

Constrained Riemannian Noncoherent Constellations for the MIMO Multiple Access Channel

Javier Álvarez-Vizoso[✉], Diego Cuevas[✉], Carlos Beltrán[✉], Ignacio Santamaria[✉], *Senior Member, IEEE*, Vít Tuček[✉], and Gunnar Peters

Abstract—We consider the design of multiuser constellations for a multiple access channel (MAC) with K users, with M antennas each, that transmit simultaneously to a receiver equipped with N antennas through a Rayleigh block-fading channel when no channel state information (CSI) is available to either the transmitter or the receiver. In full-diversity scenarios where the coherence time is at least $T \geq (K + 1)M$, the proposed constellation design criterion is based on the asymptotic expression of the multiuser pairwise error probability (PEP) derived by Brehler and Varanasi (2001). In non-full diversity scenarios, for which the previous PEP expression is no longer valid, the proposed design criteria are based on proxies of the PEP recently proposed by Ngo and Yang (2021). Although both the PEP expression and its bounds or proxies were previously considered intractable for optimization, in this work we derive their respective unconstrained gradients. These gradients are in turn used in the optimization of the proposed cost functions in different Riemannian manifolds representing different power constraints. In particular, in addition to the standard unitary space-time modulation (USTM) leading to optimization on the Grassmann manifold, we consider a more relaxed per-codeword power constraint leading to optimization on the so-called *oblique manifold*, and an average power constraint leading to optimization on the so-called *trace manifold*. Equipped with these theoretical tools, we design multiuser constellations for the MIMO MAC in full-diversity and non-full-diversity scenarios with state-of-the-art performance in terms of symbol error rate (SER).

Index Terms—Noncoherent communications, multiple-input multiple-output (MIMO) communications, multiple access channel (MAC), manifold optimization, pairwise error probability (PEP), union bound (UB).

I. INTRODUCTION

IN multiple-input multiple-output (MIMO) noncoherent wireless communications over fast fading channels, the channel state information (CSI) is assumed to be unknown at both the transmitter and receiver. It is usual to consider in the study of noncoherent communications a block-fading model in which the MIMO channel matrix with M transmit and N receive antennas remains constant during a T -symbol coherence interval, after which it changes to a new independent

realization for another T symbols. In the single-user case and under additive Gaussian noise, it was proved by Hochwald and Marzetta [1], [2] that the $T \times M$ space-time transmit matrices that achieve the ergodic noncoherent capacity for the MIMO block-fading model can be factored as the product of an isotropically distributed $T \times M$ truncated unitary matrix, also called Stiefel matrix, and a diagonal $M \times M$ matrix with real nonnegative entries. Further, when $T \gg M$ the nonzero entries of the diagonal matrix take the same value, showing that in this regime it is optimal to transmit unitary space-time codewords $\mathbf{X}^H \mathbf{X} = \mathbf{I}_M$. Using the same signal model, Zheng and Tse [3] proved that at high signal-to-noise-ratio (SNR) and when $T \geq 2M$, ergodic capacity can be achieved by transmitting isotropically distributed unitary matrices. Motivated by these information-theoretic results, numerous methods for the design of single-user constellations formed by truncated unitary signal matrices, called unitary space-time modulations (USTM), have been investigated and proposed over the last decades [4]–[14]. In MIMO noncoherent constellations, information is carried by the column span (i.e., a subspace) of the transmitted $T \times M$ matrix, $\mathbf{X}$. The problem of designing single-user noncoherent codebooks is thus closely related to finding optimal packings in Grassmann manifolds [3], [15], and the resulting constellations are referred to as Grassmannian constellations.

In the multiuser case, the design of noncoherent constellations is significantly more complex, as many of the theoretical results that exist for the single-user case (as well as the insights gained from them), such as the optimality of unitary space-time or Grassmannian constellations at high-SNR, are no longer true. In this work, we consider the design of noncoherent constellations for the MIMO multiple access channel (MAC), a problem for which there is no satisfactory solution yet. In the MAC, several users transmit information simultaneously over the same bandwidth and at the same channel use or time slot to a common receiver. A common example is the uplink channel in broadband cellular communications, where several users communicate with a base station (BS). In the case of coherent communications with perfect channel state information (CSI) at the receiver, capacity results for the MIMO MAC can be found in [16]. For instance, it is well-known that for the 2-user MAC the capacity region is a pentagon, and the Pareto optimal achievable rate pairs (R_1, R_2) at the corner points of the pentagon are reached by successive cancellation.

For noncoherent communications, however, the full capacity

J. Álvarez Vizoso, D. Cuevas, and I. Santamaria are with the Department of Communications Engineering, Universidad de Cantabria, 39005 Santander, Spain (e-mail: javier.alvarezvizoso@unican.es; diego.cuevas@unican.es; i.santamaria@unican.es).

C. Beltrán is with the Department of Mathematics, Statistics and Computing, Universidad de Cantabria, 39005 Santander, Spain (e-mail: carlos.beltran@unican.es).

V. Tuček and G. Peters are with the Department of Wireless Algorithms, Huawei Technologies, 16440 Kista, Sweden (email: vit.tucek@huawei.com; gunnar.peters@huawei.com)

region of the MIMO-MAC is unknown. For the K -user single-input multiple-output (SIMO) MAC, it was conjectured by Shamai and Marzetta in [17] that for block-fading channels with coherence time $T > 1$ the sum capacity can be achieved by no more than $K = T$ users, which is supported by asymptotic analysis and simulation results. For the two-user MIMO MAC an achievable DoF (degrees of freedom) region has been proposed in [18]. The optimal DoF region for a two-user SIMO MAC has been derived in [19]. Existing theoretical results however do not provide clear insights regarding the structure of the transmit space-time matrices for the MIMO MAC.

For the SIMO case with single-antenna users, energy-based noncoherent constellation designs have been proposed for the uplink channel in [20]–[22], and noncoherent schemes based on a differential phase-shift keying (DPSK) modulation have been recently proposed in [23], [24]. These energy-based or DPSK-based designs, however, cannot be directly extended to the MIMO case. Of particular importance for the K -user MIMO MAC is the work of Brehler and Varanasi in [25], where the authors derived an asymptotic expression of the joint pairwise probability of error (PEP) of the optimal receiver and showed that to ensure full diversity of NM for each user the coherence time must be at least $T \geq (K + 1)M$. However, the PEP expression in [25] was considered to be intractable for optimization, so none of the subsequent studies have used it as a criterion to design multiuser constellations. Most of the proposed criteria in the literature either optimize single-user Grassmannian designs with or without partitioning; that is, using independently designed single-user codebooks, or designing a large single-user codebook that is then partitioned according to some subspace distance measure into K smaller single-user codebooks [25]–[28].

As an alternative to the PEP criterion, in [28], [29] Ngo and Yang recently proposed two PEP proxies that have a very natural geometric meaning in terms of separating joint detection hypothesis. However, the proposed proxies are functions of the eigenvalues of a certain matrix and therefore their optimization is considered challenging in [28]. It is interesting to note at this point that while the exact asymptotic PEP expression in [25] is only valid in uplink channels where the full-diversity condition $T \geq (K + 1)M$ is met, which we will refer to from now on as full-diversity scenarios, the proxies proposed in [28] are valid in non-full diversity scenarios where $T < (K + 1)M$. In fact, as we will show in this work, the PEP cost function of Brehler and Varanasi and the PEP proxies of Ngo and Yang, yield two complementary designs that can be applied, respectively, to full-diversity and non-full diversity scenarios. In addition, as also shown in this work, both cost functions can be optimized on different manifolds representing the different power constraints typically employed in the K -user MIMO MAC.

The main contributions of the paper are the following:

- 1) We have developed Riemannian optimization techniques for designing multiuser noncoherent codebooks for the MIMO MAC in manifolds other than the complex Grassmannian. These manifolds correspond to alternative power normalizations to the one used in unitary space-

time modulations (USTM), which need not be optimal for noncoherent multiuser communications. In particular, in addition to the standard Grassmann manifold, we have considered the complex oblique manifold and the trace manifold, another type of oblique manifold, (Sec. II-B), resulting from a per-codeword power constraint and an average power constraint, respectively.

- 2) We have obtained, for the first time in the literature, closed-form formulas for the gradients of cost functions previously proposed for the design of multiuser noncoherent constellations for the MAC, but whose optimization was so far considered to be intractable. These functions are union bounds, i.e., sums over all joint codewords of the dominant factor (when one user is in error) of the asymptotic PEP derived by Brehler and Varanasi [25], and the b and δ functions (bounds of the PEP) proposed by Ngo and Yang [28].
- 3) Using these gradient expressions, we have developed Riemannian techniques over different manifolds to optimize the b and δ functions proposed in [28] for non-full diversity scenarios, as well as to optimize the exact PEP asymptotic expression derived by Brehler and Varanasi [25] for full-diversity scenarios.
- 4) Our results show that the best performing designs in the non-full diversity case, i.e. $T < (K + 1)M$, are those obtained using the δ cost function on the trace manifold. Whereas in the full-diversity case, i.e. $T \geq (K + 1)M$, the best performing designs are those obtained with the union bound of the asymptotic PEP on the Grassmann manifold.

The remainder of this paper is organized as follows. Sec. II-A introduces the system model for the K -user MIMO MAC and the optimal multiuser maximum likelihood detector. Sec. II-B describes the different Riemannian manifolds considered for codebook optimization, along with their projection and retraction steps. In Sec. III-A the asymptotic joint PEP is reviewed in order to introduce the optimization cost function for full-diversity designs and its Riemannian gradient is computed in Sec. III-B. In Sec. IV-A the optimization cost functions for non-full diversity designs are presented, and their gradients are obtained in Sec. IV-B. Sec. V-B presents and analyzes the simulation results for the noncoherent multiuser constellation designs in non-full diversity scenarios, while Sec. V-C discusses the results corresponding to the designs in full-diversity scenarios. The rest of Sec. V studies the initialization and complexity analysis of our methods, the comparison with pilot-based schemes and the performance under correlated channels. Finally, we present our conclusions in Sec. VI. The paper also includes an Appendix A for mathematical background on Riemannian manifolds.

Notation: In this paper, matrices are denoted by bold-faced upper case letters, column vectors are denoted by bold-faced lower case letters, and scalars are denoted by light-faced lower case letters. The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian conjugate, respectively. The trace and determinant of a matrix $\mathbf{A}$ will be denoted, respectively, as $\text{tr}(\mathbf{A})$ and $\det(\mathbf{A})$. We denote by $\text{diag}(\mathbf{a})$ a diagonal matrix

whose diagonal is $\mathbf{a}$, and $\mathbf{I}_n$ denotes the identity matrix of size n . We write $\mathcal{CN}(0, 1)$ to denote a complex proper Gaussian distribution with zero mean and unit variance, $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ denotes a complex Gaussian vector in $\mathbb{C}^n$ with zero mean and covariance matrix $\mathbf{R}$. The complex Grassmann manifold of M -dimensional subspaces of the T -dimensional complex vector space $\mathbb{C}^T$ is denoted by $\mathbb{G}(M, \mathbb{C}^T)$. The complex Stiefel manifold of unitary M -frames in $\mathbb{C}^T$ is written as $\text{St}(M, \mathbb{C}^T)$. We denote by $\mathbb{E}[\cdot]$ the mathematical expectation. Unless stated otherwise $\log$ refers to the natural logarithm. Some background material about the Stiefel and Grassmann manifolds, which is needed for the paper, is relegated to the Appendix. Additional notation is introduced as needed in the text.

II. MANIFOLD OPTIMIZATION FOR THE MIMO MAC

A. System Model

We consider a noncoherent MIMO MAC with K transmitters, or users, simultaneously transmitting to a common receiver or base station. To keep the notation simple, we assume that all users have the same number of transmit antennas M , (the extension to a different number of antennas per user is straightforward), and the receiver has N antennas. The channel of user k is $\mathbf{H}_k \in \mathbb{C}^{M \times N}$ following a Rayleigh fading distribution ($\mathbf{H}_k(i, j) \sim \mathcal{CN}(0, 1)$) and is assumed to remain constant over T symbol periods, over which communication occurs. In the next transmission block, the channels of all users change to an independent realization (block-fading channel). User k transmits at rate r_k (bits/channel use), so within a coherence block sends a matrix chosen equiprobably from a codebook $\mathcal{C}_k = \{\mathbf{X}_{k,1}, \dots, \mathbf{X}_{k,L_k}\}$ with $L_k = 2^{r_k T}$. Unlike the single-user case, for the MAC the transmitted matrices do not have to be necessarily semi-unitary or Stiefel ($\mathbf{X}_{k,i}^H \mathbf{X}_{k,i} = \mathbf{I}_M$).

Let us consider for notational simplicity the two-user MIMO MAC. We adhere to the notation in [25], [30] and define the $T \times 2M$ matrix of transmitted codewords $\mathbf{F}_i = [\mathbf{X}_{1,i_1}, \mathbf{X}_{2,i_2}]$, where i_1 and i_2 index the symbol chosen by user 1 and 2 respectively in the joint codeword, or hypothesis, $\mathbf{F}_i$. Note that even if $\mathbf{X}_{k,i_1} \in \mathbb{G}(M, \mathbb{C}^T)$ for $k = 1, 2$, the multiuser codeword $\mathbf{F}_i$ is not a Stiefel matrix anymore ($\mathbf{F}_i^H \mathbf{F}_i \neq \mathbf{I}_{2M}$). The set of multiuser codewords is

$$\mathcal{F} = \{\mathbf{F}_i = [\mathbf{X}_{1,i_1}, \mathbf{X}_{2,i_2}], \mathbf{X}_{1,i_1} \in \mathcal{C}_1, \mathbf{X}_{2,i_2} \in \mathcal{C}_2\},$$

and has cardinality $|\mathcal{F}| = |\mathcal{C}_1 \times \mathcal{C}_2| = L_1 L_2$.

Each user may have a different signal-to-noise-ratio (SNR) due to the different transmit powers and different path losses. Let $P_k = \frac{1}{TL_k} \sum_{i=1}^{L_k} \text{tr}(\mathbf{X}_{k,i}^H \mathbf{X}_{k,i})$ be the average transmit power of user k . We assume that $P_k \leq P$, where P is the maximum transmit power of all users. When the multiuser codeword $\mathbf{F}_i = [\mathbf{X}_{1,i_1}, \mathbf{X}_{2,i_2}]$ is transmitted, the signal received at the BS is

$$\mathbf{Y} = \mathbf{X}_{1,i_1} \mathbf{H}_1 + \mathbf{X}_{2,i_2} \mathbf{H}_2 + \mathbf{W}, \quad (1)$$

where the matrix $\mathbf{W} \in \mathbb{C}^{T \times N}$ represents the additive Gaussian noise, modeled as $w_{ij} \sim \mathcal{CN}(0, 1)$, so $\mathbf{W}$ is a normal matrix with independent columns generated as $\mathbf{w}_n \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_T)$,

$n = 1, \dots, N$. Unless otherwise stated, we will assume that all users have the same SNR, defined in dB as $10 \log_{10} P$.

The conditional distribution of the observations $\mathbf{Y}$ given the transmitted multiuser codeword $\mathbf{F}_i$ is

$$p(\mathbf{Y}|\mathbf{F}_i) = \frac{1}{\pi^{TN} \det(\mathbf{R}_i)^N} \exp \text{tr}(-\mathbf{R}_i^{-1} \mathbf{Y} \mathbf{Y}^H), \quad (2)$$

so each column of $\mathbf{Y}$ follows a zero-mean complex normal distribution with covariance matrix $\mathbf{R}_i = \mathbf{X}_{1,i_1} \mathbf{X}_{1,i_1}^H + \mathbf{X}_{2,i_2} \mathbf{X}_{2,i_2}^H + \mathbf{I}_T$. If the codewords are chosen with equal probability, the optimal Maximum Likelihood (ML) detector for uncorrelated Rayleigh channels is

$$\hat{\mathbf{F}}_i = \arg \min_{\mathbf{F}_i \in \mathcal{F}} \text{tr}(\mathbf{Y}^H \mathbf{R}_i^{-1} \mathbf{Y}) + N \log \det(\mathbf{R}_i). \quad (3)$$

Notice that the ML detector at the BS needs to know the SNR of all users. The SNR depends primarily on the path loss and the transmit power and therefore varies on a much slower temporal scale than the multipath fading. It is therefore feasible to have this long-term CSI available at the BS.

When correlation exists either at the transmitters or at the receiver, and the correlation matrices are known, it is necessary to modify the optimal detector as follows. First of all, it is convenient to rewrite the received signal model in (1) for more tractable analysis as follows

$$\mathbf{y} = (\mathbf{I}_N \otimes \mathbf{X}_{1,i_1}) \mathbf{h}_1 + (\mathbf{I}_N \otimes \mathbf{X}_{2,i_2}) \mathbf{h}_2 + \mathbf{w}, \quad (4)$$

where $\mathbf{y} = \text{vec}(\mathbf{Y}) \in \mathbb{C}^{NT \times 1}$, $\mathbf{h}_i = \text{vec}(\mathbf{H}_i) \in \mathbb{C}^{NM \times 1}$ for $i = 1, 2$, $\mathbf{w} = \text{vec}(\mathbf{W}) \in \mathbb{C}^{NT \times 1}$ and $\otimes$ denotes Kronecker product. We assume that $\mathbb{E}[\mathbf{h}_1 \mathbf{h}_1^H] = \mathbb{E}[\mathbf{h}_2 \mathbf{h}_2^H] = \mathbf{\Sigma}$, and that the $MN \times MN$ covariance matrix $\mathbf{\Sigma}$ can be decomposed following the standard Kronecker product model so that $\mathbf{\Sigma} = \mathbf{\Sigma}_{tx} \otimes \mathbf{\Sigma}_{rx}$, where $\mathbf{\Sigma}_{tx}$ is the $M \times M$ transmit correlation matrix and $\mathbf{\Sigma}_{rx}$ is the $N \times N$ receive correlation matrix. Conditioned on the transmit signals, the NT receive signal $\mathbf{y}$ follows a complex normal distribution with zero mean and covariance matrix

$$\begin{aligned} \bar{\mathbf{R}}_i &= (\mathbf{I}_N \otimes \mathbf{X}_{1,i_1}) (\mathbf{\Sigma}_{tx} \otimes \mathbf{\Sigma}_{rx}) (\mathbf{I}_N \otimes \mathbf{X}_{1,i_1})^H \\ &+ (\mathbf{I}_N \otimes \mathbf{X}_{2,i_2}) (\mathbf{\Sigma}_{tx} \otimes \mathbf{\Sigma}_{rx}) (\mathbf{I}_N \otimes \mathbf{X}_{2,i_2})^H + \mathbf{I}_{NT}, \end{aligned} \quad (5)$$

and therefore the optimal Maximum Likelihood (ML) detector for correlated Rayleigh channels is

$$\hat{\mathbf{F}}_i = \arg \min_{\mathbf{F}_i \in \mathcal{F}} \mathbf{y}^H \bar{\mathbf{R}}_i^{-1} \mathbf{y} + \log \det(\bar{\mathbf{R}}_i). \quad (6)$$

Clearly, when $\mathbf{\Sigma} = \mathbf{I}_{NM}$, the ML detector (6) boils down to (3). The model generalizes to K users in a straightforward manner.

