attainable for the original problem in (31). To see this, we define $\widehat{\boldsymbol{S}}_{\text{on}}, \widehat{\boldsymbol{U}}_t$, and $\widehat{\boldsymbol{\Lambda}}_t$ in the following way:

$$\widehat{\boldsymbol{S}}_{\text{on}} = \sqrt{TM}\boldsymbol{U}_{i^*}\boldsymbol{K}_S \quad \widehat{\boldsymbol{U}}_t = \boldsymbol{I}_M$$
$$\left[\widehat{\boldsymbol{\Lambda}}_t\right]_{j,j} = M\delta_{j1}, \quad \text{for } j = 1, \ldots, M \quad (38)$$

together with $\widehat{\boldsymbol{\Lambda}}_r$ as in (35). The $T \times M$ constant matrix $\boldsymbol{K}_S$ has all the entries equal to zero except the $(1, 1)$ entry which is one, and i^* is defined in (36). In that case

$$p^* = q^* = NTM^2 \max_{\boldsymbol{U}_r \in \mathcal{U}_N} \lambda_{\max}\left(\boldsymbol{F}_{i^*}\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_{i^*}^H\right). \quad (39)$$

Remark that for the choice in (38) the power constraint on the transmitted codeword is satisfied with equality, $\widehat{\boldsymbol{K}}_t$ is a diagonal matrix, and both $\widehat{\boldsymbol{K}}_t$ and $\widehat{\boldsymbol{K}}_r$ are rank one matrices.

Step 2: In the first step, we have determined $\widehat{\boldsymbol{K}}_t$ and $\widehat{\boldsymbol{\Lambda}}_r$. It remains now to compute $\widehat{\boldsymbol{S}}_{\text{on}}$ and $\widehat{\boldsymbol{U}}_r$. First, note that from (38), we can write

$$\widehat{\boldsymbol{S}}_{\text{on}} = \sqrt{TM}\left[\boldsymbol{s} \quad \boldsymbol{0}_{T\times(M-1)}\right] \quad (40)$$

where the unitary vector $\boldsymbol{s} \in \mathbb{C}^T$ is the eigenvector associated to the maximal eigenvalue of the matrix $\boldsymbol{F}_{i^*}\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_{i^*}^H$, i.e., it is the first column of the unitary matrix $\boldsymbol{U}_{i^*}$. Hence, we need to find such a unitary vector $\boldsymbol{s}$ and a unitary matrix $\boldsymbol{U}_r$ that maximize p^* in (39). Let

$$\widehat{\lambda} = \max_{\boldsymbol{U}_r \in \mathcal{U}_N} \lambda_{\max}\left(\boldsymbol{F}_{i^*}\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_{i^*}^H\right).$$

Then, (41)–(43) shown at the bottom of the page hold. Let $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{s}})$ be the solution pair of (43), i.e.,

$$(\hat{\boldsymbol{u}}, \hat{\boldsymbol{s}}) = \underset{\boldsymbol{u} \in \mathbb{C}^N, \|\boldsymbol{u}\|=1, \boldsymbol{s} \in \mathbb{C}^T, \|\boldsymbol{s}\|=1}{\arg\max} (\boldsymbol{u} \otimes \boldsymbol{s})^H \boldsymbol{\Upsilon}^{-1}(\boldsymbol{u} \otimes \boldsymbol{s}). \quad (44)$$

Then, using (40) and the fact that $\hat{\boldsymbol{u}}$ is the i^*th column of $\widehat{\boldsymbol{U}}_r$ (where i^* is defined in (36) and due to the passage from (42) to (43) we can w.l.o.g. assume that $i^* = 1$, although any other choice of i^* would just change $\widehat{\boldsymbol{U}}_r$ but not $\widehat{\boldsymbol{K}}_r$), we have

$$\widehat{\boldsymbol{S}}_{\text{on}} = \sqrt{TM}\left[\hat{\boldsymbol{s}} \quad \boldsymbol{0}_{T\times(M-1)}\right] \quad \widehat{\boldsymbol{K}}_r = N\hat{\boldsymbol{u}}\hat{\boldsymbol{u}}^H. \quad (45)$$

Equations (35), (38), (44), and (45) complete the proof.

$$\widehat{\lambda} = \max_{\boldsymbol{U}_r \in \mathcal{U}_N} \max_{i=1,\ldots,N} \max_{\boldsymbol{s} \in \mathbb{C}^T, \|\boldsymbol{s}\|=1} \boldsymbol{s}^H\left(\boldsymbol{F}_i\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_i^H\right)\boldsymbol{s} \quad (41)$$

$$= \max_{\boldsymbol{U}_r \in \mathcal{U}_N} \max_{\substack{i=1,\ldots,N \\ \boldsymbol{s} \in \mathbb{C}^T, \|\boldsymbol{s}\|=1}} \boldsymbol{s}^H\left(\boldsymbol{f}_i^T \otimes \boldsymbol{I}_T\right)\left(\boldsymbol{U}_r^H \otimes \boldsymbol{I}_T\right)\boldsymbol{\Upsilon}^{-1}(\boldsymbol{U}_r \otimes \boldsymbol{I}_T)(\boldsymbol{f}_i \otimes \boldsymbol{I}_T)\boldsymbol{s}$$

$$\left(\text{from (32), } \widehat{\boldsymbol{\Upsilon}} = \left(\boldsymbol{U}_r^H \otimes \boldsymbol{I}_T\right)\boldsymbol{\Upsilon}(\boldsymbol{U}_r \otimes \boldsymbol{I}_T) \text{ and } \boldsymbol{F}_i = \boldsymbol{f}_i^T \otimes \boldsymbol{I}_T\right)$$

$$= \max_{\boldsymbol{U}_r=[\boldsymbol{u}_1 \quad \boldsymbol{u}_2 \quad \ldots \quad \boldsymbol{u}_N] \in \mathcal{U}_N} \max_{\substack{i=1,\ldots,N \\ \boldsymbol{s} \in \mathbb{C}^T, \|\boldsymbol{s}\|=1}} (\boldsymbol{u}_i^H \otimes \boldsymbol{s}^H)\boldsymbol{\Upsilon}^{-1}(\boldsymbol{u}_i \otimes \boldsymbol{s}) \quad (42)$$

$$= \max_{\boldsymbol{u} \in \mathbb{C}^N, \|\boldsymbol{u}\|=1, \boldsymbol{s} \in \mathbb{C}^T, \|\boldsymbol{s}\|=1} (\boldsymbol{u} \otimes \boldsymbol{s})^H \boldsymbol{\Upsilon}^{-1}(\boldsymbol{u} \otimes \boldsymbol{s}). \quad (43)$$

MUTUAL INFORMATION FOR GAUSSIAN SIGNALING
IN THE LOW SNR REGIME

In this section, we show that the expression for the mutual information between $\boldsymbol{X}$ and $\boldsymbol{S}$ up to second order in ρ, for Gaussian signaling and sufficiently low SNR, is given by (10).

