Non-coherent Communication in Multiple-Antenna Systems: Receiver design and Codebook construction

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

Abstract

We address the problem of space-time codebook design for non-coherent communications in multiple-antenna wireless systems. In contrast with other approaches, the channel matrix is modeled as an unknown deterministic parameter at both the receiver and the transmitter, and the Gaussian observation noise is allowed to have an arbitrary correlation structure, known by the transmitter and the receiver. In order to handle the unknown deterministic space-time channel, a generalized likelihood ratio test (GLRT) receiver is considered. A new methodology for space-time codebook design under this non-coherent setup is proposed. It optimizes the probability of error of the GLRT receiver's detector in the high signal-to-noise ratio (SNR) regime by solving a high-dimensional nonlinear non-smooth optimization problem in a two-step approach. (i) Firstly, a convex semidefinite programming (SDP) relaxation of the codebook design problem yields a rough estimate of the optimal codebook. (ii) This is then refined through a geodesic descent optimization algorithm that exploits the Riemannian geometry imposed by the power constraints on the space-time codewords. The results obtained through computer simulations illustrate the advantages of our method. For the specific case of spatio-temporal white observation noise, our codebook constructions replicate the performance of state-of-art known solutions. The main point here is that our methodology permits to extend the codebook construction to any given correlated noise environment. The simulation results show the good performance of these new designed codes in colored noise setups.

Index Terms

Multiple-input multiple output (MIMO) systems, non-coherent communications, colored noise, space-time constellations, generalized likelihood ratio test (GLRT) receiver, semidefinite programming (SDP), geodesic descent algorithm (GDA), Grassmannian packings, equiangular tight frame (ETF).

The material in this paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Toulouse, France, 2006.

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

M. Beko, J. Xavier and V. Barroso are with the Instituto Superior Técnico – Instituto de Sistemas e Robótica, Av. Rovisco Pais, 1049-001 Lisboa, Portugal (e-mail: marko@isr.ist.utl.pt, jxavier@isr.ist.utl.pt, vab@isr.ist.utl.pt).

I. INTRODUCTION

THE main challenges in designing wireless communication systems include addressing the highly random channel conditions, which may vary rapidly, and also the additive observation noise at the receiver. Exploiting temporal and spatial diversity employing multiple antennas at the transmitter and receiver and encoding the data over several symbol intervals, known as *space-time coding*, has shown to be an efficient approach when dealing with the problems aforesaid, provided that either channel state information (CSI) is accessible at the receiver [1], [2], [3], [4], or the signal power at the receiver is significantly higher than the power of the additive observation noise [5], [6], [7]. In slowly fading scenarios, when the fading channel coefficients remain approximately constant for many symbol intervals, channel stability enables the receiver to be trained in order to acquire the CSI. This is usually referred to as *coherent detection*. Code design for the coherent systems is performed with the assumption that the CSI is available at the receiver. It is known [1], [3] that, when CSI is available at the receiver, the maximal achievable rate, referred to as *capacity* of the link, increases linearly (for rich scattering environments) with the minimum number of transmit and receive antennas. In fast fading scenarios, fading coefficient change into new, almost independent values before being learned by the receiver through training signals. Using multiple antennas at the transmitter increases the number of parameter to be estimated at the receiver which makes this problem more serious. This makes the *non-coherent* detection mode, where the receiver detects the transmitted symbols without having information about the realization of the channel, an attractive option for these fast fading scenarios.

Previous work. The capacity of non-coherent multiple antenna systems was studied in [5], [6]. Under the additive white observation Gaussian noise and Rayleigh channel assumptions, it has been shown that at high signal-to-noise ratio (SNR), or when the coherence interval, T, is much bigger than the number of transmit antennas M, capacity can be achieved by using constellations of unitary matrices as codebooks. These *unitary constellations* are capacity optimal. Furthermore, in [8] has been shown that, under the assumption of equal-energy codewords and high SNR, scaled unitary codebooks optimize the union bound (UB) on the error probability. Hence, at high SNR unitary constellations are optimal from both the capacity and symbol error probability viewpoints. Optimal unitary constellations correspond to optimal packings in Grassmann manifolds [9]. In [7], [10], a systematic method for designing unitary space-time constellations was presented. In [11], Sloane's algorithms [12] for producing sphere packings in real Grassmannian space have been extended to complex Grassmannian space. For a small number of transmit antennas, by using *chordal distance* as the design criterion, the corresponding constellations improve on the bit error rate (BER) when compared with the unitary space-time constellations presented in [7]. In [13] the problem of designing signal constellations for the multiple antenna noncoherent Rayleigh fading channel has been examined. The asymptotic UB on the probability of error has been considered, which, consequently, gave rise to a different notion of distance on the Grassmann manifold. By doing this, a method of iteratively designing signals, called successive updates, has been introduced. The signals obtained therein are, in contrast to [7], [11], guaranteed to achieve the full diversity order of the channel. Later, Borran et. al. [14], under the assumption of equally probable codewords, presented a technique that uses Kullback-Liebler (KL) divergence between the probability density functions induced at the receiver by distinct transmitted codewords as a design criterion for codebook construction. The codes thereby obtained collapse to the unitary constellations at high SNR, but at low SNR they have a multilevel structure and show improvement over unitary constellations of the same size. In [15] a family of space-time codes suited for noncoherent multi-input multi-output (MIMO) systems was presented. These codes use all the degrees of freedom of the system, and they are constructed as codes on the Grassmann manifold by the exponential map. Recently, in [16], [17] some sub-optimal simplified decodings for the class of unitary space-time codes obtained via the exponential map were presented.

The techniques aforementioned can not be readily extended to the more realistic and challenging scenario, where the Gaussian observation noise has an arbitrary correlation structure. The assumption of spatio-temporal Gaussian observation noise is common, as there are at least two reasons for making it. First, it yields mathematical expressions that are relatively easy to deal with. Second, it can be justified via the central limit theorem. Although customary, the assumption of spatio-temporal *white* Gaussian observation noise is clearly an approximation. In general, in realistic scenarios, the noise term might have very rich correlation structure, e.g, see pp. 10, 159, 171 in [18]. The generalization to arbitrary noise covariance matrices encompasses many scenarios of interest as special cases: spatially colored or not jointly with temporally colored or not observation noise, multiuser environment, etc. Intuitively, unitary space-time constellations are not the optimal ones for this scenario.

In this paper, we look for a more practical code design criterion based on error probability, rather than capacity analysis. The calculus of the exact expression for the average error probability for the general non-coherent systems seems not to be tractable. Instead, we consider pairwise error probability (PEP) in high SNR regime, and use it to find a code design criterium (a merit function) for an arbitrarily given noise correlation structure.

Contribution. Our contributions in this area are summarized in the following. (**1**) The main contribution of this paper is a new technique that systematically designs space-time codebooks for non-coherent multiple-antenna communication systems. Contrary to other approaches, the Gaussian observation noise may have an arbitrary correlation structure. In general correlated noise environments, computer simulations show that the space-time codes obtained with our method significantly outperform those already known which were constructed for spatio-temporally white noise case. We recall that codebook constructions for arbitrary noise correlation structures were not previously available and this demonstrates the interest of the codebook design methodology introduced herein. (**2**) For the special case of spatio-temporal white observation noise, our codebooks converge to the previously known unitary structure, namely the codes in [7] (in fact, our codes are marginally better). Also, for this specific scenario and $M = 1$ we show that the problem of finding good codes coincides with the very well known packing problem in the complex projective space. We compare our best configurations against the codes in [10] and the Rankin bound. We manage to improve the best known results and in some cases actually provide optimal packings in complex projective spaces which attain the Rankin upper bound. (**3**) Theoretical analysis leading to an upper bound on PEP in the high SNR scenario for the Gaussian observation noise with an arbitrary correlation structure.

Paper organization. The paper is organized as follows. In section II, we introduce the data model and formulate the problem addressed in this paper. We describe the structure of our non-coherent receiver and discuss the selection of the codebook design criterium. In section IV we propose a new algorithm that systematically designs non-coherent space-time constellations for an arbitrarily given noise covariance matrix and any M, N, K and T, respectively, number of transmitter antennas, number of receiver antennas, size of codebook, and channel coherence interval. In Section V, we present codebook constructions for several important special cases and compare their performance with state-of-art solutions. Section VI presents the main conclusions of our paper.

Throughout the 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 expectation operator is denoted by $\mathsf{E}[\,\cdot\,]$. For any matrix $\boldsymbol{A}$ we write its trace as $\mathsf{tr}(\boldsymbol{A})$. The symbol $\|\boldsymbol{A}\| = \sqrt{\mathsf{tr}(\boldsymbol{A}^H \boldsymbol{A})}$ denotes the Frobenius norm. The Kronecker product of two matrices is denoted by $\otimes$. The N-dimensional identity matrix is denoted by $\boldsymbol{I}_N$ and the $M \times N$ matrix of all zeros by $\boldsymbol{0}_{M\times N}$ (also $\boldsymbol{0}_N = \boldsymbol{0}_{N\times N}$). The minimum (maximum) eigenvalue of the Hermitean 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 operator $\mathsf{vec}(\boldsymbol{A})$ stacks all columns of the matrix $\boldsymbol{A}$ on top of each other, from left to right. The curled inequality symbol $\succeq$ represents matrix inequality between Hermitean matrices. For matrices $\boldsymbol{A}_i : n_i \times n_i$, $\mathsf{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m)$ is the $n \times n$, $n = n_1 + ... + n_m$, block diagonal matrix obtained by diagonal concatenation of the $\boldsymbol{A}_i$'s.

II. PROBLEM FORMULATION

Data model and assumptions. The communication system comprises M transmit and N receive antennas and we assume a block flat fading channel model with coherence interval T. That is, we assume that the fading coefficients remain constant during blocks of T consecutive symbol intervals, and change into new, independent values at the end of each block. In complex base band notation we have the model

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

where $\boldsymbol{X}$ is the $T \times M$ matrix of transmitted symbols (the matrix $\boldsymbol{X}$ is called hereafter a space-time codeword), $\boldsymbol{Y}$ is the $T \times N$ matrix of received symbols, $\boldsymbol{H}$ is the $N \times M$ matrix of channel coefficients, and $\boldsymbol{E}$ is the $T \times N$ matrix of zero-mean additive observation noise. In $\boldsymbol{Y}$, time indexes the rows and space (receive antennas) indexes the columns. We shall work under the following assumptions. **A1 (Channel matrix)** The channel matrix $\boldsymbol{H}$ is not known at the receiver neither at the transmitter, and no stochastic model is assumed for it. **A2 (Transmit power constraint)** The codeword $\boldsymbol{X}$ is chosen from a finite codebook $\mathcal{C} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_K\}$ known to the receiver, where K is the size of the codebook. We impose the power constraint $\mathsf{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k) = 1$ for each codeword. Furthermore, we assume that $T \geq M$ and each codeword is of full rank, i.e., $\mathsf{rank}(\boldsymbol{X}) = M$; **A3 (Noise distribution)** The observation noise at the receiver is zero mean and obeys circular complex Gaussian statistics, that is, $\mathsf{vec}(\boldsymbol{E}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{\Upsilon})$.

The noise covariance matrix $\boldsymbol{\Upsilon} = \mathsf{E}[\mathsf{vec}(\boldsymbol{E})\mathsf{vec}(\boldsymbol{E})^H]$ is known at the transmitter and at the receiver. Remark that in assumption A3, we let the data model depart from the customary assumption of spatio-temporal *white* Gaussian observation noise. Also, note that one cannot perform "pre-whitening" in order to revert the general colored case ($\boldsymbol{\Upsilon} \neq \boldsymbol{I}_{TN}$) into the spatio-temporal white noise case ($\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$) since the "vector level" pre-whitening would break down the matrix structure of the data model.