B. Riemannian Manifolds for Noncoherent Multiuser Constellation Designs

For a review of the manifold geometry needed in our optimization methods we refer the reader to Appendix A.

Under the usual USTM assumption (Grassmannian constellations) used in most previous works, the transmitted codewords are normalized as:

$$\mathbf{X}_{k,i}^H \mathbf{X}_{k,i} = \frac{P_k T}{M} \mathbf{I}_M, \text{ so that } \|\mathbf{X}_{k,i}\|_F^2 = P_k T, \forall \mathbf{X}_{k,i} \in \mathcal{C}_k. \quad (7)$$

Under this constraint the codewords are represented by (scaled) Stiefel matrices up to unitary transformations, so that optimization must be performed on the Grassmannian manifold $\mathbb{G}(M, \mathbb{C}^T)$. The constraint (7) means that the signals transmitted by the different antennas are orthogonal to each other, all of them with equal power.

A more relaxed constraint is to require the total per-codeword transmit power to be normalized but without requiring that the signals transmitted by different antennas be orthogonal; that is, without requiring each codeword to be a Stiefel matrix. Let us recall that the use of USTM is not necessarily optimal in the MIMO MAC. That is to say, one requires only that

$$\|\mathbf{X}_{k,i}\|_F^2 = P_k T, \quad \forall \mathbf{X}_{k,i} \in \mathcal{C}_k. \quad (8)$$

This realizes the codewords as points in the complex sphere of radius $\sqrt{P_k T}$: take the $T \times M$ matrix $\mathbf{X}_{k,i}$ and flatten it to a vector of length TM of Euclidean norm 1 normalizing it by $1/\sqrt{P_k T}$. Then the users' constellations $\mathcal{C}_1 \times \dots \times \mathcal{C}_K$ correspond to a point in the so-called *oblique manifold*, denoted as $\mathcal{OB}_{\mathbb{C}}(TM, \sum_{k=1}^K L_k)$, which is the product of as many complex spheres as codewords $\mathbb{S}^{TM-1} \times \dots \times \mathbb{S}^{TM-1}$. Indeed, every fixed-norm codeword corresponds to a point in a sphere of dimension TM , and the number of spheres needed is the added cardinality of all constellations $\sum_{k=1}^K L_k$. To optimize constellation points on this oblique manifold one just needs to project unconstrained gradients onto spheres, and to do the retraction, unflatten the vectors to restore $T \times M$ matrices, and renormalize each point from 1 to $\sqrt{M}$.

Finally, the least stringent power constraint normalizes the average transmit power of each user k :

$$\frac{1}{L_k} \sum_{i=1}^{L_k} \text{tr}(\mathbf{X}_{k,i}^H \mathbf{X}_{k,i}) = P_k T, \quad k = 1, \dots, K. \quad (9)$$

In this case, it is the whole constellation of every user that behaves as a point on a sphere of radius $\sqrt{L_k P_k T}$, that is to say, there is a vector that represents the constellation by flattening the concatenated matrices of the L_k codewords of a user. Simplifying this case to the situation where all the users have the same number of codewords, $L_k = L$, the multiuser constellation $\mathcal{C}_1 \times \dots \times \mathcal{C}_K$ corresponds to points in a different oblique manifold, that we call the *trace manifold* denoted as $\mathbb{T}r(K, L, M, \mathbb{C}^T) = \mathcal{OB}_{\mathbb{C}}(TML, K)$, which corresponds to the product of as many complex spheres as users. The three manifolds considered provide solutions to the same problem since they allow to satisfy the most general constraint, Eq. (9), i.e., Grassmannian codes are codes from the oblique manifold, so Eq. (7) satisfies Eq. (8), and these two are in turn constellations on the trace manifold, since they satisfy Eq. (9). However the converse is not true. This is why initializing and optimizing in less constrained manifolds may yield different constellations by reaching (local) minima or maxima of cost functions that are not accessible in the more constrained submanifold.

To optimize a function f on a general manifold $\mathcal{M} \subset \mathbb{C}^T$ we just need to compute the unconstrained complex Euclidean gradient Df and use the corresponding projector

$\mathbf{P}_{\mathbf{X}} : T_{\mathbf{X}}\mathbb{C}^T \rightarrow T_{\mathbf{X}}\mathcal{M}$, along with a reasonable retraction function $\mathbf{R}_{\mathbf{X}} : T_{\mathbf{X}}\mathcal{M} \rightarrow \mathcal{M}$. We summarize here these steps for the manifolds investigated in our present work corresponding to the different power constraints discussed above (we use the notation of standard references on matrix manifold optimization such as [31]).

- **Grassmannian manifold** $\mathbb{G}(M, \mathbb{C}^T)$: a Stiefel (or semi-unitary) matrix per user's codeword.
 - *Projection*: $\mathbf{P}_{\mathbf{X}}(\dot{\mathbf{Z}}) = (\mathbf{I}_T - \mathbf{X}\mathbf{X}^H)\dot{\mathbf{Z}}$.
 - *Retraction*: $\mathbf{R}_{\mathbf{X}}(\dot{\mathbf{Z}})$ = first M columns of the QR decomposition of $\dot{\mathbf{Z}}$.
- **Oblique manifold** $\mathcal{OB}_{\mathbb{C}}(TM, \sum_i L_i)$ (per-codeword power constraint): a complex sphere of radius $\sqrt{P_k T}$ per user's codeword.
 - *Projection*: $\mathbf{P}_{\mathbf{X}}(\dot{\mathbf{Z}}) = \dot{\mathbf{Z}} - \mathbf{X} \cdot \text{ddiag}(\mathbf{X}^H \dot{\mathbf{Z}})$, where $\text{ddiag}(\mathbf{W})$ is the diagonal matrix whose diagonal is that of $\mathbf{W}$.
 - *Retraction*: $\mathbf{R}_{\mathbf{X}}(\dot{\mathbf{Z}}) = (\mathbf{X} + \dot{\mathbf{Z}})[\text{ddiag}((\mathbf{X} + \dot{\mathbf{Z}})^H(\mathbf{X} + \dot{\mathbf{Z}}))]^{-1/2}$ and normalization by scaling from 1 to $\sqrt{P_k T}$.
- **Trace manifold** $\mathbb{T}r(K, L, M, \mathbb{C}^T) = \mathcal{OB}_{\mathbb{C}}(TML, K)$ (average per-user power constraint): a different oblique manifold with one complex sphere per user codebook.
 - *Projection*: $\mathbf{P}_{\mathbf{X}}(\dot{\mathbf{Z}}) = \dot{\mathbf{Z}} - \mathbf{X} \cdot \text{ddiag}(\mathbf{X}^H \dot{\mathbf{Z}})$. A different realization of the constraint that we also employ as it proves to be numerically useful, denoting $\mathbf{X}_C := \sum_{i=1}^{L_k} \mathbf{X}_{k,i}$, is

$$\mathbf{P}_{\mathbf{X}}(\dot{\mathbf{Z}}) = \dot{\mathbf{Z}} - \frac{\Re \langle \dot{\mathbf{Z}}, \mathbf{X}_C \rangle}{\langle \mathbf{X}_C, \mathbf{X}_C \rangle} \mathbf{X}_C.$$

- *Retraction*: $\mathbf{R}_{\mathbf{X}}(\dot{\mathbf{Z}}) = (\mathbf{X} + \dot{\mathbf{Z}})[\text{ddiag}((\mathbf{X} + \dot{\mathbf{Z}})^H(\mathbf{X} + \dot{\mathbf{Z}}))]^{-1/2}$ and renormalization by scaling as

$$\mathbf{X}_{k,i} \mapsto \sqrt{\frac{P_k T L_k}{\sum_{i=1}^{L_k} \|\mathbf{X}_{k,i}\|_F^2}} \mathbf{X}_{k,i}.$$

The projector functions are expressions that remove from a tangent vector (matrix) $\dot{\mathbf{Z}}$ at a point $\mathbf{X}$ its projection onto the normal space at $\mathbf{X}$, for example the normal component of $\dot{\mathbf{Z}}$ at $\mathbf{X}$ on the Grassmann manifold is given by $\mathbf{X}\mathbf{X}^H \dot{\mathbf{Z}}$ (this is the generalization of the unit sphere case, where the point vector $\mathbf{x}$ already has the direction of the normal space to the sphere, so that $\mathbf{x}(\mathbf{x}^H \mathbf{v})$ yields the normal component of $\mathbf{v}$). This projection is similarly generalized from this to the oblique and trace manifolds which are products of spheres.

The retraction step of a tangent vector $\dot{\mathbf{Z}}$ at $\mathbf{X}$ is however a function that one must choose so that the constraint equations that define the manifold are satisfied by its output. There is a natural choice for this retraction function given by the exponential map in Riemannian geometry, i.e., by following the geodesic from $\mathbf{X}$ of length $\|\dot{\mathbf{Z}}\|$ and direction $\dot{\mathbf{Z}}$. However, this is not needed for first-order methods, since the geodesic equation is a second-order differential equation, and thus any first-order function of $\dot{\mathbf{Z}}$ whose output yields a point on the manifold can in principle suffice. Since the Grassmann manifold consists of matrices $\mathbf{X}$ that are Stiefel or semi-unitary, taking the first M columns of a QR decomposition

of a tangent matrix yields an orthonormal basis, i.e., a point on the Grassmann manifold that is close to the tangent point. Similarly, for the oblique and trace manifold, the purpose is to project down tangent vectors back onto the spheres that comprise the manifold, so that the corresponding power constraints are satisfied. This is achieved for unit spheres by the expressions above for $\mathbf{R}_{\mathbf{X}}(\dot{\mathbf{Z}})$, which carry out the projection onto a Cartesian product of spheres; they are followed by a renormalization of the output so that the symbols $\mathbf{X}$ satisfy the value of the power constraint in place. Therefore the retraction function is the last step of every iteration in Riemannian optimization as it yields the updated constellation on the chosen manifold.

Finally, let us introduce some notation conventions in order to compute gradients or partial derivatives on manifolds in the multiuser setting. In particular, notice that a joint codeword, $\mathbf{F}_i = [\mathbf{X}_{1,i_1}, \dots, \mathbf{X}_{K,i_K}]$, in the multiuser constellation $\mathcal{C}$ consists of a choice of a single-user codeword $\mathbf{X}_{k,i_k} \in \mathcal{C}_k$ from each of the users' codebooks. This entails that a multiuser function $f(\mathcal{C})$ may depend on each of the individual codewords $\mathbf{X}_{k,i_k}$ through different pairs (transmitted $\rightarrow$ received after ML detection) of multiuser codewords ($\mathbf{F}_i \rightarrow \mathbf{F}_j$). The transmitted symbol $\mathbf{X}_{k,i_k}$ can be a subset of the columns of either $\mathbf{F}_i$ or $\mathbf{F}_j$, or both. Therefore, we will need to compute partial derivatives of functions of joint multiuser codewords with respect to single-user symbols, i.e. with respect to only a particular subset of M columns from the total KM columns of a joint codeword.

Let us consider any component x_{ab} of any codeword $\mathbf{X} \in \mathcal{C}_k$ from the k -th user constellation as the varying parameter, so that the change in $\mathbf{X}$ is given by $\mathbf{X} + t\dot{\mathbf{Z}}$, where the matrix $\dot{\mathbf{Z}}$ will usually be one of the $\dot{\mathbf{Z}}_{ab}$, which is 1 at row a and column b and zero elsewhere. Then the joint constellation $\mathcal{C}$ changes to $\mathcal{C}(t)$ by updating any joint codeword that includes $\mathbf{X}$, i.e. changing $\mathbf{F} = [\mathbf{X}_1 \dots \mathbf{X}_K]$ into $\mathbf{F}(t) = [\mathbf{X}_1 \dots \mathbf{X} + t\dot{\mathbf{Z}} \dots \mathbf{X}_K]$. Correspondingly, the value of any function that depends on the complete joint constellation $f(\mathcal{C})$ changes to $f(\mathcal{C}(t))$. We shall specify with respect to which single-user codeword the constellation is varying by writing $f(\mathcal{C}(t)|\mathbf{X} + t\dot{\mathbf{Z}})$. If a single-user codeword is present in a joint codeword we shall write $\mathbf{X} \in \mathbf{F}$, and denote by $f(\mathbf{F}_i(t)|\mathbf{X} + t\dot{\mathbf{Z}}, \mathbf{F}_j)$ the updated value of a function of several multiuser codewords, e.g. $f(\mathbf{F}_i, \mathbf{F}_j)$.

Recalling the relationship of a matrix derivative and its gradient included in the appendix, Eq. (36), we may write the directional derivative in the $\dot{\mathbf{Z}}$ direction of any function of a joint constellation, with respect to any single user codeword $\mathbf{X}$, as:

$$D_{\mathbf{X}}f(\mathcal{C})(\dot{\mathbf{Z}}) = \left. \frac{d}{dt} \right|_{t=0} f(\mathcal{C}(t)|\mathbf{X} + t\dot{\mathbf{Z}}) = \Re(\langle D_{\mathbf{X}}f(\mathcal{C}), \dot{\mathbf{Z}} \rangle_F). \quad (10)$$

Then the unconstrained or Euclidean partial derivative matrix of the function f at $\mathcal{C}$ with respect to codeword $\mathbf{X}$ is the matrix of components

$$[D_{\mathbf{X}}f(\mathcal{C})]_{ab} = \frac{\partial f}{\partial x_{ab}} = \frac{\partial f}{\partial \Re(x_{ab})} + i \frac{\partial f}{\partial \Im(x_{ab})},$$

which we will be able to find by identifying the matrix $\dot{\mathbf{X}}$ in $\Re(\langle \dot{\mathbf{X}}, \dot{\mathbf{Z}} \rangle_F)$ when taking derivatives in the direction of $\dot{\mathbf{Z}}_{ab}$, for every matrix component (a, b) of $\mathbf{X}$.

We summarize in Algorithm 1 a general Riemannian manifold optimization method for designing noncoherent MIMO MAC constellations for K users.

Algorithm 1 Riemannian optimization for K -user MAC

Input: $\sum_{k=1}^K L_k$ uniformly distributed points in $\mathbb{C}^{T \times M}$

Output: Optimized joint constellation $\mathcal{C}$

- 1) Choose cost function f and manifold $\mathcal{M} \in \{\mathbb{G}(M, \mathbb{C}^T), \mathcal{OB}_{\mathbb{C}}(TM, \sum_k L_k), \mathbb{T}r(K, L, M, \mathbb{C}^T)\}$.
- 2) Compute unconstrained gradient $D_{\mathbf{X}_{k,i}}f(\mathcal{C})$ for every codeword $\mathbf{X}_{k,i}$ in $\mathcal{C}_k$, ($k = 1, \dots, K, i = 1, \dots, L_k$).
- 3) Project down to the chosen manifold tangent space at every $\mathbf{X}$:

$$\nabla_{\mathbf{X}}f(\mathcal{C}) = \mathbf{P}_{\mathbf{X}}(D_{\mathbf{X}}f(\mathcal{C})).$$

- 4) Compute the norm of the full gradient:

$$\|\nabla f(\mathcal{C})\| = \sqrt{\sum_{k=1}^K \sum_{i=1}^{L_k} \|\nabla_{\mathbf{X}_{k,i}}f(\mathcal{C})\|_F^2}.$$

- 5) Move every codeword a step h in the direction of steepest ascent (descent) retracting back onto the manifold:

$$\mathbf{X}_{new} = \mathbf{R}_{\mathbf{X}} \left(\pm h \frac{\nabla_{\mathbf{X}}f(\mathcal{C})}{\|\nabla f(\mathcal{C})\|} \right).$$

- 6) Evaluate $f(\mathcal{C}_{new})$ and repeat step 5 with smaller h until cost function improves its value with respect to $f(\mathcal{C})$.
 - 7) Update constellation by substituting $\mathbf{X} \mapsto \mathbf{X}_{new}$ for every codeword.
 - 8) Repeat 2 – 7 until the number of iterations or improvement in f reach a threshold.
 - 9) Return constellation $\mathcal{C}_k = \{\mathbf{X}_{k,i}\}_{i=1}^{L_k}$, for every user $k = 1, \dots, K$.
-

III. FULL-DIVERSITY NONCOHERENT MULTIUSER CONSTELLATIONS FOR THE MIMO MAC

A. Noncoherent Joint Pairwise Error Probability

The study of coherent and noncoherent multiuser space-time communications was carried out extensively in [30] and [25] using the results of [26], where the asymptotic analysis of the error probability of quadratic receivers in Rayleigh fading channels was studied in detail. One of the important results of [25] is that at least $T = (K + 1)M$ temporal dimensions are necessary to achieve full-spatial diversity of MN for every user. Recall that the spatial diversity indicates the slope of the symbol error rate (SER) vs. SNR curve when $\text{SNR} \rightarrow \infty$. For $K = 2$ users, this means that the coherence time must be at least $T = 3M$ symbol periods.

Assuming full-diversity scenarios, Brehler and Varanasi derived in [25] the asymptotic joint pairwise error probability of the ML detector in the noncoherent case. However, the PEP expression in [25] has not been used as an optimization criterion so far as it was considered intractable for optimization. Further, the PEP expression was thought not to give

clear insights for constellation design, as discussed in [27] and [28]. In the present work, we prove that the asymptotic PEP formula can not only be used to optimize joint constellations but actually provides the designs of choice for full-diversity scenarios. To the best of our knowledge, in this paper we provide for the first time exact formulas of its gradient on several manifolds with respect to every single-user codeword for any number of users.

Let us introduce the following notation for the orthogonal projection matrix onto the orthogonal complement of the subspace spanned by the columns of $\mathbf{M}$:

$$\mathbf{P}_M^\perp = \mathbf{I} - \mathbf{M}(\mathbf{M}^H \mathbf{M})^{-1} \mathbf{M}^H.$$

Following [25], when comparing two joint hypothesis $\mathbf{F}_i$ vs. $\mathbf{F}_j$, the single-user codewords are to be reordered within the multiuser codeword so that the terms in error appear first, i.e., $\mathbf{F}_i = [\mathbf{F}_i^{(r)} \ \mathbf{F}^{(c)}]$ and $\mathbf{F}_j = [\mathbf{F}_j^{(r)} \ \mathbf{F}^{(c)}]$, where $\mathbf{F}^{(c)}$ are the codewords common to the two hypotheses, and $\mathbf{F}_i^{(r)}$, $\mathbf{F}_j^{(r)}$ are the codewords of the users in error between the two different hypothesis. With these conventions in place, the following proposition shows the expression derived in [25] for the asymptotic PEP $\mathcal{P}(\mathbf{F}_i \rightarrow \mathbf{F}_j)$, i.e., the error probability in a binary hypothesis test between $\mathbf{F}_i$ and $\mathbf{F}_j$.