First, we calculate $\mathrm{tr}(\mathsf{E}[\boldsymbol{Z}^2])$, where $\boldsymbol{Z} = \boldsymbol{\Upsilon}^{-1/2}(\boldsymbol{K}_r \otimes \boldsymbol{S}\boldsymbol{K}_t\boldsymbol{S}^H)\boldsymbol{\Upsilon}^{-1/2}$. Remark that $\mathrm{tr}(\mathsf{E}[\boldsymbol{Z}^2]) = \mathrm{tr}(\mathsf{E}[\boldsymbol{F}])$ where $\boldsymbol{F} = (\widehat{\boldsymbol{\Upsilon}}^{-1}(\boldsymbol{\Lambda}_r \otimes \widetilde{\boldsymbol{S}}\widetilde{\boldsymbol{S}}^H))^2$, the matrix $\widehat{\boldsymbol{\Upsilon}}$ is defined in (32), $\boldsymbol{K}_r = \boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{U}_r^H$ (as in Appendix II), and $\widetilde{\boldsymbol{S}} = \boldsymbol{S}\boldsymbol{K}_t^{1/2}$. If the vector $\widetilde{\boldsymbol{s}}_i$, for $i = 1, \ldots, M$, represents the ith column of the matrix $\widetilde{\boldsymbol{S}}$, then it holds

$$
\begin{aligned}
&\mathrm{tr}\left(\mathsf{E}[\boldsymbol{Z}^2]\right) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{M}\mathsf{E}\left[\mathrm{tr}\left(\widehat{\boldsymbol{\Upsilon}}^{-1}(\boldsymbol{\Lambda}_r \otimes \widetilde{\boldsymbol{s}}_i\widetilde{\boldsymbol{s}}_i^H)\widehat{\boldsymbol{\Upsilon}}^{-1}(\boldsymbol{\Lambda}_r \otimes \widetilde{\boldsymbol{s}}_j\widetilde{\boldsymbol{s}}_j^H)\right)\right] \\
&= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}\sum_{z=1}^{N}\lambda_k\lambda_z\mathsf{E}\left[\mathrm{tr}\left(\widehat{\boldsymbol{\Upsilon}}_{kz}\widetilde{\boldsymbol{s}}_i\widetilde{\boldsymbol{s}}_i^H\widehat{\boldsymbol{\Upsilon}}_{zk}\widetilde{\boldsymbol{s}}_j\widetilde{\boldsymbol{s}}_j^H\right)\right]
\end{aligned}
$$

where $\widehat{\boldsymbol{\Upsilon}}_{kz} = \boldsymbol{F}_k\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_z^H$, for $k, z = 1, \ldots, N$, the matrix $\boldsymbol{F}_k$ is defined in (32), and λ_k's are the eigenvalues of the matrix $\boldsymbol{\Lambda}_r$. Let $\Phi(\boldsymbol{A}, \boldsymbol{B}) = \mathsf{E}[\boldsymbol{y}^H\boldsymbol{A}\boldsymbol{y}\boldsymbol{y}^H\boldsymbol{B}\boldsymbol{y}]$ where $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C})$. Next, we need the following result (see, e.g., [29, p. 564]): if $\boldsymbol{y} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C})$, then for $\boldsymbol{A}, \boldsymbol{B}$ Hermitian matrices

$$
\mathsf{E}[\boldsymbol{y}^H\boldsymbol{A}\boldsymbol{y}\boldsymbol{y}^H\boldsymbol{B}\boldsymbol{y}] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{C})\mathrm{tr}(\boldsymbol{B}\boldsymbol{C}) + \mathrm{tr}(\boldsymbol{A}\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}). \quad (46)
$$

Now, by using (46), it is easy to check that

$$
\begin{aligned}
\Phi(\boldsymbol{A}, \boldsymbol{A}^H) &= \frac{1}{4}(\Phi(\boldsymbol{A} + \boldsymbol{A}^H, \boldsymbol{A} + \boldsymbol{A}^H) \\
&\quad + \Phi(j\boldsymbol{A} - j\boldsymbol{A}^H, j\boldsymbol{A} - j\boldsymbol{A}^H)) \\
&= \mathrm{tr}(\boldsymbol{A}\boldsymbol{C})\mathrm{tr}(\boldsymbol{A}^H\boldsymbol{C}) + \mathrm{tr}(\boldsymbol{A}\boldsymbol{C}\boldsymbol{A}^H\boldsymbol{C})
\end{aligned}
$$

for any matrix $\boldsymbol{A}$. This, and using the facts that $\widehat{\boldsymbol{\Upsilon}}_{kz} = \widehat{\boldsymbol{\Upsilon}}_{zk}^H$ and $\widetilde{\boldsymbol{s}}_i = \boldsymbol{E}_i\widetilde{\boldsymbol{s}}$ [where $\widetilde{\boldsymbol{s}} = \mathrm{vec}(\widetilde{\boldsymbol{S}})$ and $\boldsymbol{E}_i$ is defined in (11)], implies (47) shown at the bottom of the page, where $\widetilde{\boldsymbol{P}}_{ij} = \boldsymbol{E}_i\widetilde{\boldsymbol{P}}\boldsymbol{E}_j^H$ and $\widetilde{\boldsymbol{P}} = ((\boldsymbol{K}_t^T)^{1/2} \otimes \boldsymbol{I}_T)\boldsymbol{P}((\boldsymbol{K}_t^T)^{1/2} \otimes \boldsymbol{I}_T)$.