Receiver. According to the system model (1) and the assumptions above mentioned, the conditional probability density function of the received vector $\boldsymbol{y} = \mathsf{vec}(\boldsymbol{Y})$, given the transmitted matrix $\boldsymbol{X}$ and the unknown realization of the channel $\boldsymbol{h} = \mathsf{vec}\left(\boldsymbol{H}^H\right)$, is given by

$$p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{h}) = \frac{\exp\left\{-\left\|\boldsymbol{y} - (\boldsymbol{I}_N \otimes \boldsymbol{X})\boldsymbol{h}\right\|^2_{\boldsymbol{\Upsilon}^{-1}}\right\}}{\pi^{TN}\det\boldsymbol{\Upsilon}},$$

where the notation $||\boldsymbol{z}||^2_{\boldsymbol{A}} = \boldsymbol{z}^H\boldsymbol{A}\boldsymbol{z}$ was 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. We assume a generalized likelihood ratio test (GLRT) receiver which decides the index *k* of the codeword as

$$\widehat{k} = \underset{k=1,2,\ldots,K}{\arg\min} \left\|\boldsymbol{y} - (\boldsymbol{I}_n \otimes \boldsymbol{X}_k)\,\widehat{\boldsymbol{h}}_k\right\|^2_{\boldsymbol{\Upsilon}^{-1}}$$

where

$$\widehat{\boldsymbol{h}}_k = (\boldsymbol{\mathcal{X}}_k^H\boldsymbol{\mathcal{X}}_k)^{-1}\boldsymbol{\mathcal{X}}_k^H\boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{y} \quad \text{and} \quad \boldsymbol{\mathcal{X}}_k = \boldsymbol{\Upsilon}^{-1/2}\left(\boldsymbol{I}_N \otimes \boldsymbol{X}_k\right). \tag{2}$$

In the sequel, we also refer to $\boldsymbol{\mathcal{X}}_k$ as a codeword. The GLRT [21] is composed of a bank of K parallel processors where the *k*-th processor assumes the presence of the *k*-th codeword and computes a likelihood, after replacing the channel by its maximum likelihood (ML) estimate. The GLRT detector chooses the codeword associated with the processor exhibiting the largest likelihood. We note that the GLRT performs sub-optimally when compared with the ML receiver, as the latter can exploit the knowledge of channel statistics'. However, since assumption A1 is in force (i.e., channel statistics are unknown), the GLRT yields an attractive (implementable) solution in the present setup. Note also that, for the special case of unitary constellations, i.e., $\boldsymbol{X}_k^H\boldsymbol{X}_k = \frac{1}{M}\boldsymbol{I}_M$ for all k, and spatio-temporal white Gaussian noise, it is readily shown that the two receivers coincide.

Codebook design criterion. In this paper, our goal is to design a codebook $\mathcal{C} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_K\}$ of size K for the current setup. A codebook $\mathcal{C}$ is a point in the space $\mathcal{M} = \{(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K) : \mathsf{tr}(\boldsymbol{X}_k^H\boldsymbol{X}_k) = 1\}$. First, we must adopt a merit function $f : \mathcal{M} \to \mathbb{R}$ which gauges the quality of each constellation $\mathcal{C}$. The average error probability for a specific $\mathcal{C}$ would be the natural choice, but the theoretical analysis seems to be intractable. Instead, as usual [6], [7], we rely on a PEP study to construct our merit function. For the special case of unitary codebooks, spatio-temporal white Gaussian noise ($\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$) and independent identically distributed (iid) Rayleigh fading, the exact expression for the PEP has been derived in [6]. However, its computation for arbitrary matrix constellations

$\mathcal{C}$ and noise correlation matrix $\boldsymbol{\Upsilon}$ seems to be burdensome. As in [6], [7], in this paper, we focus on the high SNR regime. We resort to the asymptotic expression of the PEP in this regime, for arbitrary $\mathcal{C}$ and $\boldsymbol{\Upsilon}$. This paper completes our previous work in [19]. For a corresponding study in the low SNR regime, see [20]. To start the PEP analysis, we consider a codebook with only two codewords, i.e., $\mathcal{C} = \{\boldsymbol{X}_1, \boldsymbol{X}_2\}$. Let $P_{\boldsymbol{X}_i \to \boldsymbol{X}_j}$ be the probability of the GLRT receiver deciding $\boldsymbol{X}_j$ when $\boldsymbol{X}_i$ is sent. In Appendix I it is shown that, at sufficiently high SNR, we have the approximation

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} \simeq \mathcal{Q}\left(\frac{1}{\sqrt{2}}\sqrt{\boldsymbol{h}^H \boldsymbol{L}_{ij}(\mathcal{C})\boldsymbol{h}}\right), \tag{3}$$

where

$$\boldsymbol{L}_{ij}(\mathcal{C}) = \boldsymbol{\mathcal{X}}_i^H \left(\boldsymbol{I}_{TN} - \boldsymbol{\mathcal{X}}_j \left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-1} \boldsymbol{\mathcal{X}}_j^H\right) \boldsymbol{\mathcal{X}}_i \tag{4}$$

and $\mathcal{Q}(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$ is the $\mathcal{Q}$-function. Equation (3) shows that the probability of misdetecting $\boldsymbol{X}_i$ for $\boldsymbol{X}_j$, depends on the channel realization $\boldsymbol{h} = \mathrm{vec}\left(\boldsymbol{H}^H\right)$ and on the relative geometry of the codewords $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{X}}_j$. We can decouple the action of $\boldsymbol{h}$ and $\boldsymbol{L}_{ij}(\mathcal{C})$ as follows: using the inequality $\boldsymbol{h}^H \boldsymbol{L}_{ij}(\mathcal{C})\boldsymbol{h} \geq \lambda_{\min}(\boldsymbol{L}_{ij}(\mathcal{C}))\, ||\boldsymbol{h}||^2$ (which is an equality when $M = 1$ and $\boldsymbol{\Upsilon} = \boldsymbol{I}_{NT}$) and the fact that $\mathcal{Q}(\cdot)$ is monotonically non-increasing, we have the upper bound on the PEP for high SNR

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} \leq \mathcal{Q}\left(\frac{1}{\sqrt{2}}\, ||\boldsymbol{h}|| \sqrt{\lambda_{\min}(\boldsymbol{L}_{ij}\,(\mathcal{C}))}\right). \tag{5}$$

We cannot control the power of the channel $\boldsymbol{h}$, but we can design codebooks $\mathcal{C}$ aiming at maximizing $\lambda_{\min}(\boldsymbol{L}_{ij}\,(\mathcal{C}))$.

Problem formulation. Following a worst-case approach, we are led from (5) to define the codebook merit function $f : \mathcal{M} \to \mathbb{R}$, $\mathcal{C} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K\} \mapsto f(\mathcal{C})$ as

$$f(\mathcal{C}) = \min\left\{f_{ij}(\mathcal{C}) : 1 \leq i \neq j \leq K\right\} \tag{6}$$

where $f_{ij}(\mathcal{C}) = \lambda_{\min}(\boldsymbol{L}_{ij}(\mathcal{C}))$ with $\boldsymbol{L}_{ij}$ defined as in (4). Note that f_{ij} depends on $\mathcal{C}$ only through the two codewords $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$. Constructing an optimal codebook $\mathcal{C} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_K\}$ amounts to solving the optimization problem

$$\mathcal{C}^* = \underset{\mathcal{C} \in \mathcal{M}}{\arg\max}\quad f(\mathcal{C}). \tag{7}$$

The problem defined in (7) is a high-dimensional, non-linear and non-smooth optimization problem. As an example, for a codebook of size $K = 256$ the number of f_{ij} functions is $K(K-1) = 65280$. Also, for $T = 8$ and $M = 2$, there are $2KTM = 8192$ real variables to optimize.

III. Considerations about the new codebook merit function

Before addressing the codebook design problem (6) we draw in this section some conclusions about the codebook merit function f in (6). In subsection III-A, we show that, for the special case of spatio-temporally white noise and $K = 2$, the unitary constellations are the optimal ones with respect to f. In subsection III-B, we show that, when restricting attention to unitary codebooks, our codebook design criterion corresponds to a packing problem in the Grassmannian space with respect to *spectral distance*, for the white noise case.

A. Optimality of unitary codewords for the white noise case

Consider a codebook with two codewords $\mathcal{C} = \{\boldsymbol{X}_1, \boldsymbol{X}_2\}$. We want to maximize $f(\mathcal{C}) = \min\{\lambda_{\min}(\boldsymbol{L}_{ij}(\mathcal{C})) : i \neq j\}$ subject to $\mathsf{tr}\left(\boldsymbol{X}_k^H \boldsymbol{X}_k\right) = 1$. There are two distinct situations: $T \geq 2M$ and $T < 2M$. We start with $T \geq 2M$. We rewrite $\boldsymbol{L}_{ij}(\mathcal{C})$ as

$$\boldsymbol{L}_{ij}(\mathcal{C}) = \left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right)^{\frac{1}{2}} \left(\boldsymbol{I}_{MN} - \boldsymbol{\mathcal{U}}_i^H \boldsymbol{\mathcal{U}}_j \boldsymbol{\mathcal{U}}_j^H \boldsymbol{\mathcal{U}}_i\right) \left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right)^{\frac{1}{2}} \tag{8}$$

where $\boldsymbol{\mathcal{U}}_k = \boldsymbol{\mathcal{X}}_k \left(\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k\right)^{-\frac{1}{2}}$ contain an orthonormal basis for the subspace spanned by the columns of $\boldsymbol{\mathcal{X}}_k$. To proceed with the analysis we use an useful fact from the CS decomposition, see [23] pp. 199: there exist unitary matrices $\boldsymbol{\mathcal{W}}_i$, $\boldsymbol{\mathcal{W}}_j$ and $\boldsymbol{\mathcal{C}}_{ij} = \mathsf{diag}\left(\cos\alpha_{ij,1}, \ldots, \cos\alpha_{ij,MN}\right)$, $0 \leq \alpha_{ij,1} \leq \cdots \leq \alpha_{ij,MN} \leq \frac{\pi}{2}$ such that $\boldsymbol{\mathcal{U}}_j^H \boldsymbol{\mathcal{U}}_i = \boldsymbol{\mathcal{W}}_j^H \boldsymbol{\mathcal{C}}_{ij} \boldsymbol{\mathcal{W}}_i$. The angles $\alpha_{ij,k}$ are the *principal angles* between the subspaces spanned by $\boldsymbol{\mathcal{W}}_i$ and $\boldsymbol{\mathcal{W}}_j$ (equivalently, by $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{X}}_j$). Plugging this in (8) and invoking Ostrowski's theorem pp. 224, 225 in [24] yields the lower and upper bounds

$$\lambda_{\min}\left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right) \left(1 - \cos^2\alpha_{ij,1}\right) \leq \lambda_{\min}\left(\boldsymbol{L}_{ij}(\mathcal{C})\right) \leq \lambda_{\min}\left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right) \left(1 - \cos^2\alpha_{ij,MN}\right). \tag{9}$$

The lower-bound in (9) show that, in order to maximize f, one should simultaneously increase $\lambda_{\min}\left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right)$ and the angle $\alpha_{ij,1}$. Unfortunately, this does not offer much insight into the form of the optimal codebook for the case of arbitrary noise covariance matrix $\boldsymbol{\Upsilon}$. One of the reasons originates from the fact that pairwise error probabilities are not symmetric for this general case.