Proposition 1 (Noncoherent Asymptotic PEP [25]) *Let us assume that there is no correlation between the channel fading coefficients, equal SNR users¹, and that $\mathbf{F}_i^{(r)H} \mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)}$ has full rank (i.e. $T \geq (r + K)M$, with r the number of symbols in error). Then the total pairwise error probability of the optimal detector, for detecting $\mathbf{F}_j$ when receiving $\mathbf{F}_i$, approaches arbitrarily closely to*

$$\mathcal{P}(\mathbf{F}_i \rightarrow \mathbf{F}_j) = \frac{\sigma^{2rNM} \sum_{n=0}^{rNM} \binom{2rNM-n}{rNM} (n!)^{-1} (\hat{c}_{ij})^n}{\det(\mathbf{F}_i^{(r)H} \mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)})^N}, \quad (11)$$

where $\hat{c}_{ij} = N \log \frac{\det(\mathbf{F}_i^H \mathbf{F}_i)}{\det(\mathbf{F}_j^H \mathbf{F}_j)} \geq 0$, which can always be guaranteed by relabeling the hypothesis accordingly, and $\sigma^2 = P10^{-\frac{\text{SNR}}{10}}$ is the noise variance of the model of Eq. (1).

Notice that the denominator in $\mathcal{P}(\mathbf{F}_i \rightarrow \mathbf{F}_j)$ is the factor that encodes for the distance between joint codewords in error. This leads us to propose a multiuser union bound cost function for the design of noncoherent multiuser constellations:

$$f(\mathcal{C}) = \sum_{i \neq j} \sigma^{2rNM} \det(\mathbf{F}_i^{(r)H} \mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)})^{-N}, \quad (12)$$

where the sum is over all the joint multiuser codewords in $\mathcal{C}$.

Remark 1 The multiuser union bound (UB) criterion (12) is a natural generalization of the single-user ($K = 1$) UB defined by

$$\text{UB}(\mathbf{X}_1, \dots, \mathbf{X}_L) = \sum_{i < j} \det(\mathbf{I}_M - \mathbf{X}_i^H \mathbf{X}_j \mathbf{X}_j^H \mathbf{X}_i)^{-N}, \quad (13)$$

¹To consider users with different SNRs simply requires introducing a fixed diagonal matrix in the cost function. When there is spatial correlation at either the transmit or the receive arrays, it is necessary to include another matrix in the expression. These matrices are fixed and do not change the optimization process.

which has been proposed in [32], [14] to design single-user Grassmannian constellations. Optimized designs on the Grassmann manifold using this criterion have been obtained in our previous works [14], [33].

To see the connection between the single-user and the multiuser criteria, notice that when there is only one user present $\mathbf{F}_i = \mathbf{X}_i$ and $\mathbf{P}_{\mathbf{F}_j}^\perp = \mathbf{I} - \mathbf{X}_j \mathbf{X}_j^H$, using $\mathbf{X}_i^H \mathbf{X}_i = \mathbf{X}_j^H \mathbf{X}_j = \mathbf{I}_M$, so up to a scaling constant

$$f(\mathcal{C})_{K=1} = \text{UB}(\mathbf{X}_1, \dots, \mathbf{X}_L).$$

Notice that r in (12) may take values from 1 symbol in error to all the K users in error, which makes the number of terms in the sum increasingly large: as the size of $|\mathcal{C}| = L_1 \cdots L_K$ grows, the number of pairs of hypotheses i, j , i.e. number of terms in the sum (12), grows as $\sim |\mathcal{C}|^2$. For example, for two users $K = 2$, and $r = 1$, there are $L_1(L_1 - 1)L_2 + L_1L_2(L_2 - 1)$ terms in (12), whereas for $r = 2$ there are $L_1L_2(L_1 - 1)(L_2 - 1)$ terms, that is, the number of terms with two symbols in error grows with one order higher. This would make the multiuser optimization problem computationally unfeasible as the number of users and codewords grow. However, the contribution of the factor σ^{2rNM} is σ^{2NM} for the less numerous one-error terms and σ^{4NM} for the more numerous two-error terms. Since σ is inversely proportional to the SNR, the two-error terms are weighed two orders of magnitude less than the one-error terms. Since the number of terms in the union bound is finite, despite being more numerous, the PEP terms with higher slope or diversity $4NM$ (two-symbols-in-error terms) are dominated for sufficiently large SNR by those with lower diversity $2NM$ (one-symbol-in-error terms). Because of this, and in order for the optimization to become feasible computationally, we propose to consider only the one-symbol-in-error terms, that is

$$\begin{aligned} F(\mathcal{C}) &= \sum_{\substack{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C} \\ r=1}} \mathcal{F}_{ij}(\mathbf{F}_i^{(r)}, \mathbf{F}_j) \\ &= \sum_{\substack{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C} \\ r=1}} \det(\mathbf{F}_i^{(r)H} \mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)})^{-N}, \end{aligned} \quad (14)$$

so that the proposed design criterion Min- F for full-diversity scenarios finally becomes:

$$\underset{\mathcal{C}_1, \dots, \mathcal{C}_K}{\text{argmin}} F(\mathcal{C}). \quad (15)$$

For example, for a 2-user MAC with $r = 1$, let $\mathbf{F}_i = [\mathbf{A} \ \mathbf{C}]$ be the transmitted codeword when user 1 is in error, mistaking $\mathbf{X}_{1,i} = \mathbf{A}$ for $\mathbf{X}_{1,j} = \mathbf{B}$, whereas user 2 detects the correct symbol $\mathbf{C}$, so that $\mathbf{F}_j = [\mathbf{B} \ \mathbf{C}]$, $\mathbf{F}_i^{(r)} = \mathbf{A}$ and $\mathbf{F}^{(c)} = \mathbf{C}$. In this case, each summand $\mathcal{F}_{ij}$ in the cost function (14) can be written explicitly in terms of the single-user codewords as

$\mathcal{F}_{ij}(\mathbf{A}, \mathbf{B}, \mathbf{C})$ so that:

$$\begin{aligned} F(\mathcal{C})_{K=2} &= \sum_{\substack{\mathbf{A} \neq \mathbf{B} \in \mathcal{C}_1 \text{ or } \mathcal{C}_2 \\ \mathbf{C} \in \mathcal{C}_2 \text{ or } \mathcal{C}_1}} \mathcal{F}_{ij}(\mathbf{A}, \mathbf{B}, \mathbf{C}) \\ &= \sum_{\substack{\mathbf{A} \neq \mathbf{B} \in \mathcal{C}_1 \text{ or } \mathcal{C}_2 \\ \mathbf{C} \in \mathcal{C}_2 \text{ or } \mathcal{C}_1}} \det(\mathbf{A}^H(\mathbf{I} - [\mathbf{B} \mathbf{C}][[\mathbf{B} \mathbf{C}]^H[\mathbf{B} \mathbf{C}]]^{-1}[\mathbf{B} \mathbf{C}]^H)\mathbf{A})^{-N} \\ &= \sum_{\substack{\mathbf{A} \neq \mathbf{B} \in \mathcal{C}_1 \text{ or } \mathcal{C}_2 \\ \mathbf{C} \in \mathcal{C}_2 \text{ or } \mathcal{C}_1}} \det \left(\mathbf{A}^H \mathbf{A} - \mathbf{A}^H [\mathbf{B} \mathbf{C}] \right. \\ &\quad \left. \cdot \begin{bmatrix} \mathbf{B}^H \mathbf{B} & \mathbf{B}^H \mathbf{C} \\ \mathbf{C}^H \mathbf{B} & \mathbf{C}^H \mathbf{C} \end{bmatrix}^{-1} [\mathbf{B} \mathbf{C}]^H \mathbf{A} \right)^{-N}. \end{aligned}$$

In order to design codebooks based on minimizing the joint probability of error, we propose to perform a gradient descent algorithm over the packing $\mathcal{C}$ to minimize the cost function (14), for which we need the Riemannian gradient vector of f in the Grassmannian product manifold or the oblique and trace manifolds described in Sec. II-B. Notice that if one performs the optimization of the constellation within a submanifold other than the Grassmannian, like the oblique manifold, the single-user codewords need not be Stiefel matrices, so the terms $\mathbf{A}^H \mathbf{A}$, $\mathbf{B}^H \mathbf{B}$, $\mathbf{C}^H \mathbf{C}$ of the last equation do not necessarily simplify to the identity, therefore the gradients of these type of functions must be computed without assuming these terms are the identity matrix at every step.

B. Gradient Computation

Let us write $\mathbf{X} = \mathbf{F}[i_1(\mathbf{X}) : i_M(\mathbf{X})]$ for the extraction of the M columns in $\mathbf{F}$ running from column i_1 to column i_M corresponding to the position of the codeword $\mathbf{X}$ inside the concatenated matrix $\mathbf{F}$. Then, for any other matrix $\mathbf{A}$ with the size of $\mathbf{F}$, $\mathbf{A}[i_1(\mathbf{X}) : i_M(\mathbf{X})]$ extracts the corresponding columns of $\mathbf{A}$ located where the block of $\mathbf{X}$ is within $\mathbf{F}$. With this and all the notational conventions defined in previous sections we arrive at the following fundamental result, that is actually valid for F summed over any number of symbols in error.

Theorem 1 Let $\mathbf{M}_j := \mathbf{F}_j^H \mathbf{F}_j$ and $\mathbf{G}_{ij} := \mathbf{F}_i^{(r)H} \mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)}$. The unconstrained Euclidean gradient of F with respect to codeword $\mathbf{X}$ is:

$$D_{\mathbf{X}} F(\mathcal{C}) = \sum_{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C}} D_{\mathbf{X}} \mathcal{F}_{ij}(\mathbf{F}_i^{(r)}, \mathbf{F}_j), \quad (16)$$

where for a codeword in error, $\mathbf{X} = \mathbf{X}^{(r)}$, the gradient matrix is given by the corresponding block of M columns in the following expression

$$\begin{aligned} D_{\mathbf{X}^{(r)}} \mathcal{F}_{ij}(\mathbf{F}_i^{(r)}, \mathbf{F}_j) &= \\ &- 2N \mathcal{F}_{ij} \left[\mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \right] [i_1(\mathbf{X}^{(r)}) : i_M(\mathbf{X}^{(r)})]. \end{aligned} \quad (17)$$

And for $\mathbf{X} = \mathbf{X}^{(c)}$, a codeword not in error, we have:

$$\begin{aligned} D_{\mathbf{X}^{(c)}} \mathcal{F}_{ij} &= 2N \mathcal{F}_{ij} \left[(\mathbf{I}_T - \mathbf{F}_j \mathbf{M}_j^{-1} \mathbf{F}_j^H) \right. \\ &\quad \left. \cdot \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} \mathbf{F}_j \mathbf{M}_j^{-1} \right] [j_1(\mathbf{X}^{(c)}) : j_M(\mathbf{X}^{(c)})]. \end{aligned} \quad (18)$$

Proof: Let us vary the components of $\mathbf{F}_j$ corresponding to a codeword $\mathbf{X}^{(c)}$ that is not in error, i.e. every component within $\mathbf{F}^{(c)}$, so that $d\mathbf{F}_i^{(r)}/dt|_{t=0} = 0$. Let $\dot{\mathbf{Z}}_j$ be the matrix of the dimensions of $\mathbf{F}_j$ and with 0 everywhere except 1 at a fixed component, i.e. the unit variation of element $[\mathbf{F}_j]_{ab}$ for any row a and column b in the $\mathbf{F}^{(c)}$ part. We get the derivative of $\mathbf{P}_{\mathbf{F}_j}^\perp$ using the derivative of an inverse matrix which is

$$\left. \frac{d(\mathbf{R}(t)^{-1})}{dt} \right|_{t=0} = -\mathbf{R}(0)^{-1} \left(\left. \frac{d\mathbf{R}(t)}{dt} \right|_{t=0} \right) \mathbf{R}(0)^{-1}, \quad (19)$$

and so

$$\begin{aligned} \left. \frac{d}{dt} \right|_{t=0} \mathbf{P}_{\mathbf{F}_j}^\perp (\mathbf{F}_j + t\dot{\mathbf{Z}}_j) &= -\dot{\mathbf{Z}}_j \mathbf{M}_j^{-1} \mathbf{F}_j^H - \mathbf{F}_j \mathbf{M}_j^{-1} \dot{\mathbf{Z}}_j^H \\ &\quad + \mathbf{F}_j \mathbf{M}_j^{-1} (\dot{\mathbf{Z}}_j^H \mathbf{F}_j + \mathbf{F}_j^H \dot{\mathbf{Z}}_j) \mathbf{M}_j^{-1} \mathbf{F}_j^H. \end{aligned}$$

Thus, for the derivative of $\mathcal{F}_{ij}$ when varying only this element in the codewords of $\mathcal{C}$, and using the derivative of the determinant formula by Jacobi, one obtains

$$\begin{aligned} \left. \frac{d}{dt} \right|_{t=0} \mathcal{F}_{ij}(\mathbf{F}_i^{(r)}, \mathbf{F}_j + t\dot{\mathbf{Z}}_j) &= \\ &- N(\det \mathbf{G}_{ij})^{-N-1} \det \mathbf{G}_{ij} \operatorname{tr} \left[\mathbf{G}_{ij}^{-1} \left. \frac{d}{dt} \right|_{t=0} \mathbf{G}_{ij} (\mathbf{F}_j + t\dot{\mathbf{Z}}_j) \right] \\ &= -N \mathcal{F}_{ij} \operatorname{tr} \left[\mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} \left. \frac{d\mathbf{P}_{\mathbf{F}_j}^\perp}{dt} \right|_{t=0} \mathbf{F}_i^{(r)} \right] \\ &= -N \mathcal{F}_{ij} \operatorname{tr} \left[\mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} (-\dot{\mathbf{Z}}_j \mathbf{M}_j^{-1} \mathbf{F}_j^H - \mathbf{F}_j \mathbf{M}_j^{-1} \dot{\mathbf{Z}}_j^H) \mathbf{F}_i^{(r)} \right] \\ &\quad - N \mathcal{F}_{ij} \operatorname{tr} \left[\mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} \mathbf{F}_j \mathbf{M}_j^{-1} (\dot{\mathbf{Z}}_j^H \mathbf{F}_j + \mathbf{F}_j^H \dot{\mathbf{Z}}_j) \mathbf{M}_j^{-1} \mathbf{F}_i^{(r)H} \mathbf{F}_i^{(r)} \right]. \end{aligned}$$

As always, using the cyclic property of the trace, that $\mathbf{P}_{\mathbf{F}_j}^\perp$ and $\mathbf{G}_{ij}$ are Hermitian, and the definition of Frobenius inner product, this reduces to

$$\begin{aligned} \left. \frac{d}{dt} \right|_{t=0} \mathcal{F}_{ij}(\mathbf{F}_i^{(r)}, \mathbf{F}_j + t\dot{\mathbf{Z}}_j) &= \\ &2N \mathcal{F}_{ij} \Re \langle \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} \mathbf{F}_j \mathbf{M}_j^{-1}, \dot{\mathbf{Z}}_j \rangle_F \\ &\quad - 2N \mathcal{F}_{ij} \Re \langle \mathbf{F}_j \mathbf{M}_j^{-1} \mathbf{F}_j^H \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} \mathbf{F}_j \mathbf{M}_j^{-1}, \dot{\mathbf{Z}}_j \rangle_F, \end{aligned}$$

which implies that the partial derivative with respect to that matrix element is

$$\begin{aligned} \frac{\partial \mathcal{F}_{ij}}{\partial x_{ab}^{(c)}} &= 2N \mathcal{F}_{ij} \left[\mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} \mathbf{F}_j \mathbf{M}_j^{-1} \right. \\ &\quad \left. - \mathbf{F}_j \mathbf{M}_j^{-1} \mathbf{F}_j^H \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} \mathbf{F}_j \mathbf{M}_j^{-1} \right]_{ab}. \end{aligned}$$

Therefore, collecting terms, the unconstrained gradient of $\mathcal{F}_{ij}$ with respect to a codeword $\mathbf{X}^{(c)}$ in $\mathbf{F}^{(c)}$, corresponding to the columns $[j_1(\mathbf{X}^{(c)}) : j_M(\mathbf{X}^{(c)})]$ of $\mathbf{F}_j$, is:

$$\begin{aligned} \left[\frac{\partial \mathcal{F}_{ij}}{\partial \mathbf{X}^{(c)}} \right] &= \dot{\mathbf{X}}^{(c)} = 2N \mathcal{F}_{ij} \left[(\mathbf{I}_T - \mathbf{F}_j \mathbf{M}_j^{-1} \mathbf{F}_j^H) \right. \\ &\quad \left. \cdot \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} \mathbf{F}_j \mathbf{M}_j^{-1} \right] [j_1(\mathbf{X}^{(c)}) : j_M(\mathbf{X}^{(c)})]. \end{aligned}$$

Similarly, one can vary the components of $\mathbf{F}_i^{(r)}$ corresponding to the codeword $\mathbf{X}^{(r)}$ in error, so that $d\mathbf{F}_j/dt|_{t=0} = 0$, and therefore:

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} \mathcal{F}_{ij}(\mathbf{F}_i^{(r)} | \mathbf{X}^{(r)} + t\dot{\mathbf{Z}}, \mathbf{F}_j) &= \\ &- N\mathcal{F}_{ij} \operatorname{tr} \left[\mathbf{G}_{ij}^{-1} \left(\dot{\mathbf{Z}}^H \mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)} + \mathbf{F}_i^{(r)H} \mathbf{P}_{\mathbf{F}_j}^\perp \dot{\mathbf{Z}} \right) \right] \\ &= -N\mathcal{F}_{ij} \operatorname{tr} \left[\dot{\mathbf{Z}}^H \mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \right] - N\mathcal{F}_{ij} \operatorname{tr} \left[\mathbf{G}_{ij}^{-1} \mathbf{F}_i^{(r)H} \mathbf{P}_{\mathbf{F}_j}^\perp \dot{\mathbf{Z}} \right] \\ &= -2N\mathcal{F}_{ij} \Re \langle \mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1}, \dot{\mathbf{Z}} \rangle_F, \end{aligned}$$

which yields

$$\frac{\partial \mathcal{F}_{ij}}{\partial x_{ab}^{(r)}} = -2N\mathcal{F}_{ij} \left[\mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \right]_{ab},$$

and finally

$$\begin{aligned} \left[\frac{\partial \mathcal{F}_{ij}}{\partial \mathbf{X}^{(r)}} \right] &= \dot{\mathbf{X}}^{(r)} \\ &= -2N\mathcal{F}_{ij} \left[\mathbf{P}_{\mathbf{F}_j}^\perp \mathbf{F}_i^{(r)} \mathbf{G}_{ij}^{-1} \right] [i_1(\mathbf{X}^{(r)}) : i_M(\mathbf{X}^{(r)})]. \end{aligned}$$

□

IV. NON-FULL DIVERSITY MULTIUSER CONSTELLATIONS FOR THE MIMO MAC

A. Cost Functions

Obviously, whenever the coherence time is sufficiently long such that $T \geq (K+1)M$, it will be preferable to design and employ full-diversity multiuser constellations. However, the constraint $T \geq (K+1)M$ may be difficult to meet in fast-fading channels with high-mobility users, especially when the number of users or transmitting antennas grows.