It can readily be shown that

$$
\begin{aligned}
&\mathrm{tr}\left((\mathsf{E}[\boldsymbol{Z}])^2\right) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{M}\mathrm{tr}\left(\boldsymbol{\Upsilon}^{-1}(\boldsymbol{K}_r \otimes \boldsymbol{E}_i\widetilde{\boldsymbol{P}}\boldsymbol{E}_i^H)\boldsymbol{\Upsilon}^{-1}(\boldsymbol{K}_r \otimes \boldsymbol{E}_j\widetilde{\boldsymbol{P}}\boldsymbol{E}_j^H)\right) \\
&= \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}\sum_{z=1}^{N}\lambda_k\lambda_z\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz}\widetilde{\boldsymbol{P}}_{ii}\widehat{\boldsymbol{\Upsilon}}_{zk}\widetilde{\boldsymbol{P}}_{jj}). \quad (48)
\end{aligned}
$$

Combining (47) with (48) results in (10). This concludes the proof.

OPTIMIZATION PROBLEM FOR GAUSSIAN SIGNALLING

In this appendix, we prove that the maximum of the optimization problem defined in (12) is attained by $\widehat{\boldsymbol{P}}$, defined in (13), and by $\widehat{\boldsymbol{K}}_t$ and $\widehat{\boldsymbol{K}}_r$ presented in (5). As in Appendix II, let $\boldsymbol{U}_t\boldsymbol{\Lambda}_t\boldsymbol{U}_t^H$ $(\boldsymbol{U}_r\boldsymbol{\Lambda}_r\boldsymbol{U}_r^H)$ represent the SVD of the matrix $\boldsymbol{K}_t$ $(\boldsymbol{K}_r)$, and let $\widehat{\boldsymbol{P}}, \widehat{\boldsymbol{U}}_r, \widehat{\boldsymbol{\Lambda}}_r, \widehat{\boldsymbol{U}}_t$, and $\widehat{\boldsymbol{\Lambda}}_t$ denote the optimal values of the matrices $\boldsymbol{P}, \boldsymbol{U}_r, \boldsymbol{\Lambda}_r, \boldsymbol{U}_t$, and $\boldsymbol{\Lambda}_t$, respectively. We repeat the analysis presented in Appendix II. In the first step, $\widehat{\boldsymbol{\Lambda}}_r$ is calculated, whereas the second step determines $\widehat{\boldsymbol{P}}, \widehat{\boldsymbol{U}}_t, \widehat{\boldsymbol{\Lambda}}_t$, and $\widehat{\boldsymbol{U}}_r$.

Step 1: Note that

$$
\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}\sum_{z=1}^{N}\lambda_k\lambda_z\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz}\widetilde{\boldsymbol{P}}_{ij})\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{zk}\widetilde{\boldsymbol{P}}_{ji}) = \boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{\lambda} \quad (49)
$$

where the vector $\boldsymbol{\lambda} = [\lambda_1 \quad \lambda_2 \quad \ldots \quad \lambda_N]^T$. The entry (k, z), for $k, z = 1, \ldots, N$, of the $N \times N$ Hermitian matrix $\boldsymbol{A}$ is given by $[\boldsymbol{A}]_{k,z} = \sum_{i=1}^{M}\sum_{j=1}^{M}\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz}\widetilde{\boldsymbol{P}}_{ij})\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{zk}\widetilde{\boldsymbol{P}}_{ji})$.

Approach: We will show that the matrix $\boldsymbol{A}$ is positive semidefinite. In that case, the function $\boldsymbol{\lambda}^T\boldsymbol{A}\boldsymbol{\lambda}$ is convex on $\mathcal{D}_N$ defined in (29), and the maximum is achieved at an extreme point of $\mathcal{D}_N$ (see [28, p. 119]). To this end, let us introduce the result that will be used in the proof: if an $NT \times NT$ matrix $\boldsymbol{Q}$ is positive semidefinite, then $N \times N$ matrix $\boldsymbol{R}$, with $[\boldsymbol{R}]_{k,z} = \sum_{i=1}^{T}\sum_{j=1}^{T}\boldsymbol{Q}((k-1)T + i, (z-1)T + j)$ for $k, z = 1, \ldots, N$, is also positive semidefinite. In other words, $[\boldsymbol{R}]_{k,z}$ is equal to the sum of all elements of the (k, z)th block of the matrix $\boldsymbol{Q}$ (the matrix $\boldsymbol{Q}$ consists in N^2 disjoint blocks where each of the blocks is of size $T \times T$). Now, remark that $\boldsymbol{R} = (\boldsymbol{I}_N \otimes \boldsymbol{1}_T)^T\boldsymbol{Q}(\boldsymbol{I}_N \otimes \boldsymbol{1}_T)$ where the $T \times 1$ vector $\boldsymbol{1}_T$ is the vector of all ones [from now on, we write $\boldsymbol{R} = \mathrm{sum}(N, T, \boldsymbol{Q})$]. Hence, the matrix $\boldsymbol{R}$ is positive semidefinite. The subsequent series of results of matrix analysis proves that $\boldsymbol{A}$ is positive semidefinite.

