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. This optimizes the probability of error of the GLRT receiver's detector in the high signal-to-noise ratio (SNR) regime, thus solving a high-dimensional nonlinear non-smooth optimization problem in a two-step approach: **(i)** firstly, a convex 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.

Index Terms

Multiple-input multiple output (MIMO) systems, non-coherent communications, space-time constellations, generalized likelihood ratio test (GLRT) receiver, semidefinite programming (SDP), geodesic descent algorithm (GDA).

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 comprise of dealing with the highly random channel conditions, which may vary rapidly, and also with 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 a constellation 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, scaled unitary codebooks are optimal from an asymptotic (high SNR) union bound (UB) on the error probability minimization perspective. 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], Sloan'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 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 recover 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 $\operatorname{tr}(\boldsymbol{A})$. 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 $\mathbf{0}_{M \times N}$ (also $\mathbf{0}_N = \mathbf{0}_{N \times N}$). 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 operator $\operatorname{vec}(\boldsymbol{A})$ stacks all columns of the matrix $\boldsymbol{A}$ on the top of each other, from left to right. The curled inequality symbol $\succeq$ represents matrix inequality between Hermitian matrices.

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 $\mathcal{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_K\}$ known to the receiver, where K is the size of the codebook. We impose the power constraint $\operatorname{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k) = 1$ for each codeword. Furthermore, we assume that $T \geq M$ and each codeword is of full rank, i.e., $\operatorname{rank}(\boldsymbol{X}) = M$;

A3. (Noise distribution) The observation noise at the receiver is zero mean and obeys circular complex Gaussian statistics, that is, $\text{vec}(\boldsymbol{E}) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon})$. The noise covariance matrix $\boldsymbol{\Upsilon} = \mathsf{E}[\text{vec}(\boldsymbol{E})\ \text{vec}(\boldsymbol{E})^H]$ is known at the transmitter and at the receiver.

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 colored case ($\boldsymbol{\Upsilon} \neq \boldsymbol{I}_{TN}$) into the spatio-temporal white noise case ($\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$). To see this, let's consider two systems where system 1 is described by

$$\boldsymbol{Y}_1 = \boldsymbol{X}_1 \boldsymbol{H}^H + \boldsymbol{E}_1, \tag{2}$$

with $\boldsymbol{e}_1 = \text{vec}(\boldsymbol{E}_1) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon})$, and system 2 is given by

$$\boldsymbol{Y}_2 = \boldsymbol{X}_2 \boldsymbol{H}^H + \boldsymbol{E}_2, \tag{3}$$

with $\boldsymbol{e}_2 = \text{vec}(\boldsymbol{E}_2) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{TN})$. The systems (2) and (3) are equivalent to

$$\boldsymbol{y}_1 = \text{vec}(\boldsymbol{Y}_1) = (\boldsymbol{I}_N \otimes \boldsymbol{X}_1)\, \text{vec}(\boldsymbol{H}^H) + \boldsymbol{e}_1, \tag{4}$$

$$\boldsymbol{y}_2 = \text{vec}(\boldsymbol{Y}_2) = (\boldsymbol{I}_N \otimes \boldsymbol{X}_2)\, \text{vec}(\boldsymbol{H}^H) + \boldsymbol{e}_2 \tag{5}$$

respectively. After pre-whitening, from (4) we get

$$\widetilde{\boldsymbol{y}_1} = \boldsymbol{\Upsilon}^{-\frac{1}{2}} \boldsymbol{y}_1 = \boldsymbol{\Upsilon}^{-\frac{1}{2}} (\boldsymbol{I}_N \otimes \boldsymbol{X}_1)\, \text{vec}(\boldsymbol{H}^H) + \widetilde{\boldsymbol{e}_1} \tag{6}$$

with $\widetilde{\boldsymbol{e}_1} = \boldsymbol{\Upsilon}^{-\frac{1}{2}} \boldsymbol{e}_1 \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{TN})$. From (5) and (6) we deduce that the systems 1 and 2 are not equivalent, i.e., the unitary constellations (which are optimal for spatio-temporally white noise at high SNR) cannot be employed by performing suitable pre-whitening.

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

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

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 the index $\widehat{k}$ such that

$$\begin{aligned}
\widehat{k} &= \underset{k=1,2,\ldots,K}{\text{argmax}}\ p(\boldsymbol{y}|\boldsymbol{X}_k, \widehat{\boldsymbol{g}}_k) \\
&= \underset{k=1,2,\ldots,K}{\text{argmin}}\ ||\boldsymbol{y} - \widetilde{\boldsymbol{X}_k} \widehat{\boldsymbol{g}}_k||^2_{\boldsymbol{\Upsilon}^{-1}}
\end{aligned}$$

where

$$\widetilde{\boldsymbol{X}_k} = \boldsymbol{I}_N \otimes \boldsymbol{X}_k \quad \text{and} \quad \widehat{\boldsymbol{g}}_k = (\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k)^{-1} \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\Upsilon}^{-\frac{1}{2}} \boldsymbol{y} \tag{7}$$

with $\boldsymbol{\mathcal{X}}_k = \boldsymbol{\Upsilon}^{-\frac{1}{2}} \widetilde{\boldsymbol{X}_k}$ denoting the whitened version of $\widetilde{\boldsymbol{X}_k}$. The GLRT [21], [22], [23] is composed of a bank of K parallel processors where the k-th processor assumes the presence of the k-th 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. Due to the respective expression for the ML estimate of the channel, equation (7), we note that since each codeword of the codebook has full rank (assumption $\mathbf{A2}$), the channel estimate is well defined.

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 $\mathcal{M} = \{(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K) : \text{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k) = 1\}$. 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 in $\mathbb{R}^n$. 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 commonplace [6]- [7], we rely on a pairwise error probability study to construct our merit function. For the special case of unitary codebooks ($\boldsymbol{X}_k^H \boldsymbol{X}_k = \frac{1}{M} \boldsymbol{I}_M$), spatio-temporal white Gaussian noise ($\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$) and independent identically distributed (iid) Rayleigh fading, the exact expression and Chernoff upper bound for the pairwise error probability have been derived in [6]. However, the calculus of these expressions for the general case, i.e, arbitrary matrix constellations $\mathcal{C}$ and noise correlation matrix $\boldsymbol{\Upsilon}$, seems to be burdensome. Since the advantages of good codes are usually more impressive in high SNR scenarios we focus our attention in constructing codes that are optimal for this range of SNR. Namely, in this paper we resort to the asymptotic expression of the PEP in the high SNR 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]. 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. It can be shown (see Appendix I) that at sufficiently high SNR we have the approximation

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} \simeq \mathcal{Q}\left(\frac{1}{\sqrt{2}} \sqrt{\boldsymbol{g}^H \boldsymbol{L}_{ij} \boldsymbol{g}}\right), \tag{8}$$

with

$$\boldsymbol{L}_{ij} = \boldsymbol{\mathcal{X}}_i^H \boldsymbol{\Pi}_j^\perp \boldsymbol{\mathcal{X}}_i \quad \text{and} \quad \boldsymbol{\Pi}_j^\perp = \boldsymbol{I}_{TN} - \boldsymbol{\mathcal{X}}_j \left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-1} \boldsymbol{\mathcal{X}}_j^H$$

where $\mathcal{Q}(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ is the $\mathcal{Q}$-function and $\boldsymbol{\Pi}_j^\perp$ is the orthogonal projector onto the orthogonal complement of the column space of $\boldsymbol{\mathcal{X}}_j$.

Equation (8) shows that the probability of misdetecting $\boldsymbol{X}_i$ for $\boldsymbol{X}_j$, depends on the channel realization $\boldsymbol{g} = \text{vec}\left(\boldsymbol{H}^H\right)$ and on the relative geometry of the codewords $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{X}}_j$. We can decouple the action of $\boldsymbol{g}$ and $\boldsymbol{L}_{ij}$

as follows: using the inequality $\boldsymbol{g}^H \boldsymbol{L}_{ij} \boldsymbol{g} \geq \lambda_{\min}(\boldsymbol{L}_{ij}) \,||\boldsymbol{g}||^2$ and the fact that $\mathcal{Q}(\cdot)$ is monotonically non-increasing, we have the upper bound on the PEP for high SNR

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} \leq \mathcal{Q}\left(\frac{1}{\sqrt{2}}\,||\boldsymbol{g}||\,\sqrt{\lambda_{\min}(\boldsymbol{L}_{ij})}\right). \tag{9}$$

We cannot control the power of the channel $\boldsymbol{g} = \mathrm{vec}(\boldsymbol{H}^H)$, but we can design codebooks aiming at maximizing $\lambda_{\min}(\boldsymbol{L}_{ij})$.

Geometrical interpretation. This later objective has a clear geometric interpretation. Define $\boldsymbol{V} = \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i$. Then $\boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i$ is the orthogonal projection of $\boldsymbol{\mathcal{X}}_i$ onto the orthogonal complement of span $\{\boldsymbol{\mathcal{X}}_j\}$, see figure 1. Now, note that

$$\boldsymbol{L}_{ij} = \boldsymbol{V}^H \boldsymbol{V} = (\boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i)^H (\boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i)$$

is the corresponding Gram matrix and

$$\sqrt{\det(\boldsymbol{V}^H \boldsymbol{V})} = \sqrt{\lambda_{\min}(\boldsymbol{V}^H \boldsymbol{V}) \cdot \ldots \cdot \lambda_{\max}(\boldsymbol{V}^H \boldsymbol{V})} \geq \lambda_{\min}(\boldsymbol{V}^H \boldsymbol{V})^{\frac{MN}{2}}.$$

Hence, by maximizing $\lambda_{\min}(\boldsymbol{V}^H \boldsymbol{V})$, we are increasing $\sqrt{\det\left((\boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i)^H (\boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i)\right)}$ which is proportional to the volume of the parallelepiped spanned by the columns of the $\boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i$. That is, we are trying to place $\boldsymbol{\mathcal{X}}_i$ in the orthogonal complement of span $\{\boldsymbol{\mathcal{X}}_j\}$.