In the non-full diversity case, the full-rank condition required in Proposition 1 is not satisfied and the asymptotic PEP formula is no longer correct in this scenario. However, one can use instead proxy functions that provide bounds for the PEP valid without full diversity. Several cost functions have been proposed to design constellations for noncoherent communications in the multiple access channel, cf. [28] or [29]. In particular, the authors of [28] introduce several cost functions to design multiuser constellations on the Grassmannian, such as the functions b , δ , and $J_{1/2}$ to be presented below. These depend on pairs of joint codewords $\mathbf{F}_i, \mathbf{F}_j \in \mathcal{C}$, and are all related to the leading exponent of the joint pairwise error probability $\mathcal{P}(\mathbf{F}_i \rightarrow \mathbf{F}_j)$, providing bounds that serve as proxies for this PEP. Although the criteria in [28] were proposed for both full-diversity and non-full-diversity scenarios, our experience indicates that for full-diversity scenarios with $T \geq (K+1)M$ the criterion based on the asymptotic expression of the PEP, described in Subsection III, provides much better results. The criteria described in this section are thus specifically useful to design non-full diversity multiuser constellations.

The main geometrical motivation to study these cost functions however stems from the fact that they are related to a geometrical interpretation of δ as a Riemannian distance between Hermitian positive definite matrices defined from the joint codewords $(\mathbf{I}_T + \mathbf{F}_i \mathbf{F}_i^H)$ and $(\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)$. Thus, in [28],

design criteria are proposed that maximize these cost functions, due to their relation to the worst PEP and the intuition of separating the closest pair of joint codewords in the manifold of $T \times T$ Hermitian positive definite matrices. However, the authors in [28] only optimize the $\max -J_{1/2, \min}$ for USTM single-user codewords, i.e. they optimized $\max -J_{1/2, \min}$ in the Grassmann manifold. In the following, we work out the theoretical basis needed to optimize a union-bound-based generalization of the cost functions b and δ , and to do so on the different manifolds presented in Subsection II-B.

This leads us to propose the following cost functions for the design of noncoherent multiuser constellations for the MIMO MAC in non-full diversity scenarios.

Definition 1 The SER union bound proxy function b for a joint constellation $\mathcal{C}$ is defined as

$$b_{UB}(\mathcal{C}) := \log \left[\sum_{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C}} \exp \left(-N \sum_{l=1}^T |\log \lambda_l(\mathbf{F}_i, \mathbf{F}_j)| \right) \right], \quad (20)$$

where $\lambda_l(\mathbf{F}_i, \mathbf{F}_j)$ are the eigenvalues of $\mathbf{\Gamma}(\mathbf{F}_i, \mathbf{F}_j) := (\mathbf{I}_T + \mathbf{F}_i \mathbf{F}_i^H)(\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1}$, for $\mathbf{F}_i, \mathbf{F}_j \in \mathcal{C}$ multiuser codewords.

Notice that $\lambda_l \geq 0$, and except for a subset of measure zero in the space of matrices $\mathbf{F}_i, \mathbf{F}_j$ the eigenvalues will be positive so that the logarithm is well defined. Equivalently, we can use the pairwise hypothesis function

$$b(\mathbf{F}_i, \mathbf{F}_j) := \sum_{l=1}^T |\log \lambda_l(\mathbf{F}_i, \mathbf{F}_j)|,$$

so that

$$b_{UB}(\mathcal{C}) = \log \left[\sum_{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C}} \exp[-Nb(\mathbf{F}_i, \mathbf{F}_j)] \right].$$

Definition 2 The SER union bound proxy function δ for a joint constellation $\mathcal{C}$ is defined as

$$\delta_{UB}(\mathcal{C}) := \log \left[\sum_{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C}} \exp \left(-N \sqrt{\sum_{l=1}^T \log^2 \lambda_l(\mathbf{F}_i, \mathbf{F}_j)} \right) \right], \quad (21)$$

where $\lambda_l(\mathbf{F}_i, \mathbf{F}_j)$ are the eigenvalues of $\mathbf{\Gamma}(\mathbf{F}_i, \mathbf{F}_j)$ as above, for $\mathbf{F}_i, \mathbf{F}_j \in \mathcal{C}$.

Similar notation to the former cost function leads us to write

$$\delta(\mathbf{F}_i, \mathbf{F}_j) := \sqrt{\sum_{l=1}^T \log^2 \lambda_l(\mathbf{F}_i, \mathbf{F}_j)},$$

so that

$$\delta_{UB}(\mathcal{C}) = \log \left[\sum_{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C}} \exp[-N\delta(\mathbf{F}_i, \mathbf{F}_j)] \right].$$

The motivation for these cost functions comes from the following results from [28, Prop. 4 and 5] which relate the

pairwise function we just defined to the exponent of the joint probability of error.

Proposition 2 ([28]) *The joint PEP exponent is upper- and lower-bounded as*

$$b(\mathbf{F}_i, \mathbf{F}_j) + T \geq -\frac{1}{N} \log \mathcal{P}(\mathbf{F}_i \rightarrow \mathbf{F}_j) \geq \frac{1}{2} b(\mathbf{F}_i, \mathbf{F}_j) - T \log 2. \quad (22)$$

Proposition 3 ([28]) *The natural Riemannian distance between $\mathbf{I}_T + \mathbf{F}_i \mathbf{F}_i^H$ and $\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H$ in the manifold of Hermitian positive definite matrices is $\delta(\mathbf{F}_i, \mathbf{F}_j)$. Furthermore, $b(\mathbf{F}_i, \mathbf{F}_j)$ is bounded as*

$$\sqrt{T} \delta(\mathbf{F}_i, \mathbf{F}_j) \geq b(\mathbf{F}_i, \mathbf{F}_j) \geq \delta(\mathbf{F}_i, \mathbf{F}_j).$$

By multiplying by $-N$ and taking exponentials, (22) becomes

$$\begin{aligned} \exp[-Nb(\mathbf{F}_i, \mathbf{F}_j)] \exp[-NT] &\leq \mathcal{P}(\mathbf{F}_i \rightarrow \mathbf{F}_j) \\ &\leq \exp\left[\frac{-N}{2} b(\mathbf{F}_i, \mathbf{F}_j)\right] \exp[NT \log 2], \end{aligned}$$

Since δ is in turn a bound on b , these relations lead us to expect a leading order behavior such as

$$\mathcal{P}(\mathbf{F}_i \rightarrow \mathbf{F}_j) \sim \exp[-Nb(\mathbf{F}_i, \mathbf{F}_j)]$$

and

$$\mathcal{P}(\mathbf{F}_i \rightarrow \mathbf{F}_j) \sim \exp[-N\delta(\mathbf{F}_i, \mathbf{F}_j)],$$

and by summing over all pairs of joint codewords and taking logarithms, the inequality yields

$$\begin{aligned} e^{-NT} \sum_{\mathbf{F}_i \neq \mathbf{F}_j} \exp[-Nb(\mathbf{F}_i, \mathbf{F}_j)] &\leq \sum_{\mathbf{F}_i \neq \mathbf{F}_j} \mathcal{P}(\mathbf{F}_i \rightarrow \mathbf{F}_j) \\ &\leq e^{NT \log 2} \sum_{\mathbf{F}_i \neq \mathbf{F}_j} \exp\left[\frac{-N}{2} b(\mathbf{F}_i, \mathbf{F}_j)\right]. \end{aligned} \quad (23)$$

This provides bounds on the union bound PEP and on its leading exponent by taking logarithms, which justify our definition of b_{UB} in Eq. (20) as a figure of merit to optimize (the factor $1/2$ is irrelevant in the normalized gradients). Similarly for δ_{UB} . Therefore the criteria that we propose are Min- δ given by

$$\operatorname{argmin}_{\mathcal{C}_1, \dots, \mathcal{C}_K} \delta_{UB}(\mathcal{C}), \quad (24)$$

and Min- b given by

$$\operatorname{argmin}_{\mathcal{C}_1, \dots, \mathcal{C}_K} b_{UB}(\mathcal{C}), \quad (25)$$

where $\mathcal{C}$ is built up out of the concatenation of K codewords $\mathbf{X}_{k,i} \in \mathcal{C}_k$, from each of the users' constellations, which is a set of points in the corresponding manifold. It is worth noticing that the previous criteria using Definitions 1 and 2 resemble optimization methods using the approximation $\max_i x_i \approx \epsilon \log \sum_i \exp(x_i/\epsilon)$, which is employed in [28, Sec. VI-A] to smooth several objective functions, in particular when the optimization problem consists of maximizing the minimum value of an objective function. This implies ϵ is a free optimization parameter that needs to be specified for the numerical convergence of every setup. The union

bound functions (20) and (21) would correspond to choosing $\epsilon = -1/N$ (note that an overall constant factor would not affect the direction of the normalized gradients in a gradient ascent/descent method). This makes our proposed designs optimized for a given number of receive antennas but free of other optimization parameters.

For completeness, since it is used in the comparisons of our simulation experiments in Sec. V, we also introduce another cost function proposed in [28]:

$$\begin{aligned} J_{1/2}(\mathbf{F}_i, \mathbf{F}_j) &= \frac{1}{2} \log \det(2\mathbf{I}_T + (\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1} (\mathbf{I}_T + \mathbf{F}_i \mathbf{F}_i^H) \\ &\quad + (\mathbf{I}_T + \mathbf{F}_i \mathbf{F}_i^H)^{-1} (\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)) - T \log 2, \end{aligned} \quad (26)$$

whose optimization criterion is called Max- $J_{1/2, \min}$:

$$\operatorname{argmax}_{\mathcal{C}_1, \dots, \mathcal{C}_K} \min_{\mathbf{F}_i \neq \mathbf{F}_j} J_{1/2}(\mathbf{F}_i, \mathbf{F}_j). \quad (27)$$

Following [28], this cost function will always be optimized over the Grassmann manifold, unless stated otherwise, and using the smooth approximation discussed above it is equivalent to

$$\operatorname{argmax}_{\mathcal{C}_1, \dots, \mathcal{C}_K} \epsilon \log \sum_{i \neq j} \exp\left(-\frac{J_{1/2}(\mathbf{F}_i, \mathbf{F}_j)}{\epsilon}\right).$$

B. Gradient Computation

In order to perform a gradient descent on the cost functions defined above, we need to compute their derivatives along directions tangent to the manifolds of interest. First of all, we must ensure that the $\lambda_l(\mathbf{F}_i, \mathbf{F}_j)$ functions are smooth almost everywhere, which can be proved on any chart by using the locally Lipschitz property and Rademacher's theorem [34, Th. 3.1.6]. Even though this guarantees that the gradients are well-defined except for a subset of measure zero, at every iteration of the algorithm, the explicit computation of the derivatives for our manifolds of interest is essentially intractable using manifold charts. Instead of differentiating the restricted functions on the manifold, we use a simpler method based on differentiation on the ambient space and then projecting down the unrestricted gradient to the manifold, which is justified by the excellent optimization results it provides (see Sec. V-B). This method relies on the following fundamental result (notice that multiple eigenvalues happen only for a subset of measure zero in the space of matrices).

Lemma 1 *Let $\Gamma(t) \in \mathbb{C}^{T \times T}$ be a matrix function with eigenvalue system $\Gamma(0)\mathbf{v}_0 = \lambda_0\mathbf{v}_0$, such that λ_0 is a simple eigenvalue with associated eigenvector $\mathbf{v}_0$. Then there are functions $\lambda(t), \mathbf{v}(t)$ defined for all Γ in a neighbourhood of $\Gamma(0)$ such that $\lambda(0) = \lambda_0$, $\mathbf{v}(0) = \mathbf{v}_0$ and $\Gamma(t)\mathbf{v}(t) = \lambda(t)\mathbf{v}(t)$, with $\mathbf{v}_0^H \mathbf{v}(t) = 1$. Moreover, these functions are infinitely differentiable with derivatives:*

$$\left. \frac{d\lambda}{dt} \right|_{t=0} = \frac{1}{\mathbf{u}_0^H \mathbf{v}_0} \mathbf{u}_0^H \cdot \left. \frac{d\Gamma}{dt} \right|_{t=0} \cdot \mathbf{v}_0 \quad (28)$$

and

$$\left. \frac{d\mathbf{v}}{dt} \right|_{t=0} = (\lambda \mathbf{I}_T - \Gamma)^+ \left(\mathbf{I}_T - \frac{\mathbf{v}_0 \mathbf{u}_0^H}{\mathbf{u}_0^H \mathbf{v}_0} \right) \left. \frac{d\Gamma}{dt} \right|_{t=0} \cdot \mathbf{v}_0, \quad (29)$$

where $\mathbf{u}_0$ is the eigenvector associated to the eigenvalue λ_0^* of $\mathbf{\Gamma}^H(0)$, i.e. $\mathbf{\Gamma}^H(0)\mathbf{u}_0 = \lambda_0^*\mathbf{u}_0$, and $(\cdot)^+$ is the Moore-Penrose pseudo-inverse.

Proof: See, e.g., [35, Th. 2]. $\square$

With all these tools and the notation conventions of the previous section, we can derive the explicit gradients of any of the proposed cost functions for the K -user MIMO MAC.

Theorem 2 When $\mathbf{\Gamma}_{ij} := \mathbf{\Gamma}(\mathbf{F}_i, \mathbf{F}_j)$ satisfy the conditions of Lemma 1 for every $\mathbf{F}_i, \mathbf{F}_j \in \mathcal{C}$, the unconstrained partial derivative of the function δ_{UB} , with respect to the single-user codeword $\mathbf{X}$, is given by

$$D_{\mathbf{X}} \delta_{UB}(\mathcal{C}) = \sum_{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C}} \frac{-N \exp[-N\delta(\mathbf{F}_i, \mathbf{F}_j)]}{\exp[\delta_{UB}(\mathcal{C})]\delta(\mathbf{F}_i, \mathbf{F}_j)} \times \sum_{l=1}^T \frac{\log(\lambda_l)}{\lambda_l \mathbf{u}_l^H \mathbf{v}_l} \mathbf{u}_l^H \cdot \frac{\partial \mathbf{\Gamma}_{ij}}{\partial \mathbf{X}} \cdot \mathbf{v}_l, \quad (30)$$

where the λ_l , $\mathbf{u}_l$ and $\mathbf{v}_l$ are for $\mathbf{\Gamma}_{ij}$ as defined in Lemma 1, and $\partial \mathbf{\Gamma}_{ij} / \partial \mathbf{X}$ is a matrix of matrices that depend on whether $\mathbf{F}_i$ or $\mathbf{F}_j$ include, one or both, the single-user symbol $\mathbf{X}$ as a block of its columns. In detail, let $\mathbf{E}_{mn}$ be the canonical real matrix basis, i.e. $[\mathbf{E}_{mn}]_{ij} = [\delta_{mi}\delta_{nj}]_{ij}$, for i, j, m, n indices from 1 to T , then:

- If $\mathbf{X}$ is the symbol in $\mathbf{F}_i$ of a user in error (i.e. a differing block with respect to $\mathbf{F}_j$):

$$\left[\frac{\partial \mathbf{\Gamma}_{ij}}{\partial \mathbf{X}} \right]_{mn} = (\mathbf{E}_{mn} \mathbf{X}^H + \mathbf{X} \mathbf{E}_{mn}^H)(\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1}. \quad (31)$$

- If $\mathbf{X}$ is the symbol in $\mathbf{F}_j$ of a user in error (i.e. a differing block with respect to $\mathbf{F}_i$):

$$\left[\frac{\partial \mathbf{\Gamma}_{ij}}{\partial \mathbf{X}} \right]_{mn} = -\mathbf{\Gamma}_{ij}(\mathbf{E}_{mn} \mathbf{X}^H + \mathbf{X} \mathbf{E}_{mn}^H)(\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1}. \quad (32)$$

- If $\mathbf{X}$ is not a symbol in error (i.e. a common block between $\mathbf{F}_i$ and $\mathbf{F}_j$):

$$\left[\frac{\partial \mathbf{\Gamma}_{ij}}{\partial \mathbf{X}} \right]_{mn} = (\mathbf{I}_T - \mathbf{\Gamma}_{ij})(\mathbf{E}_{mn} \mathbf{X}^H + \mathbf{X} \mathbf{E}_{mn}^H)(\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1}. \quad (33)$$

Proof: Only the terms in the sum that correspond to joint codewords $\mathbf{F}_i$ or $\mathbf{F}_j$ containing $\mathbf{X}$ are nonzero in the derivative, hence we have

$$\frac{d}{dt} \bigg|_{t=0} \delta_{UB}(\mathcal{C}(t)|\mathbf{X} + t\dot{\mathbf{Z}}) = \sum_{\substack{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C} \\ \mathbf{X} \in \mathbf{F}_i \text{ or } \mathbf{F}_j}} \frac{-N \exp[-N\delta(\mathbf{F}_i, \mathbf{F}_j)]}{\exp[\delta_{UB}(\mathcal{C})]} \cdot \frac{d}{dt} \bigg|_{t=0} \delta(\mathbf{F}_i, \mathbf{F}_j|\mathbf{X} + t\dot{\mathbf{Z}}),$$

where

$$\exp[\delta_{UB}(\mathcal{C})] = \sum_{\mathbf{F}_i \neq \mathbf{F}_j \in \mathcal{C}} \exp[-N\delta(\mathbf{F}_i, \mathbf{F}_j)].$$

Notice that one must pay attention to whether the symbol $\mathbf{X}$ is included in the multiuser codeword $\mathbf{F}_i$, $\mathbf{F}_j$ or both, so each

term in the previous sum should actually be written $\delta(\mathbf{F}_i|\mathbf{X} + t\dot{\mathbf{Z}}, \mathbf{F}_j)$, $\delta(\mathbf{F}_i, \mathbf{F}_j|\mathbf{X} + t\dot{\mathbf{Z}})$ or a variation on both depending on the case for each summand. Now, without loss of generality

$$\frac{d}{dt} \bigg|_{t=0} \delta(\mathbf{F}_i, \mathbf{F}_j|\mathbf{X} + t\dot{\mathbf{Z}}) = \frac{1}{2} \left[\sum_{l=1}^T (\log \lambda_l)^2 \right]^{-\frac{1}{2}} \sum_{l=1}^T 2 \frac{\log \lambda_l}{\lambda_l} \frac{d}{dt} \bigg|_{t=0} \lambda_l(\mathbf{F}_i, \mathbf{F}_j|\mathbf{X} + t\dot{\mathbf{Z}}),$$

and, using the notation and discussion from Lemma 1, we know that we can differentiate the eigenvalues with respect to every component of the matrix $\mathbf{X}$, in the direction $\dot{\mathbf{Z}}_{mn}$, to obtain

$$\frac{d}{dt} \bigg|_{t=0} \lambda_l(\mathbf{F}_i, \mathbf{F}_j|\mathbf{X} + t\dot{\mathbf{Z}}_{mn}) = \frac{1}{\mathbf{u}_l^H \mathbf{v}_l} \mathbf{u}_l^H \cdot \left[\frac{\partial \mathbf{\Gamma}(\mathbf{F}_i, \mathbf{F}_j)}{\partial \mathbf{X}} \right]_{mn} \cdot \mathbf{v}_l.$$

Notice that for any joint codeword, e.g. $\mathbf{F}_i = [\mathbf{X}_{1,i_1} \dots \mathbf{X}_{K,i_K}]$, where our chosen $\mathbf{X}$ corresponds to some symbol i_r of some user r , i.e. $\mathbf{X} = \mathbf{X}_{r,i_r} \in \mathcal{C}_r$, then the factors of $\mathbf{\Gamma}_{ij}$ expand by blocks as

$$\mathbf{F}_i \mathbf{F}_i^H = \mathbf{X} \mathbf{X}^H + \sum_{k \neq r}^K \mathbf{X}_{k,i_k} \mathbf{X}_{k,i_k}^H,$$

where the block $\mathbf{X}$ is the symbol of the user of interest, and similarly for $\mathbf{F}_j$. Thus we must take derivatives of $\mathbf{\Gamma}_{ij}$ with respect to an $\mathbf{X}$ that either belongs to only one of the sums of the above expansions of $\mathbf{F}_i \mathbf{F}_i^H$ or $\mathbf{F}_j \mathbf{F}_j^H$, or to both, which then requires the derivative of a product.