1) Let $\boldsymbol{\mathcal{A}}$ represent the $NTM \times NTM$ matrix obtained from $\widehat{\boldsymbol{\Upsilon}}^{-1}$ defined in (32) as explained in the following. Seen as a block matrix with N^2 disjoint blocks, where each of them is of size $MT \times MT$, any (k, z)th block of $\boldsymbol{\mathcal{A}}$, for $k, z = 1, \ldots, N$, consists of M^2 identical disjoint sub-blocks of dimension $T \times T$. For the (k, z)th block of $\boldsymbol{\mathcal{A}}$, the corresponding subblock is $\widehat{\boldsymbol{\Upsilon}}_{kz}^*$, i.e., the complex conjugate of $\widehat{\boldsymbol{\Upsilon}}_{kz} = \boldsymbol{F}_k\widehat{\boldsymbol{\Upsilon}}^{-1}\boldsymbol{F}_z^H$, where $\boldsymbol{F}_k$ is defined in (32). Note that $\boldsymbol{\mathcal{A}}$ is positive semidefinite. For any nonzero complex vector $\boldsymbol{v} = [\boldsymbol{v}_1^T \quad \ldots \quad \boldsymbol{v}_{MN}^T]^T$, where $\boldsymbol{v}_i \in \mathbb{C}^T$ for $i = 1, \ldots, MN$, it holds $\boldsymbol{v}^H\boldsymbol{\mathcal{A}}\boldsymbol{v} = \boldsymbol{u}^H(\widehat{\boldsymbol{\Upsilon}}^{-1})^*\boldsymbol{u} \geq 0$, where $\boldsymbol{u} = [\boldsymbol{u}_1^T \quad \ldots \quad \boldsymbol{u}_N^T]^T$ with $\boldsymbol{u}_j = \sum_{i=1}^{M}\boldsymbol{v}_{(j-1)M+i}$

$$
\mathrm{tr}\left(\mathsf{E}[\boldsymbol{Z}^2]\right) = \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}\sum_{z=1}^{N}\lambda_k\lambda_z\left(\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz}\widetilde{\boldsymbol{P}}_{ij})\mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{zk}\widetilde{\boldsymbol{P}}_{ji}) + \mathrm{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz}\widetilde{\boldsymbol{P}}_{ii}\widehat{\boldsymbol{\Upsilon}}_{zk}\widetilde{\boldsymbol{P}}_{jj})\right) \quad (47)
$$

for $j = 1, \ldots, N$. Since $\widehat{\boldsymbol{\Upsilon}}^{-1} \succ \mathbf{0}$, then $(\widehat{\boldsymbol{\Upsilon}}^{-1})^* \succ \mathbf{0}$ and $\mathcal{A}$ is positive semidefinite.

2) Define $\mathcal{B} = \mathbf{1}_{N \times N} \otimes \widetilde{\boldsymbol{P}}$ where $\mathbf{1}_{N \times N}$ is the $N \times N$ matrix of all ones. Then, $\mathcal{B}$ is positive semidefinite since the Kronecker product of two positive (semi)definite matrices is positive (semi)definite (see [30, p. 245]). Also, the matrix $\mathcal{C}$ defined as $\mathcal{C} = \mathcal{A} \odot \mathcal{B}$, where the symbol $\odot$ represents the Schur (Hadamard) product, is positive semidefinite since both $\mathcal{A}$ and $\mathcal{B}$ are positive semidefinite, and the Schur product of two positive (semi)definite matrices is also positive (semi)definite (see [31, p. 458]).

3) Next, we define the matrix $\mathcal{D}$ as $\mathcal{D} = \sum(MN, T, \mathcal{C})$, which is positive semidefinite. Remark that $\operatorname{sum}(1, T, \widehat{\boldsymbol{\Upsilon}}_{kz}^* \odot \widetilde{\boldsymbol{P}}_{ij}) = \operatorname{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz}^H \widetilde{\boldsymbol{P}}_{ij})$. Now, it is easy to see that the matrix $\boldsymbol{A}$ can be presented as follows: $\boldsymbol{A} = \operatorname{sum}(N, M, \mathcal{D} \odot \mathcal{D}^*)$ since

$$\sum_{i=1}^{M} \sum_{j=1}^{M} \operatorname{tr}(\widehat{\boldsymbol{\Upsilon}}_{zk} \widetilde{\boldsymbol{P}}_{ij}) \operatorname{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz} \widetilde{\boldsymbol{P}}_{ji})$$
$$= \sum_{i=1}^{M} \sum_{j=1}^{M} \operatorname{tr}(\widehat{\boldsymbol{\Upsilon}}_{kz} \widetilde{\boldsymbol{P}}_{ij}) \operatorname{tr}(\widehat{\boldsymbol{\Upsilon}}_{zk} \widetilde{\boldsymbol{P}}_{ji})$$

for $k, z = 1, \ldots, N$ and Hermitian $\widehat{\boldsymbol{\Upsilon}}^{-1}$ and $\widetilde{\boldsymbol{P}}$. Thus, as stated, the matrix $\boldsymbol{A}$ is positive semidefinite and the maximum is achieved at an extreme point of $\mathcal{D}_N$; so the optimal value of $\boldsymbol{\Lambda}_r$ is given by

$$\left[\widehat{\boldsymbol{\Lambda}}_r\right]_{k,k} = N \delta_{kk^*} \tag{50}$$

for $k = 1, \ldots, N$, where

$$\left(k^*, \widehat{\boldsymbol{P}}\right) = \underset{\substack{k=1,\ldots,N \\ \boldsymbol{P} = \boldsymbol{P}^H \succeq \mathbf{0}, \operatorname{tr}(\boldsymbol{P}) \leq TM}}{\arg \max} \sum_{i=1}^{M} \sum_{j=1}^{M} \operatorname{tr}(\widehat{\boldsymbol{\Upsilon}}_{kk} \widetilde{\boldsymbol{P}}_{ij}) \operatorname{tr}(\widehat{\boldsymbol{\Upsilon}}_{kk} \widetilde{\boldsymbol{P}}_{ji}). \tag{51}$$

The optimization problem in (13) now becomes

$$\underset{\boldsymbol{U}_r \in \mathcal{U}_N, \boldsymbol{K}_t \in \mathcal{P}_M, \boldsymbol{P} = \boldsymbol{P}^H \succeq \mathbf{0}, \operatorname{tr}(\boldsymbol{P}) \leq TM}{\max} N^2$$
$$\times \sum_{i=1}^{M} \sum_{j=1}^{M} \operatorname{tr}(\widehat{\boldsymbol{\Upsilon}}_{k^*k^*} \widetilde{\boldsymbol{P}}_{ij}) \operatorname{tr}(\widehat{\boldsymbol{\Upsilon}}_{k^*k^*} \widetilde{\boldsymbol{P}}_{ji}). \tag{52}$$

Step 2: In the second step, we will determine the optimal values of $\boldsymbol{K}_t$, $\boldsymbol{P}$, and $\boldsymbol{U}_r$.