In the remainder of this subsection, we use the upper-bound in (9) to show that unitary codewords ($\boldsymbol{X}_k^H \boldsymbol{X}_k = \frac{1}{M}\boldsymbol{I}_M$) are optimal for the specific case of spatio-temporal white Gaussian observation noise $\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$. Let $\boldsymbol{X}_k = \boldsymbol{U}_k \boldsymbol{\Sigma}_k \boldsymbol{V}_k^H$ denote a thin singular value decomposition (SVD) of $\boldsymbol{X}_k$, where $\boldsymbol{U}_k$, $\boldsymbol{V}_k$ are unitary matrices and $\boldsymbol{\Sigma}_k$ contains the singular values. Using this SVD representation for the codewords and $\boldsymbol{\Upsilon} = \boldsymbol{I}_{NT}$ in (4), it is not difficult to obtain

$$\lambda_{\min}(\boldsymbol{L}_{ij}(\mathcal{C})) = \lambda_{\min}\left(\boldsymbol{\Sigma}_i \left(\boldsymbol{I}_T - \boldsymbol{U}_i^H \boldsymbol{U}_j \boldsymbol{U}_j^H \boldsymbol{U}_i\right) \boldsymbol{\Sigma}_i\right). \tag{10}$$

Since the matrices $\boldsymbol{V}_k$ do not appear in (10), any optimal constellation can be described in the form $\boldsymbol{X}_k = \boldsymbol{U}_k \boldsymbol{\Sigma}_k$. Further, similar to (9), we have

$$\lambda_{\min}\left(\boldsymbol{\Sigma}_i^2\right) \left(1 - \cos^2\theta_{ij,1}\right) \leq \lambda_{\min}(\boldsymbol{L}_{ij}(\mathcal{C})) \leq \lambda_{\min}\left(\boldsymbol{\Sigma}_i^2\right) \left(1 - \cos^2\theta_{ij,M}\right) \tag{11}$$

where $0 \leq \theta_{ij,1} \leq \ldots \leq \theta_{ij,M} \leq \frac{\pi}{2}$ denote the principal angles between the subspaces spanned by $\boldsymbol{U}_i$ and $\boldsymbol{U}_j$ (equivalently, by $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$). Since $\lambda_{\min}\left(\boldsymbol{\Sigma}_i^2\right) \leq \frac{1}{M}\mathsf{tr}(\boldsymbol{X}_i^H \boldsymbol{X}_i) = \frac{1}{M}$, the last inequality in (11) yields the upper bound on the codebook merit function:

$$f\left(\mathcal{C}\right) \leq \lambda_{\min}(\boldsymbol{L}_{ij}(\mathcal{C})) \leq \frac{1}{M}. \tag{12}$$

The upper-bound is attained only if $\lambda_{\min}(\boldsymbol{\Sigma}_i^2) = \frac{1}{M}$, i.e., each $\boldsymbol{X}_k = \frac{1}{\sqrt{M}}\boldsymbol{U}_k$ is an unitary codeword. Furthermore, for unitary codewords, the upper-bound is attained only if $\theta_{ij,1} = \cdots = \theta_{ij,M} = \frac{\pi}{2}$, i.e., the codewords are mutually orthogonal $\boldsymbol{X}_i^H\boldsymbol{X}_j = \boldsymbol{0}$ (this can be established using (10)). In sum, we showed that for the special case of spatio-temporally white noise and $K = 2$, the unitary constellations are the optimal ones with respect to our codebook design criterion f. We recall that the unitary structure was also shown to be optimal in [5], [6], [8] from both the capacity and asymptotic UB on the probability of error minimization viewpoints.

We now comment on the situation $M \leq T < 2M$. In this case, we would have

$$\boldsymbol{L}_{ij}(\mathcal{C}) = \left(\boldsymbol{\mathcal{X}}_i^H\boldsymbol{\mathcal{X}}_i\right)^{\frac{1}{2}}\boldsymbol{\mathcal{W}}_i\begin{bmatrix}\boldsymbol{I}_{(TN-MN)} - \boldsymbol{\mathcal{C}}_{ij}^2 & \boldsymbol{0}_{(TN-MN)\times(2MN-TN)}\\ \boldsymbol{0}_{(2MN-TN)\times(TN-MN)} & \boldsymbol{0}_{2MN-TN}\end{bmatrix}\boldsymbol{\mathcal{W}}_i^H\left(\boldsymbol{\mathcal{X}}_i^H\boldsymbol{\mathcal{X}}_i\right)^{\frac{1}{2}}, \tag{13}$$

for some unitary matrix $\boldsymbol{\mathcal{W}}_i$ and diagonal $\boldsymbol{\mathcal{C}}_{ij}$. Given that $T < 2M$, we see that the lower right block of zeros in the middle matrix in the right-hand side of (13) is non-void. Thus, $\lambda_{\min}(\boldsymbol{L}_{ij}) = 0$ and plugging this in (5) yields the upper-bound $P_{\boldsymbol{X}_i\to\boldsymbol{X}_j} \leq \mathcal{Q}(0) = 0.5$ which holds irrespective of the choice of codewords. Thus, we cannot extract a guideline for codebook construction in this case. This motivates the following assumption. **A4 (Length of channel coherence)** In this work, the length of the coherence interval T is at least as twice as large as the number of transmit antennas M: $T \geq 2M$.

The preceding assumption is not surprising since, for the special case $\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$, Rayleigh fading and in high SNR scenario, it is known that the length of the coherence interval has to be necessarily at least as twice as large as the number of transmit antennas ($2M \leq T$) to achieve full order of diversity MN [8], but also, from the capacity viewpoint it is found that there is no point in using more than $\frac{T}{2}$ transmit antenna when one wants to maximize the number of degrees of freedom [9].

B. Codebook design as a Grassmannian packing

It is instructive to compare our codebook construction criterion with the one proposed in [6], [7] defined as

$$\mathcal{C}^* = \underset{\mathcal{C}\in\mathcal{N}}{\arg\min}\ \max_{1\leq i\neq j\leq K}\ \mathsf{tr}\left(\boldsymbol{X}_i^H\boldsymbol{X}_j\boldsymbol{X}_j^H\boldsymbol{X}_i\right) \tag{14}$$

where the constraint space is the set of unitary codebooks $\mathcal{N} = \left\{(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K) : \boldsymbol{X}_k^H\boldsymbol{X}_k = \frac{1}{M}\boldsymbol{I}_M\right\}$. It is readily seen that (14) is equivalent to

$$\mathcal{C}^* = \underset{\mathcal{C}\in\mathcal{N}}{\arg\min}\ \max_{1\leq i\neq j\leq K}\ \cos^2\theta_{ij,1} + \cdots + \cos^2\theta_{ij,M} \tag{15}$$

where $0 \leq \alpha_{ij,1} \leq \cdots \leq \alpha_{ij,M} \leq \pi/2$ denote the principal angles between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$. In order to compare our approach with the one proposed in [6], [7], we must temporarily adopt the signal model assumptions in [6], [7], i.e., we consider white noise ($\boldsymbol{\Upsilon} = \boldsymbol{I}_{NT}$) and also unitary codebooks. In this setup, our codebook construction criterion in (7), simplifies to

$$\mathcal{C}^* = \underset{\mathcal{C}\in\mathcal{N}}{\arg\min}\ \max_{1\leq i\neq j\leq K}\ \cos^2\theta_{ij,1}. \tag{16}$$

It is clear that both criterions in (15) and (16) aim at building codebooks by reducing the pairwise "spatial crosstalk" between distinct codewords. The distinction lies in how this crosstalk is measured: the strategy in (15) looks at the average of the principal angles and corresponds to the Grassmannian chordal distance [7], whereas our criterion in (16) considers the worst-case and leads to the Grassmannian spectral distance [10]. We recall that, as defined in [10], the squared spectral distance of two linear subspaces of the same dimension, say $\mathcal{L}_i, \mathcal{L}_j \subset \mathbb{C}^n$, is given by $\sin^2 \theta_{ij,1}$ where $\theta_{ij,1}$ is the minimal principal angle between $\mathcal{L}_i$ and $\mathcal{L}_j$. It can be computed as follows: if the matrices $\boldsymbol{U}_i, \boldsymbol{U}_j$ contain in their columns an orthonormal basis for $\mathcal{L}_i, \mathcal{L}_j$, respectively, then $\sin^2 \theta_{ij,1} = 1 - \sigma_{ij}^2$ where σ_{ij} is the maximal singular value of $\boldsymbol{U}_i^H \boldsymbol{U}_j$. Given this definition, it follows that $\sin^2 \theta_{ij,1}$ corresponds to the squared spectral distance between the codewords $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ (more precisely, between their respective range spaces). Please refer to [10] and [22] for more details on packing problems in Grassmannian space. The reader is referred to [31] for a more in depth discussion on the geometry of complex Grassmann manifolds regarding distance, cut and conjugate locus, etc.

We note that, for this particular scenario, the criterion presented in (14) is easier to deal with mathematically. Also, from (16) we see that, for $M = 1$, the problem of finding good codes coincides with the very well known packing problem in the complex projective space [10].

IV. Codebook Construction

In this section, we address the design of a codebook of size K as stated in (7). We propose a two-phase methodology. Phase 1 solves a sequence of K convex semi-definite programs (SDP) to obtain a rough estimate of the optimal codebook. Phase 2 refines it through a geodesic descent optimization algorithm (GDA) which exploits the Riemannian geometry of the constraints. This approach is summarized in table I. Details on both phases are given below.

Input:	`System parameters` $M, N, T, K, \boldsymbol{\Upsilon}$
Phase 1)	`Choose the first codeword (randomly generated, etc)`
•	`Set` $k = 2$
loop:	`Solve a SDP to obtain the` k`-th codeword`
•	`Set` $k = k + 1$
•	`If` $k \leq K$ `goto` **loop**
Phase 2)	`Run GDA to obtain the final codebook`
Output:	`The codebook matrix` $\boldsymbol{\mathcal{C}}^* = \begin{bmatrix} \mathsf{vec}(\boldsymbol{X}_1^*) & \ldots & \mathsf{vec}(\boldsymbol{X}_K^*) \end{bmatrix}$

TABLE I

Codebook design algorithm

Phase 1: SDP relaxations. This phase constructs a sub-optimal codebook $\widehat{\mathcal{C}} = \{\widehat{\boldsymbol{X}}_1, ..., \widehat{\boldsymbol{X}}_K\}$. The codebook is constructed incrementally. Addition of a new codeword involves solving a SDP. The first codeword $\widehat{\boldsymbol{X}}_1$ can be obtained by several distinct strategies, e.g, randomly generated, choosing eigenvectors associated to the smallest eigenvalues of the noise covariance matrix, etc. Let $\widehat{\mathcal{C}}_{k-1} = \{\widehat{\boldsymbol{X}}_1, ..., \widehat{\boldsymbol{X}}_{k-1}\}$ be the current codebook. To find the new codeword $\widehat{\boldsymbol{X}}_k$, we face (7) with the first $k-1$ codewords fixed as in $\widehat{\mathcal{C}}_{k-1}$

$$\begin{aligned} \widehat{\boldsymbol{X}}_k &= \underset{\mathsf{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k)=1}{\arg\max} \quad f(\widehat{\boldsymbol{X}}_1, \ldots, \widehat{\boldsymbol{X}}_{k-1}, \boldsymbol{X}_k) \\ &= \underset{\mathsf{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k)=1}{\arg\max} \quad \min_{1 \leq m \leq k-1} \left\{ \lambda_{\min}(\boldsymbol{L}_{mk}),\ \lambda_{\min}(\boldsymbol{L}_{km}) \right\}. \end{aligned} \tag{17}$$