Problem formulation. Following a worst-case approach, we are led from (9) to define the codebook merit function

$$f \,:\, \mathcal{M} \to \mathbb{R} \quad \text{and} \quad \mathcal{C} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K\} \mapsto f(\mathcal{C})$$

as

$$f(\mathcal{C}) = \min\{f_{ij}(\mathcal{C}) : 1 \leq i \neq j \leq K\} \tag{10}$$

where $f_{ij}(\mathcal{C}) = \lambda_{\min}(\boldsymbol{L}_{ij}(\mathcal{C}))$. Constructing an optimal codebook $\mathcal{C} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_K\}$ amounts to solving the optimization problem

$$\mathcal{C}^* = \underset{\mathcal{C} \in \mathcal{M}}{\arg\max} \quad f(\mathcal{C}). \tag{11}$$

The problem defined in (11) 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. Moreover, note that we have

$$f(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_K) = f(\boldsymbol{X}_1 e^{i\theta_1}, \boldsymbol{X}_2 e^{i\theta_2}, \ldots, \boldsymbol{X}_K e^{i\theta_K})$$

for any $\theta_k \in \mathbb{R}$ and $k = 1, \ldots, K$. This means that f depends on each $\boldsymbol{X}_k$ ($\|\boldsymbol{X}_k\| = 1$) only through the line spanned by it (i.e., $\{\lambda \boldsymbol{X}_k \,:\, \lambda \in \mathbb{C}\}$). Thus, we can interpret the optimization problem in (11) as a packing problem in a product of projective spaces [10], [24].

Before deriving an algorithm to design non-coherent space-time constellations for arbitrary noise covariance matrix $\boldsymbol{\Upsilon}$ and any M, N, T and K, in the next section we draw some conclusions about our new codebook merit function f defined in (10). First, 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. Then, we show that our design problem is related to a packing problem in Grassmannian space [24]. Furthermore, we will show that packings in Grassmannian space with respect to *spectral distance* should be the natural choice for codebook constellations.

III. CONSIDERATIONS ABOUT THE NEW CODEBOOK MERIT FUNCTION

We start by rewriting the matrix $\boldsymbol{L}_{ij} = \boldsymbol{\mathcal{X}}_i^H \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i$ as

$$\boldsymbol{L}_{ij} = \left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right)^{\frac{1}{2}} \left(\boldsymbol{I}_{MN} - \boldsymbol{U}_i^H \boldsymbol{U}_j \boldsymbol{U}_j^H \boldsymbol{U}_i\right) \left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right)^{\frac{1}{2}} \tag{12}$$

where $\boldsymbol{U}_i = \boldsymbol{\mathcal{X}}_i \left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right)^{-\frac{1}{2}}$, $\boldsymbol{U}_j = \boldsymbol{\mathcal{X}}_j \left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-\frac{1}{2}}$. That is, $\boldsymbol{U}_i$ contains an orthonormal basis for the subspace spanned by the columns of $\boldsymbol{\mathcal{X}}_i$. Notice that $\boldsymbol{U}_j^H \boldsymbol{U}_j = \boldsymbol{U}_i^H \boldsymbol{U}_i = \boldsymbol{I}_{MN}$. To proceed with the analysis we use the known fact from [25] pp.199: if $\boldsymbol{U}_i$, $\boldsymbol{U}_j$ 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:

(i) If $2MN \leq TN$ ($2M \leq T$), then

$$\boldsymbol{Q}\boldsymbol{U}_i\boldsymbol{W}_1 = \begin{bmatrix} \boldsymbol{I}_{MN} \\ \boldsymbol{0}_{MN} \\ \boldsymbol{0}_{(TN-2MN)\times MN} \end{bmatrix} \quad \boldsymbol{Q}\boldsymbol{U}_j\boldsymbol{W}_2 = \begin{bmatrix} \boldsymbol{C}_{ij} \\ \boldsymbol{S}_{ij} \\ \boldsymbol{0}_{(TN-2MN)\times MN} \end{bmatrix} \tag{13}$$

where $\boldsymbol{C}_{ij}$ is a diagonal $MN \times MN$ matrix with diagonal entries $\cos\alpha_1, \ldots, \cos\alpha_{MN}$, $0 \leq \alpha_1 \leq \ldots \leq \alpha_{MN} \leq \frac{\pi}{2}$, and $\boldsymbol{S}_{ij}^2 + \boldsymbol{C}_{ij}^2 = \boldsymbol{I}_{MN}$. Now, using (13) we can write

$$\boldsymbol{W}_2^H \boldsymbol{U}_j^H \boldsymbol{Q}^H \boldsymbol{Q} \boldsymbol{U}_i \boldsymbol{W}_1 = \boldsymbol{W}_2^H \boldsymbol{U}_j^H \boldsymbol{U}_i \boldsymbol{W}_1 = \boldsymbol{C}_{ij} \Rightarrow \boldsymbol{U}_j^H \boldsymbol{U}_i = \boldsymbol{W}_2 \boldsymbol{C}_{ij} \boldsymbol{W}_1^H, \tag{14}$$

so α_i for $i = 1, \ldots, MN$ are the *principal angles* between the subspaces spanned by $\boldsymbol{U}_i$ and $\boldsymbol{U}_j$. Due to Ostrowski's theorem pp.224,225 in [26], and equations (12) and (14), it is not difficult to see that the following inequality holds

$$\lambda_{\min}\left(\boldsymbol{L}_{ij}\right) \quad \geq \quad \lambda_{\min}\left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right) \lambda_{\min}\left(\boldsymbol{S}_{ij}^2\right). \tag{15}$$

Clearly, from (15), we deduce that in order to minimize an upper bound on PEP in high SNR regime, one should simultaneously increase $\lambda_{\min}\left(\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i\right)$ and $\lambda_{\min}\left(\boldsymbol{S}_{ij}^2\right)$. Unfortunately, the right-hand side of the inequality (15) does not offer much insight into the form of the optimal codebook for the case of arbitrary noise covariance matrix $\boldsymbol{\Upsilon}$ (even for the case $K = 2$). One of the reasons originates from the fact that pairwise error probabilities are not symmetric for this general case. Hence, in the sequel, we treat the specific case of spatio-temporal white Gaussian observation noise to find out what conclusions can we draw about the form of the optimal codebook.

(a) Special case (spatio-temporal white noise): $\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$, $2M \leq T$. Remark that using (7), (9), and for $\boldsymbol{\Upsilon} = \boldsymbol{I}_{TN}$, we have

$$\boldsymbol{L}_{ij} = \boldsymbol{I}_N \otimes \left(\boldsymbol{X}_i^H \boldsymbol{X}_i - \boldsymbol{X}_i^H \boldsymbol{X}_j \left(\boldsymbol{X}_j^H \boldsymbol{X}_j\right)^{-1} \boldsymbol{X}_j^H \boldsymbol{X}_i\right).$$

Hence,

$$\lambda_{\min}(\boldsymbol{L}_{ij}) = \lambda_{\min}\left(\boldsymbol{X}_i^H\boldsymbol{X}_i - \boldsymbol{X}_i^H\boldsymbol{X}_j\left(\boldsymbol{X}_j^H\boldsymbol{X}_j\right)^{-1}\boldsymbol{X}_j^H\boldsymbol{X}_i\right). \tag{16}$$

From (16), an immediate conclusion is that the code design criterion in (11) does not depend on the number of receive antennas N. Because $T \geq 2M$ (in particular $T \geq M$), using a thin singular value decomposition (SVD), we can write $\boldsymbol{X}_i = \boldsymbol{V}_i\boldsymbol{D}_i\boldsymbol{W}_i$ and $\boldsymbol{X}_j = \boldsymbol{V}_j\boldsymbol{D}_j\boldsymbol{W}_j$ where $\boldsymbol{V}_i$ and $\boldsymbol{V}_j$ are $T \times M$ unitary (orthonormal) matrices, $\boldsymbol{W}_i$ and $\boldsymbol{W}_j$ are $M \times M$ unitary matrices, and $\boldsymbol{D}_i$, $\boldsymbol{D}_j$ are $M \times M$ real nonnegative diagonal matrices. It is not difficult to see that

$$\lambda_{\min}(\boldsymbol{L}_{ij}) = \lambda_{\min}\left(\boldsymbol{D}_i^2 - \boldsymbol{D}_i\boldsymbol{V}_i^H\boldsymbol{V}_j\boldsymbol{V}_j^H\boldsymbol{V}_i\boldsymbol{D}_i\right). \tag{17}$$

As we can see from (17), the matrices $\boldsymbol{W}_i$ and $\boldsymbol{W}_j$ do not appear in the expression. This implies that any optimal constellation can be described in the form $\boldsymbol{X}_i = \boldsymbol{V}_i\boldsymbol{D}_i$. We now show that for two symbol constellations ($K = 2$) the unitary constellations are optimal in the sense of maximizing the codebook merit function defined in (10). Toward this end, note that for $\boldsymbol{X}_1 = \boldsymbol{V}_1\boldsymbol{D}_1$ and $\boldsymbol{X}_2 = \boldsymbol{V}_2\boldsymbol{D}_2$

$$f\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right) = \min\{f_{12}\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right), f_{21}\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right)\} \leq f_{12}\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right) = \lambda_{\min}(\boldsymbol{L}_{12}), \tag{18}$$

where $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are $T \times M$ unitary (orthonormal) matrices, and $\boldsymbol{D}_1$, $\boldsymbol{D}_2$ are $M \times M$ real nonnegative diagonal matrices. Since $\boldsymbol{V}_1$, $\boldsymbol{V}_2$ are $T \times M$ matrices with orthonormal columns and $2M \leq T$, as before, we know that there exist $M \times M$ unitary matrices $\boldsymbol{\mathcal{W}}_1$ and $\boldsymbol{\mathcal{W}}_2$, and a $T \times T$ unitary matrix $\boldsymbol{\mathcal{Q}}$ with the following properties [25]:

$$\boldsymbol{\mathcal{Q}}\boldsymbol{V}_1\boldsymbol{\mathcal{W}}_1 = \begin{bmatrix} \boldsymbol{I}_M \\ \boldsymbol{0}_M \\ \boldsymbol{0}_{(T-2M)\times M} \end{bmatrix} \quad \boldsymbol{\mathcal{Q}}\boldsymbol{V}_2\boldsymbol{\mathcal{W}}_2 = \begin{bmatrix} \boldsymbol{C}_{12} \\ \boldsymbol{S}_{12} \\ \boldsymbol{0}_{(T-2M)\times M} \end{bmatrix} \tag{19}$$

where $\boldsymbol{C}_{12}$ is a diagonal $M \times M$ matrix with diagonal entries $\cos\beta_1, \ldots, \cos\beta_M$, $0 \leq \beta_1 \leq \ldots \leq \beta_M \leq \frac{\pi}{2}$, and $\boldsymbol{S}_{12}^2 + \boldsymbol{C}_{12}^2 = \boldsymbol{I}_M$. Substituting (19) in (18) yields

$$\begin{aligned} \lambda_{\min}(\boldsymbol{L}_{12}) &= \lambda_{\min}\left(\boldsymbol{D}_1\boldsymbol{\mathcal{W}}_1\boldsymbol{S}_{12}^2{\boldsymbol{\mathcal{W}}_1}^H\boldsymbol{D}_1\right) \\ &= \lambda_{\min}\left(\boldsymbol{S}_{12}{\boldsymbol{\mathcal{W}}_1}^H\boldsymbol{D}_1^2\boldsymbol{\mathcal{W}}_1\boldsymbol{S}_{12}\right) \leq \lambda_{\min}\left(\boldsymbol{D}_1^2\right)\lambda_{\max}\left(\boldsymbol{S}_{12}^2\right) \end{aligned} \tag{20}$$

where (20) is valid due to Ostrowski's theorem. Since $\lambda_{\min}\left(\boldsymbol{D}_1^2\right) \leq \frac{1}{M}\text{tr}(\boldsymbol{X}_1^H\boldsymbol{X}_1) = \frac{1}{M}$, and also using (18) and (20) we have the upper bound on the codebook merit function for $K = 2$:

$$f\left(\boldsymbol{X}_1, \boldsymbol{X}_2\right) \leq \frac{1}{M}. \tag{21}$$

Since we want to maximize the codebook merit function, from (18) and (21) we can list some of the conditions for it to happen:

1. The constellation of unitary matrices is optimal, i.e., $\boldsymbol{D}_1 = \boldsymbol{D}_2 = \frac{1}{\sqrt{M}}\boldsymbol{I}_M$ and $\boldsymbol{X}_1^H\boldsymbol{X}_1 = \boldsymbol{X}_2^H\boldsymbol{X}_2 = \frac{1}{M}\boldsymbol{I}_M$.
2. We want $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ to be separated as much as possible. The optimal scenario is when $\beta_1 = \frac{\pi}{2}$, the case when codewords $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are mutually orthogonal, i.e., $\boldsymbol{X}_2^H\boldsymbol{X}_1 = 0$.

In this case, the inequality sign in (21) can be replaced with an equality sign. Thus, 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.

(ii) For $M \leq T \leq 2M$, then

$$\boldsymbol{Q}\boldsymbol{U}_i\boldsymbol{W}_1 = \begin{bmatrix} \boldsymbol{I}_{TN-MN} & \mathbf{0}_{(TN-MN)\times(2MN-TN)} \\ \mathbf{0}_{(2MN-TN)\times(TN-MN)} & \boldsymbol{I}_{2MN-TN} \\ \mathbf{0}_{TN-MN} & \mathbf{0}_{(TN-MN)\times(2MN-TN)} \end{bmatrix}$$

$$\boldsymbol{Q}\boldsymbol{U}_j\boldsymbol{V}_1 = \begin{bmatrix} \boldsymbol{C}_{ij} & \mathbf{0}_{(TN-MN)\times(2MN-TN)} \\ \mathbf{0}_{(2MN-TN)\times(TN-MN)} & \boldsymbol{I}_{2MN-TN} \\ \boldsymbol{S}_{ij} & \mathbf{0}_{(TN-MN)\times(2MN-TN)} \end{bmatrix}$$

where $\boldsymbol{C}_{ij}$ is a diagonal $(TN - MN) \times (TN - MN)$ matrix with diagonal entries $\cos\alpha_1, \ldots, \cos\alpha_{TN-MN}$, $0 \leq \alpha_1 \leq \ldots \leq \alpha_{TN-MN} \leq \frac{\pi}{2}$, and $\boldsymbol{S}_{ij}^2 + \boldsymbol{C}_{ij}^2 = \boldsymbol{I}_{TN-MN}$.

Repeating the analysis for the case $2M \leq T$, we have

$$\boldsymbol{L}_{ij} = \left(\boldsymbol{\mathcal{X}}_i^H\boldsymbol{\mathcal{X}}_i\right)^{\frac{1}{2}}\boldsymbol{W}_1 \begin{bmatrix} \boldsymbol{S}_{ij}^2 & \mathbf{0}_{(TN-MN)\times(2MN-TN)} \\ \mathbf{0}_{(2MN-TN)\times(TN-MN)} & \mathbf{0}_{2MN-TN} \end{bmatrix} \boldsymbol{W}_1^H\left(\boldsymbol{\mathcal{X}}_i^H\boldsymbol{\mathcal{X}}_i\right)^{\frac{1}{2}}. \tag{22}$$

From (22) we conclude that the upper bound in (9) is a constant for any combination of two codewords $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{X}}_j$. This implies the following result.

Proposition: The length of the coherence interval T should be at least as twice as large as the number of transmit antennas M. In symbols, $2M \leq T$.

The result 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]. Therefore, in this work, when designing constellations for arbitrary $\boldsymbol{\Upsilon}$, we take $2M \leq T$.

Remark that

$$\lambda_{\min}(\boldsymbol{L}_{12}) = \lambda_{\min}\left(\boldsymbol{S}_{12}\boldsymbol{\mathcal{W}}_1{}^H\boldsymbol{D}_1^2\boldsymbol{\mathcal{W}}_1\boldsymbol{S}_{12}\right) \geq \lambda_{\min}\left(\boldsymbol{D}_1^2\right)\lambda_{\min}\left(\boldsymbol{S}_{12}^2\right) \tag{23}$$

where (23) is valid due to Ostrowski's theorem and (20). Since we have seen that the unitary constellations are the optimal ones with respect to f for the case of spatio-temporally white Gaussian noise, equation (23) shows that our design problem for any K is related to a packing problem in Grassmannian space [24]. Furthermore, it seems that packings in Grassmannian space with respect to *spectral distance* should be the natural choice for codebook constellations. Moreover, from (11) and (16) for M=1 we see that the problem of finding good codes coincides with the very well known packing problem in the complex projective space [10]. Please refer to [10] and [24] for more details on packing problems in Grassmannian space.

IV. Codebook Construction

We propose a two-phase methodology to tackle the optimization problem in (11). In phase one, we start by solving a convex semi-definite programming (SDP) relaxation to obtain a rough estimate of the optimal codebook. Phase two refines it through a geodesic descent optimization algorithm (GDA) which efficiently exploits the Riemannian geometry of the constraints. Suppose a codebook of size K is desired. In table I, page 11 we give the strategy that has shown to be efficient.

input:	`M, N, T, K,` $\boldsymbol{\Upsilon}$
step 1)	`Choose the first codeword;`
step 2)	`Set` $k = 2$;
step 3)	`Perform SDP relaxation to obtain` k`-th codeword;`
step 4)	`Set` $k = k + 1$;
step 5)	`if` $k \leq K$, `return to Step 3);`
step 6)	`Run the geodesic descent algorithm (GDA) to obtain the final codebook;`
output:	`The matrix` $\boldsymbol{X} = \left[\texttt{vec}(\boldsymbol{X}_1) \quad \ldots \quad \texttt{vec}(\boldsymbol{X}_K) \right]$

TABLE I

Codebook design algorithm

We now explain Steps (3) and (6), respectively, in more detail.

Phase 1: SDP relaxation. This phase constructs a sub-optimal codebook $\mathcal{C}^* = \{\boldsymbol{X}_1^*, ..., \boldsymbol{X}_K^*\}$. The codebook is constructed incrementally. We start assuming that we are in a possession of an acceptable estimate of a codebook of size $k-1$, while we are interested in a rough estimate of a codebook of size k, where $k = 2, 3, ..., K$. We obtain an estimate of a codebook of size k by retaining the first $k-1$ codewords. Hence, we solve the optimization problem in the sequel consecutively $K-1$ times. There are several strategies for choosing the first codeword $\boldsymbol{X}_1^*$, e.g, randomly generated, filling columns of the matrix with eigenvectors associated to the smallest eigenvalues of the noise covariance matrix, etc. Addition of a new codeword consists in solving a SDP. Let $\mathcal{C}_{k-1}^* = \{\boldsymbol{X}_1^*, ..., \boldsymbol{X}_{k-1}^*\}$

be the codebook at the $k-1^{\underline{th}}$ stage. The new codeword is found by solving

$$\begin{aligned} \boldsymbol{X}_k^* &= \underset{\text{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k)=1}{\arg\max} \quad f(\boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_{k-1}^*, \boldsymbol{X}_k) \\ &= \underset{\text{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k)=1}{\arg\max} \quad \min_{1 \le m \le k-1} \{\lambda_{\min}(\boldsymbol{L}_{mk}),\ \lambda_{\min}(\boldsymbol{L}_{km})\}. \end{aligned} \tag{24}$$

We can show that the optimization problem defined in (24) is equivalent to

$$(\boldsymbol{\mathfrak{X}}_k^*, \text{vec}(\boldsymbol{X}_k{}^*), t^*) = \ \arg\max \ \ t \tag{25}$$

with the following constraints

$$\begin{aligned} &\text{LMI}_{A_m}(\boldsymbol{\mathfrak{X}}_k, \text{vec}(\boldsymbol{X}_k), t) \succeq \mathbf{0},\ m = 1, ..., k-1 \\ &\text{LMI}_{B_m}(\boldsymbol{\mathfrak{X}}_k, \text{vec}(\boldsymbol{X}_k), t) \succeq \mathbf{0},\ m = 1, ..., k-1 \\ &\text{tr}(\boldsymbol{\mathfrak{X}}_k) = 1, \ \ \boldsymbol{\mathfrak{X}}_k = \text{vec}(\boldsymbol{X}_k)\text{vec}^H(\boldsymbol{X}_k) \end{aligned} \tag{26}$$

where the abbreviations $\text{LMI}_{A_m}(\boldsymbol{\mathfrak{X}}_k, \text{vec}(\boldsymbol{X}_k), t)$ and $\text{LMI}_{B_m}(\boldsymbol{\mathfrak{X}}_k, \text{vec}(\boldsymbol{X}_k), t)$ denote linear (actually, affine) matrix inequalities in the variables $\boldsymbol{\mathfrak{X}}_k$, $\text{vec}(\boldsymbol{X}_k)$, 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.