Approach: Similarly to the analysis of Appendix II, in order to calculate $\widehat{\boldsymbol{K}}_t$, $\widehat{\boldsymbol{P}}$, and $\widehat{\boldsymbol{U}}_r$, we relax the problem in (52), find an upper bound on the relaxed problem, and show that this bound is attainable for the original problem. Let p^* be the maximum of the problem in (52), and let $\overline{\boldsymbol{P}} = (\boldsymbol{I}_M \otimes \widehat{\boldsymbol{\Upsilon}}_{k^*k^*}^{1/2}) \widetilde{\boldsymbol{P}}(\boldsymbol{I}_M \otimes$

$\widehat{\boldsymbol{\Upsilon}}_{k^*k^*}^{1/2})$ with $\operatorname{tr}(\overline{\boldsymbol{P}}) \leq K = TM^2 \lambda_{\max}(\widehat{\boldsymbol{\Upsilon}}_{k^*k^*})$. Then, the relaxed problem is defined as

$$q^* = \underset{\boldsymbol{U}_r \in \mathcal{U}_N, \overline{\boldsymbol{P}} = \overline{\boldsymbol{P}}^H \succeq \mathbf{0}, \operatorname{tr}(\overline{\boldsymbol{P}}) \leq K}{\max} N^2 \sum_{i=1}^{M} \sum_{j=1}^{M} \operatorname{tr}(\overline{\boldsymbol{P}}_{ij}) \operatorname{tr}(\overline{\boldsymbol{P}}_{ji}). \tag{53}$$

where $\overline{\boldsymbol{P}}_{ij} = \boldsymbol{E}_i \overline{\boldsymbol{P}} \boldsymbol{E}_j^H = \widehat{\boldsymbol{\Upsilon}}_{k^*k^*}^{1/2} \widetilde{\boldsymbol{P}}_{ij} \widehat{\boldsymbol{\Upsilon}}_{k^*k^*}^{1/2}$, $\widetilde{\boldsymbol{P}} = ((\boldsymbol{K}_t^T)^{1/2} \otimes \boldsymbol{I}_T) \boldsymbol{P}((\boldsymbol{K}_t^T)^{1/2} \otimes \boldsymbol{I}_T)$, and the matrix $\boldsymbol{E}_i$, for $i = 1, \ldots, M$, is defined in (11). The matrix $\widehat{\boldsymbol{\Upsilon}}_{k^*k^*}^{1/2}$ is the square root of $\widehat{\boldsymbol{\Upsilon}}_{k^*k^*}$, i.e., it is the $T \times T$ matrix such that $(\widehat{\boldsymbol{\Upsilon}}_{k^*k^*}^{1/2})^2 = \widehat{\boldsymbol{\Upsilon}}_{k^*k^*}$. Clearly, $p^* \leq q^*$. Now, note that, as $\operatorname{tr}(\overline{\boldsymbol{P}}) \leq K$, it holds

$$\sum_{i=1}^{M} \operatorname{tr}^2(\overline{\boldsymbol{P}}_{ii}) + 2 \sum_{1 \leq k < j \leq M} \operatorname{tr}(\overline{\boldsymbol{P}}_{kk}) \operatorname{tr}(\overline{\boldsymbol{P}}_{jj}) \leq K^2.$$

Also, since

$$\sum_{i=1}^{M} \sum_{j=1}^{M} \operatorname{tr}(\overline{\boldsymbol{P}}_{ij}) \operatorname{tr}(\overline{\boldsymbol{P}}_{ji}) = \sum_{i=1}^{M} \operatorname{tr}^2(\overline{\boldsymbol{P}}_{ii}) + 2 \sum_{1 \leq k < j \leq M} \operatorname{tr}(\overline{\boldsymbol{P}}_{kj}) \operatorname{tr}(\overline{\boldsymbol{P}}_{jk})$$

we can write the equation shown at the bottom of the page, due to the fact proved in the sequel that $\operatorname{tr}(\overline{\boldsymbol{P}}_{kk}) \operatorname{tr}(\overline{\boldsymbol{P}}_{jj}) \geq \operatorname{tr}(\overline{\boldsymbol{P}}_{kj}) \operatorname{tr}(\overline{\boldsymbol{P}}_{jk})$ for every pair (k, j) where $k, j = 1, \ldots, M$.

Proof: We start by noting that

$$\left([\boldsymbol{e}_k \quad \boldsymbol{e}_j]^T \otimes \boldsymbol{I}_T\right) \overline{\boldsymbol{P}} ([\boldsymbol{e}_k \quad \boldsymbol{e}_j] \otimes \boldsymbol{I}_T) = \begin{bmatrix} \overline{\boldsymbol{P}}_{kk} & \overline{\boldsymbol{P}}_{kj} \\ \overline{\boldsymbol{P}}_{jk} & \overline{\boldsymbol{P}}_{jj} \end{bmatrix} \succeq \mathbf{0}$$