In Appendix II, we show that (17) can be reformulated as

$$(\boldsymbol{\mathfrak{X}}_k^*, \widehat{\boldsymbol{X}_k}, t^*) = \underset{\substack{\text{LMI}_{A_m}(\boldsymbol{\mathfrak{X}}_k, \boldsymbol{X}_k, t) \succeq \mathbf{0},\ m=1,...,k-1 \\ \text{LMI}_{B_m}(\boldsymbol{\mathfrak{X}}_k, \boldsymbol{X}_k, t) \succeq \mathbf{0},\ m=1,...,k-1 \\ \mathsf{tr}(\boldsymbol{\mathfrak{X}}_k)=1, \quad \boldsymbol{\mathfrak{X}}_k = \mathsf{vec}(\boldsymbol{X}_k)\mathsf{vec}^H(\boldsymbol{X}_k)}}{\arg\max} \quad t \tag{18}$$

where the abbreviations $\text{LMI}_{A_m}(\boldsymbol{\mathfrak{X}}_k, \boldsymbol{X}_k, t)$ and $\text{LMI}_{B_m}(\boldsymbol{\mathfrak{X}}_k, \boldsymbol{X}_k, t)$ denote linear matrix inequalities (LMI) in the variables $\boldsymbol{\mathfrak{X}}_k$, $\boldsymbol{X}_k$, and t of type A and B, respectively, for $m = 1, ..., k-1$. The proof and the meaning of the LMI's of type A and B are given in Appendix II. The nonconvex rank constraint in (18) (note that $\boldsymbol{\mathfrak{X}}_k = \mathsf{vec}(\boldsymbol{X}_k)\mathsf{vec}^H(\boldsymbol{X}_k)$ and $\mathsf{tr}(\boldsymbol{\mathfrak{X}}_k) = 1$ imply that $\mathsf{rank}\,(\boldsymbol{\mathfrak{X}}_k) = 1$) makes (18) hard to solve. However, relaxing this restriction as $\boldsymbol{\mathfrak{X}}_k \succeq \mathsf{vec}(\boldsymbol{X}_k)\mathsf{vec}^H(\boldsymbol{X}_k)$ yields the SDP

$$(\boldsymbol{\mathfrak{X}}_k^*, \widehat{\boldsymbol{X}_k}, t^*) = \underset{\substack{\text{LMI}_{A_m}(\boldsymbol{\mathfrak{X}}_k, \boldsymbol{X}_k, t) \succeq \mathbf{0},\ m=1,...,k-1 \\ \text{LMI}_{B_m}(\boldsymbol{\mathfrak{X}}_k, \boldsymbol{X}_k, t) \succeq \mathbf{0},\ m=1,...,k-1 \\ \mathsf{tr}(\boldsymbol{\mathfrak{X}}_k)=1, \quad \begin{bmatrix} \boldsymbol{\mathfrak{X}}_k & \mathsf{vec}(\boldsymbol{X}_k) \\ \mathsf{vec}^H(\boldsymbol{X}_k) & 1 \end{bmatrix} \succeq \mathbf{0}}}{\arg\max} \quad t. \tag{19}$$

The rank 1 relaxation is usually known as the Shor relaxation [25]. The optimization problem in (19) is convex in the variables $\boldsymbol{\mathfrak{X}}_k$, $\boldsymbol{X}_k$ and t. Note that $\boldsymbol{\mathfrak{X}}_k$ is of size $MT \times MT$, independent of the codebook's size K. The number of LMI constraints in (19) is of order $2(k-1)$. To solve numerically the optimization problem in (19) we used the *Self-Dual-Minimization* package SeDuMi 1.1 [26]. Once the problem defined in (19) is solved, we adopt a randomized technique similar to [27] to obtain the codeword $\widehat{\boldsymbol{X}}_k$ from the output variable $\boldsymbol{\mathfrak{X}}_k^*$. The technique consists in generating independent realizations of random vectors that follow a Gaussian distribution with zero mean and covariance matrix $\boldsymbol{\mathfrak{X}}_k$, i.e., $\boldsymbol{z}_l \overset{iid}{\sim} \mathcal{CN}(\mathbf{0}, \boldsymbol{\mathfrak{X}}_k)$, for $l = 1, 2, ..., L$, where L is a parameter to be chosen (in our simulations $L = 10000$). After forcing norm 1, $\boldsymbol{x}_l = \boldsymbol{z}_l / ||\boldsymbol{z}_l||$, we retain the best sample, i.e., $\widehat{\boldsymbol{X}}_k = \mathsf{ivec}(\boldsymbol{x}_{l^*})$ where $l^* = \arg\max\{f(\widehat{\boldsymbol{X}}_1, \widehat{\boldsymbol{X}}_2, ..., \widehat{\boldsymbol{X}}_{k-1}, \mathsf{ivec}(\boldsymbol{x}_l)) \,:\, l = 1, 2, \ldots, L\}$. Here, the operator $\mathsf{ivec}(\,\cdot\,)$ acts as an inverse for vec(·): it reshapes the TM-dimensional vector into a $T \times M$ matrix. We are clearly dealing with a suboptimal solution for a codebook.

Phase 2: Geodesic Descent Algorithm. Problem (7) involves searching codebooks over the constraint set $\mathcal{M} = \{(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K) : \mathsf{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k) = 1\}$. Note that $\mathcal{M}$, as the Cartesian product of K spheres, is a smooth manifold. The geodesic descent algorithm (GDA), presented in table II, is a locally-convergent iterative algorithm which improves codebooks by travelling along geodesics in $\mathcal{M}$ (hence, securing the power-constraints imposed on the codewords). The GDA is initialized with the codebook $\widehat{\boldsymbol{\mathcal{C}}}$ produced by phase 1. We now describe the main ideas behind the outline of the GDA presented in table II. From label **input** to the label **loop**, we simply re-arrange the initial codebook in the vector $\boldsymbol{x}$ and compute f_c, the merit function evaluated at this point. In table II, the notation $\Re \boldsymbol{Z}$ and $\Im \boldsymbol{Z}$ denote the real and imaginary part of the complex matrix $\boldsymbol{Z}$. Throughout the algorithm in table II, $\boldsymbol{x}$ always denote the current iterate for the codebook and f_c the merit function evaluated at $\boldsymbol{x}$.

The main loop starts at the label **loop**. First, we identify the set "active" constraint pairs (i, j), i.e., those satisfying $f_c \leq f_{ij}(\boldsymbol{x}) \leq f_c + \epsilon$, where ϵ denotes a small positive constant (in all simulations herein presented we have chosen $\epsilon = 10^{-5}$). We let $\mathcal{A} = \{(i_l, j_l) : l = 1, \ldots, L\}$ denote the set of active pairs. Then, we check if there is an ascent direction $\boldsymbol{d}^*$ simultaneously for all functions f_{ij} with $(i, j) \in \mathcal{A}$. If it exists $\boldsymbol{d}^*$ such that $\nabla^T f_{i_l j_l}(\boldsymbol{x})\, \boldsymbol{d}^* > 0$, for $1 \leq l \leq L$, we can try to improve our cost function locally. The gradient $\nabla f_{ij}(\boldsymbol{x})$ is computed in Appendix III. This ascent direction $\boldsymbol{d}^*$ is searched within $T_{\boldsymbol{x}}\mathcal{M}$, the tangent space to the manifold $\mathcal{M}$ at the current codebook $\boldsymbol{x}$, and consists in solving a linear program. It is the constraint $\boldsymbol{X d} = \boldsymbol{0}_{K \times 1}$ (equivalently, $\boldsymbol{x}_k^T \boldsymbol{d}_k = 0$ for $k = 1, \ldots, K$) which ensures that $\boldsymbol{d} \in T_{\boldsymbol{x}}\mathcal{M}$. The box constraint $-\boldsymbol{1}_{2KTM \times 1} \leq \boldsymbol{d} \leq +\boldsymbol{1}_{2KTM \times 1}$ is added to bound the solution of the linear program. If there is no such ascent direction, GDA stops and returns the current point. Otherwise, we perform an Armijo search along the geodesic which emanates from $\boldsymbol{x}$ in the direction $\boldsymbol{d}^*$ in order to improve our current codebook. The expression for the geodesic is given in the **new trial** step. A geodesic is nothing but the generalization of a straight line in Euclidean space to a curved surface [28]. In loose terms, GDA resembles a sub-gradient method and consequently, the algorithm usually converges slowly near local minimizers. Note however that this is not a serious drawback since codebooks can be generated off-line.

The parameter ϵ controls the complexity of the linear program. A too small ϵ implies slow convergence of the algorithm, whereas a big ϵ increases the complexity of the linear program (by increasing L, the number of active pairs). For a codebook of size K = 256, and T = 8, $M = 2$, the gradient matrix $\boldsymbol{G}$ can be of size 10000×8000 (remark that $L_{\max} = K(K-1) = 65280$). Although the matrix $\boldsymbol{G}$ is a sparse matrix, it is preferably to impose it to be of moderate size too. The choice of ϵ controls that.

As a final remark, we note that the utility of the Phase 1 in table I for large K is an open issue. We have found it quite useful for small and moderate sized codebooks. For example, for the real case, $M = 1$ and $T = 2$, Phase 1 already provides us the optimal codebook for $K = 2^p$ where $p = 1, 2, 3, \ldots$. In all simulations herein presented the procedure presented in table I has been implemented.