Due to the rank condition in (26) (note that the equations $\boldsymbol{\mathfrak{X}}_k = \text{vec}(\boldsymbol{X}_k)\text{vec}^H(\boldsymbol{X}_k)$ and $\text{tr}(\boldsymbol{\mathfrak{X}}_k) = 1$ imply that rank$(\boldsymbol{\mathfrak{X}}_k) = 1$), the design of the codewords, once again, translates into a difficult high-dimensional nonlinear optimization problem. However, relaxing this restriction as

$$\boldsymbol{\mathfrak{X}}_k \succeq \text{vec}(\boldsymbol{X}_k)\text{vec}^H(\boldsymbol{X}_k) \tag{27}$$

and rewriting (27) as

$$\begin{bmatrix} \boldsymbol{\mathfrak{X}}_k & \text{vec}(\boldsymbol{X}_k) \\ \text{vec}^H(\boldsymbol{X}_k) & 1 \end{bmatrix} \succeq \mathbf{0}$$

the optimization problem in (25) becomes

$$(\boldsymbol{\mathfrak{X}}_k^*, \text{vec}(\boldsymbol{X}_k{}^*), t^*) = \ \arg\max \ \ t \tag{28}$$

with the constraints

$$\begin{aligned} &\text{LMI}_{A_m}(\boldsymbol{\mathfrak{X}}_k, \text{vec}(\boldsymbol{X}_k), t) \succeq \mathbf{0},\ m = 1, ..., k-1 \\ &\text{LMI}_{B_m}(\boldsymbol{\mathfrak{X}}_k, \text{vec}(\boldsymbol{X}_k), t) \succeq \mathbf{0},\ m = 1, ..., k-1 \\ &\text{tr}(\boldsymbol{\mathfrak{X}}_k) = 1, \ \ \begin{bmatrix} \boldsymbol{\mathfrak{X}}_k & \text{vec}(\boldsymbol{X}_k) \\ \text{vec}^H(\boldsymbol{X}_k) & 1 \end{bmatrix} \succeq \mathbf{0}. \end{aligned}$$

The rank 1 relaxation is usually known as the Shor relaxation [27]. The optimization problem in (28) is a convex one in the variables $\mathfrak{X}_k$, vec$(\boldsymbol{X}_k)$ and t. Remark that for $K = 256$, $M = 2$, $N = 2$, $T = 8$ and in the last passage through the loop, i.e., for $k = K$, the output variable $\mathfrak{X}_k$ is of dimension 16×16 (does not depend on N and K) and the number of linear matrix inequality constraints that needs to be defined is of order K. To solve the optimization problem in (28) we used the *Self-Dual-Minimization* package SeDuMi 1.1 [28]. Once the problem defined in (28) is solved we need to extract the $k^{\underline{th}}$ codeword from the output variable $\mathfrak{X}_k$. Toward this end, we adopt a technique similar to [32]. The technique consists in generating independent realizations of random vectors that follow a Gaussian distribution with zero mean and covariance matrix $\mathfrak{X}_k$, i.e., $\boldsymbol{z}_l \overset{iid}{\sim} \mathcal{CN}\left(\boldsymbol{0}, \mathfrak{X}_k\right)$, for $l = 1, 2, ..., L$, where L is a parameter to be chosen (in all simulations herein presented we assumed L=10000). After forcing norm 1, i.e., $\boldsymbol{v}_l = \boldsymbol{z}_l / ||\boldsymbol{z}_l||$ for $l = 1, 2, ..., L$, we choose the $k - th$ codeword, $\boldsymbol{X}_k^* = \text{ivec}(\boldsymbol{v}_l^*)$ where

$$l^* = \underset{l = 1, 2, ..., L}{\arg\max} \quad f(\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, ..., \boldsymbol{X}_{k-1}^*, \text{ivec}(\boldsymbol{v})). \tag{29}$$

The operation "ivec" operates as an inverse of "vec" (reshapes the TM-dimensional vector into a $T \times M$ matrix). Note that $\boldsymbol{X}_k^*$ is a valid codeword because tr$(\boldsymbol{X}_k^{*H} \boldsymbol{X}_k^*) = 1$. We are clearly dealing with a suboptimal solution for a codebook.

Phase 2: Geodesic Descent Algorithm. Problem (28) requires the optimization of a non-smooth function over the smooth manifold $\mathcal{M}$ (Cartesian product of K spheres). After phase 1, i.e., having solved the optimization problem (28) consecutively $K - 1$ times, for k = 2,3,..,K, we are now in possession of a rough estimate of a codebook of size K. To refine this estimate we resort to an iterative algorithm, which we call GDA (geodesic descent algorithm). In table II we explain the GDA in more detail.

Let $\mathcal{C}_k$ be the $k^{\underline{th}}$ iterate (the initialization $\mathcal{C}_0$ is furnished by phase 1). Note that the power constraint tr$(\boldsymbol{X}_i^H \boldsymbol{X}_i) = 1$, for $i = 1, 2, ..., K$, can be equivalently written as

$$\boldsymbol{x}_i^T \boldsymbol{x}_i = 1,$$

where

$$\boldsymbol{x}_i = \begin{bmatrix} \Re \, \text{vec}(\boldsymbol{X}_i) \\ \Im \, \text{vec}(\boldsymbol{X}_i) \end{bmatrix} \in \mathbb{R}^{2TM},$$

and $\Re$ and $\Im$ denote the real and imaginary part of a complex quantity, respectively. In step 3 each $\boldsymbol{x}_i$, $i = 1, ..., K$ is used to construct the vector $\boldsymbol{x}$. In step 4 we identify the index set $\mathcal{A}$ of "active" constraint pairs (i, j), i.e., $\mathcal{A} = \{(i, j) : f_{ij}(\mathcal{C}_k) \leq f(\mathcal{C}_k) + \epsilon\}$ where ϵ is arbitrary small (in all simulations herein presented we have chosen $\epsilon = 10^{-5}$). In step 8 we check if there is an ascent direction $\boldsymbol{d}$ simultaneously for all functions f_{ij} with $(i, j) \in \mathcal{A}$. We know that if it exists $\boldsymbol{d}$ such that $\nabla^T f_{i_a j_a}(\boldsymbol{x}) \, \boldsymbol{d} > 0$, for $1 \leq i_a \neq j_a \leq K, a = 1, 2, ..., z$, we can try to improve our cost function locally. In order to solve the optimization problem in step 8 we need to determine the gradient $\nabla f_{i_a j_a}$. In Appendix III, we give its respective expression. This ascent direction $\boldsymbol{d}$ is searched within

input: The matrix $\boldsymbol{X} = \begin{bmatrix} \mathsf{vec}(\boldsymbol{X}_1) & \dots & \mathsf{vec}(\boldsymbol{X}_K) \end{bmatrix}$

step 1) Determine the value of the merit function, $\mathsf{cost} = f(\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_K)$;

step 2) Initialize $\epsilon = 10^{-5}$;

step 3) Construct the vector $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} = \begin{bmatrix} \Re\,\mathsf{vec}(\boldsymbol{X}_1) \\ \Im\,\mathsf{vec}(\boldsymbol{X}_1) \\ \vdots \\ \Re\,\mathsf{vec}(\boldsymbol{X}_K) \\ \Im\,\mathsf{vec}(\boldsymbol{X}_K) \end{bmatrix}$;

step 4) Determine z, the number of combinations $(\boldsymbol{X}_i, \boldsymbol{X}_j)$, $1 \leq i \neq j \leq K$, such that $f_{ij}(\mathcal{C}) = \lambda_{\min}(\boldsymbol{L}_{ij}(\mathcal{C}))$ falls into the interval $[\mathsf{cost}, \mathsf{cost} + \epsilon]$, i.e. f_{ij} attains the minimum. These combinations $(\boldsymbol{X}_i, \boldsymbol{X}_j)$ are called the active ones;

step 5) Determine the gradient, $\nabla f_{i_a j_a}(\boldsymbol{x})$, for every active combination $(\boldsymbol{X}_{i_a}, \boldsymbol{X}_{j_a})$, $1 \leq i_a \neq j_a \leq K, a = 1, 2, ..., z$;

step 6) Construct the gradient matrix

$$\boldsymbol{G} = \begin{bmatrix} \nabla^T f_{i_1 j_1}(\boldsymbol{x}) \\ \vdots \\ \nabla^T f_{i_z j_z}(\boldsymbol{x}) \end{bmatrix}_{z \times 2KTM} ;$$

step 7) Construct the matrix

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{x}_1^T & 0 & \cdots & 0 \\ 0 & \boldsymbol{x}_2^T & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \boldsymbol{x}_K^T \end{bmatrix}_{K \times 2KTM} ;$$

step 8) Solve the linear program

$$\begin{aligned} (\boldsymbol{d}^*, s^*) = \quad & \arg\max \quad s; \\ & \boldsymbol{G}\boldsymbol{d} \geq s\mathbf{1}_{z\times 1} \\ & \boldsymbol{H}\boldsymbol{d} = \mathbf{0}_{K\times 1} \\ & -\mathbf{1}_{2KTM\times 1} \leq \boldsymbol{d} \leq +\mathbf{1}_{2KTM\times 1} \end{aligned}$$

step 9) If $s \leq 0$, Go to Step (16);

step 10) Initialize $\beta = 0.9$, $c = 0$, $c_{max} = 400$ and $t = 1$;

step 11)	`Construct the geodesic` $$\Gamma(t) = \begin{bmatrix} \boldsymbol{x}_1(t) \\ \vdots \\ \boldsymbol{x}_K(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 \cos(\|\boldsymbol{d}_1\|t) + \frac{\boldsymbol{d}_1}{\|\boldsymbol{d}_1\|} \sin(\|\boldsymbol{d}_1\|t) \\ \vdots \\ \boldsymbol{x}_K \cos(\|\boldsymbol{d}_K\|t) + \frac{\boldsymbol{d}_K}{\|\boldsymbol{d}_K\|} \sin(\|\boldsymbol{d}_K\|t) \end{bmatrix};$$
step 12)	`Determine temporary value of the merit function,` `tempcost` $= f(\text{ivec}(\boldsymbol{x}_1(t)), \text{ivec}(\boldsymbol{x}_2(t)), ..., \text{ivec}(\boldsymbol{x}_K(t)))$;
step 13)	`If tempcost` $>$ `cost, then cost` $=$ `tempcost,` $\boldsymbol{x}_i = \boldsymbol{x}_i(t)$ `for` i`=1,2,..,`K`. Return to Step (3);`
step 14)	`Increment` c`, update` $t = \beta^c$;
step 15)	`If` $c \leq c_{max}$`, Return to Step (12);`
step 16)	`Return the matrix` $\boldsymbol{X} = \begin{bmatrix} \text{vec}(\boldsymbol{X}_1) & \ldots & \text{vec}(\boldsymbol{X}_K) \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_i & \ldots & \boldsymbol{x}_K \end{bmatrix}$;
output:	`The matrix` $\boldsymbol{X} = \begin{bmatrix} \text{vec}(\boldsymbol{X}_1) & \ldots & \text{vec}(\boldsymbol{X}_K) \end{bmatrix}$

TABLE II

GDA ALGORITHM

$T_{\mathcal{C}_k}\mathcal{M}$, the tangent space to $\mathcal{M}$ at $\mathcal{C}_k$, and consists in solving a linear program. To ensure that $\boldsymbol{d}$ belongs to $T_{\mathcal{C}_k}\mathcal{M}$, the constraint $\boldsymbol{Hd} = \boldsymbol{0}_{K\times 1}$ (equivalently, $\boldsymbol{x}_i^T\boldsymbol{d}_i = 0$ for i=1, 2, ..., K) in step 8 is introduced. The constraint $-\boldsymbol{1}_{2KTM\times 1} \leq \boldsymbol{d} \leq +\boldsymbol{1}_{2KTM\times 1}$ bounds the solution of the linear program in step 8. If there is no such ascent direction, the algorithm stops. Otherwise, we perform an Armijo search for $f(\mathcal{C})$ along the geodesic which emanates from $\mathcal{C}_k$ in the direction $\boldsymbol{d}$. This Armijo search determines $\mathcal{C}_{k+1}$ and we repeat the loop. From the expression for the geodesic in step 11, it is easy to see that we travel along the surface of the sphere S^{2TM-1}, i.e., $\boldsymbol{x}_i(t)^T\boldsymbol{x}_i(t) = 1$ for every i=1,2,...,K.