Let us assume first that $\mathbf{X}$ is a symbol in $\mathbf{F}_i$ of an user in error, i.e. it does not appear in $\mathbf{F}_j$, then the partial derivative of $\mathbf{\Gamma}_{ij}$ with respect to the component (m, n) of the symbol $\mathbf{X}$ is

$$\begin{aligned} \left[\frac{\partial \mathbf{\Gamma}(\mathbf{F}_i, \mathbf{F}_j)}{\partial \mathbf{X}} \right]_{mn} &= \frac{d}{dt} \bigg|_{t=0} \mathbf{\Gamma}(\mathbf{F}_i|\mathbf{X} + t\dot{\mathbf{Z}}_{mn}, \mathbf{F}_j) \\ &= \frac{d}{dt} \bigg|_{t=0} \left(\mathbf{I}_T + (\mathbf{X} + t\dot{\mathbf{Z}}_{mn})(\mathbf{X}^H + t\dot{\mathbf{Z}}_{mn}^H) \right. \\ &\quad \left. + \sum_{k \neq r}^K \mathbf{X}_{k,i_k} \mathbf{X}_{k,i_k}^H \right) (\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1} \\ &= (\dot{\mathbf{Z}}_{mn} \mathbf{X}^H + \mathbf{X} \dot{\mathbf{Z}}_{mn}^H)(\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1}, \end{aligned}$$

which yields the first of the formulas of the theorem. Similarly, when $\mathbf{X}$ is a symbol in $\mathbf{F}_j$ of an user in error, i.e. not appearing

in $\mathbf{F}_i$, one obtains

$$\begin{aligned}
 \left[\frac{\partial \Gamma(\mathbf{F}_i, \mathbf{F}_j)}{\partial \mathbf{X}} \right]_{mn} &= \frac{d}{dt} \bigg|_{t=0} \Gamma(\mathbf{F}_i, \mathbf{F}_j | \mathbf{X} + t\dot{\mathbf{Z}}_{mn}) \\
 &= \frac{d}{dt} \bigg|_{t=0} (\mathbf{I}_T + \mathbf{F}_i \mathbf{F}_i^H) \left(\mathbf{I}_T + (\mathbf{X} + t\dot{\mathbf{Z}}_{mn})(\mathbf{X}^H + t\dot{\mathbf{Z}}_{mn}^H) \right. \\
 &\quad \left. + \sum_{k \neq r}^K \mathbf{X}_{k,j_k} \mathbf{X}_{k,j_k}^H \right)^{-1} \\
 &= (\mathbf{I}_T + \mathbf{F}_i \mathbf{F}_i^H)(-1)(\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1} \\
 &\quad \cdot \frac{d}{dt} \bigg|_{t=0} \left(\mathbf{I}_T + (\mathbf{X} + t\dot{\mathbf{Z}}_{mn})(\mathbf{X}^H + t\dot{\mathbf{Z}}_{mn}^H) + \sum_{k \neq r}^K \mathbf{X}_{k,j_k} \mathbf{X}_{k,j_k}^H \right) \\
 &\quad \cdot (\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1} \\
 &= -\Gamma_{ij}(\dot{\mathbf{Z}}_{mn} \mathbf{X}^H + \mathbf{X} \dot{\mathbf{Z}}_{mn}^H)(\mathbf{I}_T + \mathbf{F}_j \mathbf{F}_j^H)^{-1},
 \end{aligned}$$

which provides the second formula of the theorem. In the final case, $\mathbf{X}$ is a common block between $\mathbf{F}_i$ and $\mathbf{F}_j$ corresponding to a symbol of a user not in error, and therefore it appears in both terms of Γ_{ij} so that

$$\begin{aligned}
 \left[\frac{\partial \Gamma(\mathbf{F}_i, \mathbf{F}_j)}{\partial \mathbf{X}} \right]_{mn} &= \frac{d}{dt} \bigg|_{t=0} \Gamma(\mathbf{F}_i | \mathbf{X} + t\dot{\mathbf{Z}}_{mn}, \mathbf{F}_j | \mathbf{X} + t\dot{\mathbf{Z}}_{mn}) \\
 &= \frac{d}{dt} \bigg|_{t=0} \left(\mathbf{I}_T + (\mathbf{X} + t\dot{\mathbf{Z}}_{mn})(\mathbf{X}^H + t\dot{\mathbf{Z}}_{mn}^H) + \sum_{k \neq r}^K \mathbf{X}_{k,i_k} \mathbf{X}_{k,i_k}^H \right) \\
 &\quad \cdot \left(\mathbf{I}_T + (\mathbf{X} + t\dot{\mathbf{Z}}_{mn})(\mathbf{X}^H + t\dot{\mathbf{Z}}_{mn}^H) + \sum_{k \neq r}^K \mathbf{X}_{k,j_k} \mathbf{X}_{k,j_k}^H \right)^{-1},
 \end{aligned}$$

which by the product rule results in the sum of the two formulas of the other cases obtained above, yielding the last equation of the theorem. $\square$

Theorem 3 *Using the same conventions and assumptions as in Theorem 2, the gradient of the union bound proxy function b_{UB} is given by:*

$$\begin{aligned}
 D_{\mathbf{X}} b_{UB}(\mathcal{C}) &= \sum_{\mathbf{X} \neq \mathbf{Y} \in \mathcal{C}} \frac{-N \exp[-Nb(\mathbf{F}_i, \mathbf{F}_j)]}{\exp[b_{UB}(\mathcal{C})]} \\
 &\quad \times \sum_{i=1}^T \frac{\text{sign}(\log \lambda_i)}{\lambda_i \mathbf{u}_i^H \mathbf{v}_i} \mathbf{u}_i^H \cdot \frac{\partial \Gamma(\mathbf{F}_i, \mathbf{F}_j)}{\partial \mathbf{X}} \cdot \mathbf{v}_i.
 \end{aligned} \quad (34)$$

Proof: The proof is exactly analogous to the previous theorem, simply taking into account a different chain rule that results in different terms of the gradient of Γ . It is worth noticing that we can make the expression with the absolute value of the logarithm become differentiable by considering b given in terms of $|\log \lambda_i|^{1+\epsilon}$ instead, for small values of ϵ . $\square$

V. RESULTS

A. Initialization and Complexity Analysis

We obtain all our numerical results by carrying out Monte Carlo simulations to estimate the SER performance versus the

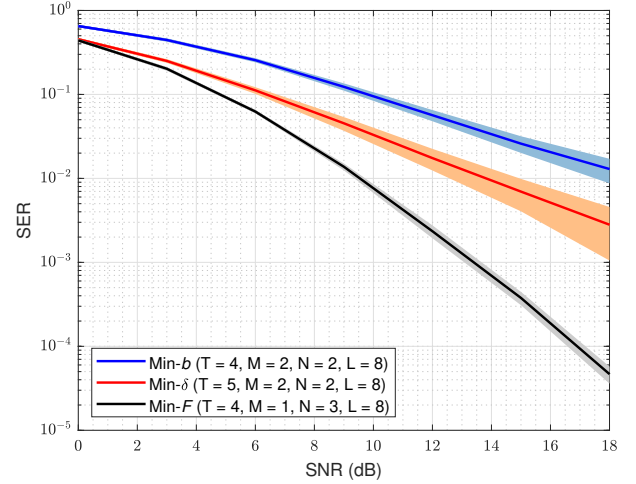


Fig. 1. Confidence intervals of the SER performance starting from 100 random initial points for different design criteria (the Min- F in the Grassmannian, and Min- b and Min- δ in the $\mathbb{T}r$ manifold), for $K = 2$ and $L = 16$ codewords per user.

SNR of the constellations designed by the different proposed criteria using Algorithm 1.

For an ensemble of 100 different initializations, Fig. 1 shows the 68% confidence interval of the SER performance of the different cost functions for a randomized initialization of the algorithm. The initialization process randomly generates normally distributed complex matrices $\mathbf{X}_{k,i}$ for every user $k = 1, \dots, K$, and codeword $i = 1, \dots, L_k$, and then applies the retraction step of the manifold of choice in order to get an initial constellation satisfying the corresponding power constraint. The optimized constellation of the Brehler-Varanasi asymptotic PEP-based design on the Grassmann manifold, Min- F , Eq. (15), shows almost no variation in its performance whereas the Min- b , Eq. (25), and Min- δ , Eq. (24), on the $\mathbb{T}r$ manifold show greater dependence on the random initial constellation. This could be due to the fact that numerical optimization based on the eigenvalue derivatives is likely to be more sensitive to initial conditions, and that the non-USTM manifolds are less constrained spaces in which to choose the initial codebook. In every analysis of this section, we have chosen the best-performing packing of an ensemble of runs of the optimization process. The rest of the parameters of the Riemannian gradient ascent/descent method (e.g., number of iterations or line-search threshold) remain fixed along all optimization runs.

Table I shows a complexity analysis of the complexity order of computing the objective functions and their Riemannian gradients for the different proposed criteria and manifolds. We assume that $|C_k| = 2^B$ for every user $k = 1, \dots, K$, and write $M_{tot} := KM$. The order of complexity of computing the unconstrained gradient of, for example, Eq. (30), is $\mathcal{O}(K2^{(2K+1)B}(3T^3 + T^2 M_{tot} + 2T^2 M + T^2 + T))$. The projection and retraction operations in the oblique manifold do not contribute to leading order since they contribute an additional $\mathcal{O}(K2^B(T^2 + 3TM^2))$. The projection and QR decomposition retraction onto the Grassmannian manifold do contribute to

TABLE I
THE COMPLEXITY ORDER OF COMPUTING THE OBJECTIVE FUNCTION AND ITS RIEMANNIAN GRADIENT FOR DIFFERENT CRITERIA

Criterion	Complexity of computing $f(\mathcal{C})$	Complexity of computing $\nabla f(\mathcal{C})$
Min- F ($\mathbb{G}, \mathcal{OB}, \text{Tr}$)	$\mathcal{O}(2^{KB}(T^2 M_{tot} + 2TM_{tot}^2 + M_{tot}^3))$	$\mathcal{O}(K2^{(K+1)B}(2T^2 M_{tot} + 3TM_{tot}^2 + M_{tot}^3))$
Min- b ($\mathbb{G}$)	$\mathcal{O}(2^{KB}(3T^3 + 2T^2 M_{tot}))$	$\mathcal{O}(K2^{(2K+1)B}(4T^3 + T^2 M_{tot}))$
Min- b ($\mathcal{OB}, \text{Tr}$)	$\mathcal{O}(2^{KB}(3T^3 + 2T^2 M_{tot}))$	$\mathcal{O}(K2^{(2K+1)B}(3T^3 + T^2 M_{tot}))$
Min- δ ($\mathbb{G}$)	$\mathcal{O}(2^{KB}(3T^3 + 2T^2 M_{tot}))$	$\mathcal{O}(K2^{(2K+1)B}(4T^3 + T^2 M_{tot}))$
Min- δ ($\mathcal{OB}, \text{Tr}$)	$\mathcal{O}(2^{KB}(3T^3 + 2T^2 M_{tot}))$	$\mathcal{O}(K2^{(2K+1)B}(3T^3 + T^2 M_{tot}))$

leading order by an additional $\mathcal{O}(K2^B(T^3 + 2T^2 M))$. Notice however that on the trace manifold, the users' constellations are represented at once by a matrix of size $TM2^B \times K$ and the projection and retraction functions are applied to all codewords in one step yielding an additional $\mathcal{O}(4K2^B TM_{tot})$. Therefore, even though the matrix representation of the constellations grows when relaxing the manifold, the contributions to the complexity order of the projection and retraction operations do not happen at leading order. Similar arguments apply to Min- F where in this case we show the complexity of the union bound sum over the one-symbol-in-error terms, following criterion (15).

B. Noncoherent Multiuser Constellation Designs in Non-full Diversity Scenarios

We first assess the performance of the multiuser designs for the MIMO MAC proposed in Sec. IV, Min- b and Min- δ , which are obtained by optimizing the cost functions b_{UB} and δ_{UB} , see Eq. (20) and Eq. (21), and criteria (25) and (24), respectively. We also take into account the different power normalization constraints mentioned in Sec. II-B, so that the optimization is carried out over the corresponding submanifolds, labeled by (Gr), (OB) and (Tr) corresponding to $\mathbb{G}(M, \mathbb{C}^T)$, $\mathcal{OB}_{\mathbb{C}}(TM, \sum_k L_k)$, $\text{Tr}(K, L, M, \mathbb{C}^T)$, a feature that yields outstanding differences in the case of non-full diversity. In every case the SER performance refers to the joint performance of the two users after applying at the BS the ML multiuser detector.

In Fig. 2 the performance of constellations designed over different manifolds is shown for $T = 5$ symbol periods, $M = 2$ transmit antennas, $N = 3$ receive antennas, and $L = 16$ codewords for each user. One can immediately appreciate that both cost functions perform better when the power constraint is relaxed: the USTM codewords from the Grassmann manifold are the worst performing designs, whereas those that constrain the average power are the best performing ones, with the per-codeword power constrained designs yielding an intermediate performance. The oblique and trace manifold constraints result in constellations with performance improvement ranging from half an order to an order of magnitude better SER. Moreover, the Min- δ criterion clearly outperforms the Min- b criterion significantly, except for the Grassmann manifold. Overall, Fig. 2 shows that in the non-full diversity case the manifold on which the optimization is carried out plays a very significant role. In other words, using codewords with different powers may produce codebooks with much better performance in multiuser scenarios.

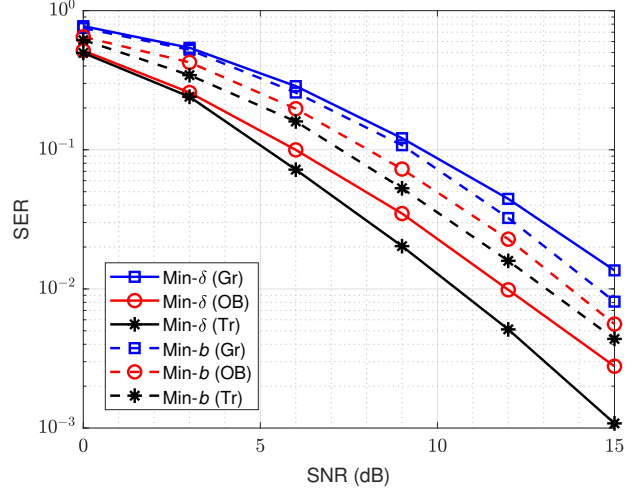


Fig. 2. Multiuser codebook performance of Min- b and Min- δ optimization designs in different Riemannian manifolds. The scenario represents a 2-user MIMO MAC with $T = 5$, $M = 2$, $N = 3$ and $L = 16$.

Using the same settings as in the previous case, Fig. 3 shows how the trace manifold design (the best performing design from the manifold comparison), outperforms the single-user designs and the multiuser Max- $J_{1/2, \min}$ criterion studied in [29] on the Grassmann manifold. In particular, we have used single-user packings optimized on the Grassmann manifold using the minimum chordal distance, Max- $d_{\min}$, and the coherence criterion [36], Max- $c_{\min}$, as explained in [14] and [13]. The figure depicts another coherence constellation, Max- c_{split} , created by optimizing a concatenated single-user constellation of double size, and then splitting it up into two, one half for each user. One would expect that this would help to optimize the cross terms in the union bound of the single-user packings, but the result shows that this type of splitting performs even slightly worse than the plain single-user coherence optimization. Finally, for this setup we are not able to distinguish the $J_{1/2, \min}$ performance from the performance of single-user designs, since the former only outperforms slightly the latter at low SNR.

In Fig. 4 and Fig. 5 the same type of analysis was carried out but for a different scenario, with $T = 4$ symbol periods, $M = 2$ transmit antennas, and a high number of receive antennas $N = 10$. In Fig. 4 the improvement in performance by using the oblique and trace manifolds is outstanding. Moreover, the optimization in the Grassmannian produces designs which are indistinguishable between both cost functions and even show a noise floor at high SNR.

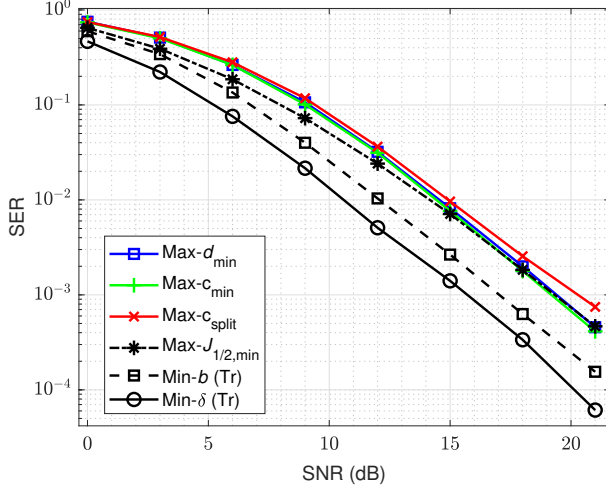


Fig. 3. Multiuser codebook performance for jointly optimized designs, for a 2-user MIMO MAC with $T = 5$, $M = 2$, $N = 3$ and $L = 16$, vs. single user designs based on the chordal and coherence criteria.

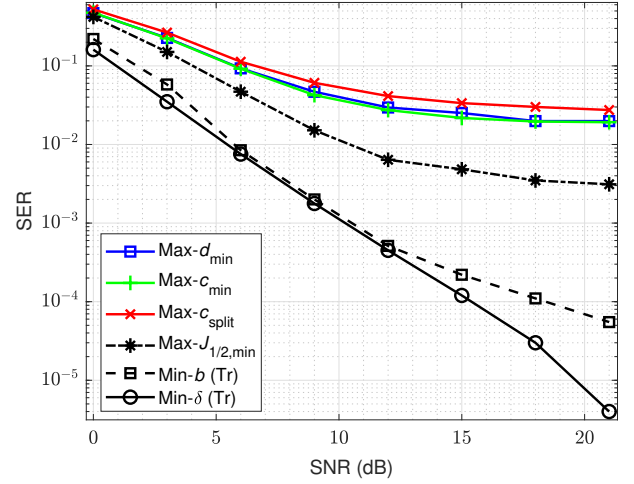


Fig. 5. Performance of jointly optimized designs for a 2-user MIMO MAC with $T = 4$, $M = 2$, $N = 10$ and $L = 16$, vs. single-user designs.