Input:	Initial codebook $\mathcal{C} = \left[\mathrm{vec}(\boldsymbol{X}_1) \quad \dots \quad \mathrm{vec}(\boldsymbol{X}_K)\right]$
•	Initialize $\epsilon = 10^{-5}$
•	Construct $\boldsymbol{x} = [\,\boldsymbol{x}_1^T \ \cdots \ \boldsymbol{x}_K^T\,]^T$ where $\boldsymbol{x}_k^T = [\,\Re\mathrm{vec}(\boldsymbol{X}_k)^T \ \Im\mathrm{vec}(\boldsymbol{X}_k)^T\,]$
•	Evaluate current codebook $f_c = f(\boldsymbol{x})$
loop:	Identify the active (i,j)'s: those satisfying $f_c \leq f_{ij}(\boldsymbol{x}) \leq f_c + \epsilon$
•	Let $\mathcal{A} = \{(i_l, j_l)\,:\, l = 1, \dots, L\}$ denote the set of active pairs
•	Construct the $L \times 2KTM$ gradient matrix $$\boldsymbol{G} = \begin{bmatrix} \nabla^T f_{i_1 j_1}(\boldsymbol{x}) \\ \vdots \\ \nabla^T f_{i_L j_L}(\boldsymbol{x}) \end{bmatrix}$$
•	Construct the $K \times 2KTM$ matrix $\boldsymbol{X} = \mathrm{diag}(\boldsymbol{x}_1^T, \cdots, \boldsymbol{x}_K^T)$
•	Solve the linear program $$(\boldsymbol{d}^*, s^*) = \begin{array}{c} \arg\max \\ \boldsymbol{G}\boldsymbol{d} \geq s\mathbf{1}_L \\ \boldsymbol{X}\boldsymbol{d} = \mathbf{0}_{K\times 1} \\ -\mathbf{1}_{2KTM} \leq \boldsymbol{d} \leq +\mathbf{1}_{2KTM} \end{array} \quad s$$
•	If $s^* \leq 0$ goto **Output**
•	Initialize $\beta = 0.9$, $c = 0$, $c_{max} = 400$ and $t = 1$
•	Partition $\boldsymbol{d}^* = (\boldsymbol{d}_1^{*T}, \dots, \boldsymbol{d}_K^{*T})^T$ where $\boldsymbol{d}_k \,:\, 2TM \times 1$
new trial:	Construct the point $$\boldsymbol{x}(t) = \begin{bmatrix} \boldsymbol{x}_1(t) \\ \vdots \\ \boldsymbol{x}_K(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 \cos(\|\boldsymbol{d}_1^*\| t) + \frac{\boldsymbol{d}_1^*}{\|\boldsymbol{d}_1^*\|} \sin(\|\boldsymbol{d}_1^*\| t) \\ \vdots \\ \boldsymbol{x}_K \cos(\|\boldsymbol{d}_K^*\| t) + \frac{\boldsymbol{d}_K^*}{\|\boldsymbol{d}_K^*\|} \sin(\|\boldsymbol{d}_K^*\| t) \end{bmatrix}$$
•	Evaluate $f_t = f(\mathrm{ivec}(\boldsymbol{x}_1(t)), \dots, \mathrm{ivec}(\boldsymbol{x}_K(t)))$
•	If $f_t > f_c$ then: $f_c = f_t$, $\boldsymbol{x}_k = \boldsymbol{x}_k(t)$ for $k = 1, \dots, K$, and goto **loop**
•	Increment $c = c + 1$, update $t = \beta^c$
•	If $c \leq c_{max}$ goto **new trial**
Output:	Return codebook $\mathcal{C}^* = \left[\mathrm{vec}(\boldsymbol{X}_1^*) \quad \dots \quad \mathrm{vec}(\boldsymbol{X}_K^*)\right]$ where $\mathrm{vec}[\,\Re\boldsymbol{X}_k^* \ \Im\boldsymbol{X}_k^*\,] = \boldsymbol{x}_k$

TABLE II

GDA ALGORITHM

V. RESULTS

We have constructed codes for three special categories of noise covariance matrices $\boldsymbol{\Upsilon}$. In all simulations we assumed a Rayleigh fading model for the channel matrix, i.e., $h_{ij} \overset{iid}{\sim} \mathcal{CN}\left(0, \sigma^2\right)$.

a) First category: spatio-temporal white observation noise: In the first category the spatio-temporal white observation noise case is considered, i.e., $\boldsymbol{\Upsilon} = \mathsf{E}[\mathsf{vec}(\boldsymbol{E})\,\mathsf{vec}(\boldsymbol{E})^H]= \boldsymbol{I}_{NT}$. We compared our codes with the best known found in [7]. We considered scenario with coherence interval $T = 8$, $M = 3$ transmit antennas, $N = 1$ receive antennas and a codebook with $K = 256$ codewords. Let dist $= \frac{1}{K}\sum_{k=1}^{K}\sqrt{\mathsf{tr}\left(\left(\boldsymbol{X}_k^H\boldsymbol{X}_k - \frac{1}{M}\boldsymbol{I}_M\right)^2\right)}$ denote the average distance of our codebook from the constellation of unitary matrices. For $M = 3$, $T = 8$ and $K = 256$, the average distance was dist $= 1.3 \cdot 10^{-2}$. As expected (see section III), the algorithm converged to constellations of unitary matrices. In figure 1 we show the symbol error rate (SER) versus SNR $= \mathsf{E}[||\boldsymbol{X}_k\boldsymbol{H}^H||^2]/\mathsf{E}[||\boldsymbol{E}||^2] = N\sigma^2/\mathsf{tr}(\boldsymbol{\Upsilon})$. The solid-plus and dashed-circle curves represent performances of codes constructed by our method, and unitary codes respectively. As we can see, our codebook constructions replicate the performance of [7] for these particular cases, with just marginal improvements. Note that, for unitary constellations and white spatio-temporal observation noise, the GLRT and the Bayesian receiver in [7] coincide. This, in conjunction with the fact that our codebook is almost-unitary, explains the comparable performance of the two approaches. For $M = 1$, in figure 2 and table-III we compare our results with [10] for $T = 2, 3, \ldots, 6$. We manage to improve the best known results and in some cases actually provide optimal packings which attain the Rankin upper bound.

b) Second category: spatially white-temporally 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 an Hermitean Toeplitz matrix $\Sigma(\boldsymbol{\rho})$. To the best of our knowledge, we are not aware of any work that treats the problem of codebook constructions in the presence of spatially white-temporally colored observation noise. Hence, we compare our codes designed (adapted) to this specific scenario with unitary codes [7]. The goal here is to demonstrate the increase of performance obtained by matching the codebook construction to the noise statistics. In figures 3–4 the solid curves represent the performance of codes constructed by our method, while the dashed curves represent the performance of unitary codes. In either case, the plus sign indicates that the GLRT receiver is implemented. The square sign indicates that the Bayesian receiver is implemented. Figure 3 plots the result of the experiment for $T = 8$, $M = 2$, $N = 1$, $K = 67$ and $\boldsymbol{\rho} =$ [1; 0.85; 0.6; 0.35; 0.1; zeros(3,1)]. It can be seen that for SER $= 10^{-3}$, our codes demonstrate a gain of 3dB when compared with the unitary codes. Figure 4 plots the result of the experiment for $T = 8$, $M = 2$, $N = 1$, $K = 32$ and $\boldsymbol{\rho}$ =[1; 0.8; 0.5; 0.15; zeros(4,1)]. For SER $= 10^{-3}$, our codes demonstrate gain of 3dB when compared with the unitary codes.

c) Third category: $\boldsymbol{E} = \boldsymbol{s}\,\boldsymbol{\alpha}^T + \boldsymbol{E}_{temp}$: In the third category, 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 for the second category, we compare our codes adapted to this particular scenario with unitary codes. In figures 5–6 the solid curves represent performance of codes constructed by our method, while the dashed curves represent performance of unitary codes [7]. Figure 5 plots the result of the experiment for $T = 8$, $M = 2$, $N = 2$, $K = 32$, $\boldsymbol{s}$ =[1; 0.7; 0.4; 0.15; zeros(4,1)], $\boldsymbol{\rho} =$ [1; 0.8; 0.5; 0.15; zeros(4,1)] and $\boldsymbol{\alpha} =$ [-1.146 + 1.189i;1.191- 0.038i]. For SER $= 10^{-3}$, once again our codes demonstrate a gain of more than 2dB gain when compared with the unitary codes. Figure 6 plots the result of the experiment for $T = 8$, $M = 2$, $N = 2$, $K = 67$, $\boldsymbol{\rho}$ =[1; 0.7; 0.4; 0.15; zeros(4,1)], $\boldsymbol{s} =$ [1; 0.8; 0.5; 0.15; zeros(4,1)] and $\boldsymbol{\alpha} =$ [-0.4534 + 0.0072i; 0.4869 + 1.9728i]. For SER $= 10^{-3}$, our codes demonstrate a gain of more than 1.5dB gain when compared with the unitary codes.

VI. Conclusions

We addressed the problem of codebook construction for non-coherent communication in multiple-antenna wireless systems. In contrast with other related approaches, the Gaussian observation noise may have an arbitrary correlation structure. The non-coherent receiver operates according to the GLRT principle. A methodology for designing space-time codebooks for this non-coherent setup, taking the probability of error of the detector in the high SNR regime as the code design criterion, is proposed. We have presented a two-phase approach to solve the resulting high-dimensional, nonlinear and non-smooth optimization problem. The first phase solves a sequence of convex semi-definite programs to obtain a rough estimate of the optimal codebook. The second phase refines it through a geodesic descent optimization algorithm which exploits the Riemannian geometry of the constraint set. Computer simulations show that our codebooks are marginally better than state-of-art known solutions for the special case of spatio-temporal white Gaussian observation noise but significantly outperform them in the correlated noise environments. This shows the relevance of the codebook construction tool proposed herein.

Appendix I
Pairwise Error Probability for Fast Fading in the high SNR regime

In this appendix, we derive the expression for the asymptotic (high SNR regime) pairwise error probability for fast fading presented in (3). If $\boldsymbol{X}_i$ is transmitted, then the probability that the receiver decides in favor of $\boldsymbol{X}_j$ is

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} = P\left(\|\boldsymbol{z}_i\|^2 > \|\boldsymbol{z}_j\|^2 \right) \tag{20}$$

where

$$\boldsymbol{z}_k = \boldsymbol{\Upsilon}^{-1/2}\left(\boldsymbol{y} - (\boldsymbol{I}_N \otimes \boldsymbol{X}_k)\,\widehat{\boldsymbol{h}}_k \right) = \boldsymbol{\Upsilon}^{-1/2}\boldsymbol{y} - \boldsymbol{\mathcal{X}}_k \widehat{\boldsymbol{h}}_k, \tag{21}$$

where $\widehat{\boldsymbol{h}}_k$ and $\boldsymbol{\mathcal{X}}_k$ are defined in (2). Since $\boldsymbol{X}_i$ was transmitted, we have

$$\boldsymbol{y} = (\boldsymbol{I}_N \otimes \boldsymbol{X}_i)\,\boldsymbol{h} + \boldsymbol{e} \tag{22}$$

where $\boldsymbol{h} = \mathsf{vec}(\boldsymbol{H}^H)$ and $\boldsymbol{e} = \mathsf{vec}(\boldsymbol{E})$. Plugging (21) and (22) in (20) yields

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} = P\left(\boldsymbol{w}^H \left(\boldsymbol{\Pi}_i^{\perp} - \boldsymbol{\Pi}_j^{\perp} \right) \boldsymbol{w} - 2\Re\left(\boldsymbol{w}^H \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i \boldsymbol{h} \right) > \lambda \right) \tag{23}$$

where $\boldsymbol{w} = \boldsymbol{\Upsilon}^{-1/2}\boldsymbol{e}$ denotes zero-mean white Gaussian noise $\boldsymbol{w} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{NT})$, $\boldsymbol{\Pi}_k^{\perp} = \boldsymbol{I}_{TN} - \boldsymbol{\mathcal{X}}_k \left(\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k\right)^{-1} \boldsymbol{\mathcal{X}}_k^H$ is the orthogonal projector onto the orthogonal complement of the column space of $\boldsymbol{\mathcal{X}}_k$ and $\lambda = \boldsymbol{h}^H \boldsymbol{\mathcal{X}}_i^H \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i \boldsymbol{h}$. Also, $\Re z$ stands for the real part of the complex number $z \in \mathbb{C}$. The expression (23) cannot be easily simplified analytically. We shall analyze (2) in the high SNR regime where the linear term of $\boldsymbol{w}$ is dominant, i.e., the quadratic term of $\boldsymbol{w}$ is negligible (see pp. 56 in [18]). Therefore, $P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} \simeq P(x > \lambda)$ where we defined $x = -2\Re\left(\boldsymbol{w}^H \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i \boldsymbol{h}\right)$. It is easy to see that x is a zero-mean real-valued Gaussian variable with variance $\sigma^2 = 2\lambda$. Thus, $P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} \simeq P(x > \lambda) = \mathcal{Q}\left(\frac{\lambda}{\sigma}\right) = \mathcal{Q}\left(\frac{1}{\sqrt{2}}\sqrt{\boldsymbol{h}^H \boldsymbol{L}_{ij}(\mathcal{C}) \boldsymbol{h}}\right)$ where $\mathcal{Q}(s) = \int_s^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$ and $\boldsymbol{L}_{ij}(\mathcal{C}) = \boldsymbol{\mathcal{X}}_i^H \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i$. This completes the proof.