A geodesic is nothing but the generalization of a straight line in Euclidean space to a curved surface [33]. 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 in step 2 controls the complexity of the optimization problem in step 8. A too small ϵ implies slow convergence of the algorithm, whereas a big ϵ increases the complexity of the linear program (by increasing z, the number of active functions f_{ij}). 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 $\boldsymbol{z}_{\mathsf{max}} = K(K-1) = 65280$). Although the matrix G is a sparse matrix, it is preferably to impose it to be of moderate size too. The choice of ϵ made in step 2 controls that.

Remark: The utility of the step 3 (SDP) 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 $T = 2$, the step 3 provides us the optimal codebook for $K = 2^p$ where $p = 1, 2, \ldots$. In this case there is no need to use step 6 of the algorithm. In all simulations herein presented the procedure presented in table I has been implemented.

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., $\boldsymbol{\Upsilon} = \mathsf{E}[\text{vec}(\boldsymbol{E})\,\text{vec}(\boldsymbol{E})^H] = \boldsymbol{I}_{NT}$. We compared our codes with the best known found in [7]. We considered scenarios with coherence interval T=8, M=1, 2 and 3 transmit antennas, N=1 receive antennas and a codebook with K=256 codewords. Let

$$\text{dist} = \frac{1}{K}\sum_{k=1}^{K}\sqrt{\text{tr}\left(\left(\boldsymbol{X}_k^H\boldsymbol{X}_k - \frac{1}{M}\boldsymbol{I}_M\right)^2\right)}$$

denote the average distance of our codebooks from the constellation of unitary matrices. For $M = 2$, $T = 8$ and $K = 256$, the average distance obtained was $\text{dist} = 1.6 \cdot 10^{-3}$, while for $M = 3$, $T = 8$ and $K = 256$, the average distance was $\text{dist} = 1.3 \cdot 10^{-2}$. As it was expected, the algorithm converged to constellations of unitary matrices. In that case and for spatio-temporal observation noise, our GLRT receiver corresponds to the Bayesian receiver (takes in account the statistics of the channel). In figures 2–4, we show the symbol error rate (SER) versus

$$\text{SNR} = \mathsf{E}\{||\boldsymbol{X}_k\boldsymbol{H}^H||^2\}/\mathsf{E}\{||\boldsymbol{E}||^2\} = N\sigma^2/\text{tr}(\boldsymbol{\Upsilon}).$$

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 are only marginally better for these particular cases. For M=1, in figure 5 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 Hermitian 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 (adopted) 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 6–8 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 6 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 $\text{SER} = 10^{-3}$, our codes demonstrate a gain of 3dB when compared with the unitary codes. Figure 7 plots the result of the experiment for T=8, M=2, N=1, K=256 and $\boldsymbol{\rho}$=[1; 0.8; 0.5; 0.15; zeros(4,1)]. For $\text{SER} = 10^{-3}$ our codes demonstrate gain of 2dB when compared with unitary codes. Figure 8 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 adopted to this particular scenario with unitary codes. In figures 9–10 the solid curves represent performance of codes constructed by our method, while the dashed curves represent performance of unitary codes [7]. Figure 9 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 10 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 convex semi-definite programming (SDP) relaxation to obtain a rough estimate of the optimal codebook. The second phase refines it through a geodesic descent optimization algorithm which efficiently exploits the Riemannian geometry of the constraints. 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 (8).

If $\boldsymbol{X}_i$ is transmitted, then the probability that the receiver decides in favor of $\boldsymbol{X}_j$ is:

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} = P(\boldsymbol{z}_i^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{z}_i > \boldsymbol{z}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{z}_j) \tag{30}$$

where for $k \in \{i,j\}$

$$\boldsymbol{z}_k = \boldsymbol{y} - \widetilde{\boldsymbol{X}_k}\widehat{\boldsymbol{g}}_k,$$

$$\widetilde{\boldsymbol{X}_k} = \boldsymbol{I}_N \otimes \boldsymbol{X}_k,\ \boldsymbol{\mathcal{X}}_k = \boldsymbol{\Upsilon}^{-\frac{1}{2}}\widetilde{\boldsymbol{X}_k}$$

$$\boldsymbol{y} = \text{vec}(\boldsymbol{Y}) = \widetilde{\boldsymbol{X}_i}\boldsymbol{g} + \boldsymbol{e}, \boldsymbol{e} = \text{vec}(\boldsymbol{E}),$$

and

$$\widehat{\boldsymbol{g}}_k = (\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k)^{-1}\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{y}$$

is the maximum likelihood (ML) estimate of the channel when $\boldsymbol{X}_k$ is transmitted. The unknown realization of the channel is denoted by $\boldsymbol{g} = \text{vec}(\boldsymbol{H}^H)$.

Let $\boldsymbol{S}_i = (\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\mathcal{X}}_i)^{-1}\boldsymbol{\mathcal{X}}_i^H \boldsymbol{\Upsilon}^{-\frac{1}{2}}$. Thus,

$$\begin{aligned} \boldsymbol{z}_i &= \boldsymbol{y} - \widetilde{\boldsymbol{X}_i}\widehat{\boldsymbol{g}}_i \\ &= \underbrace{(\boldsymbol{I}_{TN} - \widetilde{\boldsymbol{X}_i}\boldsymbol{S}_i)}_{\boldsymbol{P}_i}(\widetilde{\boldsymbol{X}_i}\boldsymbol{g} + \boldsymbol{e}) \\ &= \boldsymbol{P}_i\boldsymbol{e}. \end{aligned} \tag{31}$$

Similarly, it can be shown that

$$\boldsymbol{z}_j = \boldsymbol{\Delta}\boldsymbol{g} + \boldsymbol{P}_j\boldsymbol{e} \tag{32}$$

where $\boldsymbol{P}_j = \boldsymbol{I}_{TN} - \widetilde{\boldsymbol{X}_j}\boldsymbol{S}_j$ and $\boldsymbol{\Delta} = \boldsymbol{P}_j\widetilde{\boldsymbol{X}_i}$. Hence, substituting (31) and (32) in (30) we have

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} = P(\boldsymbol{e}^H(\boldsymbol{P}_i^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_i - \boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j)\boldsymbol{e} - \boldsymbol{e}^H\boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g} - \boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j\boldsymbol{e} > \underbrace{\boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g}}_{\lambda}). \tag{33}$$

Unfortunately, the expression (33) cannot be simplified analytically. We shall analyze (33) in the high SNR regime where the linear term of $\boldsymbol{e}$ is dominant, i.e., the quadratic term of $\boldsymbol{e}$ is negligible. Therefore,

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} \simeq P(\underbrace{-\boldsymbol{e}^H\boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g} - \boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j\boldsymbol{e}}_{z} > \lambda). \tag{34}$$

Define $z = -\boldsymbol{e}^H\boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g} - \boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j\boldsymbol{e}$. We see that z is a real Gaussian variable with zero mean (because $\mathsf{E}[\boldsymbol{e}] = 0$) and unknown variance σ^2, i.e., $z \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right)$, which will be calculated in sequel:

$$\begin{aligned} \sigma^2 &= \mathsf{E}[\boldsymbol{z}^2] = \mathsf{E}[\left(\boldsymbol{e}^H\boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g} + \boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j\boldsymbol{e}\right)^2] \\ &= \mathsf{E}[\boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j\boldsymbol{e}\boldsymbol{e}^H\boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g} + \boldsymbol{e}^H\boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g}\boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j\boldsymbol{e}]. \end{aligned}$$

Continuing with analysis,

$$\begin{aligned} \mathsf{E}[\boldsymbol{e}^H\boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g}\boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j\boldsymbol{e}] &= \text{tr}\left(\mathsf{E}[\boldsymbol{e}\boldsymbol{e}^H\boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g}\boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j]\right) \\ &= \boldsymbol{g}^H\boldsymbol{\Delta}^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{P}_j\boldsymbol{\Upsilon}\boldsymbol{P}_j^H\boldsymbol{\Upsilon}^{-1}\boldsymbol{\Delta}\boldsymbol{g} \end{aligned}$$

which implies

$$\sigma^2 = 2\boldsymbol{g}^H \boldsymbol{\Delta}^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_j \boldsymbol{\Upsilon} \boldsymbol{P}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Delta} \boldsymbol{g}$$
$$= 2\boldsymbol{g}^H \widetilde{\boldsymbol{X}}_i^H \underbrace{\boldsymbol{P}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_j \boldsymbol{\Upsilon} \boldsymbol{P}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_j}_{\boldsymbol{R}} \widetilde{\boldsymbol{X}}_i \boldsymbol{g}. \tag{35}$$

It is known that, if $z \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right)$, then

$$P(z > \lambda) = \mathcal{Q}\left(\frac{\lambda}{\sigma}\right) \tag{36}$$

where $\mathcal{Q}(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

The matrix $\boldsymbol{R}$ in (35) can be simplified and it can be easily shown that

$$\boldsymbol{R} = \boldsymbol{P}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_j = \boldsymbol{\Upsilon}^{-\frac{1}{2}} \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\Upsilon}^{-\frac{1}{2}}. \tag{37}$$

where $\boldsymbol{\Pi}_j^{\perp} = \boldsymbol{I}_{TN} - \boldsymbol{\mathcal{X}}_j \left(\boldsymbol{\mathcal{X}}_j^H \boldsymbol{\mathcal{X}}_j\right)^{-1} \boldsymbol{\mathcal{X}}_j^H$ is the orthogonal projector onto the orthogonal complement of the column space of $\boldsymbol{\mathcal{X}}_{\boldsymbol{j}}$.