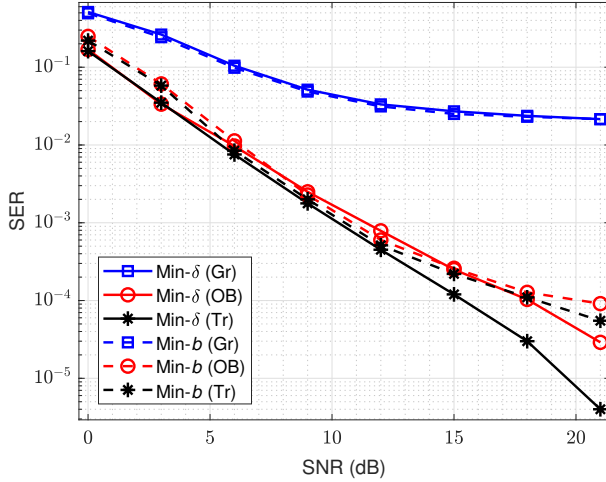


Fig. 4. Performance of b and δ optimization designs in different Riemannian manifolds, for a 2-user MIMO MAC with $T = 4$, $M = 2$, $N = 10$ and $L = 16$.

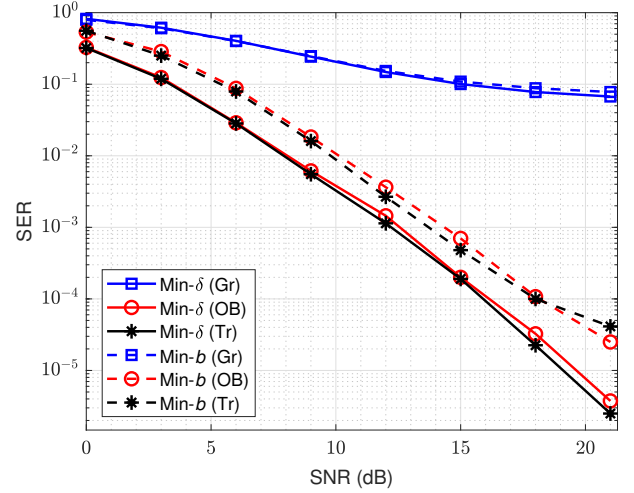


Fig. 6. Performance of multi-user optimization designs in different Riemannian manifolds, for $T = 6$, $M = 3$, $N = 3$ and $L = 16$.

With this high number of receive antennas, the single-user designs also show a noise floor of similar magnitude as for the multiuser USTM designs. Even the multiuser cost function $J_{1/2}$ does not perform well, although the floor is lower than in the other cases. However, the performance of the δ_{UB} and b_{UB} constellations is orders of magnitude better, without showing any noise floor, and it seems to attain a big portion of the diversity. Since these multiuser designs are union bounds of b and δ functions, weighting the exponential on the number of receive antennas N , it seems natural that this parameter plays a role in the designs, as these figures confirm in comparison with the previous two.

In Fig. 6 and 7 the setup considers the same number of antennas at the transmitters and the receiver, $M = N = 3$, with $T = 6$ symbol periods. The same analysis of the previous scenario applies as well, with the single-user designs showing

a noise floor at a SER value of around 10^{-1} , whereas the delta function designs perform down to below 10^{-5} SER at 20 dB. Nevertheless, we can see in Fig. 6 that in this case the optimization on the oblique manifold performs very close to the trace manifold, suggesting that for this configuration the extra degrees of freedom by optimizing the average transmit power instead of using per-codeword power constraints does not significantly affect performance. The $\text{Max-}J_{1/2,min}$ constellation seems to start developing a noise floor near 20 dB, more than an order of magnitude below the single-user noise floor. But just like in the previous cases, the δ function designs consistently outperform all the other packings considered.

For $K = 3$ users, Fig. 8 shows that relaxing the power constraint continues to make a very significant impact on the SER, where the $\text{Min-}\delta$ criterion on the $\mathbb{Tr}$ manifold is again the best performing constellation, as shown in Fig. 9.

We can conclude that the impact of allowing less restrictive constraints on the codeword powers results in a better perfor-

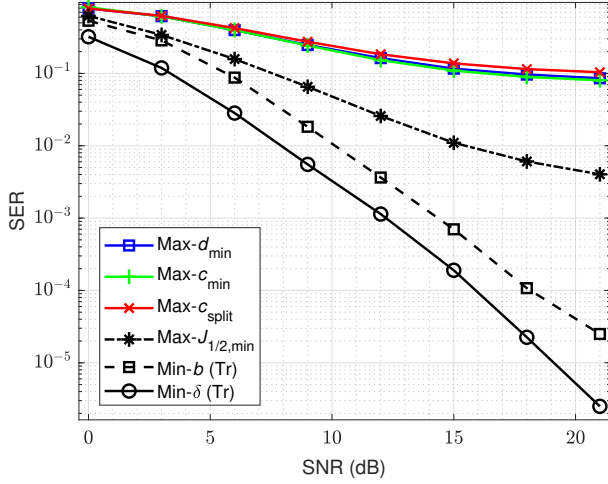


Fig. 7. Performance of jointly optimized designs ($J_{1/2}, b, \delta$) with $T = 6$, $M = 3$, $N = 3$ and $L = 16$, vs. single user designs.

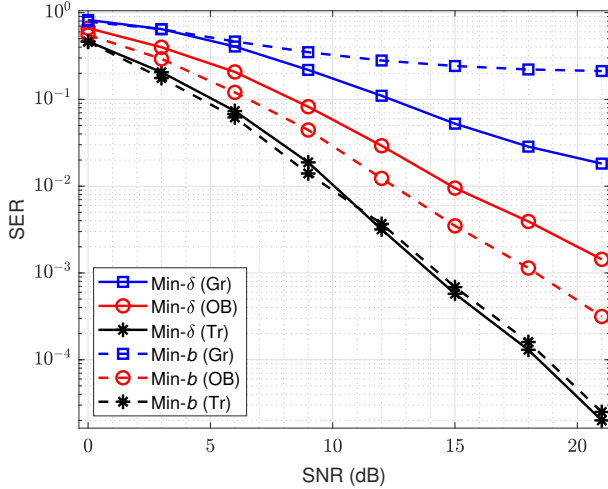


Fig. 8. The joint SER of the Min-b and Min- δ constellations compared for different power constraint manifolds, in a non-full diversity case with $T = 5$, $K = 3$, $L = 8$, $M = 2$, $N = 4$.

mance of the designed constellations when there is no full-diversity in the MAC. This is reasonable since the manifolds on which the codewords are represented and updated during the optimization, have higher dimension the less constrained the power is, i.e. there is more space to approach possible (local) minima of the cost functions. However, this conclusion only holds for non-full-diversity designs, as we shall see next for full-diversity scenarios.

C. Noncoherent Multiuser Constellation Designs in Full-diversity Scenarios

In this subsection, we study the SER performance of the multiuser designs proposed in Sec. III obtained by optimizing the union bound of the dominant term of the Brehler-Varanasi asymptotic PEP formula, Eq. (14), labeled Min- F , (unless stated otherwise this is optimized over the Grassmann manifold). We consider a 2-user MIMO MAC and work under the

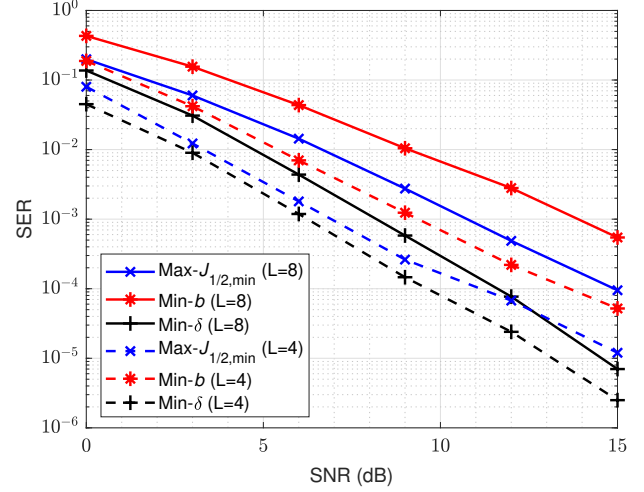


Fig. 9. The joint SER of the cost functions defined in the non-full diversity scenario, optimized over the Tr manifold, with $L = 4$ and $L = 8$ codewords per user, with $T = 7$, $K = 3$, $M = 2$, $N = 4$.

following assumptions: i) the two users have the same average SNR, and ii) there is no correlation between the channel fading coefficients. Moreover, the formula of interest is only valid in the full-diversity case, meaning that only scenarios with $T \geq (K + 1)M$ shall be analyzed here. Moreover, only the terms in the union bound corresponding to a single user in error are considered. There are two reasons for this simplification: first, the terms with only one user in error dominate the PEP expression; and second, this reduces the computational complexity dramatically, as explained in Sec. III.

In Fig. 10 the case of $T = 3$ symbol periods, $M = 1$ transmitting antennas, $N = 3$ receiving antennas, and $B = 4$ bits per symbol is studied and compared versus single-user constellations. Two multiuser designs outperform these single-user constellations: the proposed union bound optimization of criterion (15) and a min-max criterion (labeled Min- F_{max}), minimizing the dominant term of the F function using the $\max_i x_i \approx \epsilon \log \sum_i \exp(x_i/\epsilon)$ approximation. One can understand that improving at every iteration the dominant term out of the possible pairwise probability errors ought to yield performance gains, which indeed is the case as shown by the dashed curve vs. the single user constellations. However, the union bound optimization clearly outperforms this by around 2.5 dB at $\text{SER} = 10^{-3}$. This is expected since a union bound method minimizes all terms of the possible error probabilities at the same time.

A very similar configuration is shown in Fig. 11, where we consider $B = 5$ bits per codeword, $N = 4$ receive antennas, and $T = 4$. In this case, the gap between the min-max method and the union bound reduces, but the latter still provides the best results. It is interesting to note that the coherence criterion [14], [36] for single-user constellations outperforms the chordal distance criterion in the multiuser scenario, a behavior that was not so evident in the previous figure.

We may conclude from the previous two figures that single-

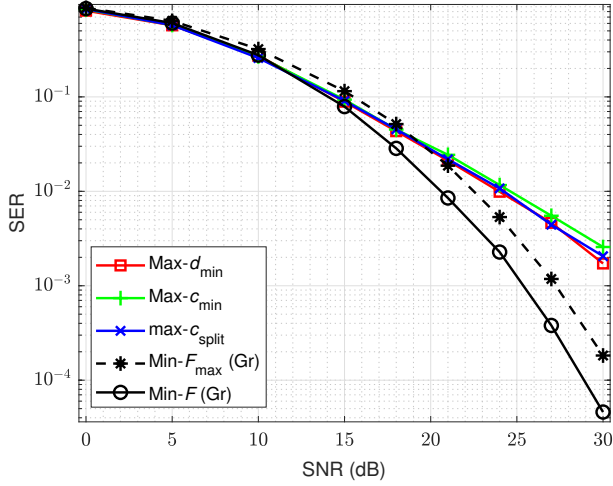


Fig. 10. Performance of a jointly optimized design based on the Brehler-Varanasi asymptotic PEP union bound, with $T = 3$, $M = 1$, $N = 3$ and $L = 16$ vs. single user designs and compared to minimizing the dominant asymptotic PEP.

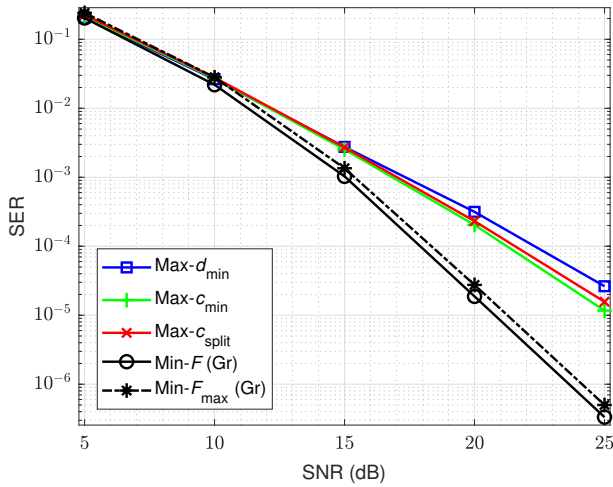


Fig. 11. Performance of jointly optimized designs based on the PEP union bound, with $T = 4$, $M = 1$, $N = 4$ and $L = 32$, vs. single user designs.

user codebooks do not only perform worse in term of SER at any given SNR, but also they do not achieve the same slope as the multiuser constellations. On the other hand, the multiuser codebooks designed with either the UB or a max-min approach attain the full-diversity of the system MN for both users. In comparison to the min-max approach the UB criterion provides some coding gain, a shift to the left of the SER vs. SNR curve.

The impact of using different manifolds in the joint union bound criterion can be seen in Fig. 12. Essentially there are not significant differences in performance. Moreover, for full-diversity scenarios the Grassmannian constellations seem to perform slightly better the higher the spectral efficiency is.

It is important to point out that the performance of a given optimized constellation in the MAC depends on many parameters: number of users, number of antennas, coherence

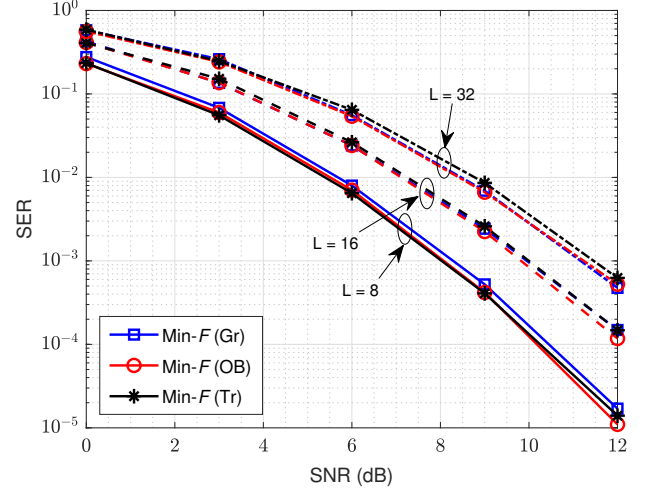


Fig. 12. Comparison of the multi-user constellation designs optimizing the Brehler-Varanasi union bound on different manifolds, for $T = 6$, $M = 2$, $N = 4$ and $L \in \{8, 16, 32\}$.

time, etc. In particular, the number of receive antennas affects the diversity (the slope of the SER vs. SNR curve) and hence can lead to important differences in performance. This is illustrated in Fig. 13, which compares the performance of all joint constellation design methods proposed in this report. The full-diversity scenario is a 2-user MAC with $T = 3$, $M = 1$, $N = 5$ and codebooks of cardinality $L = 32$. It is remarkable that the behaviour of the union bound codebook is extremely good at high SNR, as expected from the theoretical result it rests upon, and much better than any other method, including the optimization of the $\text{Min-}F_{\max}$. We can also compare in this plot the performance of the $\text{Min-}b$ and $\text{Min-}\delta$ designs, providing evidence that they are not suitable criteria for full-diversity scenarios. In fact, they do not even reach the single-user designs' performance. For full-diversity scenarios, the $\text{Min-}b$ and $\text{Min-}\delta$ designs appear to develop a noise floor at very high SNR, whereas the union bound criterion does not show any noise floor and attains the full diversity of the system MN .

In Fig. 14 we compare the multiuser design criteria in a full-diversity scenario with $T = 6$, $M = 2$ and different number of receive antennas and bit rates. It is worth mentioning that the gap in performance between the Brehler-Varanasi asymptotic PEP and the Ngo-Yang proxy functions seems to get reduced when increasing the number of bits per codeword. Still, the former outperforms all designs studied so far. Moreover, since $\text{Max-}J_{1/2, \min}$ is the best-performing criterion out of five shown in [29], we can conclude that our designs provide state-of-the-art multiuser constellations for the MAC.

Finally, in the case of three users, $K = 3$, we see in Fig. 15 that the same qualitative and quantitative results hold: the union bound optimization $\text{Min-}F$ on the Grassmannian manifold outperforms all other designs, which are more and more similar for increasing bit rate per user.

From these results and those of the previous subsection, we can conclude two major findings: i) in the non-full diversity

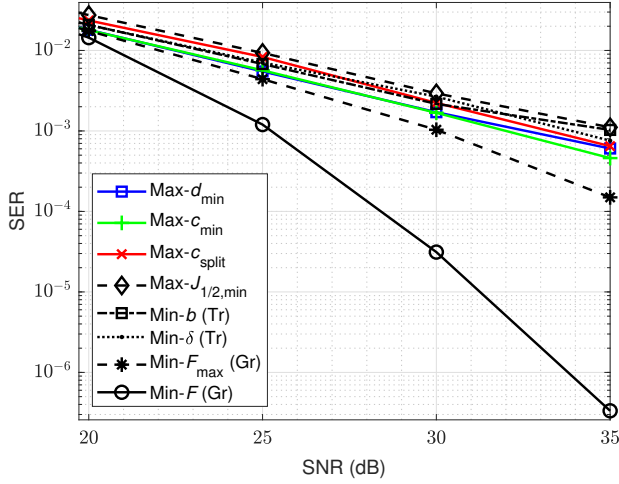


Fig. 13. Comparison of all the different constellation designs based on the proposed multi-user cost functions, for $T = 3$, $M = 1$, $N = 5$ and $L = 32$, vs. the single user designs.

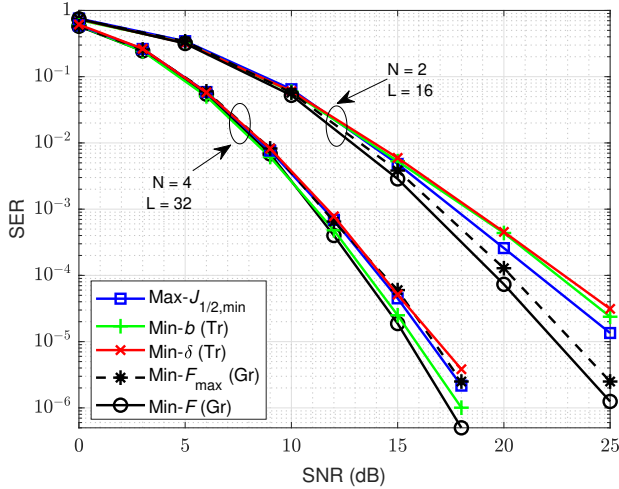


Fig. 14. Comparison of the different multi-user constellation designs based on the proposed multi-user cost functions, for $T = 6$, $M = 2$, $N \in \{2, 4\}$ and $L \in \{16, 32\}$.

case the criterion $\text{Min-}\delta$ on the trace manifold is the most suitable design, whereas in the full-diversity case the criterion $\text{Min-}F$ on the Grassmann manifold is the best performing one; ii) relaxing the Riemannian constraint on the per-user codewords, from the Grassmannian manifold to the oblique and trace manifolds, improves the error rate only in the non-full diversity case. The first conclusion can be understood by noticing that the $\text{Min-}F$ criterion is motivated by an exact formula for the asymptotic PEP that seems to be a tighter bound than the bounds of the $\text{Min-}b$ and $\text{Min-}\delta$ criteria. The second conclusion suggests that in full-diversity scenarios the diversity slope of the SER as a function of the SNR is already achieved by Grassmannian codewords, i.e. USTM, so allowing for changes among the symbol powers subject to a total power constraint has little, if any, effect on an already good full-diversity constellation, whereas power variations in non-

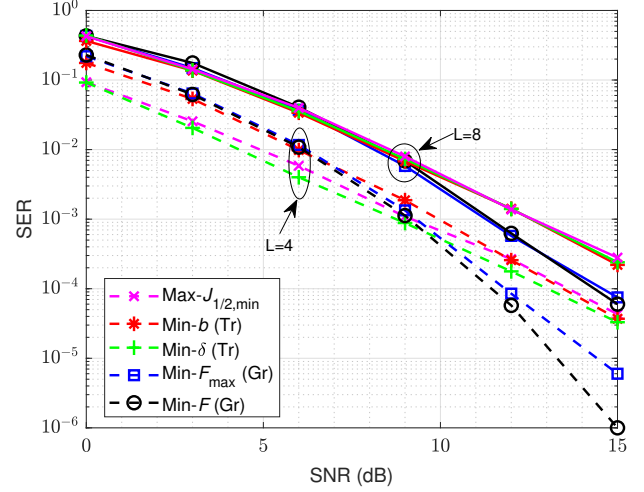


Fig. 15. The joint SER of all the considered cost functions in the three-user case $K = 3$, for two full diversity cases with $L = 4$ and $L = 8$, and $T = 8$, $M = 2$, $N = 3$.

full diversity offer extra degrees of freedom to optimize the constellation performance when the terms of the asymptotic PEP are not full rank.