APPENDIX II
OPTIMIZATION PROBLEM

In this appendix, we prove that the equivalent formulation of the optimization problem (17) is given by (18). The optimization problem in (17) can be rewritten as

$$
\begin{aligned}
(\boldsymbol{X}_k^*, t^*) = \quad & \arg\max \quad t \\
& \lambda_{\min}(\boldsymbol{L}_{mk}) \geq t, m = 1, ..., k-1 \quad (A) \\
& \lambda_{\min}(\boldsymbol{L}_{km}) \geq t, m = 1, ..., k-1 \quad (B) \\
& \mathsf{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k) = 1,
\end{aligned}
\tag{24}
$$

where $\boldsymbol{L}_{ij} = \boldsymbol{\mathcal{X}}_i^H \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i$, $\boldsymbol{\Pi}_j^{\perp} = \boldsymbol{I}_{TN} - \boldsymbol{\mathcal{X}}_j(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j)^{-1} \boldsymbol{\mathcal{X}}_j^H$, $\boldsymbol{\mathcal{X}}_i = \boldsymbol{\Upsilon}^{-\frac{1}{2}} \widetilde{\boldsymbol{X}_i}$ and $\widetilde{\boldsymbol{X}_i} = \boldsymbol{I}_N \otimes \boldsymbol{X}_i$. Define $\boldsymbol{\mathfrak{X}}_k = \mathsf{vec}(\boldsymbol{X}_k)\mathsf{vec}^H(\boldsymbol{X}_k)$. We are going to show that both (A) and (B) can be written as linear matrix inequalities (LMI's) in $\boldsymbol{\mathfrak{X}}_k$, $\boldsymbol{X}_k$ and t.

(A) Note that $\lambda_{\min}(\boldsymbol{L}_{mk}) \geq t \Leftrightarrow \boldsymbol{L}_{mk} - t\boldsymbol{I}_{MN} \succeq \boldsymbol{0}$. Since $\boldsymbol{L}_{mk} - t\boldsymbol{I}_{MN} = \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_m - \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_k (\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k)^{-1} \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_m - t\boldsymbol{I}_{MN}$ is the Schur complement [29] of $\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k$ in $\begin{bmatrix} \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k & \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_m \\ \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_k & \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_m - t\boldsymbol{I}_{MN} \end{bmatrix}$ we have the following equivalence (we assumed that $\boldsymbol{X}_k$ is of full column rank):

$$
\lambda_{\min}(\boldsymbol{L}_{mk}) \geq t \Leftrightarrow \begin{bmatrix} \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k & \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_m \\ \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_k & \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_m - t\boldsymbol{I}_{MN} \end{bmatrix} \succeq \boldsymbol{0}. \tag{25}
$$

Let $[\boldsymbol{M}]_{ij}$ denotes the ij-th element of the matrix $\boldsymbol{M}$ and $\boldsymbol{e}_i$ represents the i-th column of the identity matrix $\boldsymbol{I}_{MN}$. Then,

$$
[\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k]_{ij} = \boldsymbol{e}_i^H \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k \boldsymbol{e}_j = \boldsymbol{e}_i^H \widetilde{\boldsymbol{X}_k}^H \boldsymbol{\Upsilon}^{-1} \widetilde{\boldsymbol{X}_k} \boldsymbol{e}_j = \mathsf{tr}\left(\boldsymbol{\Upsilon}^{-1} \widetilde{\boldsymbol{X}_k} \boldsymbol{e}_j \left(\widetilde{\boldsymbol{X}_k} \boldsymbol{e}_i\right)^H\right) \tag{26}
$$

As $\widetilde{\boldsymbol{X}_k} = \boldsymbol{I}_N \otimes \boldsymbol{X}_k$, there exists matrix $\boldsymbol{K}$ of size $TMN^2 \times TM$ such that $\mathsf{vec}(\widetilde{\boldsymbol{X}_k}) = \boldsymbol{K}\,\mathsf{vec}(\boldsymbol{X}_k)$, see [30]. Hence,

$$
\widetilde{\boldsymbol{X}_k} \boldsymbol{e}_j = \mathsf{vec}\left(\widetilde{\boldsymbol{X}_k} \boldsymbol{e}_j\right) = \left(\boldsymbol{e}_j^T \otimes \boldsymbol{I}_{TN}\right) \mathsf{vec}(\widetilde{\boldsymbol{X}_k}) = (\boldsymbol{e}_j^T \otimes \boldsymbol{I}_{TN}) \boldsymbol{K} \mathsf{vec}(\boldsymbol{X}_k). \tag{27}
$$

Substituting (27) in (26) we have

$$[\boldsymbol{\mathcal{X}}_k^H\boldsymbol{\mathcal{X}}_k]_{ij} \quad = \quad \mathsf{tr}\left(\boldsymbol{B}_{ij}\left(\boldsymbol{I}_{TN}\right)\,\boldsymbol{\mathfrak{X}}_k\right), \tag{28}$$

where we define $\boldsymbol{B}_{ij}(\boldsymbol{\Phi}) = \boldsymbol{K}^H(\boldsymbol{e}_i\otimes\boldsymbol{I}_{TN})\boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{\Phi}\boldsymbol{\Upsilon}^{-\frac{1}{2}}(\boldsymbol{e}_j^T\otimes\boldsymbol{I}_{TN})\boldsymbol{K}$.

Similarly,

$$[\boldsymbol{\mathcal{X}}_m^H\boldsymbol{\mathcal{X}}_k]_{ij} = \boldsymbol{e}_i^H\boldsymbol{\mathcal{X}}_m^H\boldsymbol{\mathcal{X}}_k\boldsymbol{e}_j = \boldsymbol{e}_i^H\boldsymbol{\mathcal{X}}_m^H\boldsymbol{\Upsilon}^{-\frac{1}{2}}\widetilde{\boldsymbol{X}_k}\boldsymbol{e}_j = \boldsymbol{e}_i^H\boldsymbol{\mathcal{X}}_m^H\boldsymbol{\Upsilon}^{-\frac{1}{2}}(\boldsymbol{e}_j^T\otimes\boldsymbol{I}_{TN})\boldsymbol{K}\,\mathsf{vec}(\boldsymbol{X}_k). \tag{29}$$

(B) By repeating the analysis for the case (A) we have:

$$[\boldsymbol{L}_{km}]_{ij} \quad = \quad \boldsymbol{e}_i^H\boldsymbol{\mathcal{X}}_k^H\boldsymbol{\Pi}_m^\perp\boldsymbol{\mathcal{X}}_k\boldsymbol{e}_j = \boldsymbol{e}_i^H\widetilde{\boldsymbol{X}_k}^H\boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{\Pi}_m^\perp\boldsymbol{\Upsilon}^{-\frac{1}{2}}\widetilde{\boldsymbol{X}_k}\boldsymbol{e}_j = \mathsf{tr}\left(\boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{\Pi}_m^\perp\boldsymbol{\Upsilon}^{-\frac{1}{2}}\widetilde{\boldsymbol{X}_k}\boldsymbol{e}_j\left(\widetilde{\boldsymbol{X}_k}\boldsymbol{e}_i\right)^H\right).$$

Using (27) we obtain $[\boldsymbol{L}_{km}]_{ij} = \mathsf{tr}\left(\boldsymbol{K}^H(\boldsymbol{e}_i\otimes\boldsymbol{I}_{TN})\boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{\Pi}_m^\perp\boldsymbol{\Upsilon}^{-\frac{1}{2}}(\boldsymbol{e}_j^T\otimes\boldsymbol{I}_{TN})\boldsymbol{K}\mathsf{vec}(\boldsymbol{X}_k)\mathsf{vec}^H(\boldsymbol{X}_k)\right)$. Hence,

$$[\boldsymbol{L}_{km}]_{ij} \quad = \quad \mathsf{tr}\left(\boldsymbol{B}_{ij}(\boldsymbol{\Pi}_m^\perp)\,\boldsymbol{\mathfrak{X}}_k\right). \tag{30}$$

Combining (25), (26), (28), (29) and (30) we conclude that both (A) and (B) can be written as LMI's with respect to the variables $\boldsymbol{\mathfrak{X}}_k$, $\boldsymbol{X}_k$ and t. Consequently, the optimization problems (17) and (18) are equivalent. This concludes the proof.

APPENDIX III
CALCULATING GRADIENTS

In this section, we calculate the gradient to be used in (20). Although the function f_{ij} assumes complex valued entries, that is $f_{ij}\ :\ \underbrace{\mathbb{C}^{T\times M}\times ...\times\mathbb{C}^{T\times M}}_{K}\rightarrow\mathbb{R}\qquad f_{ij}(\boldsymbol{X}_1,\boldsymbol{X}_2,...,\boldsymbol{X}_K) = \lambda_{\min}(\boldsymbol{L}_{ij})$ where $\boldsymbol{L}_{ij} = \boldsymbol{\mathcal{X}}_i^H\boldsymbol{\Pi}_j^\perp\boldsymbol{\mathcal{X}}_i$, $\boldsymbol{\Pi}_j^\perp = \boldsymbol{I}_{TN} - \boldsymbol{\mathcal{X}}_j(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j)^{-1}\boldsymbol{\mathcal{X}}_j^H$, $\boldsymbol{\mathcal{X}}_i = \boldsymbol{\Upsilon}^{-\frac{1}{2}}\widetilde{\boldsymbol{X}_i}$ and $\widetilde{\boldsymbol{X}_i} = \boldsymbol{I}_N\otimes\boldsymbol{X}_i$, we shall treat f_{ij} as a function of the real and imaginary components of $\boldsymbol{X}_1,\boldsymbol{X}_2,...,\boldsymbol{X}_K$, i.e.

$$f_{ij}\ :\ \underbrace{\mathbb{R}^{T\times M}\times ...\mathbb{R}^{T\times M}}_{2K}\rightarrow\mathbb{R}\quad f_{ij}(\Re\boldsymbol{X}_1,\Im\boldsymbol{X}_1,\Re\boldsymbol{X}_2,\Im\boldsymbol{X}_2...,\Re\boldsymbol{X}_K,\Im\boldsymbol{X}_K) = \lambda_{\min}(\boldsymbol{L}_{ij}).$$

Let $\lambda_{\min}$ be a simple eigenvalue of the Hermitean matrix $\boldsymbol{L}_{ij}(\mathcal{C}_0)$, and let $\boldsymbol{u}_0$ be an associated unit-norm eigenvector, so that $\boldsymbol{L}_{ij}(\mathcal{C}_0)\boldsymbol{u}_0 = \lambda_{\min}\left(\boldsymbol{L}_{ij}(\mathcal{C}_0)\right)\boldsymbol{u}_0$. The differential df_{ij}, computed at the point $\mathcal{C}_0$, is given by, pp. 162 in [30], $df_{ij} = \boldsymbol{u}_0^H d\boldsymbol{L}_{ij}\boldsymbol{u}_0$ where $d\boldsymbol{L}_{ij}$ denotes the differential of the map $\mathcal{C}\mapsto\boldsymbol{L}_{ij}(\mathcal{C})$ computed at the point $\mathcal{C}_0$.