Using (37) and substituting it in (35) we have

$$\sigma^2 = 2\boldsymbol{g}^H \widetilde{\boldsymbol{X}}_i^H \boldsymbol{P}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_j \widetilde{\boldsymbol{X}}_i \boldsymbol{g} = 2\boldsymbol{g}^H \boldsymbol{\Delta}^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Delta} \boldsymbol{g}$$

which implies

$$\sigma = \sqrt{2\boldsymbol{g}^H \boldsymbol{\Delta}^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Delta} \boldsymbol{g}}\,. \tag{38}$$

Equations (33), (34), (36) and (38) result in

$$P_{\boldsymbol{X}_i \to \boldsymbol{X}_j} = \mathcal{Q}\left(\frac{\lambda}{\sigma}\right) = \mathcal{Q}\left(\frac{\boldsymbol{g}^H \boldsymbol{\Delta}^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Delta} \boldsymbol{g}}{\sqrt{2\boldsymbol{g}^H \boldsymbol{\Delta}^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Delta} \boldsymbol{g}}}\right) = \mathcal{Q}\left(\frac{1}{\sqrt{2}} \sqrt{\boldsymbol{g}^H \boldsymbol{\Delta}^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Delta} \boldsymbol{g}}\right). \tag{39}$$

Let $\boldsymbol{L}_{ij} = \boldsymbol{\Delta}^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Delta}$. Thus,

$$\boldsymbol{L}_{ij} = \boldsymbol{\Delta}^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Delta} = \widetilde{\boldsymbol{X}}_i^H \boldsymbol{P}_j^H \boldsymbol{\Upsilon}^{-1} \boldsymbol{P}_j \widetilde{\boldsymbol{X}}_i. \tag{40}$$

Hence, due to (37) and (40) it holds

$$\boldsymbol{L}_{ij} = \widetilde{\boldsymbol{X}}_i^H \boldsymbol{\Upsilon}^{-\frac{1}{2}} \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\Upsilon}^{-\frac{1}{2}} \widetilde{\boldsymbol{X}}_i = \boldsymbol{\mathcal{X}}_i^H \boldsymbol{\Pi}_j^{\perp} \boldsymbol{\mathcal{X}}_i. \tag{41}$$

Equations (39), (40) and (41) result in (8). This completes the proof.

APPENDIX II

OPTIMIZATION PROBLEM

In this section, we prove that the equivalent formulation of the optimization problem (24) is given by (25).

The optimization problem in (24) can be rewritten in the following way

$$(\boldsymbol{X}_k^*, t^*) = \quad \arg\max \quad t \tag{42}$$
$$\lambda_{\min}(\boldsymbol{L}_{mk}) \geq t, m = 1, ..., k-1 \quad (A)$$
$$\lambda_{\min}(\boldsymbol{L}_{km}) \geq t, m = 1, ..., k-1 \quad (B)$$
$$\text{tr}(\boldsymbol{X}_k^H \boldsymbol{X}_k) = 1,$$

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 $\widetilde{\boldsymbol{X}_i} = \boldsymbol{I}_N \otimes \boldsymbol{X}_i$.

Approach: Define $\boldsymbol{\mathfrak{X}}_k = \text{vec}(\boldsymbol{X}_k)\text{vec}^H(\boldsymbol{X}_k)$. We are going to show that both (A) and (B) can be written as linear matrix inequalities (LMI's) with respect to $\boldsymbol{\mathfrak{X}}_k$, $\text{vec}(\boldsymbol{X}_k)$ and t.

(A) Note that

$$\lambda_{\min}(\boldsymbol{L}_{mk}) \geq t \Leftrightarrow \boldsymbol{L}_{mk} - t\boldsymbol{I}_{MN} \succeq \boldsymbol{0}.$$

Since the matrix $\boldsymbol{L}_{mk} - t\boldsymbol{I}_{MN} = \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_m - \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_k (\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k)^{-1} \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_m - t\boldsymbol{I}_{MN}$ is the Schur complement [34] 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):

$$\lambda_{\min}(\boldsymbol{L}_{mk}) \geq t \Leftrightarrow \begin{bmatrix} \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k & \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_m \\ \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_k & \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_m - t\boldsymbol{I}_{MN} \end{bmatrix} \succeq \boldsymbol{0}. \tag{43}$$

• 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,

$$[\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k]_{ij} \;=\; \boldsymbol{e}_i^H \boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k \boldsymbol{e}_j = \boldsymbol{e}_i^H \widetilde{\boldsymbol{X}_k}^H \boldsymbol{\Upsilon}^{-1} \widetilde{\boldsymbol{X}_k} \boldsymbol{e}_j = \text{tr}\left(\boldsymbol{\Upsilon}^{-1} \widetilde{\boldsymbol{X}_k} \boldsymbol{e}_j \left(\widetilde{\boldsymbol{X}_k} \boldsymbol{e}_i\right)^H\right) \tag{44}$$

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 $\text{vec}(\widetilde{\boldsymbol{X}_k}) = \boldsymbol{K}\,\text{vec}(\boldsymbol{X}_k)$, see [36]. Hence,

$$\widetilde{\boldsymbol{X}_k} \boldsymbol{e}_j = \text{vec}\left(\widetilde{\boldsymbol{X}_k} \boldsymbol{e}_j\right) = \left(\boldsymbol{e}_j^T \otimes \boldsymbol{I}_{TN}\right) \text{vec}(\widetilde{\boldsymbol{X}_k}) = (\boldsymbol{e}_j^T \otimes \boldsymbol{I}_{TN}) \boldsymbol{K} \text{vec}(\boldsymbol{X}_k). \tag{45}$$

Substituting (45) in (44) we have

$$\begin{aligned} [\boldsymbol{\mathcal{X}}_k^H \boldsymbol{\mathcal{X}}_k]_{ij} &= \text{tr}\left(\boldsymbol{\Upsilon}^{-1} (\boldsymbol{e}_j^T \otimes \boldsymbol{I}_{TN}) \boldsymbol{K} \text{vec}(\boldsymbol{X}_k) \left((\boldsymbol{e}_i^T \otimes \boldsymbol{I}_{TN}) \boldsymbol{K} \text{vec}(\boldsymbol{X}_k)\right)^H\right) \\ &= \text{tr}\left(\boldsymbol{B}_{ij}\left(\boldsymbol{I}_{TN}\right) \boldsymbol{\mathfrak{X}}_k\right), \end{aligned} \tag{46}$$

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,

$$[\boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_k]_{ij} = \boldsymbol{e}_i^H \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\mathcal{X}}_k \boldsymbol{e}_j = \boldsymbol{e}_i^H \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\Upsilon}^{-\frac{1}{2}} \widetilde{\boldsymbol{X}_k} \boldsymbol{e}_j = \boldsymbol{e}_i^H \boldsymbol{\mathcal{X}}_m^H \boldsymbol{\Upsilon}^{-\frac{1}{2}} (\boldsymbol{e}_j^T \otimes \boldsymbol{I}_{TN}) \boldsymbol{K}\, \text{vec}(\boldsymbol{X}_k). \tag{47}$$

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

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

Using (45) we obtain

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

Hence,

$$[\boldsymbol{L}_{km}]_{ij} \;=\; \text{tr}\left(\boldsymbol{B}_{ij}(\boldsymbol{\Pi}_m^{\perp})\,\boldsymbol{\mathfrak{X}}_k\right). \tag{48}$$

Combining (43), (44), (46), (47) and (48) we conclude that both (A) and (B) can be written as LMI's with respect to the variables $\boldsymbol{\mathfrak{X}}_k$, vec$(\boldsymbol{X}_k)$ and t. Consequently, the optimization problems (24) and (25) are equivalent. This concludes the proof.

APPENDIX III
CALCULATING GRADIENTS

In this section, we calculate gradient to be used in (30). Although the function f_{ij} assumes complex valued entries, that is

$$f_{ij}\;:\;\underbrace{\mathbb{C}^{T\times M}\times ... \times \mathbb{C}^{T\times M}}_{K} \to \mathbb{R} \qquad f_{ij}(\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_K) = \lambda_{\mathsf{min}}(\boldsymbol{L}_{ij})$$

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.

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

Let $\lambda_{\min}$ be a simple eigenvalue of a hermitian matrix $\boldsymbol{L}_{ij}$, and let u_0 be an associated eigenvector, so that $\boldsymbol{L}_{ij}u_0 = \lambda_{\min}u_0$. Hence, $f_{ij} = u_0^H\boldsymbol{L}_{ij}u_0$ and differential df_{ij} is given by, [36]

$$df_{ij} = u_0^H d\boldsymbol{L}_{ij}u_0.$$

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

$$\begin{aligned} df_{ij} = u_0^H d\boldsymbol{L}_{ij}u_0 \;&=\; u_0^H[(d\widetilde{\boldsymbol{X}}_i)^H\boldsymbol{K}_j\widetilde{\boldsymbol{X}}_i + \widetilde{\boldsymbol{X}}_i^H\boldsymbol{K}_j d\widetilde{\boldsymbol{X}}_i + \widetilde{\boldsymbol{X}}_i^H d\boldsymbol{K}_j\widetilde{\boldsymbol{X}}_i]u_0 \\ &=\; \Re[\text{tr}[(d\widetilde{\boldsymbol{X}}_i)^H\underbrace{2\boldsymbol{K}_j\widetilde{\boldsymbol{X}}_i u_0 u_0^H}_{\boldsymbol{C}_i}]] + u_0^H\widetilde{\boldsymbol{X}}_i^H d\boldsymbol{K}_j\widetilde{\boldsymbol{X}}_i u_0. \end{aligned} \tag{49}$$

Continuing with analysis,

$$d\boldsymbol{K}_j = \underbrace{-\boldsymbol{\Upsilon}^{-\frac{1}{2}}\left(d\boldsymbol{\mathcal{X}}_j(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j)^{-1}\boldsymbol{\mathcal{X}}_j^H + \boldsymbol{\mathcal{X}}_j(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j)^{-1}(d\boldsymbol{\mathcal{X}}_j)^H\right)\boldsymbol{\Upsilon}^{-\frac{1}{2}}}_{\boldsymbol{K}_{j1}} - \underbrace{-\boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{\mathcal{X}}_j d\left(\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}\right)\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\Upsilon}^{-\frac{1}{2}}}_{\boldsymbol{K}_{j2}}. \quad (50)$$

Using the equality $d\left(\boldsymbol{A}^{-1}\right) = -\boldsymbol{A}^{-1}d\boldsymbol{A}\boldsymbol{A}^{-1}$ [36], we can write

$$d\left(\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}\right) = -\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}\left((d\boldsymbol{\mathcal{X}}_j)^H\boldsymbol{\mathcal{X}}_j + \boldsymbol{\mathcal{X}}_j^H d\boldsymbol{\mathcal{X}}_j\right)\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}. \quad (51)$$