where $\boldsymbol{e}_k$ represents the kth column of $\boldsymbol{I}_M$. Now, let a_i, b_i, b_i^*, and c_i, for $i = 1, \ldots, T$, represent the diagonal entries of the matrices $\overline{\boldsymbol{P}}_{kk}$, $\overline{\boldsymbol{P}}_{kj}$, $\overline{\boldsymbol{P}}_{jk}$, and $\overline{\boldsymbol{P}}_{jj}$, respectively (i.e., $\left[\overline{\boldsymbol{P}}_{kk}\right]_{i,i} = a_i$, $\left[\overline{\boldsymbol{P}}_{kj}\right]_{i,i} = b_i$, $\left[\overline{\boldsymbol{P}}_{jk}\right]_{i,i} = b_i^*$, and $\left[\overline{\boldsymbol{P}}_{jj}\right]_{i,i} = c_i$). It is not difficult to see that $(\boldsymbol{I}_2 \otimes \boldsymbol{k}_i^T) \begin{bmatrix} \overline{\boldsymbol{P}}_{kk} & \overline{\boldsymbol{P}}_{kj} \\ \overline{\boldsymbol{P}}_{jk} & \overline{\boldsymbol{P}}_{jj} \end{bmatrix} (\boldsymbol{I}_2 \otimes \boldsymbol{k}_i) = \begin{bmatrix} a_i & b_i \\ b_i^* & c_i \end{bmatrix} \succeq \mathbf{0}$ for $i = 1, \ldots T$, where $\boldsymbol{k}_i$ represents the ith column of $\boldsymbol{I}_T$. Let $\boldsymbol{S}_i = \begin{bmatrix} a_i & b_i \\ b_i^* & c_i \end{bmatrix}$. Hence, $\sum_{i=1}^{T} \boldsymbol{S}_i = \begin{bmatrix} \operatorname{tr}(\overline{\boldsymbol{P}}_{kk}) & \operatorname{tr}(\overline{\boldsymbol{P}}_{kj}) \\ \operatorname{tr}(\overline{\boldsymbol{P}}_{jk}) & \operatorname{tr}(\overline{\boldsymbol{P}}_{jj}) \end{bmatrix} \succeq \mathbf{0}$ from which the desired inequality is readily obtained. Remark that we have used the fact that the sum of positive-semidefinite matrices is a positive-semidefinite matrix. Thus, we have $p^* \leq q^* \leq N^2 K^2$. We now prove that an upper bound on the relaxed problem in (53) is attainable for the original problem in (52). To see this, we define $\widehat{\boldsymbol{P}}$ and $\widehat{\boldsymbol{K}}_t$ on the following way:

$$\widehat{\boldsymbol{P}} = TM\boldsymbol{K}_{\boldsymbol{P}} \otimes \boldsymbol{U}_{k^*} \boldsymbol{K}_{\boldsymbol{P}_1} \boldsymbol{U}_{k^*}^H, \quad \left[\widehat{\boldsymbol{K}}_t\right]_{j,j} = M\delta_{j1} \tag{54}$$

$$q^* \leq \underset{\substack{\boldsymbol{U}_r \in \mathcal{U}_N \\ \overline{\boldsymbol{P}} = \overline{\boldsymbol{P}}^H \succeq \mathbf{0}, \operatorname{tr}(\overline{\boldsymbol{P}}) \leq K}}{\max} N^2 K^2 - 2N^2 \left(\sum_{1 \leq k < j \leq M} \operatorname{tr}(\overline{\boldsymbol{P}}_{kk}) \operatorname{tr}(\overline{\boldsymbol{P}}_{jj}) - \sum_{1 \leq k < j \leq M} \operatorname{tr}(\overline{\boldsymbol{P}}_{kj}) \operatorname{tr}(\overline{\boldsymbol{P}}_{jk}) \right) \leq N^2 K^2$$

where we assume that $\boldsymbol{U}_{k^*} \boldsymbol{D}_{k^*} \boldsymbol{U}_{k^*}^H$ represents the SVD of $\widehat{\boldsymbol{\Upsilon}}_{k^* k^*}$ (the matrix $\boldsymbol{D}_{k^*}$ is the diagonal matrix of the corresponding eigenvalues sorted in the decreasing order), the $M \times M$ $(T \times T)$ constant matrix $\boldsymbol{K}_{\boldsymbol{P}}$ $(\boldsymbol{K}_{\boldsymbol{P}_1})$ has all the entries equal to zero except the $(1, 1)$ entry which is one, and k^* is defined in (51). Note that $\widehat{\boldsymbol{K}}_t$ is a diagonal matrix. In that case

$$p^* = q^* = N^2 T^2 M^4 \max_{\boldsymbol{U}_r \in \mathcal{U}_N} \lambda_{\max}^2 (\widehat{\boldsymbol{\Upsilon}}_{k^* k^*}) \qquad (55)$$

where, from (51) and (54)

$$k^* = \arg \max_{k=1,\ldots,N} \lambda_{\max}(\widehat{\boldsymbol{\Upsilon}}_{k^* k^*}).$$

By repeating the analysis of the ON–OFF signaling in Appendix II [see (41)–(43)], we easily find

$$\widehat{\boldsymbol{P}} = TM \boldsymbol{K}_{\boldsymbol{P}} \otimes \hat{\boldsymbol{s}} \hat{\boldsymbol{s}}^H, \quad \widehat{\boldsymbol{K}}_r = N \hat{\boldsymbol{u}} \hat{\boldsymbol{u}}^H \qquad (56)$$

where $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{u}}$ are the solutions of the optimization problem in (6). Equations (54) and (56) conclude the proof.

APPENDIX V
PEP FOR FAST FADING IN THE LOW SNR REGIME

In this appendix, we derive the expression for the low SNR regime PEP presented in (15). In [21], it has been shown that if $\boldsymbol{S}_i$ is transmitted, then the probability that the receiver decides in favor of $\boldsymbol{S}_j$ is

$$P_{\boldsymbol{S}_i \to \boldsymbol{S}_j} = P(\boldsymbol{e}^H (\boldsymbol{P}_i^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_i - \boldsymbol{P}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_j) \boldsymbol{e}$$
$$- 2 \Re \left\{ \boldsymbol{e}^H \boldsymbol{P}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Delta} \boldsymbol{g} \right\} > \boldsymbol{g}^H \boldsymbol{L}_{ij} \boldsymbol{g}) \quad (57)$$

where, for $k \in \{i, j\}$, $\boldsymbol{P}_k = \boldsymbol{I}_{TN} - \widetilde{\boldsymbol{S}_k} \boldsymbol{Z}_k$, $\widetilde{\boldsymbol{S}_k} = \boldsymbol{I}_N \otimes \boldsymbol{S}_k$, $\boldsymbol{Z}_k = (\boldsymbol{\mathcal{S}}_k^H \boldsymbol{\mathcal{S}}_k)^{-1} \boldsymbol{\mathcal{S}}_k^H \boldsymbol{\Upsilon}^{-1/2}$, $\boldsymbol{\mathcal{S}}_k = \boldsymbol{\Upsilon}^{-1/2} \widetilde{\boldsymbol{S}_k}$, $\boldsymbol{\Delta} = \boldsymbol{P}_j \widetilde{\boldsymbol{S}_i}$, $\boldsymbol{e} = \text{vec}(\boldsymbol{E})$, $\boldsymbol{L}_{ij} = \boldsymbol{\mathcal{S}}_i^H \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{S}}_i$, and $\boldsymbol{\Pi}_j^{\perp} = \boldsymbol{I}_{TN} - \boldsymbol{\mathcal{S}}_j (\boldsymbol{\mathcal{S}}_j^H \boldsymbol{\mathcal{S}}_j)^{-1} \boldsymbol{\mathcal{S}}_j^H$. The unknown realization of the channel is denoted by $\boldsymbol{g} = \text{vec}(\boldsymbol{H})$, whereas $\Re\{a\}$ represents the real part of the complex number a. Unfortunately, the expression (57) cannot be simplified analytically. Hence, we will analyze (57) in the low SNR regime where the linear term of $\boldsymbol{e}$ is negligible (see [21] for the analysis of (57) in the high SNR regime). Therefore, at sufficiently low SNR