D. Performance Analysis under Correlated Channels

In this subsection, we analyze the performance of the proposed designs under Rayleigh correlated channels. As a representative scenario, we consider an uplink channel with two users having one antenna, $M = 1$, transmitting information to a BS equipped with $N = 4$ antennas such that there is correlation only at the receiver side. The single-input multiple-output (SIMO) uplink channels are distributed as $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \Sigma)$, $i = 1, 2$. We consider the well-known exponential correlation model $\Sigma = \{\rho_c^{|i-j|}\}_{i,j=1,\dots,N}$, where $0 \leq \rho_c \leq 1$ is the correlation coefficient measuring the correlation between consecutive antennas in the BS array [37]. Note that the objective is to assess the robustness of the designed noncoherent MAC constellations, which assume that the channels are uncorrelated, when in fact there is correlation at the receiver side. MAC constellations could be designed for a given, fixed, correlation matrix, Σ . This problem, however, is considered a topic for future research. Fig. 16 shows the impact of correlation on the performance of the constellation designed with the $\text{Min-}F$ criterion. It should be noted that although the constellation has been designed considering an uncorrelated channel (therefore $\rho_c = 0$), the BS applies the optimal multiuser detector assuming Σ known (cf. (6)). It is observed that a correlation value of $\rho_c = 0.5$ has a minor impact on the SER. For the degradation to be significant, the uplink channels must be highly correlated with $\rho_c = 0.9$, a scenario that is considered unrealistic. There is therefore a certain robustness against correlation of the designed constellations.

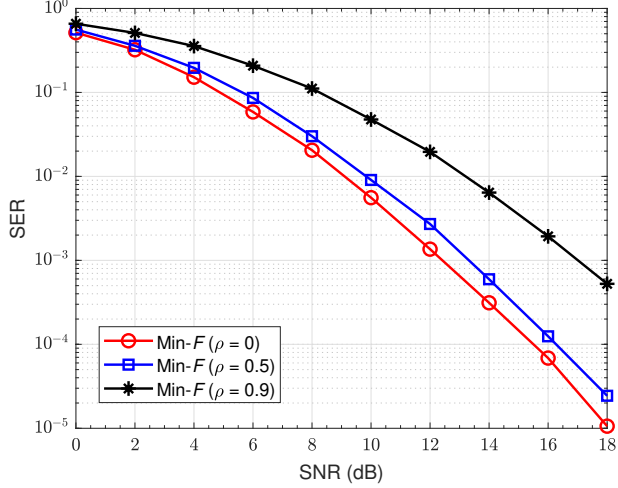


Fig. 16. Performance under correlated Rayleigh channels of the MAC constellation in a full-diversity case (optimized over the Grassmannian manifold). The parameters of the full-diversity scenario are $T = 4$, $K = 2$, $M = 1$, $N = 4$ and $L = 16$ symbols per user, and ρ_c denotes the correlation coefficient of the exponential model.

E. Comparison with Pilot-based Schemes

In this section, we compare the SER performance of different multiuser designs for the MIMO MAC for full diversity (Min- F design in the Grassmann manifold) and non-full diversity scenarios (Min- δ design in the $\mathbb{T}r$ manifold) with a coherent pilot-based scheme.

The transmitted signals for the pilot-based scheme contain orthogonal pilot sequences followed by spatially multiplexed QAM data symbols. The cardinality of the QAM constellation from which the data symbols are taken is chosen so that the coherent scheme has the same spectral efficiency as the noncoherent designs. The transmitted signals are normalized so that the average transmit power of the pilot-based scheme is the same as that of the noncoherent schemes.

Fig. 17 shows the SER curves of the best-performing constellation in the full-diversity case, which is the Brehler-Varanasi asymptotic PEP-based design in the Grassmann manifold, in comparison with the pilot-based scheme described above for $T = 4$, $M = 1$, $N = 4$, $K = 2$ users and different constellation sizes L_1 and L_2 . As we can observe, in all cases the Min- F design clearly outperforms the pilot-based scheme, with a power gain of about 2.5 dB for $L_1 = L_2 = 4$, 2 dB for $L_1 = L_2 = 16$, and 5 dB for $L_1 = 64$ and $L_2 = 4$.

In Fig. 18, we compare the SER curves of the best-performing constellation in the non-full diversity case, which is the Min- δ design in the $\mathbb{T}r$ manifold, with the pilot-based scheme described above for $T = 5$, $M = 2$, $N = 3$, $K = 2$ users and different constellation sizes L_1 and L_2 . We can see again that the noncoherent multiuser design has a significantly better SER performance than the pilot-based scheme. We also notice that the power gains are much higher in this case than in the full-diversity scenario, with about 6 dB for $L_1 = L_2 = 4$ and $L_1 = L_2 = 8$, and 7.5 dB for $L_1 = 64$ and $L_2 = 4$.

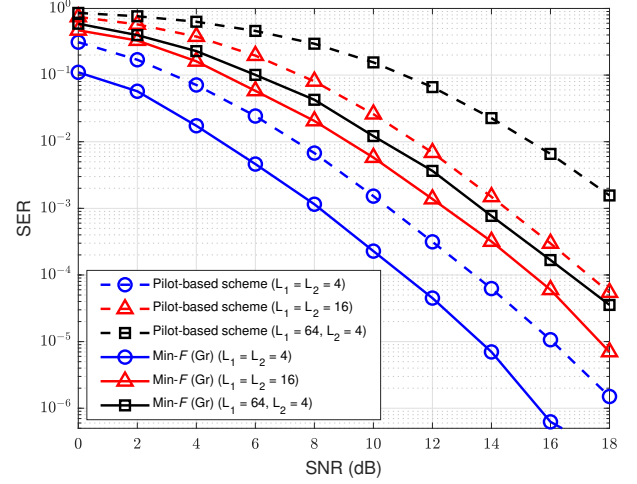


Fig. 17. Comparison of the best-performing constellation in the full-diversity case (Brehler-Varanasi asymptotic PEP-based design in the $\mathbb{G}r$ manifold) vs. pilot-based schemes, for $T = 4$, $K = 2$, $M = 1$, $N = 4$, and $L_1 \in \{4, 16, 64\}$, $L_2 \in \{4, 16\}$.

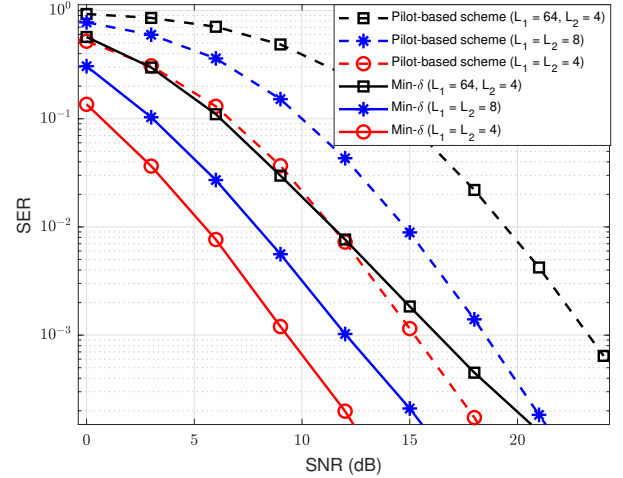


Fig. 18. Comparison of the best-performing constellation in the non-full diversity case (Min- δ design in the $\mathbb{T}r$ manifold) vs. pilot-based schemes, for $T = 5$, $K = 2$, $M = 2$, $N = 3$, and $L_1 \in \{4, 8, 64\}$, $L_2 \in \{4, 8\}$.

F. Achievable Rate Analysis

In this subsection, we study the achievable rates obtained by the proposed designs in full-diversity and non-full-diversity scenarios. Let us consider for simplicity a 2-user MAC channel where the two users have the same SNR. The codewords sent by user 1 and user 2 are selected equally likely from their respective codebooks $\mathcal{C}_1 = \{\mathbf{X}_{1,1}, \dots, \mathbf{X}_{1,L_1}\}$ and $\mathcal{C}_2 = \{\mathbf{X}_{2,1}, \dots, \mathbf{X}_{2,L_2}\}$, designed according to one of the criteria described in this paper. The following pairs of rates in

b/s/Hz (R_1, R_2) are achievable

$$\begin{aligned} R_1 &\leq \frac{1}{T} I(\mathbf{X}_1; \mathbf{Y} | \mathbf{X}_2), \\ R_2 &\leq \frac{1}{T} I(\mathbf{X}_2; \mathbf{Y} | \mathbf{X}_1), \\ R_1 + R_2 &\leq \frac{1}{T} I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}), \end{aligned} \quad (35)$$

where $I(\mathbf{X}; \mathbf{Y}) = E \left[\log_2 \frac{p(\mathbf{Y} | \mathbf{X})}{p(\mathbf{Y})} \right]$ denotes the mutual information between the channel input and output. The unconditional probability density function of the observations is a mixture of Gaussian densities

$$\begin{aligned} p(\mathbf{Y}) &= \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} p(\mathbf{X}_{1,i}, \mathbf{X}_{2,j}) p(\mathbf{Y} | \mathbf{X}_{1,i}, \mathbf{X}_{2,j}) \\ &= \frac{1}{L_1 L_2} \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} \frac{\exp(-\text{tr}(\mathbf{Y}^H \mathbf{R}_{ij}^{-1} \mathbf{Y}))}{(\pi^T \det(\mathbf{R}_{ij}))^N}, \end{aligned}$$

where $\mathbf{R}_{ij} = \mathbf{X}_{1,i} \mathbf{X}_{1,i}^H + \mathbf{X}_{2,j} \mathbf{X}_{2,j}^H + \mathbf{I}_T$ is the covariance matrix of the (independent) columns of the complex Gaussian matrix $\mathbf{Y}$ conditioned on $\mathbf{X}_{1,i}, \mathbf{X}_{2,j}$. The expressions for the conditional densities can be derived in a similar way. The expectations to compute the mutual information in (35) unfortunately do not have a closed form; therefore we resort to Monte Carlo simulations to estimate R_1, R_2 , and the sum-rate $R_1 + R_2$. Fig. 19 shows the achievable rate for user 1, R_1 , (user 2 achieves the same rate since the two users have the same SNR), and the achievable sum rate, $R_1 + R_2$, for constellations designed over different manifolds with the Min- δ criterion. This first example considers a full-diversity scenario with parameters $T = 6, M = 2, N = 2, K = 2$, and $L_1 = L_2 = 16$. As we observed in the SER curves, Fig. 19 shows that in full-diversity scenarios, optimizing on manifolds with more degrees of freedom than the Grassmann manifold is not advantageous in terms of achievable rate. A different behavior is observed in scenarios without full-diversity, as shown in Fig. 20 representing the achievable rates of noncoherent constellations optimized over the Grassmann, oblique, and trace manifolds when $T = 5, M = 2, N = 3, K = 2$, and $L_1 = L_2 = 16$. In non-full-diversity scenarios, there may be a significant increase in achievable rates when the optimization is performed on manifolds of higher dimensionality than the Grassmannian. These results suggest that SER improvements provided by the δ function optimized over the trace manifold also translate into achievable rate improvements.

VI. CONCLUSION

In this paper we have developed Riemannian optimization techniques for designing noncoherent constellations for the MIMO MAC. In particular, we have developed optimized multiuser space-time codebooks for full-diversity ($T \geq (K+1)M$) and non-full diversity scenarios ($T < (K+1)M$). For full-diversity scenarios, the cost function is a union bound of the dominant terms (i.e., those terms corresponding to the case where only one of the users of the MAC channel is in error) of the asymptotic PEP. For non-full diversity scenarios the PEP expression is no longer valid and therefore we use

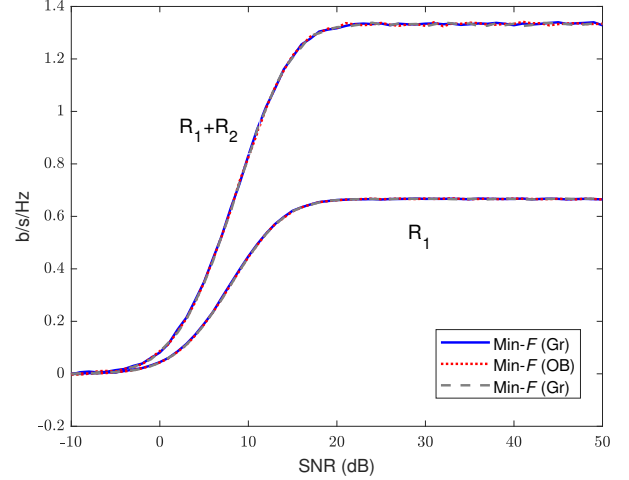


Fig. 19. Achievable rate for user 1 (R_1) and achievable sum-rate ($R_1 + R_2$) vs. SNR for full-diversity constellations optimized over the different Riemannian manifolds for $T = 6, M = 2, N = 2, K = 2, L_1 = L_2 = 16$.

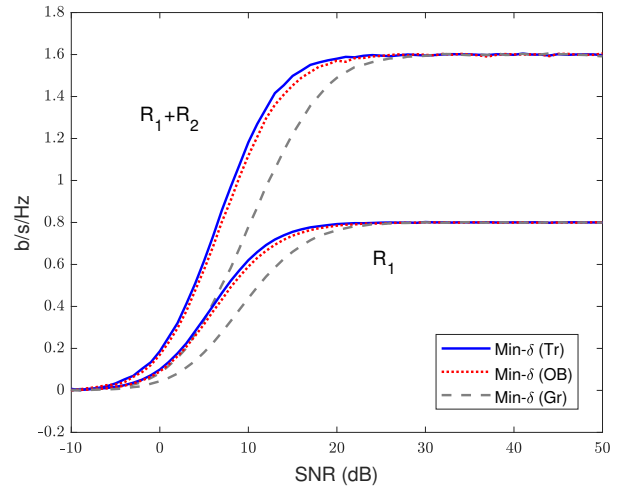


Fig. 20. Achievable rate for user 1 (R_1) and achievable sum-rate ($R_1 + R_2$) vs. SNR for non-full diversity constellations optimized over the different power constraint manifolds for $T = 5, M = 2, N = 3, K = 2, L_1 = L_2 = 16$.

union bounds of some recently proposed proxies of the PEP, called the δ and b functions, as design criteria. The proposed cost functions and the corresponding Riemannian optimization techniques are valid for any number of users.

In addition to the traditional Grassmann manifold, which is optimal only in the single-user case, we consider the optimization of multiuser codebooks in other Riemannian manifolds corresponding to different power constraints on the codewords. We show that the manifold on which the optimization is performed can have a significant impact on performance, especially in non-full diversity scenarios. Our results suggest that in non-full diversity case the δ cost function optimized on the trace manifold, corresponding to an average power constraint, is the best performing design. Whereas in the full-diversity case the best performing constellations in terms of

symbol error rate are those designed using the dominant factor of the asymptotic joint PEP on the Grassmann manifold. Future lines of work include the development of an asymptotic PEP formula for non-full diversity scenarios that would avoid the use of proxies in this case and the study of noncoherent schemes for the broadcast channel.

APPENDIX

A. Riemannian Manifolds

The complex Grassmannian $\mathbb{G}(M, \mathbb{C}^T)$ is the set of M -dimensional complex subspaces of $\mathbb{C}^T$, with $T > M$, that is a complex manifold of dimension $M(T - M)$. Elements in $\mathbb{G}(M, \mathbb{C}^T)$ are represented by matrices in the Stiefel manifold $\mathbf{A} \in \mathbb{S}_t(M, \mathbb{C}^T)$, that is $\mathbf{A} \in \mathbb{C}^{T \times M}$, $\mathbf{A}^H \mathbf{A} = \mathbf{I}_M$. This representation is not unique, since $\mathbf{A}$ and $\mathbf{A}\mathbf{U}$ with $\mathbf{U}$ a unitary $M \times M$ matrix represent the same element in $\mathbb{G}(M, \mathbb{C}^T)$, so formally we should denote elements of the Grassmannian as $[\mathbf{A}]$ where $\mathbf{A} \in \mathbb{S}_t(M, \mathbb{C}^T)$ is a unitary basis for that subspace, $\mathbf{P}_\mathbf{A} = \mathbf{A}\mathbf{A}^H$ denotes the orthogonal projection onto $[\mathbf{A}]$ and $[\mathbf{A}]$ is the class of $\mathbf{A}$ under the quotient by the set of $M \times M$ unitary matrices $\mathcal{U}_M$. Mathematically, this defines a Riemannian structure on the Grassmannian given by the *Riemannian submersion*

$$\begin{aligned} \pi : \mathbb{S}_t(M, \mathbb{C}^T) &\rightarrow \mathbb{G}(M, \mathbb{C}^T) = \mathbb{S}_t(M, \mathbb{C}^T)/\mathcal{U}_M \\ \mathbf{A} &\mapsto [\mathbf{A}]. \end{aligned}$$

Sometimes we will consider the optimization of a real function φ whose argument can be either a complex matrix in the ambient space $\mathbf{X} \in \mathbb{C}^{T \times M}$, a Stiefel matrix $\mathbf{X} \in \mathbb{S}_t(M, \mathbb{C}^T)$, or a point in the Grassmannian $[\mathbf{X}] \in \mathbb{G}(M, \mathbb{C}^T)$. We will denote the function generically as $\varphi(\mathbf{X})$, meaning for the Grassmann that $\varphi(\mathbf{X}\mathbf{U}) = \varphi(\mathbf{X})$ for any $M \times M$ unitary matrix $\mathbf{U}$, and employ the notation $D\varphi(\mathbf{X})$ to denote the unconstrained derivative of the function in the ambient space, and $\nabla\varphi(\mathbf{X})$ to denote the gradient of the function on the tangent space of the Grassmannian. In both cases it will be understood that the derivative or the gradient is evaluated at $\mathbf{X}$ or $[\mathbf{X}]$, respectively. In particular, these derivatives play two roles: on the one hand, the unconstrained derivative $D\varphi(\mathbf{X})$, depending only on a point $\mathbf{X}$, is the Jacobian matrix of φ with respect to the components X_{mn} of $\mathbf{X}$, for $m = 1, \dots, T$, $n = 1, \dots, M$, i.e., as a matrix it has complex components given by

$$D\varphi(\mathbf{X})_{mn} = \frac{\partial\varphi}{\partial X_{mn}} = \frac{\partial\varphi}{\partial \Re(X_{mn})} + i \frac{\partial\varphi}{\partial \Im(X_{mn})}.$$

On the other hand, these derivatives, when depending both on a point $\mathbf{A}$ and a tangent vector $\dot{\mathbf{A}}$, are to be understood as directional derivatives in their respective tangent spaces, for example

$$D\varphi(\mathbf{A})(\dot{\mathbf{A}}) = \lim_{t \rightarrow 0} \frac{\varphi(\mathbf{A} + t\dot{\mathbf{A}}) - \varphi(\mathbf{A})}{t} = \left. \frac{d}{dt} \right|_{t=0} \varphi(\mathbf{A} + t\dot{\mathbf{A}}).$$

With this interpretation we can define partial derivatives of φ with respect to the real and imaginary part of every direction in the tangent space and thus arrive at the Jacobian matrix again. The relationship between both objects in the ambient space can be verified to be:

$$D\varphi(\mathbf{A})(\dot{\mathbf{A}}) = \Re(\langle D\varphi(\mathbf{A}), \dot{\mathbf{A}} \rangle_F), \quad (36)$$

a property which will serve as requirement for the definition of gradient vector $\nabla\varphi$ on a general manifold (see Corollary 1).