Substituting (51) and (50) in (49) we get

$$u_0^H\widetilde{\boldsymbol{X}_i}^H d\boldsymbol{K}_j\widetilde{\boldsymbol{X}}_i u_0 = u_0^H\widetilde{\boldsymbol{X}_i}^H\boldsymbol{K}_{j1}\widetilde{\boldsymbol{X}}_i u_0 + u_0^H\widetilde{\boldsymbol{X}_i}^H\boldsymbol{K}_{j2}\widetilde{\boldsymbol{X}}_i u_0$$

with

$$\begin{aligned} u_0^H\widetilde{\boldsymbol{X}_i}^H\boldsymbol{K}_{j1}\widetilde{\boldsymbol{X}}_i u_0 &= -2\Re[u_0^H\boldsymbol{\mathcal{X}}_i^H\boldsymbol{\mathcal{X}}_j\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}(d\boldsymbol{\mathcal{X}}_j)^H\boldsymbol{\mathcal{X}}_i u_0] \\ &= \Re[\mathrm{tr}[\left(d\widetilde{\boldsymbol{X}}_j\right)^H\underbrace{-2\boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{\mathcal{X}}_i u_0 u_0^H\boldsymbol{\mathcal{X}}_i^H\boldsymbol{\mathcal{X}}_j\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}}_{\boldsymbol{C}_{j1}}]] \end{aligned}$$

and

$$\begin{aligned} u_0^H\widetilde{\boldsymbol{X}_i}^H\boldsymbol{K}_{j2}\widetilde{\boldsymbol{X}}_i u_0 &= 2\Re[u_0^H\boldsymbol{\mathcal{X}}_i^H\boldsymbol{\mathcal{X}}_j\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}(d\boldsymbol{\mathcal{X}}_j)^H\boldsymbol{\mathcal{X}}_j\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_i u_0] \\ &= \Re[\mathrm{tr}[\left(d\widetilde{\boldsymbol{X}}_j\right)^H\underbrace{2\boldsymbol{\Upsilon}^{-\frac{1}{2}}\boldsymbol{\mathcal{X}}_j\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_i u_0 u_0^H\boldsymbol{\mathcal{X}}_i^H\boldsymbol{\mathcal{X}}_j\left(\boldsymbol{\mathcal{X}}_j^H\boldsymbol{\mathcal{X}}_j\right)^{-1}}_{\boldsymbol{C}_{j2}}]] \end{aligned}$$

Define $\boldsymbol{C}_j = \boldsymbol{C}_{j1} + \boldsymbol{C}_{j2}$. Thus,

$$df_{ij} = \Re[\mathrm{tr}[(d\widetilde{\boldsymbol{X}}_i)^H\boldsymbol{C}_i]] + \Re[\mathrm{tr}[(d\widetilde{\boldsymbol{X}}_j)^H\boldsymbol{C}_j]].$$

Note that $d\widetilde{\boldsymbol{X}}_i = \boldsymbol{I}_N \otimes d\boldsymbol{X}_i$, then

$$df_{ij} = \Re[\mathrm{tr}[(d\boldsymbol{X}_i)^H\overline{\boldsymbol{C}_i}]] + \Re[\mathrm{tr}[(d\boldsymbol{X}_j)^H\overline{\boldsymbol{C}_j}]]$$

where $\overline{\boldsymbol{C}_i} = \sum_{k=1}^N \boldsymbol{C}_{ik}$ where $\boldsymbol{C}_{ik}$ is a diagonal block of the matrix $\boldsymbol{C}_i$ of size $T \times M$., i.e.,

$$\boldsymbol{C}_i = \begin{bmatrix} \boldsymbol{C}_{i1} & & & \\ & \boldsymbol{C}_{i2} & & \\ & & \ddots & \\ & & & \boldsymbol{C}_{iN} \end{bmatrix}.$$

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

$$\nabla f_{ij}(\boldsymbol{x}) = \begin{bmatrix} \mathbf{0}_{(i-1)c\times 1} \\ \Re\text{vec}(\overline{\boldsymbol{C}_i}) \\ \Im\text{vec}(\overline{\boldsymbol{C}_i}) \\ \mathbf{0}_{(j-i-1)c\times 1} \\ \Re\text{vec}(\overline{\boldsymbol{C}_j}) \\ \Im\text{vec}(\overline{\boldsymbol{C}_j}) \\ \mathbf{0}_{(K-j)c\times 1} \end{bmatrix}_{2KTM\times 1}$$

for $1 \leq i \neq j \leq K$ and $c = 2TM$, where $\boldsymbol{x} = \begin{bmatrix} \Re\,\text{vec}(\boldsymbol{X}_1) \\ \Im\,\text{vec}(\boldsymbol{X}_1) \\ \vdots \\ \Re\,\text{vec}(\boldsymbol{X}_K) \\ \Im\,\text{vec}(\boldsymbol{X}_K) \end{bmatrix}$.

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] H. L. Van Trees, *Detection, Estimation, and Modulation Theory.*, Part I New York: Wiley, 1968.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing.*, Vol. II – Detection Theory, Prentice Hall, 1998.

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

[24] 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.

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

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

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

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

[29] I. Polik, "Addendum to the Sedumi user guide Version 1.1," http://sedumi.mcmaster.ca

[30] D. Peaucelle, "Users's Guide for SeDuMi Interface 1.04," http://www.laas.fr/peaucell/SeDuMiInt.html

[31] Y. Labit, D. Peaucelle and D. Henrion, "SeDuMi Interface 1.02 : a Tool for Solving LMI Problems with SeDuMi," *Proceedings of the CACSD Conference*, Glasgow, September 2002.

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

[33] 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.

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

[35] W. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry.* New York: Academic Press.

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

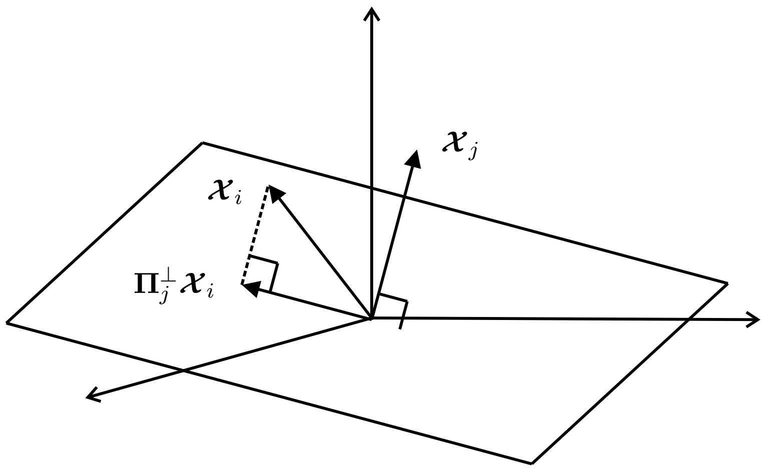

Fig. 1. Geometrical interpretation of $\mathbf{\Pi}_j^{\perp}\boldsymbol{\mathcal{X}}_i$.

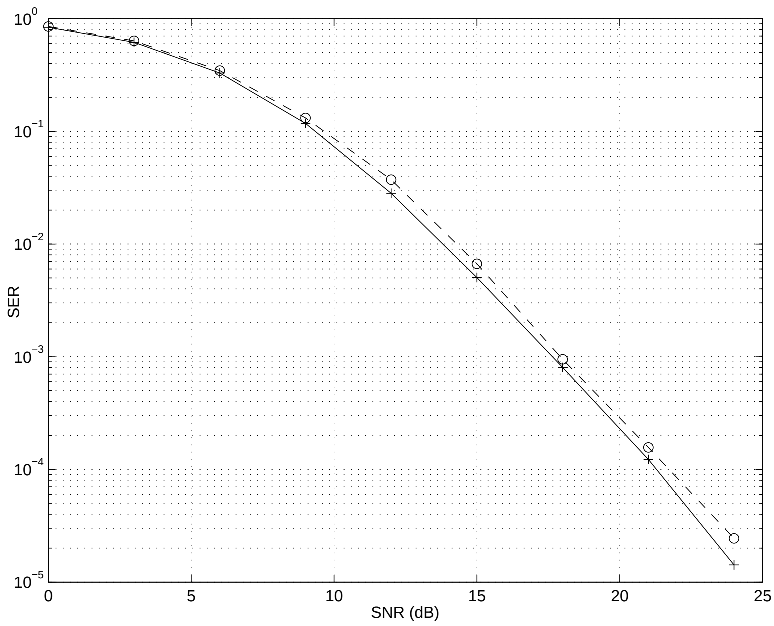

Fig. 2. Category 1 - spatio-temporally white observation noise: T=8, M=3, N=1, K=256, $\mathbf{\Upsilon}$= $\boldsymbol{I}_{NT}$. Plus-solid curve:our codes; circle-dashed curve:unitary codes.

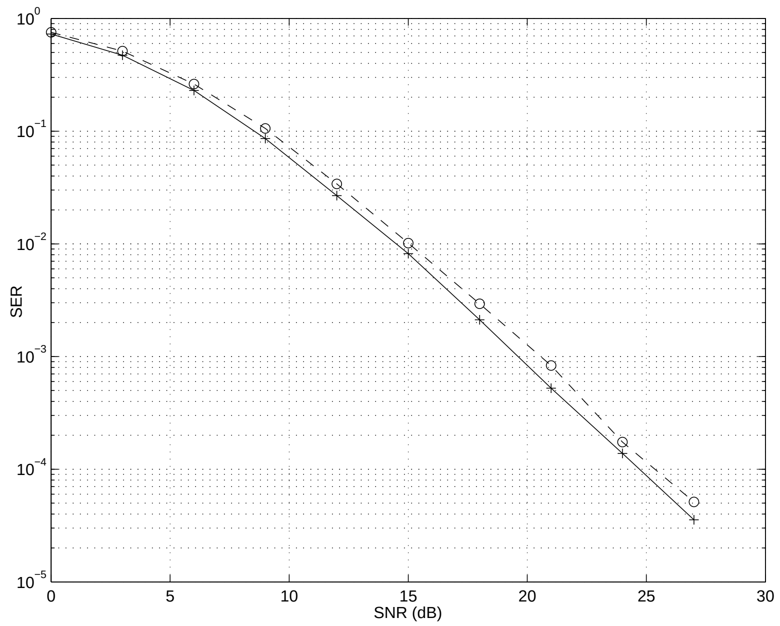

Fig. 3. Category 1 - spatio-temporally white observation noise: T=8, M=2, N=1, K=256, $\boldsymbol{\Upsilon}$= $\boldsymbol{I}_{NT}$. Plus-solid curve: our codes; circle-dashed curve: unitary codes.