$$P_{\boldsymbol{S}_i \to \boldsymbol{S}_j} \approx P \left(\boldsymbol{e}^H (\boldsymbol{P}_i^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_i - \boldsymbol{P}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_j) \boldsymbol{e} > \boldsymbol{g}^H \boldsymbol{L}_{ij} \boldsymbol{g} \right). \quad (58)$$

It is easy to see that, for $k \in \{i, j\}$, $\boldsymbol{e}^H \boldsymbol{P}_k^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_k \boldsymbol{e} \stackrel{d}{=} \boldsymbol{z}^H \boldsymbol{\Pi}_k^{\perp} \boldsymbol{z}$, where $\boldsymbol{z} = \boldsymbol{\Upsilon}^{-1/2} \boldsymbol{e} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{TN})$ and $\stackrel{d}{=}$ means equal in distribution. Then, from (58), it holds

$$P_{\boldsymbol{S}_i \to \boldsymbol{S}_j} \approx P \left(\boldsymbol{z}^H \left(\boldsymbol{U}_j \boldsymbol{U}_j^H - \boldsymbol{U}_i \boldsymbol{U}_i^H \right) \boldsymbol{z} > \boldsymbol{g}^H \boldsymbol{L}_{ij} \boldsymbol{g} \right) \quad (59)$$

where, for $k \in \{i, j\}$, $\boldsymbol{U}_k = \boldsymbol{\mathcal{S}}_k (\boldsymbol{\mathcal{S}}_k^H \boldsymbol{\mathcal{S}}_k)^{-1/2}$. That is, $\boldsymbol{U}_k$ contains an orthonormal basis for the subspace spanned by the columns of $\boldsymbol{\mathcal{S}}_k$. Notice that $\boldsymbol{U}_k^H \boldsymbol{U}_k = \boldsymbol{I}_{MN}$. To proceed with the analysis we use the known fact from [33, p. 199]: if $\boldsymbol{U}_i$ and $\boldsymbol{U}_j$ are $TN \times MN$ matrices with orthonormal columns $(T \geq M)$,

then there exist $MN \times MN$ unitary matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, and a $TN \times TN$ unitary matrix $\boldsymbol{Q}$ with the following properties.
If $2MN \leq TN$ $(2M \leq T)$, then

$$\boldsymbol{Q} \boldsymbol{U}_i \boldsymbol{W}_1 = \begin{bmatrix} \boldsymbol{I}_{MN} & \boldsymbol{0}_{MN} & \boldsymbol{0}_{MN \times (TN-2MN)} \end{bmatrix}^T$$
$$\boldsymbol{Q} \boldsymbol{U}_j \boldsymbol{W}_2 = \begin{bmatrix} \boldsymbol{C}_{ij} & \boldsymbol{S}_{ij} & \boldsymbol{0}_{MN \times (TN-2MN)} \end{bmatrix}^T \quad (60)$$

where $\boldsymbol{C}_{ij}$ is a diagonal $MN \times MN$ matrix with diagonal entries $\cos \alpha_i$, $i = 1, \ldots, MN$, $0 \leq \alpha_1 \leq \cdots \leq \alpha_{MN} \leq \pi/2$, and $\boldsymbol{S}_{ij}^2 + \boldsymbol{C}_{ij}^2 = \boldsymbol{I}_{MN}$. Now, using (60), we have $\boldsymbol{W}_2^H \boldsymbol{U}_j^H \boldsymbol{Q}^H \boldsymbol{Q} \boldsymbol{U}_i \boldsymbol{W}_1 = \boldsymbol{C}_{ij} \Rightarrow \boldsymbol{U}_j^H \boldsymbol{U}_i = \boldsymbol{W}_2 \boldsymbol{C}_{ij} \boldsymbol{W}_1^H$, so α_i, for $i = 1, \ldots, MN$, are the *principal angles* between the subspaces spanned by $\boldsymbol{U}_i$ and $\boldsymbol{U}_j$. It can be readily shown that $\boldsymbol{z}^H (\boldsymbol{U}_j \boldsymbol{U}_j^H - \boldsymbol{U}_i \boldsymbol{U}_i^H) \boldsymbol{z} \stackrel{d}{=} \boldsymbol{c}^H \boldsymbol{D} \boldsymbol{c}$, where $\boldsymbol{D} = \begin{bmatrix} -\boldsymbol{S}_{ij}^2 & \boldsymbol{C}_{ij} \boldsymbol{S}_{ij} \\ \boldsymbol{C}_{ij} \boldsymbol{S}_{ij} & \boldsymbol{S}_{ij}^2 \end{bmatrix}$ and $\boldsymbol{c} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{2MN})$. Let $\boldsymbol{c} = \begin{bmatrix} a_1 & b_1 & \ldots & a_{MN} & b_{MN} \end{bmatrix}^T$ where a_m, $b_m \stackrel{iid}{\sim} \mathcal{CN}(0, 1)$ for $m = 1, \ldots, MN$. Hence, using the fact that $\pm \sin \alpha_i$, for $i = 1, \ldots, MN$, are the eigenvalues of the matrix $\boldsymbol{D}$, from (59), we easily get

$$P_{\boldsymbol{S}_i \to \boldsymbol{S}_j} \approx P \left(\sum_{m=1}^{MN} \sin \alpha_m \left(|a_m|^2 - |b_m|^2 \right) > \boldsymbol{g}^H \boldsymbol{L}_{ij} \boldsymbol{g} \right)$$

as required.