The tangent space to the Stiefel manifold at $\mathbf{A} \in \mathbb{S}_t(M, \mathbb{C}^T)$ is easy to describe from the defining equation $\mathbf{A}^H \mathbf{A} = \mathbf{I}_M$:

$$\begin{aligned} T_\mathbf{A} \mathbb{S}_t(M, \mathbb{C}^T) &= \\ &= \left\{ \dot{\mathbf{A}} \in \mathbb{C}^{T \times M} : \left. \frac{d}{dt} \right|_{t=0} ((\mathbf{A} + t\dot{\mathbf{A}})^H (\mathbf{A} + t\dot{\mathbf{A}})) = 0 \right\} \\ &= \{ \dot{\mathbf{A}} \in \mathbb{C}^{T \times M} : \dot{\mathbf{A}}^H \mathbf{A} + \mathbf{A}^H \dot{\mathbf{A}} = \mathbf{0} \}. \end{aligned}$$

The Riemannian submersion π allows us to identify the tangent space to the Grassmannian with the orthogonal to the kernel of π ; in other words,

$$\begin{aligned} T_{[\mathbf{A}]} \mathbb{G}(M, \mathbb{C}^T) &\equiv \{ \dot{\mathbf{A}} \in T_\mathbf{A} \mathbb{S}_t(M, \mathbb{C}^T) : \dot{\mathbf{A}} \perp \mathbf{A}\dot{\mathbf{U}}, \\ &\quad \text{for all } \dot{\mathbf{U}} \in T_{\mathbf{I}_M} \mathcal{U}_M \} \\ &= \{ \dot{\mathbf{A}} \in \mathbb{C}^{T \times M} : \dot{\mathbf{A}}^H \mathbf{A} + \mathbf{A}^H \dot{\mathbf{A}} = \mathbf{0}, \langle \dot{\mathbf{A}}, \mathbf{A}\dot{\mathbf{U}} \rangle_F = 0, \\ &\quad \forall \dot{\mathbf{U}} : \dot{\mathbf{U}} + \dot{\mathbf{U}}^H = \mathbf{0} \} \\ &= \{ (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H) \dot{\mathbf{B}} : \dot{\mathbf{B}} \in \mathbb{C}^{T \times M} \}. \end{aligned}$$

For this note that both spaces have the same (complex) dimension $M(T - M)$ and that the latter is included in the former since for $\dot{\mathbf{B}} \in \mathbb{C}^{T \times M}$, taking $\dot{\mathbf{A}} = (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H) \dot{\mathbf{B}}$, we have:

$$\begin{aligned} \dot{\mathbf{A}}^H \mathbf{A} + \mathbf{A}^H \dot{\mathbf{A}} &= \\ \dot{\mathbf{B}}^H (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H) \mathbf{A} + \mathbf{A}^H (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H) \dot{\mathbf{B}} &= \mathbf{0}, \\ \langle \dot{\mathbf{A}}, \mathbf{A}\dot{\mathbf{U}} \rangle_F &= \langle (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H) \dot{\mathbf{B}}, \mathbf{A}\dot{\mathbf{U}} \rangle_F \\ &= \langle \mathbf{A}^H (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H) \dot{\mathbf{B}}, \dot{\mathbf{U}} \rangle_F = 0. \end{aligned}$$

We thus obtain

$$T_{[\mathbf{A}]} \mathbb{G}(M, \mathbb{C}^T) \equiv \{ (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H) \dot{\mathbf{B}} : \dot{\mathbf{B}} \in \mathbb{C}^{T \times M} \}.$$

The last set obviously does not depend on the chosen representative for $[\mathbf{A}]$.

The following lemma is fundamental for the computation of gradients in the Grassmannian that will be used in the UB optimization algorithm.

Corollary 1 *Let $\varphi : \mathbb{C}^{T \times M} \rightarrow \mathbb{R}$ be a C^1 mapping, defined at least in some open neighborhood of the Stiefel manifold $\mathbb{S}_t(M, \mathbb{C}^T) \subseteq \mathbb{C}^{T \times M}$, and assume that φ can be defined as a function on $\mathbb{G}(M, \mathbb{C}^T)$, that is, we have:*

$$\varphi(\mathbf{A}) = \varphi(\mathbf{A}\mathbf{U}) \text{ for } \mathbf{A} \in \mathbb{S}_t(M, \mathbb{C}^T), \mathbf{U} \in \mathcal{U}_M.$$

Then, the gradient of φ at $\mathbf{A} \in \mathbb{S}_t(M, \mathbb{C}^T)$ as a Grassmannian mapping is:

$$\nabla\varphi(\mathbf{A}) = (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H) D\varphi(\mathbf{A}),$$

where $D\varphi$ is the unconstrained gradient of φ as a function on the ambient space $\mathbb{C}^{T \times M}$.

Proof: By definition, the gradient $\nabla\varphi(\mathbf{A})$ is the unique element of $T_{[\mathbf{A}]} \mathbb{G}(M, \mathbb{C}^T)$ such that for all $\dot{\mathbf{A}} \in T_{[\mathbf{A}]} \mathbb{G}(M, \mathbb{C}^T)$:

$$D\varphi(\mathbf{A})(\dot{\mathbf{A}}) = \Re\langle \nabla\varphi(\mathbf{A}), \dot{\mathbf{A}} \rangle.$$

Let $\dot{\mathbf{A}} = (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H)\dot{\mathbf{B}} \in T_{[\mathbf{A}]} \mathbb{G}(M, \mathbb{C}^T)$ and note that

$$\begin{aligned} \Re\langle (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H)D\varphi(\mathbf{A}), \dot{\mathbf{A}} \rangle &= \\ &= \Re\langle (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H)D\varphi(\mathbf{A}), (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H)\dot{\mathbf{B}} \rangle \\ &= \Re\langle D\varphi(\mathbf{A}), (\mathbf{I}_T - \mathbf{A}\mathbf{A}^H)\dot{\mathbf{B}} \rangle = \Re\langle D\varphi(\mathbf{A}), \dot{\mathbf{A}} \rangle \\ &= D\varphi(\mathbf{A})(\dot{\mathbf{A}}), \end{aligned}$$

and since $(\mathbf{I}_T - \mathbf{A}\mathbf{A}^H)D\varphi(\mathbf{A})$ is an element of $T_{[\mathbf{A}]} \mathbb{G}(M, \mathbb{C}^T)$, it satisfies the definition of gradient. $\square$

ACKNOWLEDGMENTS

This work was supported by Huawei Technologies, Sweden under the project GRASSCOM. The work of D. Cuevas was also partly supported under grant FPU20/03563 funded by Ministerio de Universidades (MIU), Spain. The work of Carlos Beltrán was also partly supported under grant PID2020-113887GB-I00 funded by MCIN/AEI/10.13039/501100011033. The work of I. Santamaria was also partly supported under grant PID2019-104958RB-C43 (ADELE) funded by MCIN/AEI/10.13039/501100011033.

REFERENCES

- [1] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, 1999.
- [2] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communication in Rayleigh flat-fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 543–564, 2000.
- [3] L. Zheng and D. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, 2002.
- [4] B. Hochwald, T. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1962–1973, 2000.
- [5] B. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 2041–2052, 2000.
- [6] B. Hughes, "Differential space-time modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2567–2578, 2000.
- [7] M. Beko, J. Xavier, and V. A. N. Barros, "Noncoherent communications in multiple-antenna systems: receiver design and codebook construction," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5703–5715, 2007.
- [8] R. H. Gohary and T. N. Davidson, "Noncoherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1176–1205, 2009.
- [9] I. Kammoun, A. M. Cipriano, and J. Belfiore, "Non-coherent codes over the Grassmannian," *IEEE Trans. Wireless Comm.*, vol. 6, no. 10, pp. 3657–3667, 2007.
- [10] W. Zhao, G. Leus, and G. B. Giannakis, "Orthogonal design of unitary constellations for uncoded and trellis-coded noncoherent space-time systems," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1319–1327, 2004.
- [11] G. Han and J. Rosenthal, "Geometrical and numerical design of structured unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3722–3735, 2006.
- [12] K. Ngo, A. Decurninge, M. Guillaud, and S. Yang, "Cube-split: A structured Grassmannian constellation for non-coherent SIMO communications," *IEEE Trans. Wireless Comm.*, vol. 19, no. 3, pp. 1948–1964, 2020.
- [13] D. Cuevas, C. Beltrán, I. Santamaria, V. Tucek, and G. Peters, "A fast algorithm for designing Grassmannian constellations," in *25th International ITG Workshop on Smart Antennas (WSA 2021)*, (EURECOM, France), Nov. 2021.
- [14] J. Álvarez-Vizoso, D. Cuevas, C. Beltrán, I. Santamaria, V. Tucek, and G. Peters, "Coherence-based subspace packings for MIMO noncoherent communications," *30th Eur. Sig. Proc. Conf. (EUSIPCO 2022)*, (Belgrade, Serbia), Sept. 2022.
- [15] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experimental Mathematics*, vol. 5, no. 2, pp. 139–159, 1996.
- [16] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communication*, vol. 21, pp. 684–702, June 2003.
- [17] S. Shamai and T. L. Marzetta, "Multiuser capacity in block fading with no channel state information," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 938–942, 2002.
- [18] Z. Utkovski, D. Ilik, and L. Kocarev, "An achievable pre-log region for the non-coherent block fading MIMO multiple access channel," *10th Int. Symp. Wireless Communication System*, Aug. 2013.
- [19] K. H. Ngo, S. Yang, and M. Guillaud, "The optimal DoF region for the two-user noncoherent SIMO multiple-access channel," *IEEE Information Theory Workshop (ITW)*, Nov. 2018.
- [20] A. Manolakis, M. Chowdhury, and A. Goldsmith, "Energy-based modulation for noncoherent SIMO systems," *IEEE Trans. Wireless Commun.*, vol. 11, pp. 7831–7846, Nov. 2016.
- [21] M. Chowdhury, A. Manolakis, and A. Goldsmith, "Scaling laws for noncoherent energy-based communications in the SIMO MAC," *IEEE Trans. Inf. Theory*, vol. 62, pp. 1980–1992, Apr. 2016.
- [22] X.-C. Gao, J.-K. Zhang, H. Chen, Z. Dong, and B. Vucetic, "Energy-efficient and low-latency massive SIMO using noncoherent ML detection for industrial IoT communications," *IEEE Internet of Things Journal*, vol. 6, pp. 6247–6261, Aug. 2019.
- [23] V. Baeza and A. G. Armada, "Non-coherent massive simo system based on M-DPSK for Rician channels," *IEEE Trans. Vehicular Tech.*, vol. 68, pp. 849–853, Mar. 2019.
- [24] V. Baeza, A. G. Armada, W. Zhang, M. El-Hajjar, and L. Hanzo, "A noncoherent multiuser large-scale SIMO system relying on M-ary DPSK and BICM-ID," *IEEE Trans. Vehicular Tech.*, vol. 67, pp. 1809–1814, Feb. 2018.
- [25] M. Brehler and M. K. Varanasi, "Noncoherent multiuser space-time communications: optimum receivers and signal design," in *Proc. of the 35th Annual Conference on Information Sciences and Systems*, (The John Hopkins University, Baltimore (MD), USA), Mar. 2001.
- [26] M. Brehler and M. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh fading channels with application to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2383–2399, 2001.
- [27] K.-H. Ngo, S. Yang, M. Guillaud, and A. Decurninge, "Noncoherent MIMO multiple-access channels: A joint constellation design," in *2020 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2021.
- [28] K.-H. Ngo, S. Yang, M. Guillaud, and A. Decurninge, "Joint constellation design for noncoherent MIMO multiple-access channels," *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7281–7305, 2022.
- [29] K.-H. Ngo and S. Yang, "A Riemannian metric for noncoherent constellation design and its application to multiple access channel," in *WSA 2021; 25th International ITG Workshop on Smart Antennas*, pp. 1–6, 2021.
- [30] M. Brehler and M. K. Varanasi, "Coherent multiuser space-time communications: optimum receivers and signal design," in *Proc. of the 35th Annual Conference on Information Sciences and Systems*, (The John Hopkins University, Baltimore (MD), USA), Mar. 2001.
- [31] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. USA: Princeton University Press, 2007.
- [32] M. L. McCloud, M. Brehler, and M. Varanasi, "Signal design and convolutional coding for noncoherent space-time communication on the block-Rayleigh-fading channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1186–1194, 2002.
- [33] D. Cuevas, J. Álvarez-Vizoso, C. Beltrán, I. Santamaria, V. Tucek, and G. Peters, "Union bound minimization approach for designing Grassmannian constellations," *accepted in IEEE Trans. Comm.*, 2023.
- [34] H. Federer, *Geometric measure theory*. Berlin–Heidelberg–New York: Springer-Verlag, 1969.
- [35] J. R. Magnus, "On differentiating eigenvalues and eigenvectors," *Econometric Theory*, vol. 1, no. 2, pp. 179–191, 1985.
- [36] D. Ramírez, I. Santamaria, and L. Scharf, *Coherence: In Signal Processing and Machine Learning*. Springer Nature, 2023.
- [37] M. Chiani, M. Z. Wing, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2363–2371, Oct. 2003.

Javier Álvarez-Vizoso received his Licenciatura degree in Physics at Universidad de Oviedo, Spain, in 2013, and graduated with a Ph.D. in mathematics at Colorado State University, USA, in 2018. He has held a postdoctoral position in applied mathematics at the Max Planck Institute of Solar System Research, Germany. In October 2021 he joined the Department of Communications Engineering, Universidad de Cantabria, as postdoctoral fellow, where he is currently working for the Advanced Signal Processing Group. His research interests range from Riemannian optimization to machine learning and their applications to communications.

Diego Cuevas received the B.E. degree in telecommunications engineering from the Universidad de Cantabria (UC), Spain, in 2018, and the M.Sc. degree in telecommunications engineering from the Universidad de Cantabria, Spain, in 2020. In October 2020, he joined the Department of Communications Engineering, Universidad de Cantabria, where he is currently working toward the Ph.D. degree in electrical engineering. His current research interests include noncoherent wireless communications, signal processing algorithms for multiple-antenna wireless communications and machine learning.

Carlos Beltrán received the Ph.D. degree in mathematics from the Universidad de Cantabria (UC), Spain, in 2006. He held a Postdoctoral Fellowship at the University of Toronto during 2007 and 2008, and is currently Full Professor on Mathematical Analysis at the Universidad de Cantabria. He has published over 40 research papers including some which describe a probabilistic solution to Smale's 17th problem and other well-known open problems in the frontiers of Mathematics and Computation. His research interests include numerical analysis, complexity and approximation theory, as well as applied problems. He received the Jose Luis Rubio de Francia 2010 prize (Real Sociedad Matemática Española) and the Stephen Smale 2014 prize (Society for the Foundations of Computational Mathematics) for his contributions. He has delivered more than 50 talks including several plenaries at different international conferences.

Ignacio Santamaria (M'96–SM'05) received the Telecommunication Engineer degree and the Ph.D. degree in electrical engineering from the Universidad Politécnica de Madrid (UPM), Spain, in 1991 and 1995, respectively. Since 2007 he is Full Professor at the Universidad de Cantabria (UC). He has co-authored the book "Coherence in Signal Processing and Machine Learning", Springer, 2022, and more than 200 publications in refereed journals and international conference papers. His research interests include signal processing algorithms and information-theoretic aspects of multiuser multi-antenna wireless communication systems, multivariate statistical techniques and machine learning theories. He has been involved in numerous national and international research projects on these topics. He has been a visiting researcher at the University of Florida (in 2000 and 2004), at the University of Texas at Austin (in 2009), and at the Colorado State University (in 2015). Prof. Santamaria was General Co-Chair of the 2012 IEEE Workshop on Machine Learning for Signal Processing (MLSP 2012). He has been a member of the IEEE Machine Learning for Signal Processing Technical Committee (2009–2014), member of the IEEE Signal Processing Theory and Methods (SPTM) Technical Committee (2020–2022), and member of the IEEE Data Science Initiative (DSI) steering committee (2020–2022). He has been Associate Editor of the IEEE Transactions on Signal Processing (2011–2015), and Senior Area Editor of the IEEE Transactions on Signal Processing (2013–2015). Prof. Santamaria was a co-recipient of the 2008 IEEE/COM Innovation Award for the development of a new RF-MIMO transceiver, coauthor of a paper that received the 2012 IEEE Signal Processing Society Young Author Best Paper Award, and he has recently received the 2022 IHP International Wolfgang Mehr Fellowship Award.

Vít Tuček received his PhD degree in differential geometry and global analysis at the Charles University in Prague, Czech Republic in 2017. In 2018 he joined the Scientific Center of Excellence for Quantum and Complex Systems, and Representations of Lie Algebras (QuantiXLie) at the University of Zagreb, Croatia. Currently he is affiliated with Huawei Sweden R&D where he works on non-coherent communication schemes. His main focus are homogeneous spaces and related representation theory.

Gunnar Peters received the Ph.D. degree in mathematics and the Master of Engineering degree in applied physics from the Royal Institute of Technology. He is a Senior Expert of Network Algorithms with Huawei, Sweden. He has worked in telecom for 19 years and started with Ericsson in 2000, where he was an Expert in Radio Performance. In 2010, he joined the Wireless Organization, Huawei, where he is the Director of Wireless Network Algorithm Lab. He has also worked with the University of South Carolina, Uppsala University, and the Royal Institute of Technology. His research covers stochastic state mechanics, signal processing, and radio resource management. He has received the Individual Gold Medal Award and the Team Gold Medal Award for his work in massive MIMO, advanced receivers, and the application of machine learning to RAN.