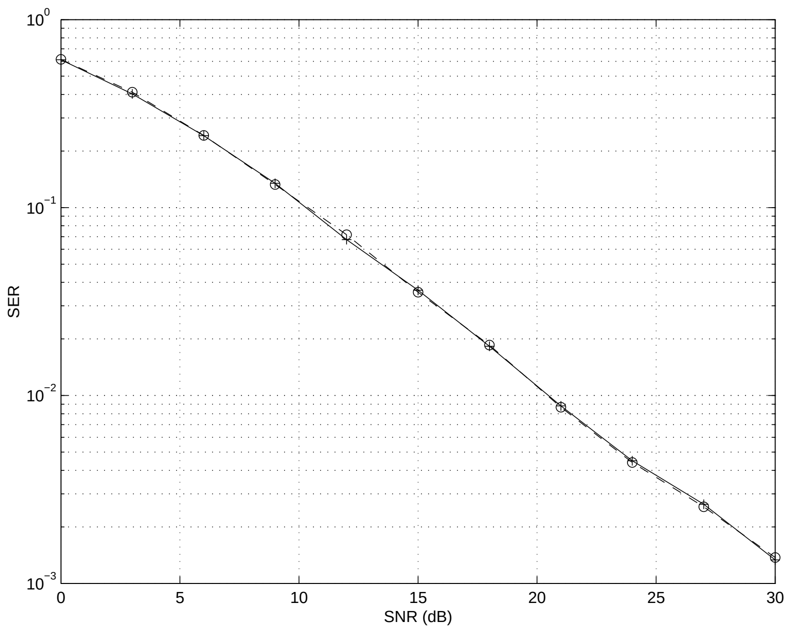

Fig. 4. Category 1 - spatio-temporally white observation noise: T=8, M=1, N=1, K=256, $\boldsymbol{\Upsilon}$= $\boldsymbol{I}_{NT}$. Plus-solid curve: our codes; circle-dashed curve: unitary codes.

		PACKING RADII (DEGREES)		
T	K	MB	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. 5. 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).

		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.

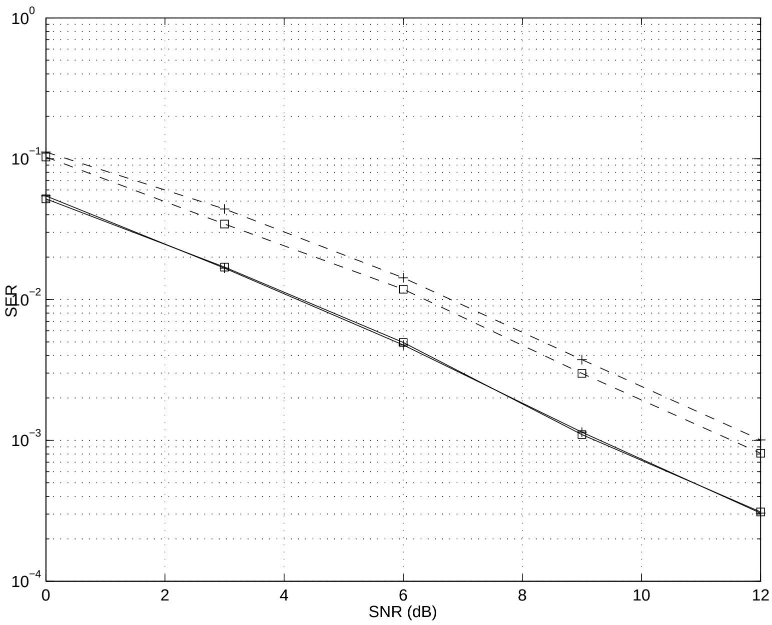

Fig. 6. Category 2 - spatially white - temporally colored: T=8, M=2, $N = 1$, K=67, $\boldsymbol{\rho}$=[1; 0.85; 0.6; 0.35; 0.1; zeros(3,1)]. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

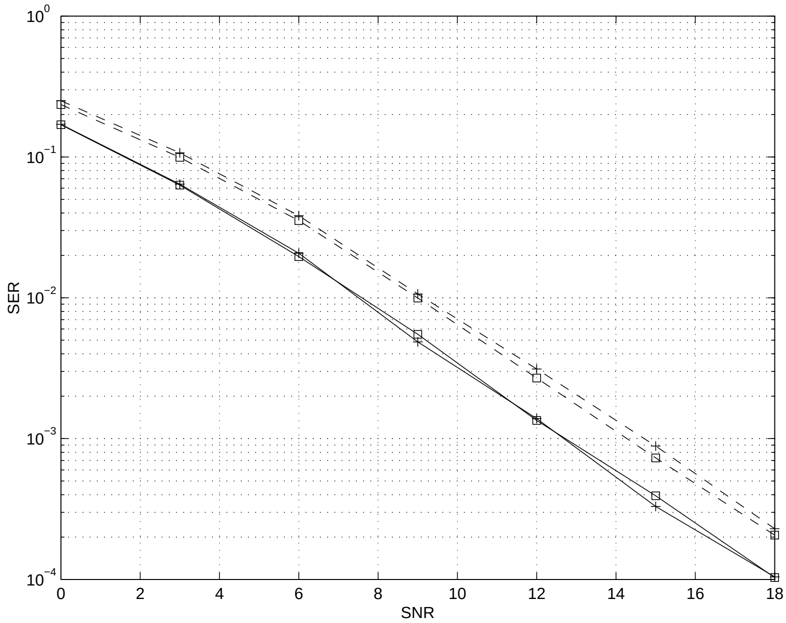

Fig. 7. Category 2 - spatially white - temporally colored: T=8, M=2, $N = 1$, K=256, $\boldsymbol{\rho}$=[1; 0.8; 0.5; 0.15; zeros(4,1)]. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

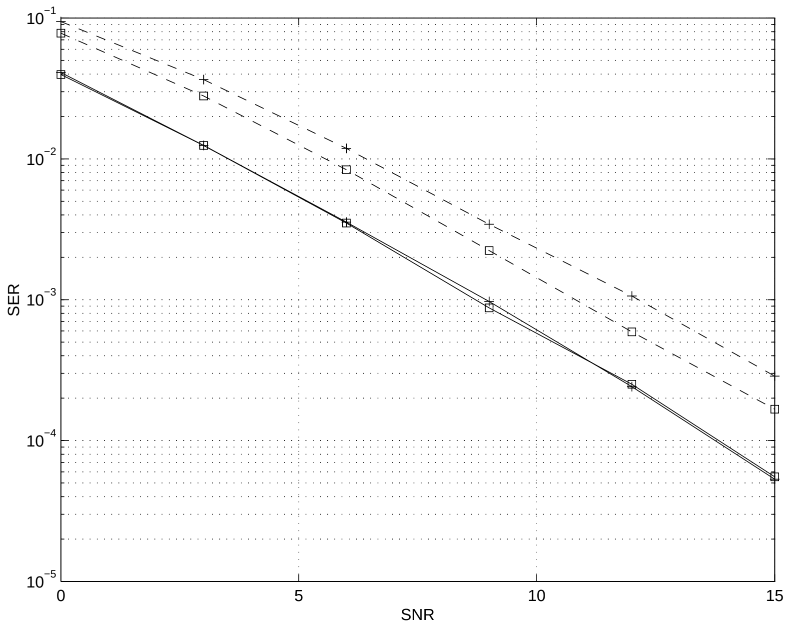

Fig. 8. Category 2 - spatially white - temporally colored: T=8, M=2, $N = 1$, K=32, $\boldsymbol{\rho}$=[1; 0.8; 0.5; 0.15; zeros(4,1)]. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

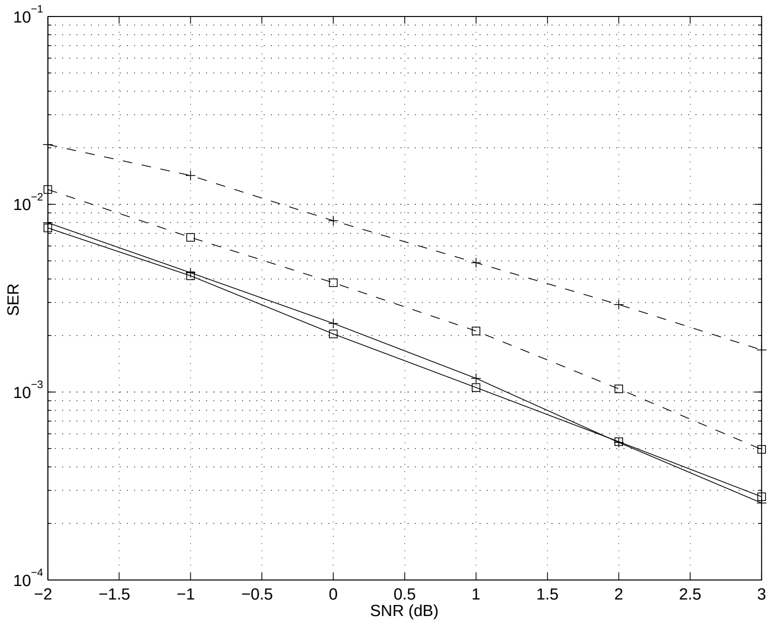

Fig. 9. Category 3: T=8, M=2, $N = 2$, K=32. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.

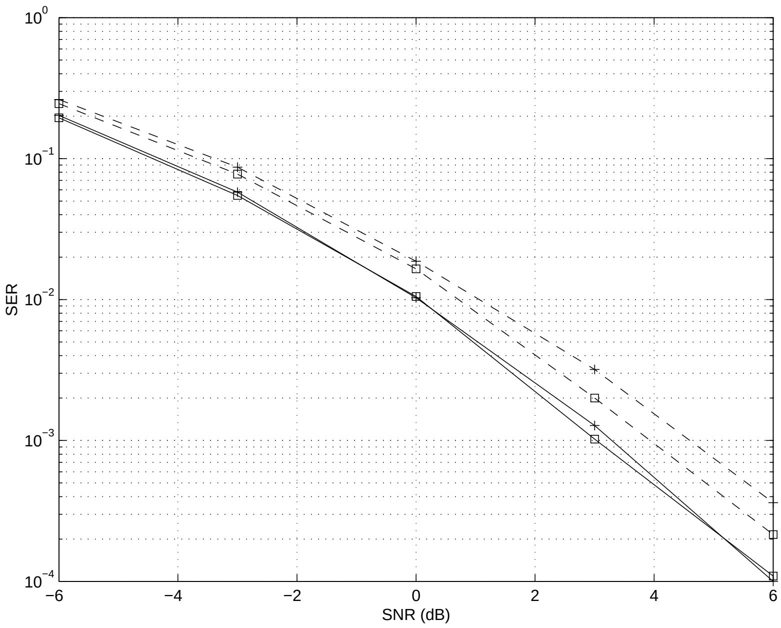

Fig. 10. Category 3: T=8, M=2, $N = 2$, K=67. Solid curves: our codes; dashed curves: unitary codes; plus signed curves: GLRT receiver; square signed curves: Bayesian receiver.