Define $\boldsymbol{K}_j = \boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{\Pi}_j^\perp\boldsymbol{\Upsilon}^{-\frac{1}{2}}$, hence $d\boldsymbol{L}_{ij} = (d\widetilde{\boldsymbol{X}_i})^H\boldsymbol{K}_j\widetilde{\boldsymbol{X}_i} + \widetilde{\boldsymbol{X}_i}^H\boldsymbol{K}_j d\widetilde{\boldsymbol{X}_i} + \widetilde{\boldsymbol{X}_i}^H d\boldsymbol{K}_j\widetilde{\boldsymbol{X}_i}$ and

$$df_{ij} = \boldsymbol{u}_0^H d\boldsymbol{L}_{ij}\boldsymbol{u}_0 \quad = \quad \Re[\mathsf{tr}[(d\widetilde{\boldsymbol{X}_i})^H\underbrace{2\boldsymbol{K}_j\widetilde{\boldsymbol{X}_i}\boldsymbol{u}_0\boldsymbol{u}_0^H}_{\boldsymbol{C}_i}]] + \boldsymbol{u}_0^H\widetilde{\boldsymbol{X}_i}^H d\boldsymbol{K}_j\widetilde{\boldsymbol{X}_i}\boldsymbol{u}_0. \tag{31}$$

Continuing with analysis,

$$d\boldsymbol{K}_j = \underbrace{-\boldsymbol{\Upsilon}^{-\frac{1}{2}}\left(d\boldsymbol{\mathcal{X}}_j(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j)^{-1}\boldsymbol{\mathcal{X}}_j^H + \boldsymbol{\mathcal{X}}_j(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j)^{-1}(d\boldsymbol{\mathcal{X}}_j)^H\right)\boldsymbol{\Upsilon}^{-\frac{1}{2}}}_{\boldsymbol{K}_{j1}} -$$

$$\underbrace{-\boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{\mathcal{X}}_j d\left(\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}\right)\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\Upsilon}^{-\frac{1}{2}}}_{\boldsymbol{K}_{j2}}. \tag{32}$$

Using the equality $d\left(\boldsymbol{A}^{-1}\right) = -\boldsymbol{A}^{-1} d\boldsymbol{A}\boldsymbol{A}^{-1}$ [30], we can write

$$d\left(\left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-1}\right) = -\left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-1}\left(\left(d\boldsymbol{\mathcal{X}}_j\right)^H \boldsymbol{\mathcal{X}}_j + \boldsymbol{\mathcal{X}}_j^H d\boldsymbol{\mathcal{X}}_j\right)\left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-1}. \tag{33}$$

Substituting (33) and (32) in (31) we get $\boldsymbol{u}_0^H \widetilde{\boldsymbol{X}_i}^H d\boldsymbol{K}_j \widetilde{\boldsymbol{X}}_i \boldsymbol{u}_0 = \boldsymbol{u}_0^H \widetilde{\boldsymbol{X}_i}^H \boldsymbol{K}_{j1} \widetilde{\boldsymbol{X}}_i \boldsymbol{u}_0 + \boldsymbol{u}_0^H \widetilde{\boldsymbol{X}_i}^H \boldsymbol{K}_{j2} \widetilde{\boldsymbol{X}}_i \boldsymbol{u}_0$ with

$$\boldsymbol{u}_0^H \widetilde{\boldsymbol{X}_i}^H \boldsymbol{K}_{j1} \widetilde{\boldsymbol{X}}_i \boldsymbol{u}_0 \quad = \quad \Re[\mathsf{tr}[\left(d\widetilde{\boldsymbol{X}}_j\right)^H \underbrace{-2\boldsymbol{\Upsilon}^{-\frac{1}{2}} \boldsymbol{\mathcal{X}}_i \boldsymbol{u}_0 \boldsymbol{u}_0^H \boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_j \left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-1}}_{\boldsymbol{C}_{j1}}]]$$

and $\boldsymbol{u}_0^H \widetilde{\boldsymbol{X}_i}^H \boldsymbol{K}_{j2} \widetilde{\boldsymbol{X}}_i \boldsymbol{u}_0 = \Re[\mathsf{tr}[\left(d\widetilde{\boldsymbol{X}}_j\right)^H \underbrace{2\boldsymbol{\Upsilon}^{-\frac{1}{2}} \boldsymbol{\mathcal{X}}_j \left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-1} \boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_i \boldsymbol{u}_0 \boldsymbol{u}_0^H \boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_j \left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-1}}_{\boldsymbol{C}_{j2}}]]$. Define $\boldsymbol{C}_j = \boldsymbol{C}_{j1} + \boldsymbol{C}_{j2}$. Thus, $df_{ij} = \Re[\mathsf{tr}[(d\widetilde{\boldsymbol{X}}_i)^H \boldsymbol{C}_i]] + \Re[\mathsf{tr}[(d\widetilde{\boldsymbol{X}}_j)^H \boldsymbol{C}_j]]$. Note that $d\widetilde{\boldsymbol{X}}_i = \boldsymbol{I}_N \otimes d\boldsymbol{X}_i$, then $df_{ij} = \Re[\mathsf{tr}[(d\boldsymbol{X}_i)^H \overline{\boldsymbol{C}_i}]] + \Re[\mathsf{tr}[(d\boldsymbol{X}_j)^H \overline{\boldsymbol{C}_j}]]$ where $\overline{\boldsymbol{C}_i} = \sum_{k=1}^N \boldsymbol{C}_{ik}$ and $\boldsymbol{C}_{ik}$ is the k-th diagonal block of the matrix $\boldsymbol{C}_i$ of size $T \times M$. Remark that the matrix $\boldsymbol{C}_i$ is of size $TN \times MN$. Now, it is straightforward to identify the gradient. Hence, the gradient is given by

$$\nabla f_{ij}(\boldsymbol{x}) = \left[\left(\boldsymbol{0}_{(i-1)c\times 1}\right)^T \quad \boldsymbol{a}_i^T \quad \left(\boldsymbol{0}_{(j-i-1)c\times 1}\right)^T \quad \boldsymbol{a}_j^T \qquad \left(\boldsymbol{0}_{(K-j)c\times 1}\right)^T\right]_{2KTM\times 1}^T$$

for $1 \leq i \neq j \leq K$, $c = 2TM$, where $\boldsymbol{x} = \left[(\Re\,\mathsf{vec}(\boldsymbol{X}_1))^T \quad (\Im\,\mathsf{vec}(\boldsymbol{X}_1))^T \quad \ldots \quad (\Re\,\mathsf{vec}(\boldsymbol{X}_K))^T \quad (\Im\,\mathsf{vec}(\boldsymbol{X}_K))^T\right]^T$ and $\boldsymbol{a}_k = \left[\left(\Re\mathsf{vec}(\overline{\boldsymbol{C}_k})\right)^T \quad \left(\Im\mathsf{vec}(\overline{\boldsymbol{C}_k})\right)^T\right]^T$.

REFERENCES

[1] I. Telatar, "Capacity of multi-antenna Gaussian channels," *Technical Memorandum, AT&T Bell Laboratories*, 1995.

[2] S. A. Alamouti, "A simple transmit diversity technique for wireless communication," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451-1458, Oct. 1998.

[3] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41-59, 1996.

[4] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744-765, Mar. 1998.

[5] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139-157, Jan. 1999.

[6] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat-fading,"*IEEE Trans. Inform. Theory*, vol. 46, pp. 543-564, Mar. 2000.

[7] 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.

[8] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh fading channels with applications to a unified analysis of coherent and noncoherent spacetime receivers," *IEEE Trans. Inform. Theory*,vol. 47, pp. 2383-2399, Sept. 2001.

[9] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359-383, Feb. 2002.

[10] J. A. Tropp, "Topics in sparse approximation", *Ph.D. dissertation: Univ. Texas at Austin*, 2004.

[11] D. Agrawal, T. J. Richardson, and R. L. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2618-2626, Sept. 2001.

[12] N. J. A. Sloane, "Packing planes in four dimensions and other mysteries," in *Proc. Conf. Algebraic Combinatorics and Related Topics*, Yamagata, Japan, Nov. 1997.

[13] M. L. McCloud, M. Brehler, and M. K. Varanasi, "Signal design and convolutional coding for noncoherent spacetime communication on the block-Rayleigh-fading channel,"*IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1186-1194, May 2002.

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

[15] I. Kammoun and J. -C. Belfiore, "A new family of Grassmann spacetime codes for noncoherent MIMO systems," *IEEE Commun. Lett.*, vol. 7, no. 11, pp. 528-530, Nov. 2003.

[16] A. M. Cipriano, I. Kammoun and J. -C. Belfiore, "Simplified decoding for some non-coherent codes over the Grassmannian," in *Proc. of IEEE International Conference on Communications (ICC)*, 2005.

[17] A. M. Cipriano and I. Kammoun, "Linear Approximation of the Exponential Map with Application to Simplified Detection of Noncoherent Systems," in *Proc. of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2006.

[18] E. G. Larsson, P. Stoica, *Space-Time Block Codes for Wireless Communications.*, Cambridge University Press, 2003.

[19] M. Beko, J. Xavier and V. Barroso, "Codebook design for non-coherent communication in multiple-antenna systems," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2006.

[20] M. Beko, J. Xavier and V. Barroso, "Codebook design for the non-coherent GLRT receiver and low SNR MIMO block fading channel," in *Proc. of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2006.

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

[22] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc . . .: packing in Grassmannian spaces," *Experiment. Math.*, vol. 5, pp. 139-159, Sept. 1996.

[23] Rajendra Bhatia, *Matrix Analysis.* Springer-Verlag, 1997.

[24] Roger A. Horn, Charles R. Johnson *Matrix Analysis.* Cambridge University Press, 1985.

[25] N. Shor, " Dual quadratic estimates in polynomial and Boolean programming,"*Annals of Operation Research*, vol. 25, pp. 163-168, 1990.

[26] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones (Updated for Version 1.05)," http://sedumi.mcmaster.ca

[27] M. X. Goemans, "Semidefinite programming in combinatorial optimization," *Mathematical Programming*, Vol. 79, pp. 143-161, 1997.

[28] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303-353, 1998.

[29] S. Boyd, L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

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

[31] S. Berceanu, "On the geometry of complex Grassmann manifold, its noncompact dual and coherent states," in *Bull. Belg. Math. Soc.*, 4, pp. 205-243, 1997.

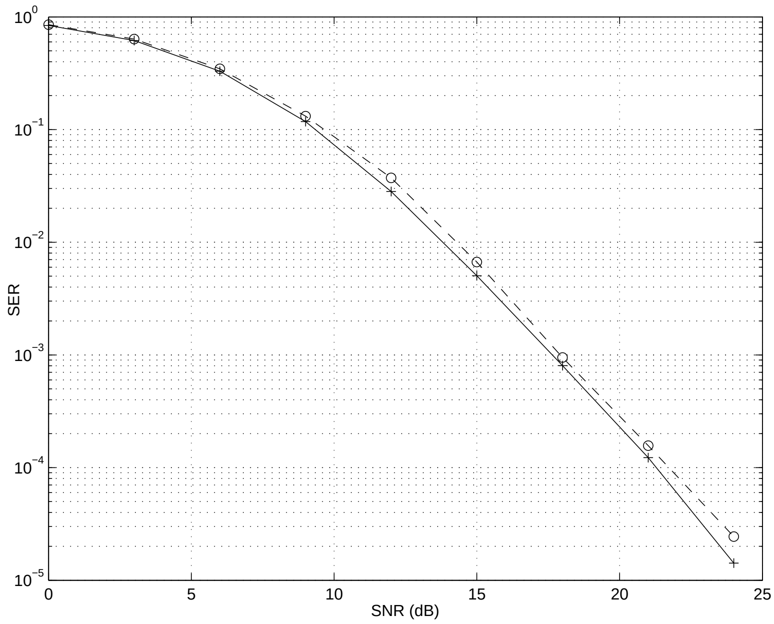

Fig. 1. Category 1 - spatio-temporally white observation noise: $T = 8$, $M = 3$, $N = 1$, $K = 256$, $\boldsymbol{\Upsilon} = \boldsymbol{I}_{NT}$. Plus-solid curve:our codes; circle-dashed curve:unitary codes.