ACKNOWLEDGMENT

REFERENCES

[1] A. Lozano, A. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.

[2] M. J. Borran, A. Sabharwal, and B. Aazhang, "On design criteria and construction of non-coherent space-time constellations," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2332–2351, Oct. 2003.

[3] S. A. Aldosari and J. M. F. Moura, "Detection in sensor networks: The saddlepoint approximation," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 327–340, Jan. 2007.

[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. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[6] V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[7] B. Hajek and V. Subramaniam, "Capacity and reliability function for small peak signal constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, 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. IEEE Inf. Theory Workshop*, 2004, pp. 428–433.

[10] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1962–1973, Sep. 2000.

[11] S. G. Srinivasan and M. K. Varanasi, "Constellation design with unequal priors and new distance criteria for the low SNR noncoherent Rayleigh fading channel," presented at the Conf. Inf. Sci. Syst., Baltimore, MD, Mar. 2005.

[12] S. G. Srinivasan and M. K. Varanasi, "Code design for the low SNR noncoherent MIMO block Rayleigh fading channel," in *Proc. IEEE Inf. Theory*, Sep. 2005, pp. 2218–2222.

[13] M. Beko, J. Xavier, and V. Barroso, "Codebook design for the non-coherent GLRT receiver and low SNR MIMO block fading channel," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Cannes, France, Jul. 2–5, 2006, DOI: 10.1109/SPAWC.2006.346468.

[14] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[15] S. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990–997, May 2005.

[16] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, "Multiple input multiple output measurements and modeling in Manhattan," in *Proc. IEEE Veh. Technol. Conf.*, Vancouver, BC, Canada, 2002, pp. 107–110.

[17] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[18] W. Weichselberger, M. Herdin, H. Ozcelik, and 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.

[19] E. G. Larsson and P. Stoica, *Space-Time Block Codes for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[20] H. Bolcskei, D. Gesbert, C. B. Papadias, and A. -J. van der Veen, Eds., *Space-Time Wireless Systems: From Array Processing to MIMO Communications*. Cambridge, U.K.: , Cambridge Univ. Press, 2006.

[21] M. Beko, J. Xavier, and V. Barroso, "Codebook design for non-coherent communication in multiple-antenna systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Toulouse, France, 2006, vol. 4, pp. IV–VI, DOI: 10.1109/ICASSP.2006.1661024.

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

[23] M. Beko, J. Xavier, and V. Barroso, "Capacity and error probability analysis of non-coherent MIMO systems in the low SNR regime," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Honolulu, HI, 2007, vol. 3, pp. III-405–III-408, DOI: 10.1109/ICASSP.2007.366558.

[24] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[25] J. A. Tropp, "Topics in sparse approximation," Ph.D. dissertation, Comput. Appl. Math. Dept., Univ. Texas, Austin, TX, 2004.

[26] X. Wu, "Wireless communications in the energy-limited regime," Ph.D. dissertation, Electr. Eng. Dept., Univ. Illinios, Urbana-Champaign, IL, 2004.

[27] G. Barriac and U. Madhow, "Space-time communication for OFDM with implicit channel feedback," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3111–3129, Dec. 2004.

[28] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1972.

[29] S. M. Kay, "Fundamentals of statistical signal processing," in *Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[30] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[31] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[32] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. New York: Addison-Wesley, 1990.

[33] R. Bhatia, *Matrix Analysis*. New York: Springer-Verlag, 1997.

[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

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

[36] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Rev. ed. New York: Wiley, 1999.

Marko Beko (S'06) was born in Belgrade, Serbia, on November 11, 1977. He received the Dipl. Eng. degree from the University of Belgrade, Belgrade, Serbia, in 2001 and the Ph.D. degree in electrical and computer engineering from Instituto Superior Tecnico (IST), Lisbon, Portugal, in 2008.

Currently, he is an Adjunct Professor at the Polytechnic Institute of Leiria, Portugal. He is also a Researcher at the Institute of Systems and Robotics (ISR), Lisbon, Portugal. His current research interests include coding and modulation for wireless MIMO communications and corresponding performance limits.

João Xavier (S'97–M'03) received the Ph.D. degree in electrical and computer engineering from Instituto Superior Tecnico (IST), Lisbon, Portugal, in 2002.

Currently, he is an Assistant Professor in the Department of Electrical and Computer Engineering, IST. He is also a Researcher at the Institute of Systems and Robotics (ISR), Lisbon, Portugal. His current research interests are in the area of statistical signal processing on manifolds and nonsmooth optimization.

Victor A. N. Barroso (M'89–SM'00) was born in Lisbon, Portugal, on December 22, 1952. He received the E.E. degree in 1976 and the Ph.D. degree in electrical and computer engineering in 1990, both from Instituto Superior Técnico (IST), Lisbon, Portugal. He received the title of "Professor com Agregação" in Electrical and Computer Engineering in 2003 from Universidade Técnica de Lisboa—IST.

In 1976, he joined the faculty of the Department of Electrical and Computer Engineering, IST. He was promoted to Assistant Professor in 1990, to Associate Professor in 1993, and to Full Professor in 2004. He has taught courses in systems and signal theory, control systems, signal and array processing, computer networks, and communications. He was Vice-President of the IST's Scientific Council from November 2001 to March 2007. Currently, he is the Director of the Portuguese pole of the Information and Communication Technology Institute (ICTI@Portugal), institution responsible for the management of the collaboration program involving several Portuguese Universities, research and development institutions, and companies with the Carnegie Mellon University. He is also the Director of Institute for Systems and Robotics (ISR), Lisbon, Portugal. His research interests are in statistical signal/image and array processing and wireless communications. He has published over 100 journal and conference papers.

Dr. Barroso was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and for the IEEE SIGNAL PROCESSING LETTERS. He was the representative of the IEEE Signal Processing Society at the IEEE Wireless Periodicals Working Group in 2000 and the General Chair of the IEEE International Workshop on Signal Processing Advances in Wireless Communications in 2004. He is affiliated with several IEEE societies.