		PACKING RADII (DEGREES)	
T	K	MB	Rankin
6	7	80.41	80.41
6	8	77.06	77.40
6	9	75.52	75.52
6	10	74.20	74.21
6	11	73.22	73.22
6	12	72.45	72.45
6	13	71.82	71.83
6	14	71.31	71.32
6	15	70.87	70.89
6	16	70.53	70.53
6	17	70.10	70.21
6	18	69.73	69.94
6	19	69.40	69.70

TABLE III

PACKING IN COMPLEX PROJECTIVE SPACE: WE COMPARE OUR BEST CONFIGURATIONS (MB) OF K POINTS IN $\mathbb{P}^{T-1}(\mathbb{C})$ AGAINST RANKIN BOUND. THE PACKING RADIUS OF AN ENSEMBLE IS MEASURED AS THE ACUTE ANGLE BETWEEN THE CLOSEST PAIR OF LINES.

		PACKING RADII (DEGREES)		
T	K	MB	JAT	Rankin
2	3	60	60	60
2	4	54.74	54.74	54.74
2	5	45.00	45.00	52.24
2	6	45.00	45.00	50.77
2	7	38.93	38.93	49.80
2	8	37.43	37.41	49.11
2	9	35.26	–	48.59
2	10	33.07	–	48.19
2	11	31.72	–	47.87
2	12	31.72	–	47.61
2	13	28.24	–	47.39
2	14	27.83	–	47.21
2	15	26.67	–	47.05
2	16	25.97	–	46.91
3	4	70.53	70.53	70.53
3	5	64.26	64.00	65.91
3	6	63.43	63.43	63.43
3	7	61.87	61.87	61.87
3	8	60.00	60.00	60.79
3	9	60.00	60.00	60.00
3	10	54.74	54.73	59.39
3	11	54.74	54.73	58.91
3	12	54.74	54.73	58.52
3	13	51.38	51.32	58.19
3	14	50.36	50.13	57.92
3	15	49.80	49.53	57.69
3	16	49.60	49.53	57.49
3	17	49.13	49.10	57.31
3	18	48.12	48.07	57.16

		PACKING RADII (DEGREES)		
T	K	MB	JAT	Rankin
4	5	75.52	75.52	75.52
4	6	70.89	70.88	71.57
4	7	69.29	69.29	69.30
4	8	67.79	67.78	67.79
4	9	66.31	66.21	66.72
4	10	65.74	65.71	65.91
4	11	64.79	64.64	65.27
4	12	64.68	64.24	64.76
4	13	64.34	64.34	64.34
4	14	63.43	63.43	63.99
4	15	63.43	63.43	63.69
4	16	63.43	63.43	63.43
5	6	78.46	78.46	78.46
5	7	74.55	74.52	75.04
5	8	72.83	72.81	72.98
5	9	71.33	71.24	71.57
5	10	70.53	70.51	70.53
5	11	69.73	69.71	69.73
5	12	69.04	68.89	69.10
5	13	68.38	68.19	68.58
5	14	67.92	67.66	68.15
5	15	67.48	67.37	67.79
5	16	67.08	66.68	67.48
5	17	66.82	66.53	67.21
5	18	66.57	65.87	66.98
5	19	66.57	65.75	66.77

Fig. 2. PACKING IN COMPLEX PROJECTIVE SPACE: We compare our best configurations (MB) of K points in $\mathbb{P}^{T-1}(\mathbb{C})$ against the Tropp codes (JAT) and Rankin bound [10]. The packing radius of an ensemble is measured as the acute angle between the closest pair of lines. Minus sign symbol (-) means that no packing is available for specific pair (T, K).

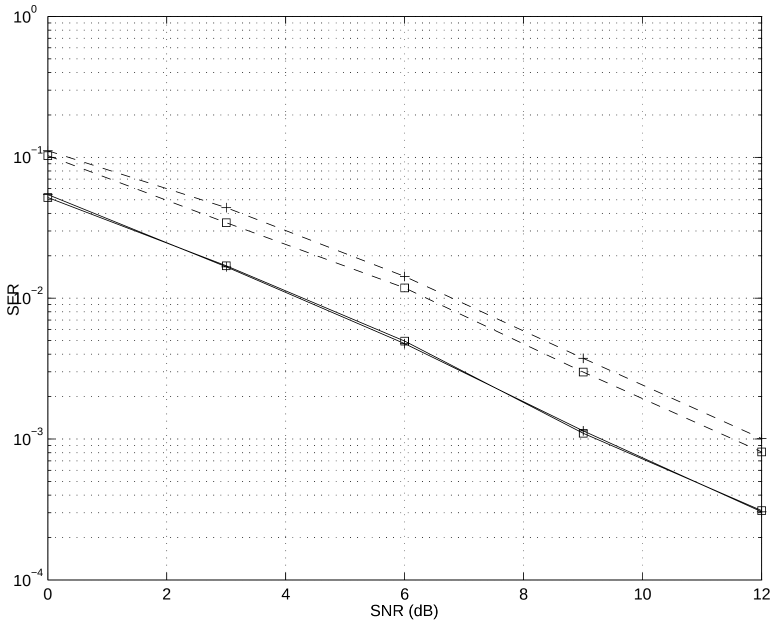

Fig. 3. Category 2 - spatially white - temporally colored: $T = 8$, $M = 2$, $N = 1$, $K = 67$, $\boldsymbol{\rho}$ =[1; 0.85; 0.6; 0.35; 0.1; zeros(3,1)]. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

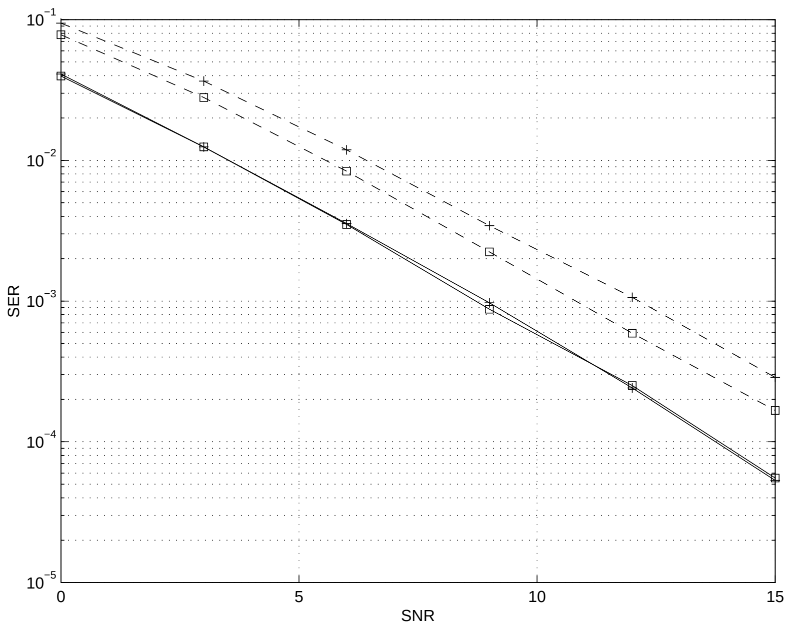

Fig. 4. Category 2 - spatially white - temporally colored: $T = 8$, $M = 2$, $N = 1$, $K = 32$, $\boldsymbol{\rho}$ =[1; 0.8; 0.5; 0.15; zeros(4,1)]. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

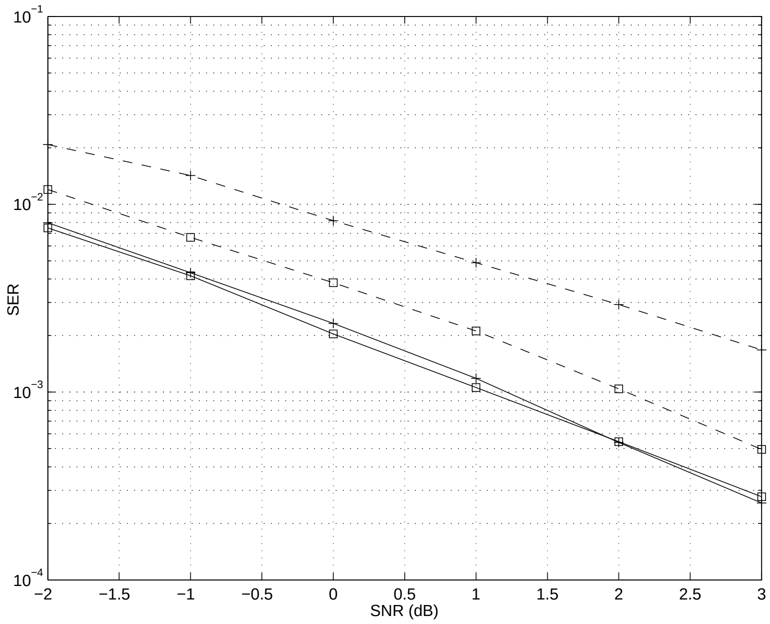

Fig. 5. Category 3: $T = 8$, $M = 2$, $N = 2$, $K = 32$. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

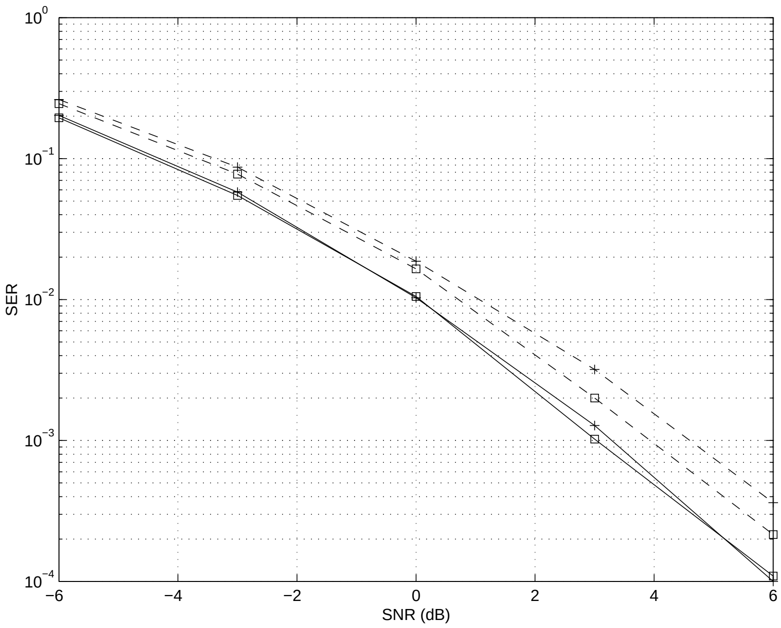

Fig. 6. Category 3: $T = 8$, $M = 2$, $N = 2$, $K = 67$. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.